EDUCATIONAL FRAMERWORK ON CLOUD COMPUTING

Part 2

A-CCT

Maja Pucelj, Annmarie Gorenc Zoran, Nadia Molek, Ali Gökdemir, Ioan Ganea, Christina Irene Karvouna, Petter Grøttheim, Leo Mršić, Maja Brkljačić, Monika Rohlik Tunjić, Alojz Hudobivnik

EDUCATIONAL FRAMEWORK ON CLOUD COMPUTING
Second part

Novo mesto, 2023

EDUCATIONAL FRAMEWORK ON CLOUD COMPUTING - SECOND PART

Maja Pucelj, Annmarie Gorenc Zoran, Nadia Molek, Ali Gökdemir, Ioan Ganea, Christina Irene Karvouna, Petter Grøttheim, Leo Mršić, Maja Brkljačić, Monika Rohlik Tunjić, Alojz Hudobivnik

Content

FIGURE CONTENT

TABLE CONTENT

1 INTRODUCTION

Inside our project team all kinds of data are respected and carefully managed as one of the company's most valuable asset. Our emotion for data can be compared with great music or, as we like to say, with great dish. On first notice, all cooks can look the same but when you taste their works you can easily tell which one is great.

We wish you a warm welcome to A-CCT Project.

In this document, you will find several value propositions which we find best suited to getting started with cloud services.

We focus our special attention to convergence in industries today, therefore we combine best practices from different industries to provide tailor made solutions with maximum efficiency for our clients.

2 APPLICATIONS

2.1 Automated Facilities Management

Methodology

Integrated facility management has been quick to adapt to new technologies and trends and improve on existing deliverables. The use of emerging technologies in facility management systems has been a boon for bringing about better and more precise delivery of service. One key element in the use of these technologies has been to automate the entire system so there is instant data and information transfer between the various units and faster mobilization of resources.

An intelligent network of electronic devices is formed and then manipulated by computerized control systems. This control system is designed to manage and monitor all the mechanical and electronic aspects, send signals for plumbing errors and even security breaches. It also controls the lighting system and interior temperature, keeps tabs on all devices — their functioning and failure — and sends immediate notifications through emails to parties concerned.

Automated facilities management involves the use of various methodologies to streamline and automate building and facility management processes. Here are some of the main methodologies used for automated facilities management:
1. IoT sensors: Internet of Things (IoT) sensors and devices can be deployed throughout a building to collect data on various systems, such as HVAC, lighting, and energy usage. This data can then be analysed to identify inefficiencies, optimize performance, and automate processes.
2. Smart building systems: Smart building systems enable remote monitoring and control of various building systems, such as HVAC, lighting, security, and occupancy. This allows for real-time optimization of these systems, reducing costs and improving occupant comfort.

3. Predictive maintenance: Predictive maintenance uses data analytics to identify maintenance issues before they occur. This allows organizations to address maintenance issues before they turn into more costly problems.
4. Building automation systems: Building automation systems (BAS) are used to control and monitor various building systems, including HVAC, lighting, and security. By automating these processes, organizations can reduce energy usage, save costs, and optimize system performance.
5. Energy management software: Energy management software analyses data from energy sources to optimize energy usage and reduce costs. This involves monitoring and controlling energy consumption and implementing strategies to reduce it over time.

Overall, the methodology used for automated facilities management involves the use of various technologies to collect, analyse, and automate building management processes. By using these methodologies, organizations can reduce costs, improve efficiency, and enhance the occupant experience of their buildings and facilities.

Value

Here are some potential benefits of automated facilities management:
1. Cost savings: Automated facilities management can reduce energy costs and optimize facility usage, leading to significant cost savings over time.
2. Improved efficiency: Automation can streamline many FM processes, such as work order management, inventory tracking, and preventive maintenance scheduling, freeing up time for staff to focus on more critical tasks.
3. Enhanced safety: Automated systems can monitor and control hazardous conditions in real-time, ensuring that facilities are secure and safe for occupants.
4. Increased accuracy: Automated systems are less prone to human error, ensuring that tasks are performed accurately and consistently.
5. Better data insights: Automation generates a wealth of data that can be used to make informed decisions about FM strategies, improve operations, and identify opportunities for further cost savings.

System Architecture

A high-level overview of the architecture of an automated facilities management system:
1. Data Collection: This is the first layer of the architecture, which involves collecting data from various sources across the facility, such as sensors, building automation systems, and other smart devices.

2. Data Processing: Once the data is collected, it is processed and analysed using machine learning algorithms to extract meaningful insights about facility performance and occupant behavior.

3. Decision Making: Based on the insights gathered from the data processing layer, the system generates automated decisions on tasks such as scheduling preventive maintenance, predicting potential issues and take preventative actions, and managing work orders.
4. Communication: The system then communicates these decisions to relevant personnel or devices, such as maintenance technicians, building automation systems, or other automated systems.
5. Reporting: The final layer of the architecture involves reporting on key facility metrics, providing stakeholders with real-time information on facility performance.

The different layers of the automated facilities management architecture work together to create an intelligent, data-driven ecosystem that optimizes facility operations and enhances the occupant experience.

2.2   Access to a database using a person's fingerprint as a password.

Value

Databases that contain important data that must be accessed by a small number of people have a high level of security. As a result, to access these databases, elements must be used that belong exclusively to the persons entitled to access the data stored in the database.
A specific element of a person is the fingerprint. Fingerprint-based access requires a hard device that reads the fingerprint of the person in question. The access method greatly reduces the risk of unauthorized persons penetrating the database.

Application architecture

- The database is located on a server or computer. The database is accessed by an application that uses the data from the database to solve some problems.
- Fingerprint reader
- It is a hard assembly provided with a sensor that reads the fingerprint of the person who wants to access the database. This montage is connected to the computer that contains the application that accesses the database.
- The application that uses the database. Entry into the application can be done based on the fingerprint read by the digital password sensor of the installation connected to the computer.

2.3   Active Directory Server

Value

The Active Directory (AD) Server provides significant value to an organization by improving the efficiency and security of its network and IT infrastructure. Here are some of the key areas where an AD server can provide value:

- Centralized Network Management: AD provides a centralized management location for all network resources, simplifying network administration and reducing administrative overhead. This centralization also enables system administrators to enforce security policies across the organization, ensuring that only authorized users can access specific resources.
- User Management: AD enables system administrators to manage user accounts, permissions, and access across the network, including remote locations. This centralization of user management helps to streamline IT tasks and reduce the likelihood of errors or inconsistencies in user management.
- Authentication and authorization: AD provide a secure and reliable method for user authentication and authorization, allowing users to securely access network resources with appropriate permissions and privileges. This helps to maintain the integrity of sensitive data and reduce the risk of security breaches.
- Group Policy Management: AD provides a comprehensive group policy management system that enables administrators to manage security settings, desktop configurations, and other policies across the organization from a central location.
- Scalability: AD offers a scalable infrastructure for authentication and user management, providing a foundation to support the growth of the organization over time. This scalability helps reduce costs while providing an efficient and adaptable infrastructure.

Overall, the Active Directory Server provides significant value to an organization by simplifying network administration, improving security, ensuring consistency in user and file management, and providing a scalable infrastructure for growth.

System Architecture

AD is divided into two layers: physical and logical. The physical layer describes and controls how AD works within the Windows® operating system architecture (for example which low-level operating system services and components it can access). The logical layer is more conceptual, allowing description of the organization and how it operates.

The system architecture of an Active Directory (AD) Server can vary based on the organization's requirements, but generally follows a similar structure:

1. Domain Controllers: These are the core component of the AD infrastructure, responsible for managing authentication and authorization for users and devices. Multiple domain controllers are typically deployed across the network to provide redundancy and ensure availability.
2. Forests: A forest is a collection of one or more domains, which are connected by trust relationships. Forests help to organize and simplify the management of complex organization structures, such as those with multiple departments or regional offices.
3. Domains: A domain is a logical grouping of network resources, such as computer and user accounts, printers, and other devices. Domains provide a centralized point of management for these resources and enable security policy enforcement across the network.

4. Organizational Units (OUs): OUs are containers within domains that allow for more granular management of resources, such as users, devices, and groups. OUs can be used to apply specific policies, delegate administrative tasks to specific users or groups, or to help organize resources by department or location.
5. Trust Relationships: Trust relationships define the level of access and permissions allowed between domains or forests. Trust relationships are used to provide secure and efficient access to resources across different domains or forests.
6. Sites: Sites are used to group domain controllers based on their physical location, allowing for faster logon times and more efficient replication of data between domain controllers.

Overall, an AD infrastructure is designed to provide a centralized and secure authentication and authorization service for users and devices within the organization. The system architecture of an Active Directory Server is designed to be flexible, scalable, and robust, providing high availability and management efficiency.

2.4   Active Directory Server

Explanation

Active Directory (AD) is a directory service developed by Microsoft for Windows domain networks. It is included in most Windows Server operating systems as a set of processes and services. Initially, Active Directory was used only for centralized domain management. However, Active Directory eventually became an umbrella title for a broad range of directory-based identity-related services.

A server running the Active Directory Domain Service (AD DS) role is called a domain controller. It authenticates and authorizes all users and computers in a Windows domain type network, assigning and enforcing security policies for all computers, and installing or updating software. For example, when a user logs into a computer that is part of a Windows domain, Active Directory checks the submitted username and password and determines whether the user is a system administrator or normal user. Also, it allows management and storage of information, provides authentication and authorization mechanisms and establishes a framework to deploy other related services: Certificate Services, Active Directory Federation Services, Lightweight Directory Services, and Rights Management Services.

The purpose of Active Directory is **to enable organizations to keep their network secure and organized without having to use up excessive IT resources**. For example, with AD, network administrators don't have to manually update every change to the hierarchy or objects on every computer on the network. **What does Active Directory do?**

Active Directory **stores information about objects on the network and makes this information easy for administrators and users to find and use**. Active Directory uses a structured data store as the basis for a logical, hierarchical organization of directory information.

There are many reasons why enterprises use directory services like Active Directory. The main reason is convenience. Active Directory enables users to log on to and manage a variety of resources from one location. Login credentials are unified so that it is easier to manage multiple devices without having to enter account details to access each individual machine.

The Active Directory Portfolio

In simplistic terms AD is often likened to a form of company phone book for the computer systems: providing a centralized directory which stores information about resources on the network, so that users can look them up and access them securely with the correct authority. So, for example, a user can easily find their nearest printer and be given access to use it.

In fact, this is only one aspect, and AD is a portfolio of technologies that provide the following broad-brush authentication, identification, and security facilities:

- The systems directory – Active Directory® Domain Services (AD DS)
- Managing users' rights to access and use content – Active Directory® Rights Management Services (AD RMS)
- Federation of user identity across, and between, organizations – Active Directory® Federation Services (AD FS)
- Handling digital certificates – Active Directory® Certificate Services (AD CS)

AD provides a centralized way to handle all these issues. It makes system and resource management more efficient and secure, increases user productivity, protects intellectual property, and helps with corporate policy and compliance issues.

The Architecture of Active Directory

AD is divided into two layers: physical and logical. The physical layer describes and controls how AD works within the Windows® operating system architecture (for example which low-level operating system services and components it can access). The logical layer is more conceptual, allowing description of the organization and how it operates (Fig. 2.1.).
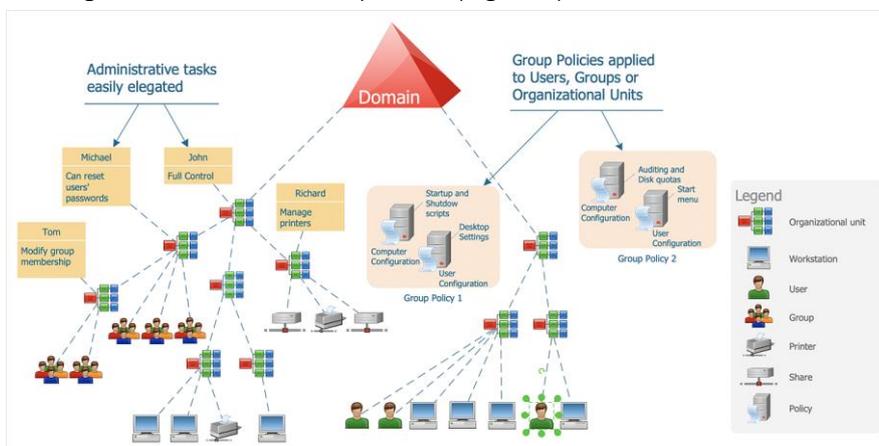
Figure 2.1. The Architecture of Active Directory. Adapted from "*Active Directory*", by Medium, 2022, p. 101.

How to Setup Active Directory (Fig. 2.2.)
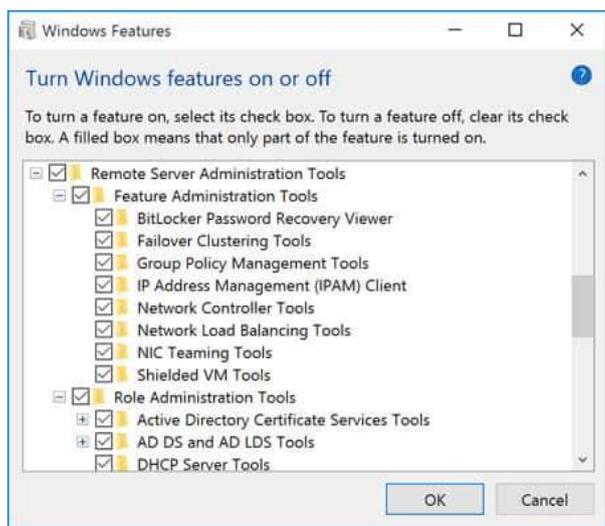
Figure 2.2. Windows feature

To begin you will need to first make sure that you have Windows Professional or Windows Enterprise installed otherwise you won't be able to install Remote Server Administration Tools.

Then do the following:

For Windows 10 Version 1809 and Windows 11:

1. Right-click on the Start button and go to Settings > Apps > Manage optional features > Add feature.
2. Now select RSAT: Active Directory Domain Services and Lightweight Directory Tools.
3. Finally, select Install then go to Start > Windows Administrative Tools to access Active Directory once the installation is complete.

For Windows 8 (And Windows 10 Version 1803)

1. Download and install the correct version of Server Administrator Tools for your device: Windows 8, Windows 10.
2. Next, right-click the Start button and select Control Panel > Programs > Programs and Features > Turn Windows features on or off.
3. Slide down and click on the Remote Server Administration Tools option.
4. Now click on Role Administration Tools.
5. Click on AD DS and AD LDS Tools and verify AD DS Tools has been checked.
6. Press Ok.
7. Go to Start > Administrative Tools on the Start menu to access Active Directory.

2.5   AI Behavior analysis Systems

Methodology: Select goals, KPIs, and metrics.

To determine whether users are reaching the right goals (e.g., purchase or conversion), select the KPIs and metrics that indicate progress toward those goals.
Examples:

- A fitness app that makes money through monthly subscriptions tracks paid subscriber growth.
- An enterprise resource planning (ERP) software that relies on annual contracts tracks users that complete the onboarding sequence.

Note that behavioral analytics goals might be a bit different than you're used to; we tend to conflate goals and results, and it's much better to start from a goal centered on measurement rather than a goal centered on a particular result.

For example: "Increase sign-up conversion from 5% to 10%" is a wonderful goal in general—but it's not an analytics goal. It's a result.

Why? Because measuring something is not the same as changing it. The macro goal of any behavioral analytics project is to accurately capture behaviors, so we can analyse those behaviors and then make informed decisions about what we may change to make an impact.

A revised version of this goal might be: "track all user journeys from Landing to Sign Up and find the best/worst conversion paths." The revised version is a better goal because it focuses on measurement rather than a result.

Map out the critical paths.
Critical paths are the series of events that need to occur or actions your users need to take for your product to be successful.

To determine your critical paths, ask, "What are the most common paths for users to reach their goals based on this service or app's design?" If the product has already been launched, you can use actual user data to answer this question. If the product is pre-launch, use wireframes of the suspected or intended flow.

Critical path examples:

- An e-commerce website tracks a user from their first page visit to adding an item to their shopping cart to checkout because that flow leads to purchases.
- A streaming music app can track users as they move from its homepage to playing a song and, hopefully, purchasing that song.

Create a tracking plan

Based on the user flow, decide which events you need to track. Tracking everything is a mistake too much data clutters the analytics and makes useful information harder to find.

Some events contain multiple properties. For example, the event for playing a song within a music app could contain properties for the song title, genre, and artist.

To keep events and properties organized, companies typically create a tracking plan in a spreadsheet. This directory of all events serves as a map for implementing the analytics tool (see Step 5, below). A tracking plan should be revised and updated as the product, team, and goals change.

(To reduce the burden of trying to share and control access to the spreadsheet, Mix panel's Lexicon feature stores the event name taxonomy for all to see.)

Involve all teams—analytics, product, marketing, and engineering—in drafting the tracking plan. They'll all need to understand how the users and events are named and organized if they're going to run reports and understand the results.

Use unique identifiers to understand behavior across platforms.

Most digital products today exist across multiple platforms, which makes it difficult to track unique users and understand their behavior across these platforms. One user can appear to be multiple people unless assigned a unique identifier—either an email or string of characters—that persists across platforms and devices and connects the touch points along their journey.

Teams should ensure their behavioral analytics platform vendor provides a unique identifier that won't change over time so that they can track and understand their behavior across platforms.

Implement analytics and begin event tracking.

Once the tracking plan is complete, you can deploy behavioral data analytics software and use its SDK or API to integrate it with your products. Assign a unique identifier for users and set up user and event properties as outlined in the tracking plan.

Sometimes you'll discover additional events you want to track during implementation. This isn't an issue as long as you update both the tracking plan and the analytics service.

Before the tracking system goes live, use test devices to verify that the event and user tracking is firing properly. Mixpanel's schemas API and data audit features can help streamline this process.

When all your tests pass, you're ready to start collecting real data and analyzing users.

Value

AI behavior analysis systems can create significant value by providing insights into human decision-making and behavior patterns that were previously difficult to obtain. Some potential benefits of these systems include:

1. Improved decision-making: By analyzing large sets of data to identify patterns and anomalies, these systems can help decision-makers make more informed choices and take preventative action early.
2. Risk mitigation: By identifying unusual behavior patterns, AI behavior analysis systems can help prevent fraud, cyberattacks, insider threats, and other risks.
3. Improved customer experience: Companies can use AI behavior analysis systems to increase their understanding of customer behavior to better personalize services, reduce customer churn, and increase satisfaction.
4. Enhanced security: By monitoring activity, AI behavior analysis systems can help increase safety measures by detecting potential threats or breaches in real-time.
5. Better healthcare outcomes: Behavior analysis systems can help identify early warning signs of health issues before they become severe, allowing for more proactive and targeted intervention.

Overall, AI behavior analysis systems can provide significant value to organizations by improving decision-making, risk mitigation, customer experience, security, and healthcare outcomes. As the capabilities of these systems continue to develop, they have the potential to create even greater value across a wide range of industries.

System Architecture

AI behavior analysis systems typically consist of three main components: data acquisition and management, analytics and modelling, and presentation and visualization.

1. Data acquisition and management: This component focuses on collecting and managing the data used to build machine learning models. This component typically includes software and hardware components such as data warehouses, data pipelines, data lakes, and data quality tools.
2. Analytics and modelling: This component focus on building machine learning models that uncover patterns in data and detect anomalies and correlations. This component typically uses algorithms such as clustering, classification, and regression to identify patterns in data.
3. Presentation and visualization: This component of an AI behavior analysis system aims to present the insights generated by the analytics and modelling to the end-users in an accessible and easily understandable way. This component may include dashboards, visualizations, and alerts to provide actionable insights to decision-makers.

In addition to these three main components, an AI behavior analysis system may also incorporate artificial intelligence and machine learning techniques to automate decision-making processes and to enable machine learning models to adapt and improve over time.

Overall, the architecture of an AI behavior analysis system is driven by the need to collect, clean, model, and present data in ways that enable users to gain insights into human behavior patterns. By leveraging these insights, organizations can make more informed decisions, mitigate risks, and improve overall performance.

2.6 Application for monitoring autonomous room cleaning equipment (vacuum cleaners) at the headquarters of small and medium-sized companies or in private homes

Value

Classic vacuum cleaners used for cleaning rooms are operated by a person who carries the suction nozzle on the floor of the room and moves the vacuum cleaner by pulling or pushing it. Vacuum cleaners have appeared that move autonomously (without human help) and the suction nozzle slides on the surface to be vacuumed as the vacuum cleaner moves.

The movement of these vacuum cleaners can be done step by step by pressing some buttons on a remote control so that the vacuum cleaner performs linear movements or turns with an angle. This requires the presence of a person to remotely control the vacuum cleaner with the help of the remote control. In this way, the person's physical effort is eased and the suction efficiency increases.

Another mode of operation of autonomous vacuum cleaners is to operate according to a certain algorithm established by the manufacturer. This algorithm is selected by the person supervising the vacuum cleaner by pressing a key on the remote control. The use of these algorithms does not always lead to efficient suction.

If an application is created that communicates with the vacuum cleaner through infrared signals or radio signals as it communicates with the remote control, personalized work regimes can be created depending on the places where the obstacles that must be replaced by the vacuum cleaner during the vacuuming operation are located.

In this way, a high quality of cleaning can be obtained because through the application, the path of movement of the vacuum cleaner can be established more precisely and the situation in which unvacated spaces can be avoided.

Also, one person can supervise several vacuum cleaners that work simultaneously and that can be operated by the application. The duties of the person supervising the operation remain only those of intervening when there are blockages in the movement of the vacuum cleaner, cleaning the bag with the vacuumed impurities and storing the vacuum cleaner.
The advantages of the application are the following:
- substantially reduces the physical effort made by the personnel responsible for cleaning;
- reducing the cleaning time of a room and increasing the quality of cleaning;

- the possibility of simultaneous use of several vacuum cleaners for cleaning the rooms of a building and reducing the number of people supervising the vacuuming operation.

System architecture

The application has the following components:
Hard application based on the Arduino board that creates communication with the vacuum cleaner and which takes over the IoT technology the commands for the vacuum cleaner from the application that manages the mode of action of the vacuum cleaner.

It is necessary to create a Cartesian or polar coordinate system that specifies the position of the vacuum cleaner on the surface of the work area. This coordinate system is created with respect to a fixed point in the room marked by a device with which the vacuum cleaner communicates via infrared or ultrasonic.

There is continuous communication between the reference point, the vacuum cleaner and the application implemented on a computer. The application monitors the position of the vacuum cleaner at any time and based on an algorithm, orders the movement of the vacuum cleaner on the surface to be cleaned.

The application can be written in C# or Python. In order for the application to be functional, a digital map of the room must be created so as to establish the coordinates of the obstacles in the respective room that must be bypassed by the vacuum cleaner.

This map can be stored in a database and used every time a certain room is vacuumed. Obviously, the map can only be useful if the structure of the obstacles in the room has not changed in the meantime (the position of the cupboards, sofas, desks). Also, the application can establish the route of the vacuum cleaner from one room to another if they are located at the same level.

For each room, it is necessary to use the room map. The reference point of the position of the vacuum cleaner on the surface to be vacuumed can be one for several rooms or there can be one point for one room.

Also, the application has a graphic interface that marks the path the vacuum cleaner took and how many times it passed on the same route. This interface provides important clues about how the surface was cleaned and whether it was completely cleaned.

The implementation stages of the application are:
- Establishing the signals emitted by the remote control for the respective vacuum cleaner. Communication with the vacuum cleaner is done through a sequence of rectangular signals. The number of periods and their duration for a certain command signal depends on the microcontroller used in the construction of the vacuum cleaner.
- Establishing the type of coordinates to which the position of the vacuum cleaner is related. To specify the position of the vacuum cleaner on the work surface, Cartesian coordinates in the plane

or polar can be chosen. At this point, the origin of the chosen coordinate system is also established. In this system, a mounting is placed that receives a signal from a mounting mounted on the vacuum cleaner based on which the coordinates where the vacuum cleaner is located on the work surface are established.

- Writing the application for the Arduino board that communicates with the vacuum cleaner. Communication with the vacuum cleaner is done through infrared signals emitted by a mounting located at the origin of the coordinate system. This assembly transmits to the computer the coordinates of the vacuum cleaner and transmits to the vacuum cleaner the commands for moving it along a certain route.
- Writing the application that commands the movement of the vacuum cleaner. This application can be written in C#, python or Java. Functioning as a digital map of the room, the application transmits the commands for moving the vacuum cleaner on the surface so that the surface is effectively covered.
- Establishing the digital map of each room in which it is to be cleaned.
- This map is stored in a database or in a .csv file. It is specific to each room.
- Checking and adjusting the application consists in checking the way in which the application commands the operation of the vacuum cleaner. Necessary adjustments will be made.

2.7 Application for managing the activity of renting tools and equipment from a company to natural persons

Value

Renting the necessary equipment or tools for different activities is an acceptable solution in terms of costs. This solution is applicable when the equipment is expensive, and it is not justified to buy them for the respective activity.

Managing the equipment rental activity can become a difficult activity for a small or medium-sized company. The evidence can be kept in excel or word files, but it is not the best solution.
Managing the equipment rental activity can become a difficult activity for a small or medium-sized company. The evidence can be kept in excel or word files, but it is not the best solution.
Creating a relational database that includes all data about equipment, about their rental activity and about customers is a much better solution.

The proposed application accesses the database and, through a graphic interface with several pages, keeps track of the entries and exits of equipment in the company. The application provides data to customers about the equipment available to be rented.

Access to the application pages is based on the password in a different way. The personnel who deal with the equipment rental do not have access to the pages through which data is entered in the database.

Application architecture

The application consists of:

- Database

The database stores the data related to the equipment intended for rent, to the customers who rented and to the rental and return operations of the equipment.

- Graphical user interface.

It consists of several pages and each page has a specific destination.

Equipment can be inserted or deleted in a page.

On another page you can view the equipment and their characteristics (Fig. 2.3. and 2.4.)

In another page, you can enter customer data.

The following pages manage the rental and return activity.

Moving from one page to another is done by pressing some buttons with the mouse.

Figure 2.3. Page displaying tools for rent
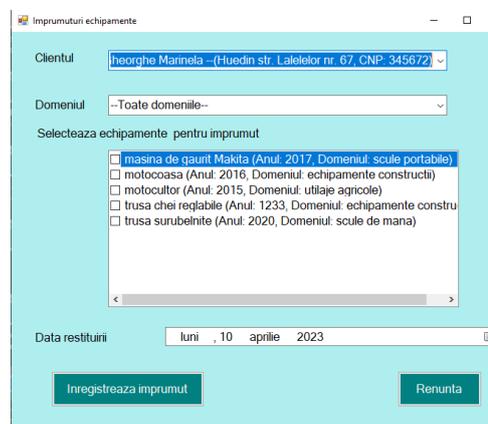
Figure 2.4. The tool rental page

2.8   Asset Tracking

Methodology

Asset tracking systems use various methodologies and technologies to monitor and manage assets as they move through the supply chain. Here are some of the main methodologies used in asset tracking:

1. RFID (Radio Frequency Identification) - RFID tags are attached to assets which are then tracked using radio waves. This system provides real-time tracking and is used to automate inventory management processes.

2. GPS (Global Positioning System) - GPS technology is used to track the movement of assets in real-time, enabling organizations to monitor the location of their assets around the clock.

3. Barcode - Barcodes are used to track and identify assets. Each asset is assigned a unique barcode, which is scanned to monitor movements and locations.

4. IoT (Internet of Things) - IoT technology enables assets to be tracked and monitored using sensors that are embedded in the assets. These sensors can send data to the cloud, which is then analysed to provide useful insights into asset movements.

2021-1-SI01-KA220-VET-000034641

5. BLE (Bluetooth Low Energy) - BLE is used to track the location of an asset within a specific area. This method is cost-effective and is useful for warehouse management.
6. Asset management software - This is software that manages and tracks assets. These systems can use the above technologies and methodologies to keep track of assets.

Overall, the methodology used for asset tracking depends on the nature of the assets being tracked, the locations they move through, and the level of detail required for the monitoring process. By leveraging these tracking methodologies, organizations can better manage their assets and optimize their supply chain operations.

Value

Asset tracking provides significant value to organizations by improving their visibility into the location, condition, and movement of assets within their supply chain. Some of the key values that asset tracking offers include:

1. Enhanced inventory management: Asset tracking provides real-time visibility into inventory levels, enabling organizations to optimize their inventory management processes and reduce waste.
2. Improved asset utilization: Asset tracking provides data on asset usage patterns, enabling organizations to optimize their asset utilization and reduce downtime and waste.
3. Increased security: Asset tracking provides real-time information on the location and movement of assets, reducing the risk of theft, loss, or damage to valuable assets.
4. Compliance: Asset tracking enables organizations to comply with regulations by providing reliable data on the movement and handling of regulated assets such as pharmaceuticals and hazardous materials.
5. Improved supply chain efficiency: Asset tracking enables organizations to optimize their supply chain operations, reducing costs, improving delivery times, and increasing customer satisfaction.
6. Better decision-making: Asset tracking provides real-time data that organizations can use to make informed decisions, such as optimizing supply chain operations, forecasting future demand, and identifying inefficiencies.

Overall, asset tracking provides significant value to organizations by improving inventory management, asset utilization, security, compliance, supply chain efficiency, and decision-making capabilities. By leveraging these insights, organizations can enhance their operations, improve their customer experience, and gain a competitive advantage in their industry.

System Architecture

Asset tracking systems can vary in architecture depending on the specific application and needs of the organization. However, there are several common components that are typically present in an asset tracking solution:

1. Hardware: This includes the physical sensors and devices used to track the assets and collect data. Examples include RFID readers, GPS trackers, barcode scanners, and IoT sensors.
2. Communication networks: These networks transmit data from the tracking devices to the backend systems. Examples include cellular networks, Wi-Fi, and Bluetooth.
3. Cloud-based software: This software is used to process and analyse the data collected from the tracking devices. This component may include data storage, database management systems, and APIs for integration with other systems.
4. Mobile Applications: These are used to enable field personnel to access asset tracking information and update the location and status of assets in real-time.
5. Analytics and reporting: These tools enable organizations to analyse the data collected from the tracking system to identify trends and opportunities for improvement.

Overall, the architecture of an asset tracking system is driven by the need to track the movement of assets accurately and efficiently through the supply chain, while also providing real-time data and insights to users. By leveraging these insights, organizations can gain better control over their assets and optimize their supply chain operations. The number and type of components included in an asset tracking system can vary depending on the specific requirements of the organization.

2.9 Attendance tracker for students

Methodology

Successful schools begin by engaging students and making sure that they will come to school regularly, so the attendance rate become very important. Attendance system is a system that is used to track the attendance of a particular person and is applied in the industries, schools, universities or working places. The attendance rate will be calculated based to the average percentage of students attending school in every class of the course. The attendance rate is important because students are more likely to succeed in academics when they attend class consistently. It's difficult for the lecturer and the class to build their skills and progress if a large number of students are frequently absent. Moreover, the students have given the right to have their own time management in university. This will cause the attendance rate of the class become a major problem because some student may choose to absent from the class. Therefore, students from university in Malaysia are required to attend the class not less than 80% per semester otherwise student will be barred from taking any examinations.

The traditional way for taking attendance has drawback, which is the data of the attendance list hard to reuse. If the lecturer wants to calculate the percentage of the students that attend to the class, he/she must calculate manually or input by typing. This also easy lead to human error such as the lecturer may wrongly. The technology-based attendance system will reduce the human involvement and decrease the human error. There are various types of attendance systems that are applied in different fields. Mostly, the working places are still using the punch card system. But some of them had integrated their system into biometric attendance system. The biometric attendance system is based on fingerprint identification using extraction of minutiae techniques and it is very reliable and convenient to verify the

identity of people. Human fingerprint is read by the reader to take the attendance as the uniqueness of human's fingerprints.

Another technology is Radio Frequency Identification (RFID) based attendance system that consists overfed Reader, RFID Tag, LCD displays and microcontroller unit. RFID can be interfaced to microcontroller through Universal Synchronous, Asynchronous Receiver Transmitter (USART). Data is transferred from RFID cards to reader and from there to the microcontroller. These attendance systems are important for largescale organizations for them to process a large number of workers' attendances rapidly. It makes the work more efficient and produces accurate results.

The NFC based attendance system is another means to tackle conventional attendance system problems above. Because the installation cost of NFC based attendance system is lower than the other advance attendance system likes the fingerprint attendance system. The main advantages of the NFC are the simple and quick way of using it and the speed of connection establishment is fast. Besides that, other important advantages of NFC technology have also included the transmission range of NFC devices. The transmission range is so short, when the user separates the two devices more than the limited range, then communication is broken. The NFC based attendance system can process the data collected in a quicker way compared to manual system which need to enter the data one by one. Besides, all the data will be saved on the server, and this can avoid of losing any students 'attendance. Students can also check their attendance rate using their smartphones through the login system from time to time to avoid any miss entering of attendance. There has some research that develop technology-based attendance system. Basically technology-based attendance system can divide into two groups: i) Biometric-based Attendance System and ii) Sensor-based Attendance System.

2.10 Automation of tasks using cloud-based services: recommendation engine

Methodology

As a first step, therefore, market basket analysis can be used in deciding the location and promotion of goods inside a store. If, as has been observed, purchasers of product A are more likely to buy product B, then high-margin candy can be placed near to the product A display. Customers who would have bought product B with their product A had they thought of it will now be suitably tempted. But this is only the first level of analysis. Differential market basket analysis can find interesting results and can also eliminate the problem of a potentially high volume of trivial results. In differential analysis, we compare results between different stores, between customers in different demographic groups, between different days of the week, different seasons of the year, etc. If we observe that a rule holds in one store, but not in any other (or does not hold in one store, but holds in all others), then we know that there is something interesting about that store. Perhaps its clientele is different, or perhaps it has organized its displays in a novel and more lucrative way. Investigating such differences may yield useful insights which can improve company sales.

Value

Although Market Basket Analysis conjures up pictures of shopping carts and supermarket shoppers, it is important to realize that there are many other areas in which it can be applied, like: analysis of credit card purchases, analysis of telephone calling patterns, identification of fraudulent medical insurance claims, analysis of telecom service purchases.

Note that despite the terminology, there is no requirement for all the items to be purchased at the same time. The algorithms can be adapted to look at a sequence of purchases (or events) spread out over time. A predictive market basket analysis can be used to identify sets of item purchases (or events) that generally occur in sequence - something of interest to direct marketers but also many others.

System Architecture

One of the key techniques used by the large retailers is called Market Basket Analysis (MBA), which uncovers associations between products by looking for combinations of products that frequently co-occur in transactions. In other words, it allows the supermarkets to identify relationships between the products that people buy. For example, customers that buy a pencil and paper are likely to buy a rubber or ruler.

"Market Basket Analysis allows retailers to identify relationships between the products that people buy."
Retailers can use the insights gained from MBA in a few ways, including:
1.   Grouping products that co-occur in the design of a store's layout to increase the chance of cross-selling.
2.   Driving online recommendation engines ("customers who purchased this product also viewed this product"); and
3.   Targeting marketing campaigns by sending out promotional coupons to customers for products related to items they recently purchased.
Given how popular and valuable MBA is, we thought we'd produce the following step-by-step guide describing how it works and how you could go about undertaking your own Market Basket Analysis.

To carry out an MBA you'll first need a data set of transactions. Each transaction represents a group of items or products that have been bought together and often referred to as an "itemset". For example, one itemset might be: {pencil, paper, staples, rubber} in which case all these items have been bought in a single transaction. In an MBA, the transactions are analysed to identify rules of association. For example, one rule could be: {pencil, paper} => {rubber}. This means that if a customer has a transaction that contains a pencil and paper, then they are likely to be interested in also buying a rubber.

Before acting on a rule, a retailer needs to know whether there is sufficient evidence to suggest that it will result in a beneficial outcome. We therefore measure the strength of a rule by calculating the following three metrics (note: other metrics are available, but these are the three most used):

- **Support:** the percentage of transactions that contain all of the items in an itemset (e.g., pencil, paper and rubber). The higher the support the more frequently the itemset occurs. Rules with a high support are preferred since they are likely to be applicable to many future transactions.

- **Confidence:** the probability that a transaction that contains the items on the left-hand side of the rule (in our example, pencil and paper) also contains the item on the right-hand side (a rubber). The higher the confidence, the greater the likelihood that the item on the right-hand side will be purchased or, in other words, the greater the return rate you can expect for a given rule.

- **Lift:** the probability of all the items in a rule occurring together (otherwise known as the support) divided by the product of the probabilities of the items on the left- and right-hand side occurring as if there was no association between them. For example, if pencil, paper and rubber occurred together in 2.5% of all transactions, pencil and paper in 10% of transactions and rubber in 8% of transactions, then the lift would be: 0.025/ (0.1*0.08) = 3.125. A lift of more than 1 suggests that the presence of pencil and paper increases the probability that a rubber will also occur in the transaction. Overall, lift summarizes the strength of association between the products on the left- and right-hand side of the rule; the larger the lift the greater the link between the two products.

To perform a Market Basket Analysis and identify potential rules, a data mining algorithm called the "Apriori algorithm" is commonly used, which works in two steps:

1. Systematically identify *item sets* that occur frequently in the data set with a support greater than a pre-specified threshold.

2. Calculate the confidence of all possible rules given the frequent *item sets* and keep only those with a confidence greater than a pre-specified threshold.

The thresholds at which the support and confidence are user-specified are likely to vary between transaction data sets. It is recommended to experiment with these to see how they affect the number of rules returned (more on this below). Finally, although the Apriori algorithm does not use lift to establish rules, lift can be used for exploring the rules returned by the algorithm.

2.11 Back-Up / Disaster relief

Methodology

Create a back-up vault in the azure console by going to the backup center and initiate a new vault, selecting backup vault as the option.

Select the resource groups and fill in the details about the vault such as name and region.

When the vault has been set up and connected to the resource group, create a policy, defining how often backups should be done, and how long the backups should be stored.

When the backup vault and the policy has been defined, connect the database that should be backed up and configure the backup by adding the correct policy and vault to the data and which data should be stored.

Value

The value of adequate redundancy and backup can be immeasurable. In a catastrophic data loss event, the losses can be extreme. And even small happenings can become extremely costly, both in terms of money lost, but also the loss of reputation.

System architecture

System architecture is presented in the Figure 2.5., Azure backup service.
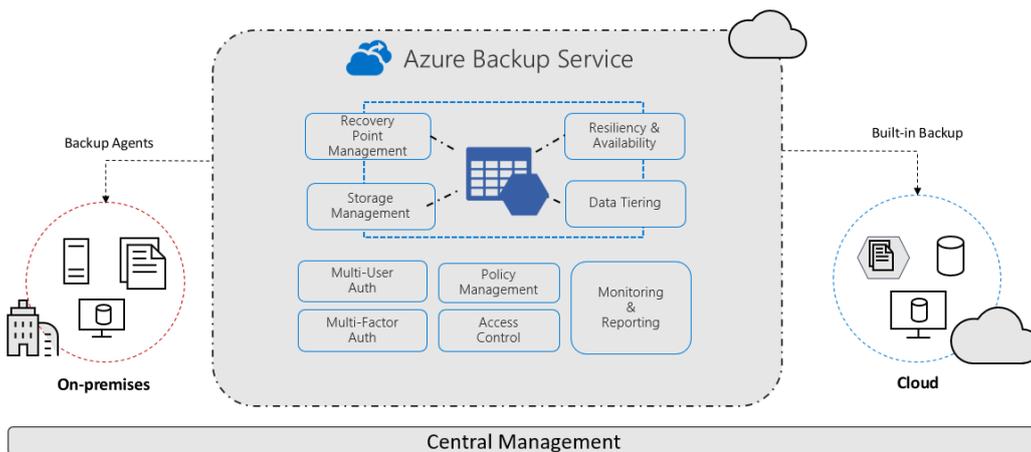
Figure 2.5. Azure Backup Service.

2.12 Chatbot for indicating free places in public parking lots in a city

Value

Both the inhabitants of a city and the people who visit that city when they travel by car to solve different problems in an area of a city face serious problems related to the places where they can park their car. Finding such a place depends to a large extent on the ability of the driver to move, on the level in which he knows the city and not least on the chance he has at a given moment.

Drivers often spend time looking for a parking spot and often find a parking spot much further from where they need to go to solve their problems.

An application that informs the driver about the number of free parking spaces in the parking lots located near the area he wants to reach is welcome. Such an application helps the driver to make a correct decision regarding the parking place where he wants to arrive and to stop wasting time looking for free spaces in the parking lots around the point where he wants to arrive to solve his problems.

An important feature is that the application communicates with the car driver by voice because he is in traffic and cannot write on his mobile phone. The application also guides the driver to the parking lot where he wants to arrive after choosing the parking lot. In addition, the application informs the driver

in real time about the situation of the parking spaces in the parking lot he has chosen.

It is possible that even if there are several free places in a parking lot chosen by the driver, they will be occupied in the meantime. Then the driver can make another decision ahead of time.

Application architecture

The chatbot application is on the driver's mobile phone or on a laptop in the car that needs to be parked. It collects data about free places in city parking lots. These data are made available by the municipality that administers the city. In this sense, there must be sensors in every public parking lot that communicate to a central application whether the respective parking lot is occupied or not.

A central application located on the municipality's servers permanently shows the state of the places in the public parking lots. The local application on the driver's device (mobile phone, laptop) accesses this data from the central application. Depending on where the driver wants to go, the local application analyses all possibilities using specialized algorithms from artificial intelligence and offers the driver the optimal option.

This application can offer more options, leaving it up to the driver to choose. After the driver has chosen a parking lot, he considers optimal, the chatbot type application directs the driver to the chosen parking lot, keeping the driver up to date with the number of free spaces in the chosen parking lot. The fact that the communication between the driver and the application is done by voice helps the driver to save time and drive the car while making the decision where to park.

2.13 Chatbot for students in EDU institution

Methodology

Profiling can be described as the act of using data to describe or profile a group of customers or prospects. It can be performed on an entire database or distinct sections of the database. The more or less distinct sections are known as segments. Segmentation can be described as the act of splitting a database into distinct sections or segments, as seen on Figure 2.6.
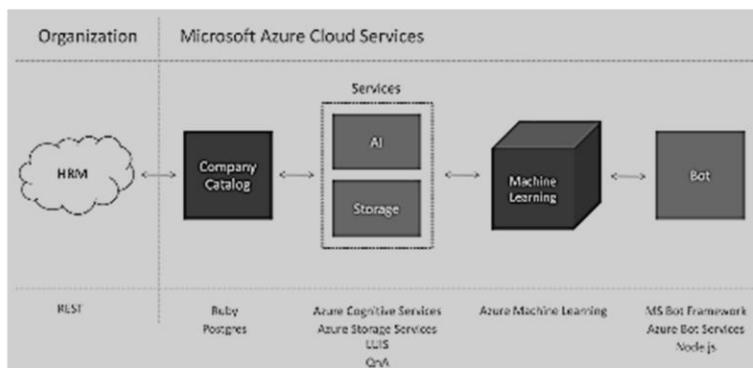
Figure 2.6. Chatbot Architecture with Technology Stack

There are two basic approaches to segmentation: market driven, and data driven. Market-driven approaches allow managers to use characteristics that they determine to be important drivers of their business. In other words, they preselect the characteristics that define the segments. This is why defining the objective is so critical. The ultimate plans for using the segments will determine the best method for creating them. On the other hand, data-driven approaches use techniques such as cluster analysis or factor analysis to find homogenous groups. This might be useful if companies are working with data about which they have little knowledge (see Figure 2.7.).
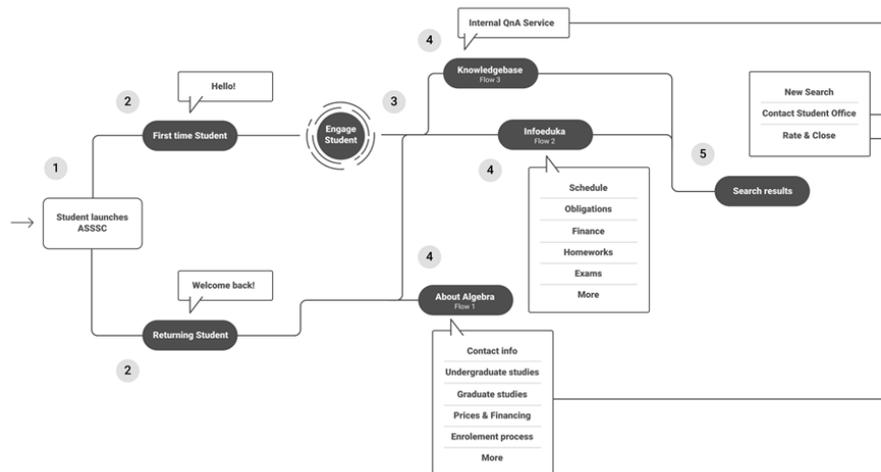
Figure 2.7. Conversation flow

Value

Data is only an asset that should be used for some higher purposes. Like any other resources like soil or coil it should be refined and on it should be added some added value to become something valuable. Knowledge and especially hidden knowledge have great potential to become, in combination of AI and driven by cognitive services, really valuable asset. One of the interfaces to that knowledgebases are Chatbots as one of last technology approaches to interaction between human and machine.

Since messaging platforms are in some moment already overtaken social network platforms, that is the open road for technologies like Chatbots. Chatbots can be used for many purposes but one of the main use-cases are in specific Bot domain of Support Bots. Powered by AI and Cognitive services they have potential to replace at least of 50% of any Support or Helpdesk services. Business case developed in this paper is Chatbot that could be trained to be used on any University to successfully decrease impact on Student Office. Research of data given by Algebra University before COVID19 says that almost 90% of commonly asked questions could be founded in open sources.

That is great opportunity for Student Service Office digital transformation. Chatbot could architecturally cover three scenarios like search for common information based on open-sourced data, search in secured internal knowledgebase and ask for student personal information about student status, obligation, schedule etc. which is enough to cover at least 40% of common student needs for Student Service engagement.

System Architecture

Software bots has the same application lifecycle as any software product but from technology perspective there are the same concerns which should be considered. Choosing right technology vendor, programming language and hosting provider is the crucial in any software product lifecycle. Whichever vendor will be chosen there should be considered all potential building blocks from the Chatbot Communication Flow.

There are some great potential solutions based on which is possible to develop truly enterprise level chatbot solutions like SAP Recast.AI, Microsoft Bot Framework, Google Dialog flow, IBM Watson and/or Amazon Lex. Every of these solutions has something that is better than other vendors have to offer but at the end as a development platform is chosen Microsoft Boot Framework because of three main reasons:

1.   MS Bot Framework can cover all targeted platforms as FB Messenger, Skype, and Slack. With the same codebase by having few considerations on mind it is possible do code once and serve on at least two targeted platforms,

2.   Bot could be hosted on MS Azure platform which could serve the bot and the bot backend services. From the security reasons Azure is great option since Skype is also Microsoft product, which is natively connected with Azure,

3.   Other vendors have great services but at this moment Microsoft is the only vendor who has everything what is needed for that kind of service. Bot Framework, Cognitive Services like Sentiment analysis, Azure Storage, Azure Active Directory, LUIS as Microsoft Language Understanding service, Hosting for both Bot and Backend services, Bot Analytics and at the end knowledge database such is QnA service (see Figure 2.8.).
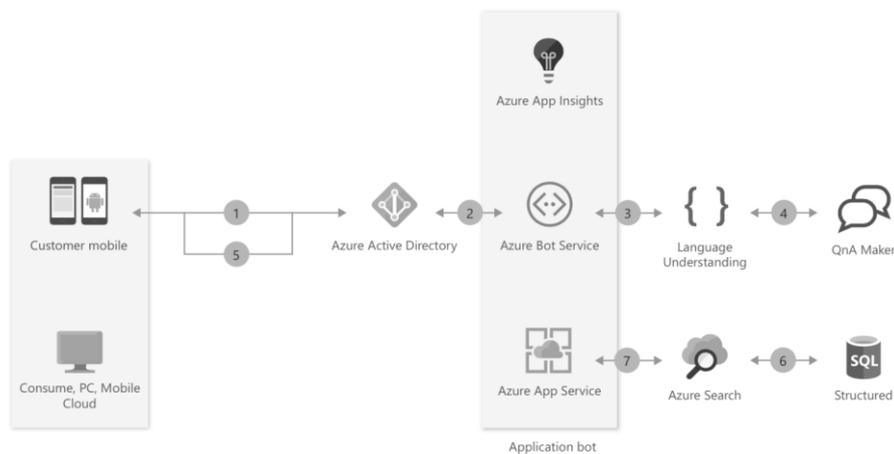
Figure 2.8. ChatBot Architecture: Azure Bot service: Microsoft Corporation

2.14 Chatbot to personalize the learning activity of students in vocational high school education

Value

The classic way of learning a lesson requires the student to go through a text written on paper or in electronic format. Solving a set of questions accompanying the lesson to be learned is the only way to check how well the student has learned the lesson. Many times, the set of questions is not accompanied by an explanation to help the student compare his own answer with the correct answer.

The only way is to return to the text of the lesson and repeat it in its entirety or on the fragments where it assumes that it can find the correct answer. The application offers the possibility for the student to learn interactively, reducing the learning effort and increasing the efficiency of the process of accumulating the information transmitted by the text of the lesson.

The application allows the student to ask questions related to the lesson he is learning, and the application gives answers relative to these aspects. Also, the application asks the student questions and provides the student's answers.

The advantages of the application are:
- The interaction between the application and the student is done in natural language using artificial intelligence.
- The student is not under stress due to the emotion generated by the fact that the answer is wrong.
- He can repeat the lesson as many times as he wants and at his own pace.
- The learning process is comfortable because the student has the feeling that he is talking to a friendly person who helps him unconditionally.
- The application together with the answer also indicates the paragraph or paragraphs in the text where the student finds the correct answer.
- The student can shade the questions in order to find the correct answers to the aspects that concern him within the lesson(s)
- The elements that lead to a misunderstanding of the lesson can also be identified. This is done by storing the questions asked by the student and their subsequent examination by specialized teachers with good training in the didactic and pedagogical field. In this way, the quality of the application is improved.

The disadvantages of the application are related to the insufficient amount of information stored in the database. The student may not find answers to all the questions he asks related to the lesson. This requires that the application is always updated. Every question asked by the student regarding the content of the lesson to be learned is important for updating the information and raising the quality level.

The conversation with the student can be done in writing or by voice. Voice conversation is closer to the natural level and requires less effort on the part of the student, but from a technical point of view it is more complicated.

2021-1-SI01-KA220-VET-000034641

For good efficiency, the application should be used for a set of several lessons. This is necessary because the student can ask questions to which the answer involves knowledge from several lessons. So, you must refer to the paragraphs in all the lessons where you can find the necessary information for a correct answer.

Application architecture

As with every chatbot there is a standard structure:
- Graphical user interface
- The application itself
- The database in which the data used for the conversation with the student are stored.
- The graphic interface with the user realizes the conversation of the human person with the actual application.

Depending on the imagination of the one who creates the application, the interface can be more complex or simpler. If the dialogue is done in writing, the user writes the message addressed to the chatbot type application.

At the same time, the application (the chatbot robot) also displays the response message on the graphic interface. If the dialogue is carried out by voice, the graphic interface displays the image of a person who simulates the movement of the lips and other elements of facial expressions to create as much as possible the feeling of having a dialogue with a real person behind the interface, similar to the way we converse via videophone.

The application itself, made up of several files, analyses the messages received from the human person through the graphic interface and prepares the answers based on the data stored in the database or in files specially created for this purpose.

The database stores the data used in the algorithms that are used in the conversation (Fig. 2.9.).
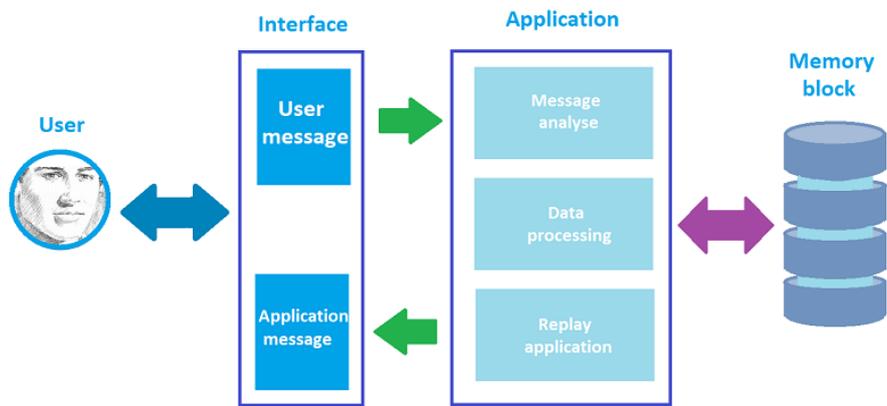
Figure 2.9. Block diagram for the application

2.15 Cloud-based e-learning

Methodology

In this paper we propose the architecture that we have designed by modifying previous architectures that we used as references. Proposed architecture consists of five layers, namely: (1) infrastructure layer; (2) platform layer; (3) application layer; (4) access layer; and (5) user layer. First layer is infrastructure layer. This layer contains architecture supporting infrastructure, such as: Cloud platform, virtual machine, virtual repositories, and physical infrastructure such as servers, network devices, storage, buildings, and other physical facilities.

The infrastructure layer shares IT infrastructure resources and connects the system huge system pool together to provide services. Cloud computing enable the hardware layer to run more like the internet, to make the hardware resources shared and accessed the data resources in secure and scalable way. The second layer is platform layer. In this layer running the operating system where e-learning application will be running. Besides the operating system, this layer also consists of variety of software that support the application layer so that it can run properly. The third layer is application layer. This layer is a specific e-learning application that is utilized for sharing learning resources and interaction among users that includes synchronous or asynchronous discussion and chatting. We added the access layer in our architecture. This access layer is the fourth layer in our proposed architecture. This layer oversees managing access to cloud e-learning services which is available on the architecture such as: types of access devices and presentation models.

 This study adopts the concept of multi-channel access which enables a variety of available services that accessible through a variety of devices (such as mobile phones, smartphones, computer, etc.) and a variety of presentation models (such as mobile applications, desktop applications, and others). The purpose of the adoption of this concept is to increase the availability of devices that access the cloud service e-learning can be found in the architecture used untrammeled access devices. Besides the

addition of the access layer, the architecture we propose the user layer consists of various educational institutions (Fig. 2.10.).
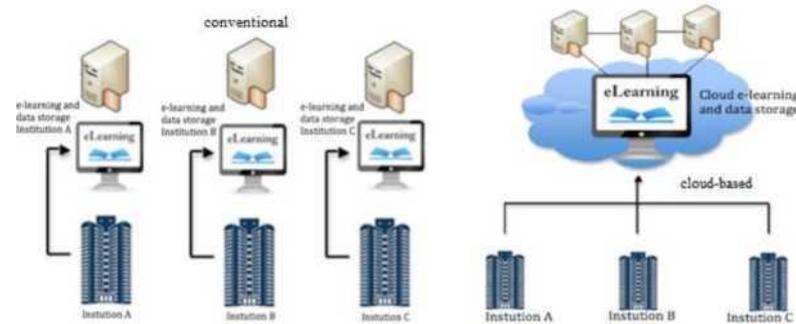
Figure 2.10. Conventional E- Learning Towards Cloud Based E-Learning

Value

The main advantages of the adoption of cloud computing is efficient in terms of cost, this is an interesting point of view that can be adapted to develop e-learning based on cloud computing. Conventional eLearning commonly used by the university developed by the university itself tend to cause lots of problems such as time to designing e-learning systems will be developed, costs for infrastructure, selecting commercial or open-source e-learning platform, the cost to hire professional staff to maintain and upgrade the system of e-learning, and so on. This process is more likely need more time.

By introducing cloud computing adopted by e-learning, institutions can use a single e-learning based on the cloud provided by a cloud provider of eLearning. This model can reduce the initial costs incurred by the institution for the implementation of e-learning by using cloud computing services, because institutions do not need to pay for the purchase of infrastructure, both in terms of procurement of servers and storage. With cloud computing, as an institution of the client can rent the infrastructure to cloud computing service providers. Likewise with the human resources for the development stage, the cloud environment of e-learning has been provided by the cloud service provider, as well as maintenance of the e-learning.

The expected advantages by adopting the cloud-based e-learning model are as follows: (1) Large capacity, these criteria could address on-demand self-service characteristic from could computing. Large scale storage in cloud environments provide advantages to the consumer to determine the storage capacity they intend to use that are adjusted to their needs and capabilities of the institution as a consumer of cloud-based e-learning; (2) Short implementation process, by using cloud-based e-learning services, educational institution could minimize their expenditure to develop the e-learning system and shorten the implementation process because the e-learning system already developed and maintained by the cloud e-learning provider; (3) High Availability, by utilizing large storage and high performance computing power, cloud e-learning could provide a high quality of service. This may happen because of the support system that supports cloud e-learning can detect the node failure and can be immediately diverted to another node. Besides the high level of availability system, with a large

storage so that many learning resources can be gathered by combining learning resources from any educational institution who joined the cloud e-learning by integrating the learning resources with integrated database system mechanism; (4) Just in time learning, using cloud computing for e-learning system encourages the use of e-learning more dynamic with added services through mobile devices, of course, by adding an integrated mobile learning services in a cloud-based e-learning. With adding mobile learning features, cloud-based e-learning become more powerful so the users could access the learning material any-time and any-where and just utilize their mobile devices like smartphones as an example.

System Architecture

The paradigm shift in the implementation of e-learning is an innovation that can help any institution in implementing e-learning. In general, the implementation conventional e-learning, e-learning web-based design, system development and maintenance as well as by internal governance institutions. It had a lot of problems, both in terms of flexibility, scalability, and accessibility. One of the main important features that can be presented in the use of e-learning in the cloud is scalability, which allows virtualization provide infrastructure layer provided by the cloud service provider. Virtualization helps solve the problem of the physical barriers that are generally inherent in the lack of resources and infrastructure to automate the management of these resources as if they were a single entity through hypervisor technologies such as virtual machine (VM) (Fig. 2.11).
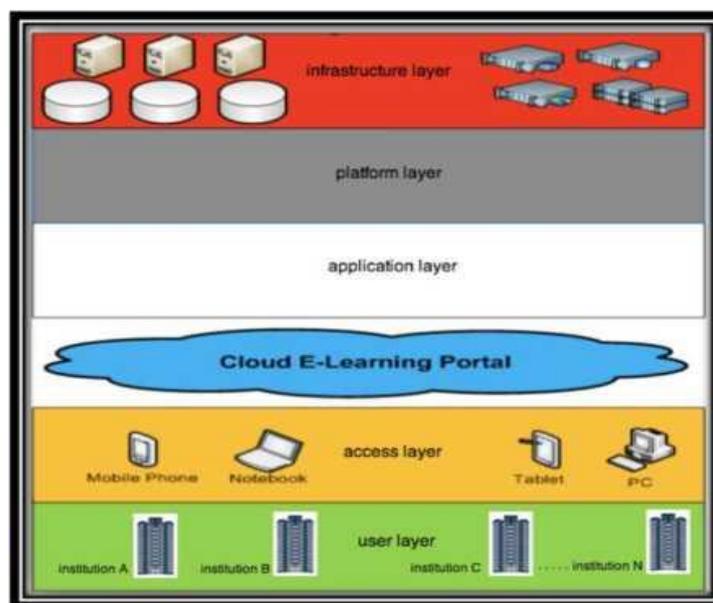
Figure 2.11. Proposed Cloud E-Learning Architecture

2.16 Communication/ Information Exchange Application/ Channels

Methodology

The methodology for building communication/information exchange applications can vary depending on the specific requirements of the application and the needs of the users. However, there are some common practices and steps that developers usually follow:

1. Defining functional and non-functional requirements: The first step is to define the functionalities that the application should have, such as messaging, voice or video calling, screen sharing, file sharing, and others. Also, non-functional requirements like performance, security, scalability, and usability should be considered.
2. Selecting the technology stack: Based on the requirements and specifications, developers select suitable technology stacks, frameworks, and programming languages to develop the application.
3. Developing the application architecture: During this step, the developers design the applications' architecture that includes the user interface, server-side, and database design.
4. Developing the core functionalities: The core functionalities of an information exchange application may include messaging, file sharing, voice and video calling, and screen sharing. Developers build these features using the defined technology stack.
5. Testing and quality assurance: After each development milestone, quality assurance tests are conducted to ensure that the application is functioning as intended.
6. Deployment: The final step is deploying the application to the production environment where users can access it.
7. Maintenance and support: After deployment, the application requires continuous maintenance and support to ensure that it remains up-to-date, secure, and functioning correctly.

Overall, the methodology used to develop communication/information exchange applications requires a multidisciplinary team that includes UX/UI designers, developers, testers, security, and support experts. A well-defined and executed methodology can ensure that the application is reliable, user-friendly, and meets the expectations of the users.

Value

Communication/information exchange applications can add value to individuals and organizations in several ways:

1. Improved communication: These applications offer users a platform to communicate in real-time, regardless of their location, facilitating efficient collaboration and problem-solving.
2. Enhanced productivity: Communication/information exchange applications streamline workflows, enable real-time access to information, and foster a sense of teamwork and collaboration, leading to enhanced productivity.
3. Cost efficiency: By enabling remote work and reducing the need for in-person meetings, communication/information exchange applications can cut down on travel costs and increase efficiency.

4. Enhanced customer service: These applications can help organizations provide better customer service by enabling real-time access to customer information and facilitating swift response to queries and complaints.

5. Competitive advantage: Efficient communication and collaboration through information exchange applications can give companies a competitive edge that enables them to stay ahead of the competition.

6. Business growth: When used correctly, information exchange applications can facilitate closer collaboration between teams and enable companies to scale and grow more quickly.

Overall, communication/information exchange applications facilitate effective communication, better collaboration, and provide easy access to information, leading to enhanced productivity, reduced costs, and a competitive advantage for individuals and businesses.

System Architecture

The architecture of communication/information exchange applications can be complex, incorporating several components. Here are some of the core components typically found in communication/information exchange applications:

1. User interface: This is the component that interacts with users, providing them with functionalities to communicate, collaborate and share information.

2. Backend servers: These servers are responsible for managing the application's core functionalities, such as messaging, file sharing, and audio/video calling.

3. Database: Applications store data, user preferences, and application logs in databases that allow for quick retrieval of information.

4. API integrations: These allow applications to connect to third-party tools and services such as email, customer relationship management (CRM), and document storage services.

5. Security: Information exchange applications must have built-in security protocols to ensure users' privacy and protect against data breaches.

6. Scalability: Information exchange applications must be able to accommodate rapid growth by being horizontally scalable across multiple servers or deployable on cloud service providers.

7. AI and machine learning: Some communication/information exchange applications integrate AI and machine learning to facilitate natural language processing, response prompts, speech recognition, sentiment analysis, and document translation.

Overall, the architecture of communication/information exchange applications is a critical aspect of their design. An efficient and well-organized architecture ensures that the application delivers a seamless and secure experience to users, reducing downtime and optimizing performance.

2.17 Continuous monitoring of the operation of some industrial installations using cloud computing and IoT technologies

Value

Liquefied gases that can produce dangerous explosive mixtures or are flammable are stored in open spaces. These spaces are located at a great distance from habitable areas in order to avoid the loss of human lives in case of extreme situations where explosions occur. The tanks located in these spaces must be constantly monitored to avoid destructive events.

The classic situation that is used is to mount sensors of pressure, temperature, concentration of the gas-air mixture, etc. These sensors are adjusted so that exceeding a certain threshold signals a local station and the alarm is triggered.

In general, at these warehouses there are security personnel who are not specialized and who do not know how to act in case of potentially dangerous situations.

In cases where the bell mounted on the local monitoring station rings, the security personnel can notify the specialized intervention teams. The intervention of these teams, which are far from the warehouse in question, can sometimes be ineffective.

The alternative is to install sensors that continuously measure the status values of the tanks and these values to be transmitted via the internet network to the company headquarters where software applications that use artificial intelligence continuously analyse the values of the status values of the tanks and inform the intervention teams about the potential danger which may appear at that warehouse.

At the headquarters of the company there are people who can make decisions depending on the situation presented by the monitored installation.

The application can notify other people who can intervene in time via mobile phone.

In all cases, there are elements of execution foreseen from the design phase to be acted upon in extreme cases to remove the danger. The application can remotely actuate the execution elements in order to prevent the total failure until the on-site movement of the teams to remedy the failure.

In case of accidental leakage of liquefied gas (propane - butane), a dangerous mixture is created between the atmospheric air and the propane-butane mixture drained from the tank. This mixture (air and gas in the tank) can become explosive even by the appearance of a spark due to a local electrostatic discharge.

A measure to keep the situation under control is monitoring the percentage of gas in the outside air and turning on some fans to exhaust the mixture into the atmosphere.

In this way, the gas concentration in the air is diluted and the mixture is no longer dangerous. In this case, the presence of the intervention team is required to remedy the defect. Starting the fans can be done directly by the application.

System architecture

The system consists of:
- sensors for measuring state quantities,
- execution elements
- the interface between the sensors, the execution elements, and the internet network
- internet network,
- the actual application implemented at the company headquarters.

The sensors continuously measure the state variables that characterize the installation. In the presented case, the sensors can measure the gas pressure in each tank, the amount of gas, the temperature of the gas in the tank, the concentration of gas in the air near the tanks.

The execution elements act in the installation to maintain its operation within normal parameters.

The interface between the sensors, the execution elements and the internet network ensure the continuous transmission of data from the sensor output to the application that monitors the operation of the installation. This interface receives the command to actuate the execution elements that adjust in the installation so that it is maintained within normal parameters.

The internet network is the support through which the data reaches from the monitored installation to the application at the company's headquarters.

The application implemented at the company's headquarters continuously analyses the data transmitted by the interface with sensors in the area of the super vegetated installation.

Data analysis is done using AI algorithms that can predict the occurrence of potentially dangerous situations in advance.

The application has a desktop component and a database.

The data provided by the sensors can be stored in a database for a longer or shorter period of time or in .csv files.

The storage of the data received from the sensors provides information about possible static errors that may appear during their operation. It is also possible to analyse the operating regimes of the installation and improve the application.

Sometimes hidden defects may appear that can only be detected easily through a detailed analysis of

the data transmitted by the sensors prior to the appearance of the defects.

The sensor interface can be made with an Arduino or Raspberry board. The application can be written in python using artificial intelligence methods. Actuators can also be activated by an Arduino-based application using IoT.

Application implementation stages:
- Establishing the quantities to be monitored and their range of variation.
- Choice of sensors and execution elements.
- Establishing their location in the installation.
- Creating the interface between sensors, execution elements and the Internet using an Arduino or Raspberry Pi board.
- Writing the application for monitoring the status variables.
- The data collected from the installation and stored either in the database or in .csv files constitute set type data used by the algorithms based on which the application is written in the analysis of the operation regimes of the installation.

The application is implemented on the computer.
- The way in which the application monitors the installation is checked and the remedies imposed by the situation are made.

2.18 Continuous patient monitoring

Methodology

Define the device template models for the sensors and controllers etc. that will be used. Defining certain parameters such properties, commands (Actions the device can take) and what kind of data it uses (telemetry).
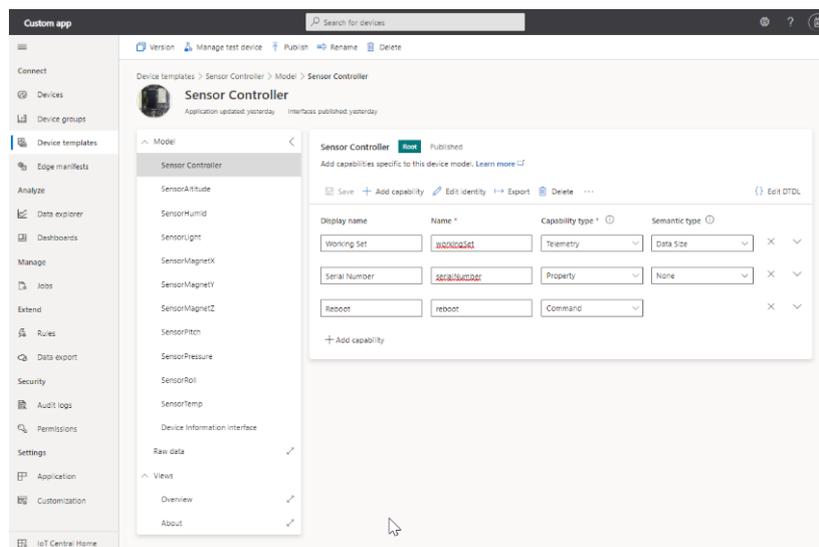
Figure 2.12. Device Template Model

When the device templates have been defined (see Figure 2.12.), the devices can be given a template and then connected to the IoT central Hub.

When the device has been added and starts generating data, there must be an analysis of this data. This is created using the IoT central hub data explorer, and building queries based on the incoming data. (see Fig. 2.13)
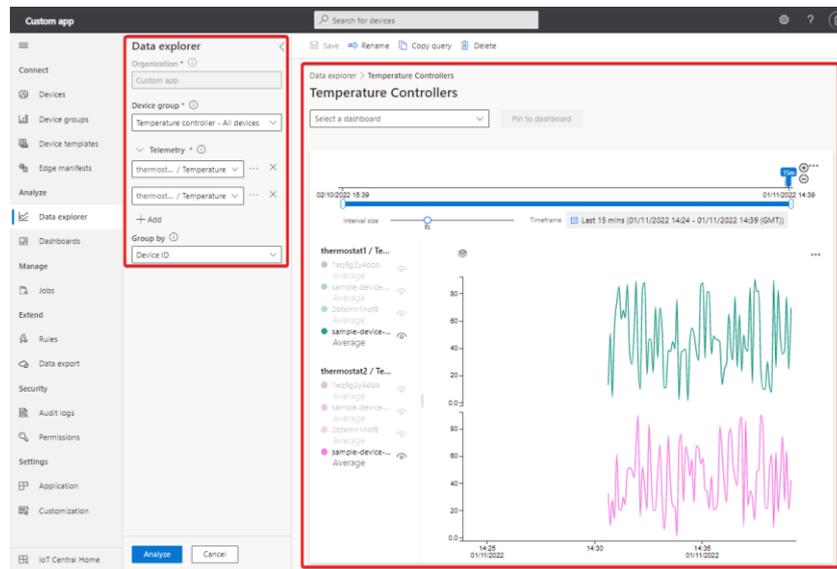
Figure 2.13. Reaction of the Device

Using the queries, rules must be defined so that the device reacts whenever certain thresholds are crossed. Either running one of the commands defined on the device, or a function on the back end, or both.

Value

Using the Internet of Things to monitor patient health, it's possible to extend patient care beyond the hospital walls, reduce re-admissions, and manage disease through in-patient or remote patient monitoring. Actions to improve patient health can be taken automatically whenever an issue arises, and medical staff can monitor, and receive immediate notice whenever changes happen with the patient.

System architecture

System architecture is presented in the Figure 2.14. below.
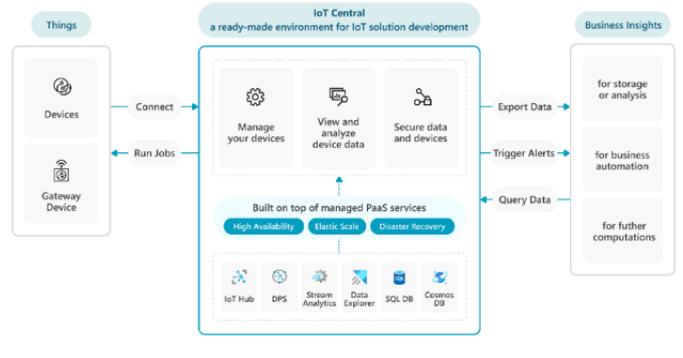
Figure 2.14. System Architecture

2.19 Creating a didactic application to help students learn a foreign language

Value

Learning a foreign language by a student meant knowing the meaning of each word in the language to be learned in relation to the word in the student's mother tongue and knowing how to write and pronounce this word correctly.

The sounds that make up a spoken word are recorded in writing through graphic signs that are generally called letters or more generally graphemes. In some words, the same grapheme is used for the same sound, regardless of the position of the sound within the word, or different graphemes or groups of graphemes are used for the same sound, depending on its position within the word or the sounds it belongs to.

In some foreign languages there are rules with many exceptions that establish for each sound what kind of graphemes are used, in others such as English there are no rules that establish for each sound what graphemes to use. The international phonemic alphabet was developed so that for each sound there is a graphic sign to be used for recording this sound in writing.

The letters of the international phonemic alphabet are not used in writing in communication between people, but only in textbooks and dictionaries to indicate how a word written in the alphabet of the respective language is pronounced.

Foreign students learning English face a difficult situation because there are very few rules for reading a word in English. The same sound is represented in writing by one, two or three groups of letters.

As a result, to learn an English word, one must know how it is written, pronounced and the meaning of this word in relation to the student's mother tongue.

Dictionaries offer all the necessary elements. In an English-Romanian dictionary, for example, it is specified how the word is written, the pronunciation of the word is specified by using symbols from the international phonemic alphabet and the meaning of the English word in the Romanian language.

Reading a phrase written in English requires knowing the pronunciation of each word that makes up the phrase. Both writing and pronunciation must be learned for each individual word. In French there are many rules for reading a written word.

It is much easier to read a word written in French or German because there are rules that say how to pronounce a group of letters, or a letter surrounded by other letters. The word can be read easily even if its meaning is not known in the reader's native language.

An application that converts a phrase written in the alphabet of the respective language into the graphic symbols of the international alphabet, helps the student to check and improve his pronunciation. Obviously, there are textbooks that provide information to facilitate learning a foreign language.

But a phrase in which unknown words appear is difficult to read and requires consulting a dictionary and extracting from it the data about the pronunciation of the words, which involves a job that requires effort and time.

This application writes the pronunciation of each word in a phrase in the international phonemic alphabet and offers the pronunciation of each word in the phrase by linking the words in the phrase. In this way, the student can learn the pronunciation of each word in the phrase by reading its transcription in the international phonetic alphabet and listening to the pronunciation of each word in the phrase.

This leads to more efficient learning of a foreign language because it is important for the student to listen to phrases pronounced in the language he is learning.

Application architecture

The structure of the application is as follows:
A text editor that allows writing the phrase that the student is studying from the point of view of pronunciation and even the meaning of its words.

A relational database in which words from the respective language are included together with the transcription of each word in the international phonetic alphabet. At the same time, files with the sounds for each graphic sign in the international alphabet must be included in the database. The actual application that converts each word of the phrase written in the letters of the alphabet of the respective language into the international phonetic alphabet.

The application can be written in C#, Python or Java.

The phrase is written in the text editor and when a button is pressed, each word in the written phrase is transcribed using the international phonetic alphabet. In this way, the student can read the phrase correctly if he knows how to pronounce each graphic symbol that makes up the international phonetic alphabet. At the same time, the application emits a sequence of sounds that are the sound correspondent of the phrase written in the alphabet of the respective language.

By confronting the written symbols with the sounds, the student will learn the foreign language more effectively. The text editor has all the facilities offered by modern editors, i.e., you can copy a phrase, you can paste a phrase copied from another application. At the same time, the expression written in the international phonetic language respects the punctuation of the phrase written in the alphabet of the respective language.

Application phases

- Creating the dictionary in the database

The database is relational and contains the word written in the alphabet of the respective language. Also, each word is written in the international phonetic alphabet. Each graphic sign of the international phonetic alphabet is related to a file that emits the corresponding sound of this graphic sign.

- Creation of files for the pronunciation of each sound that has a symbol in the international phonetic alphabet.

- Writing the actual application

The application that can be written in one of the modern languages Java, C#, Python contains a graphic interface in which the user accesses a text editor in which the text to be spoken is written. After writing the text, by pressing a button on the graphic interface, the entire or partial phrase can be read.

- Application testing and application adjustment.

It is the stage in which it is checked whether the application works correctly. Any problems are fixed.

2.20 Create test environments

Methodology

Create a virtual private cloud (VPC) on AWS, navigate to the VPC dashboard in the AWS Management Console. Here you can create a VPC and subnets that match the network architecture of your existing infrastructure.

Launch EC2 instances: Launch Amazon Elastic Compute Cloud (EC2) instances that correspond to your existing servers. You can use Amazon Machine Images (AMIs) to quickly replicate your existing servers in the cloud.

Set up load balancers: Use Elastic Load Balancing (ELB) to distribute traffic to your EC2 instances. You can set up a load balancer with the same rules as your existing infrastructure.

Configure databases: If you have databases that need to be replicated, set up Amazon Relational Database Service (RDS) instances that match your existing database servers. You can also use Amazon Aurora to create a highly available, scalable database.

Set up security groups and ACLs: Configure security groups and network ACLs to match your existing infrastructure.

Test your infrastructure: Once you have replicated your infrastructure on AWS, thoroughly test it to ensure that it works as expected.

Migrate your data: Once you are satisfied that your replicated infrastructure works correctly, migrate your data to the new infrastructure.

Value

Having a replicated test environment based on your current production infrastructure lets you create new functions and add new features to your existing applications and test them in an environment as similar as possible to the actual production configuration as possible without the risk of inadvertently changing the running systems during the testing.

System architecture

The system architecture would be whatever the original infrastructure looks like, e.g., EC2 stacks, Load balancers, Private clouds, and databases.

2.21 Data backups and archiving

Methodology

Cloud backup works by replicating company data on cloud-based servers. This can be done in two ways:
- Continuous replication: With continuous replication, the cloud provider copies your company's data to its servers as it changes. This is the most common type of cloud backup and is used by companies that need to always keep an up-to-date copy of their data.
- Scheduled replication: With scheduled replication, the cloud provider copies your company's data on a set schedule. This is often used by companies that don't need to always keep an up-to-date copy of their data.

Once your company's data has been replicated to the cloud, it can be accessed from anywhere in the world using an internet connection.

Value

In general, the value of creating data backups and archives can be achieved immediately as it provides assurance that your important data is secured and can be easily recovered in the event of a disaster or system failure. It also allows for better management of data and can aid in decision-making processes.

System Architecture

The architecture for data backups and archiving typically includes several components.

First, there is the storage hardware, such as disk arrays or tape libraries, that holds the backed-up or archived data.

Next, there is the backup and archive software that manages the creation, tracking, and restoration of backups and archives.

There is also the backup and archive server, which can be a dedicated physical or virtual machine or a set of machines that coordinate backup and archive processes across an organization's network.

In addition to these core components, there may be supplementary components such as data deduplication technology to optimize backup storage, and encryption technology to ensure the confidentiality of valuable data.

Overall, the architecture of an organization's backup and archive system will depend on its specific needs, including the size of its data sets, the volume of backup and archiving operations, and the level of risk it is willing to tolerate in terms of data loss or corruption.

2.22 Data loss prevention cloud-based system

Methodology

Analysis of content

The product should analyse the context and find it. Simple form of email with plain text, but views within binary files become a bit more complicated process. For example, it is not uncommon for a DLP tool to read Excel tables embedded in a Word file that is compressed with a ZIP tool. The product should open a file, read the Word document, analyse it, find Excel data within the original document, read it, and analyse it. Many products on the market today support about 300 file types, embedded content, multiple languages, and retrieving plain text from unidentified file types.

Content analysis techniques

After access to the content there are seven major analysis techniques used to find the violation of the rules, each with its own advantages and disadvantages.

Rules and regular expressions
This is the most common analysis available in DLP solutions. Analyses content for specific rules - such as a 16-digit number that meets credit check counts or other textual analysis. It is primarily used as a filter or for detecting easily recognizable parts of structured data such as credit card numbers, OIBs, and the like. The advantage is that rules can be quickly and easily configured. Most products come with pre-defined rules. The technology is easy to understand and easy to install into a variety of products. The

disadvantage of this technique is the tendency to high rates of false positive results. It offers very little protection for unstructured content such as sensitive intellectual property.

Fingerprinting database
This technique takes or "dumps" a database or stores live data (via ODBC connection) from the database and asks for exact matches. An example is generating rules that only require credit card numbers in their user base, ignoring their own employees who use their data for the internet store. Primarily applied for structured data from databases. The advantage is a very low number of false positive results. It allows the protection of user sensitive data and ignores other, similar data that are their own employee data within the organization. The disadvantage is that large databases affect product performance.

Accurate file matching
This technique takes the file clip ("hash") and monitors all files that match exactly. It can be considered a contextual analysis technique because the content of the file itself is not analysed. It is best to apply for multimedia files and other files where textual analysis is not necessarily possible. The advantage is that it works with any type of file and provides a low number of false positive results. The disadvantage is that it is easy to bypass match. Invalid for edited content, such as standard office documents and edited media files, as this changes the hash value.

Partial document matching
This technique requires full or partial match for the protected content. This way, you can create a policy for sensitive document protection, and the DLP solution will look for either the entire text of the document or extract from several sentences. It is best used to protect sensitive documents or similar content and source code. Unstructured content that is known as susceptible. Benefits are the ability to protect unstructured data. Mostly low false positive results. It does not rely on full compatibility of large documents and violations of rules, as it may also detect partial parallels. The disadvantages are performance constraints on a large overall volume of content that can be protected. Common phrases in a protected document can cause false positive results. It is necessary to know exactly which documents are to be protected and can be avoided.

Statistical analysis
Use machine learning and other statistical techniques to analyse content and find policy violations in content that resembles a protected content. This category includes a wide range of statistical techniques that differ greatly in implementation and efficiency. Some techniques are very similar to those used to block unwanted mail. It is best to apply for unstructured content were deterministic technique, such as partial duplication of documents, will be ineffective. Examples are documents that are impractical to partially match documents due to large documents or large volumes. The advantage is that it can work with unclear content where you may not be able to define exact match papers. The disadvantage is the tendency to false positive and false negative results.

Keywords and lexicons
This technique uses a combination of dictionaries, rules, and other analysis to protect vague content that resembles unwanted communication or information exchange. It is best to apply to totally

unstructured communications or exchanges that cannot easily be categorized by matching well-known documents, databases, or other registered sources. The advantage is that it can be used where all corporate policy or content cannot be described with specific examples. Conceptual analysis can find a breach of badly defined rules that other techniques can not follow. A disadvantage may be that in some cases these are rules that a user cannot customize, rather than being defined by a DLP manufacturer. Due to the nature of rules that can cover a wide area, the technique is very prone to false positive and false negative results.

Categories

Pre-built categories with rules and dictionaries for common types of sensitive data, such as credit card numbers and PCI protection. Best to apply to areas that fit into a particular category. Typically, easy to describe content related to privacy, policies, or guidelines specific to each business. The advantage is easier configuration and significant timesaving for policy generation. Categorization rules can be the basis for more advanced, specific business policies. For many organizations, categories can meet a large percentage of their data protection needs. A disadvantage may be that sometimes it only suits the rules and content that are easily categorized. The above-mentioned technique is the basis for most DLP products on the market. All products do not cover all techniques, and significant differences can occur between implementations.

System Architecture

Protecting Data in-Motion, at-Rest, and in-Use
DLP's goal is to protect content throughout the entire life cycle. As far as the DLP is concerned, this includes three main aspects:

• Data at-Rest - includes storage scans and other content repositories to determine where sensitive content is located. This is called content discovery. An example is the use of a DLP product for server scanning and document recognition with credit card numbers. If the server is not authorized for such data, the file may be removed, or a warning may be sent to the file owner.

• Data in-Motion - network traffic monitoring (passively or via a proxy server) to identify content that is sent over specific communication channels. An example is email monitoring and web traffic for sensitive source code snippets. Tools such traffic can often be blocked based on central policy, depending on the type of traffic.

• Data in-Use - usually resolved by workstation solutions that track data while the user is using it. An example is a detection when trying to transfer a sensitive document to USB and block copying (as opposed to blocking the full use of USB media). Tools that are used can also detect things like copying and pasting or using sensitive data in an unauthorized app.

Data in-Motion

For monitoring data on the move, it is possible to use various techniques among which the most used are listed below.

Network traffic monitoring

Most DLP solutions have the ability to passively monitor the network and perform packet capture in full and analyse content in real time.

Integration with Email

The next major component is email integration. Because e-mail works on the system's storage and forwarding, many options can be obtained, including quarantine, encryption, and filtering integration.

Filter, block, and integrate with a proxy server.

After implementing the DLP solution, and a certain period of observation of the use of sensitive data coming out of the organization over the Internet, they will want to start blocking the traffic. Blocking can be a complex venture, especially since it is trying to allow the right traffic and block only sensitive traffic and decide by analysing the content in real-time.

Internal networks

Although technically capable of monitoring internal networks, DLP is rarely used on internal traffic. Gateway solutions provide convenient monitoring points, but internal control significantly increases costs and complexity, affects performance and policy management, and increases false positive results.

Data at-Rest

While capturing online data leaks is productive, it's just a small part of the problem. Many users think it is equally important, if not more important, to find out where all of these data is stored. This is called content discovery. The biggest advantage of detecting content with the DLP tool is the ability to apply a policy to data no matter where they are stored, shared, or used.

Content discovery includes three areas:

1.   Workstation (Endpoint Discovery) - scanning content on workstations and laptops.

2.   Storage Discovery: Mass storage scanning, including file servers.

3.   Server Discovery - scan data stored on mail servers, document management systems, and databases.

Discovery Techniques

There are the following basic content discovery techniques:

- Remote scanning - a connection is made to a server or device that uses a file or application sharing protocol, and scanning is performed remotely. It is essentially mounting a remote disk and scanning from a server that takes rules with and sends results to a central server rule.

- Agent scanning - the agent is installed on the system (server) scanned and scanned locally. Agents are platform-specific and use local resources but may potentially have significantly better performance than remote scanning, especially for larger amounts of data. For endpoints, this should be the same feature of the same agent used to enforce data control in use.

Performing rules on idle data

When a policy violation is detected, the DLP tool can take various actions:

- Warning / Report - Creating an incident on the central management server.
- User Warning: Notifying the user that he might conflict with the rules.
- Quarantine / Notification - move the file to the central server to manage and leave instructions on how to request the return of the file.
- Quarantine- move the file to the quarantine location, with the ability to leave a plain text file that describes how to retrieve the file.

The combination of different implementation architecture, detection techniques, and execution options creates a powerful combination of data protection in idle mode and supports synchronization initiatives.

Data in-Use

DLP usually starts online because it is the most cost-effective way to get the widest coverage. Network monitoring is unobtrusive (unless SSL is required). Filtering is more difficult but relatively easy on the Web (especially for email) and covers all systems connected to the network. Data protection on the network does not protect data when someone goes out of business with a laptop and cannot prevent people from copying data to a portable storage such as a USB medium. To increase the scope of data outbound protection, products should be extended not only to stored data, but also to workstations where data is used. Adding a DLP agent to a workstation provides the ability to detect stored content, but also protects data that is actively used, and potentially protecting systems that are no longer on the network.

Usage Cases
Endpoint DLP is developing to support several critical cases of use:
- Perform offline rules outside the managed network or modify the rules for use outside the managed network.
- Restrict sensitive content to portable storage devices, including USB media, CD / DVD drives, home storage.
- Limit copy and paste of sensitive content.
- Restrict applications that are allowed to use sensitive content.
- Integration with Enterprise Digital Rights Management for automatic application access control access control.
- Reviewing the Use of Sensitive Content for Conformity Reporting.

Additional endpoint options
The following features are most desirable when implementing DLP at the endpoint:
- Agents and policy (s) should be managed centrally by the same DLP management server that controls motion data and idle data.
- Creating and managing rules should be fully integrated with other DLP rules on one interface.
- Incidents must be reported and managed by the central control server.
- The agent should use the same methods and content analysis rules as network servers / devices.
- Rules should be adapted based on the place where the end point (on the network or offline) is

located.

- Distributing agents should be integrated with the existing software installation tool in the organization.

Workstation limitations

Workstation storage performance and capacity limitations will limit supported types of content analysis, as well as the number and type of policies that are implemented locally. For some businesses, this may not be important, depending on the types of policies to be implemented, but in many cases, workstations impose certain restrictions on policies related to data in use.

Central management, policy management and workflow

All current DLP solutions include a central server for policy creation and management, event flow and reporting, and administration of content detection points.

User interface

Unlike other security tools, DLP tools often use non-technical staff ranging from HR to executive management. As such, the user interface should consider this mixture of technical and non-technical staff and must be easily customizable to meet the needs of a particular user group. The DLP user interface should include the following elements:

- The dashboard - a good dashboard will have elements and defaults that the user can choose for technical and non-technical users. Individual elements may only be accessible to authorized users or groups, usually groups stored in the company's directory. The dashboard should focus on information valuable to a particular user, not just a generic view. Obvious elements include the number and distribution of breaches based on weight and channel and other relevant information, in order to sum up the overall risk for the enterprise.
- The incident management screen - the incident management panel is the most important component of the user interface. This is a tool that users use to track and manage policy violations. The order should be concise, adaptable, and easy to review.
- An overview of a particular incident - When it comes to an incident, the screen should clearly and concisely state the reason for the violation, user, criticality, severity, associated incidents and any other information needed to resolve the incident.
- System Administration - Standard System Status and Administration Interface, including administration of users and groups.
- Hierarchical Administration - Status and Administration for Remote Components of DLP Solutions, such as execution points, remote offices, and endpoints, including a comparison of rules that are active there.
- Reporting: a mix of customizable pre-made reports and tools to facilitate ad-hoc reporting.
- Policy Management: This is also one of the most important elements of the central management server. This includes policy making and management.

2.23 Data management system about a company's employees

Value

Data about a company's employees is important to the company. That is why safe data storage is necessary and important. Small and medium-sized companies often fail to access specialized software with many facilities due to their high price.

Creating a simple application with a friendly and suggestive graphic interface solves the problem. This application can be written in C# or Python and the data it manages can be stored in a database. For data security, the application allows data access with password and user account.

Storing data in a database has the advantage that when changing the data format, the information is not lost. The application stores the data in a local database that can be created on a local server that can be accessed by several people who deal with employee problems.

The database can be easily updated by the staff of the HR office and can be accessed by any person from the company's management who needs data about subordinate personnel. The advantages of the application compared to the management of data on separate documents contained in excel or word files is that:

- all data are kept in one place (on a local server or in the network); in the classic system, the files can be distributed on several computers and some people cannot have easy access to these data files; the application allows access to the respective data much easier because the information is stored in one place.
- data access can be prioritized.
- some employees can have access to the data only to read the data and others to read and modify them; in addition, access to some sensitive data can be limited using passwords and user accounts.
- the application can generate different reports and statistics on demand in a very short time
- in the old management system, to create a report or a statistic, several files must be checked, the information must be extracted manually from these files.

Application architecture

The application consists of a graphic interface and a database. The graphic interface contains an entry page where the user enters the access password and the user account.

Depending on the access level, the user can have access to the pages where data can be entered or can have access to the pages where data can be read. The application has pages where tables and statistics can be generated.

The stages of using the application are the following:

Creating the database

To create the database, you must know the data stored in this database. Based on this data, the tables in the database and the correlations between the tables are created and the format of the data that will be stored in the tables is established.

The actual creation of the application

In this phase, the application code is written in a language such as C# or another language.
The application consists in creating the graphic interface and the software that makes the connection between the graphic interface and the database.

This software allows querying the database to display data in the graphic interface or write data to the database.

In order for the application to be user-friendly, a windows form is used that is populated with different elements such as: menus, access buttons, drop-down lists, image display elements, etc.
The graphic interface has several pages that can be easily navigated between. Each page has a certain specificity. Thus, there are pages through which the user can enter data into the database, pages where the user can display information from the database or pages where the user can generate reports or statistics (Fig. 2.15. & 2.16.).
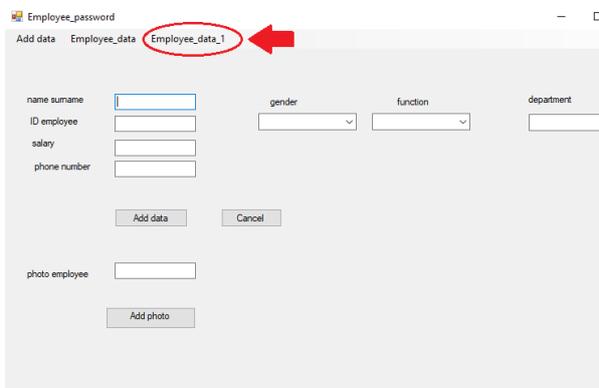
Figure 2.15. Wisdom form One
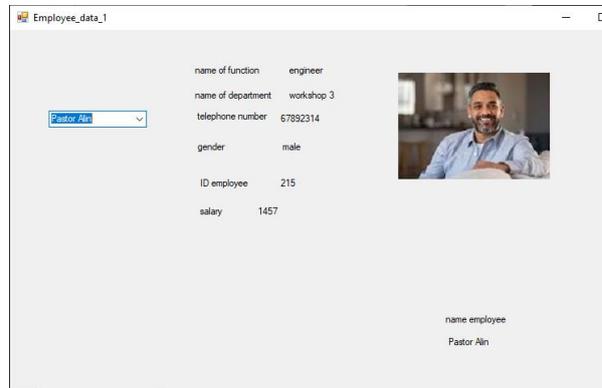
Figure 2.16. Wisdom form Two

Application testing

After the application has been written, it is tested to see if the information entered in the database through the application can be found in the tables in the database and if the information was recorded correctly in the tables (it was not truncated).

During this stage, it is checked whether the reports and statistics to be generated are correct.

Database and interface adjustments

During the testing period, the user can request that the application have new facilities or request the modification of some elements of the graphic interface for an easier use of it. The color of the graphical interface can also be modified. In this case, some codes in the written program are modified to correspond to the user's needs. You can also add, modify, or delete certain tables from the database.

2.24 Digital asset certification using distributed ledger/blockchain.

Methodology

The main features of blockchain are transparency and decentralization, which today's systems cannot boast. Digital identity combined with blockchain technology will enable people to perform tasks that are faster, simpler, and safer, including proof of identity, facts, status and data. Incredibly, the fact is that searching for new employees, checking candidate data, and job application itself could be a process that would take just a couple of mouse clicks on the computer, with the utmost certainty of the data being obtained.

But blockchain is just offering it. By placing all the information on our identity on it, with cryptography that makes the whole thing safe and transparent and always accessible through the internet, we spend all the time spent on proving identity, data, facts, and state of affairs on the most important things. Imagine that we can also enclose 3 cryptographic keys with the application for business, so that the employer can easily check with the absolute certainty that we have completed the college we have stated in his CV, whether we are unhappy and whether we are at all a person who claims to be. This process would take about a few minutes, while the same process lasts for several days, if not weeks, as the data verification is done by writing queries in each of these systems from which data comes. Blockchain got its name by the way it stores the transaction data that is happening. It stores them in blocks that together form a chain as seen on Figure 2.17.
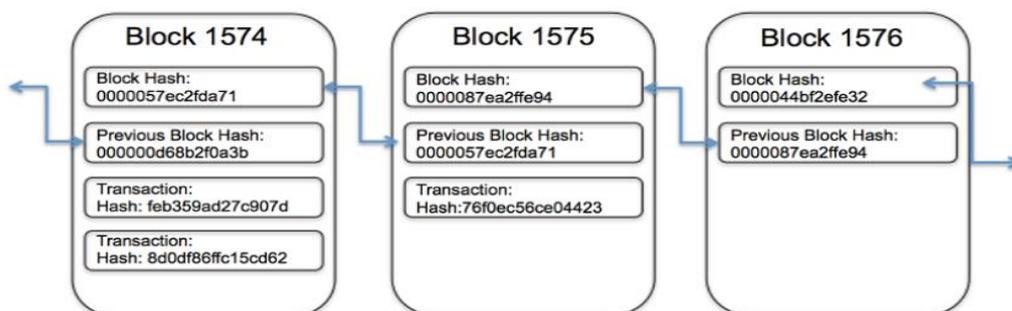
Figure 2.17. Transaction stored in blocks that connect each other to a chain

By increasing the number of transactions being made, the size of the chain in which they arise increases. The blocks record the sequence and time of the transactions that are then recorded in the network chain according to certain security rules agreed between the participants. Each block contains a hash, i.e., a digital imprint or unique identifier, then time-tagged valid transactions and hash of the previous

block. Hash of the previous block mathematically links the blocks to the chain and prevents any change of data and information in the previous blocks or inserting new blocks between the existing ones. Thus, each of the following blocks increases the security of the entire chain and reduces the already small chance of manipulation and change of value or data in the chain.

There are several types of blockchains, in this paper we will mention the 2 most common types:

A public blockchain, such as Bitcoin blockchain (the first and most known crypto value based on this technology), is a large, distributed network that runs with the release of a native token. The public blockchain is visible and open to everyone to use at all levels. An open code maintained by the developer community.

Private blockchain is smaller in volume and usually does not run with token issuance. Membership in this type of blockchain is highly controlled and is often used by organizations that have confidential members or traded with confidential information.

All types of blockchains use cryptography to enable each participant to use the network in a safe manner, and most importantly, without the need for a central authority that applies the rules. Because of this, blockchain is considered revolutionary because it is the first way to gain confidence in sending and writing digital data.

The most commonly used cryptographic algorithm is the SHA-256 algorithm. As an input, any quantity and type of data (text, document, etc.) can be used as an input, and the algorithm receives a unique 32-digit fixed size data.

An example, text "Goran", passing through the SHA-256 algorithm gives the result dbe08c149b95e2b97bfcfc4b593652adbf8586c6759bdff47b533cb4451287fb. The word "Goran" will always result in an identical hash value. Adding any character or letter at the input changes the complete hash appearance, but of course, the mentioned 32-character length always remains the same. An example, word Gordan, gives the result 48fa1be7c33664e5a0c61a006d21592cf20272aab7228b09add728aa0f11ffc7. In addition to the mentioned blocks and chains that are interconnected, there is another very important segment, which is a network. The network consists of nodes and full nodes. The device that connects and uses a blockchain network becomes a node, but if this device becomes a complete node, it must retrieve a complete record of all transactions from the very beginning of the creation of that chain and adhere to the security rules that define the chain. A complete node can lead anyone and anywhere, only the computer and the Internet are needed. But that's not so simple as it sounds.

Value

Many people mix Bitcoin and Blockchain concepts or misuse them. Those are two different things. Blockchain technology was introduced in 2008, but it was only one year later launched in the form of cryptocurrency Bitcoin. Bitcoin is therefore a cryptocurrency that has its blockchain. This blockchain is

a protocol that enables secure transmission and monitoring of cryptocurrency Bitcoin, all from the emergence of its first block (genesis block) and the first transaction. Bitcoin is designed solely as a criterion of vision that one day completely replaces fiat (paper) money and crushes the money transfer barriers that are present today. Through the years that passed, the community found that blockchain is more powerful than it originally thought, so if Bitcoin as a cryptocurrency does not live globally in everyday life, it will leave behind a revolutionary invention that potentially can change the technological world we are currently familiar with. Blockchain through its mechanism of consensus eliminates the central authorities that we know today, and which are based on today's technology.

System Architecture

MultiChain was selected as the platform for creating an application concept for entering, issuing and verifying educational certificates. MultiChain is an open-source platform that allows you to create or block your own blockchain and manage its capabilities. It is optimized for creating licensed chains (permissioned blockchains). MultiChain is compatible with Linux, Windows and Mac operating systems. Currently optimized for Linux operating systems, here referred as Linux's 64-bit Ubuntu 18.04.1 operating system with virtual machine on a single physical server using Oracle VM VitrualBox.

Virtualization of operating systems enables us to have multiple operating systems on one server, workstation, or computer, and we use them at the same time. All operating systems share computer resources. The number of virtual machines is unlimited, i.e., it depends on the amount of disk space and the memory of the computer hosted on.

In the Oracle VM VirtualBox workstation there is a File, Machine, Help, icons for the most important activities on virtual machines (New, Settings, Discard, Show), and below them there is a window where show installed virtual machines including Ubuntu's virtual machine, with assigned 100GB of disk space and 3GB of work memory. Before installing the operating system, it was necessary to create a virtual disk of the specified size that the machine would use. The complete work console and virtual machine installation is very intuitive and simple. VirtualBox offers the ability of detachable virtual machine start-ups, enabling the entire process to run without open windows and a graphical user interface (GUI).

2.25 Digital identity

Methodology

Civic is a company that develops an identification system that allows users to selectively share identifying information with companies. Their platform has a mobile application where users enter their personal information and then store them in encrypted format. The company's goal is to establish partnerships with state governments and banks, i.e., all those who can validate user identity data, and then leave a verification stamp in blockchain. The system encrypts the hash of all verified data and stores it in the blockchain and deletes all personal information of the user from their own servers.

As the company has written in its White Paper, the Civic Ecosystem is designed to encourage the participation of trusted authentication bodies called "validators". "Validators" can be the aforementioned state governments, banks, various financial institutions, etc. As Civic currently validates user identity information through its application, "validators" have the ability to verify the identity of an individual or a company that is "user" of the application. They then affix the certificate and place it in a blockchain in the form of a record known as attestation. This "verification" is actually a user's hash of personal information. Parties known as 'Service Providers' wanting to verify the same user identity data should no longer be able to independently verify that information but rather use the verified information valid for those validators of that information.

The goal is to remain a "ruler" of your identity and to have full control over personal information so that it must give prior consent to each transaction of information about its identity between the validator and the service provider. By smart deals, validators have the ability to sell their approvals to service providers, but also to service providers to see at what prices different validators offer their approvals. Each validator can declare the price it is willing to sell personal user information. After the user, validator and provider confirms the transactions through the smart deal system, the service provider pays the validator the required amount in the form of CVC tokens. After that, a clever contract will allocate CVC tokens and the user will get their share of the participation. The user can use their tokens to purchase products and services on the Civic platform.

As we mentioned, the user is the one who is responsible for their data and stores them on some of their personal devices using the Civic app, and it is also recommended to back up a personal account on the cloud system. Since user identity data is not centralized, i.e., not on Civic servers, there is no possibility of massive identity theft since the data of each user is actually on their devices and that data will be stolen, so it is necessary to break it into each device separately. This information largely helps to suppress the black market for personal information. E.g., Black credit card market is quite widespread because transactions can only be done by knowing these data without the knowledge of the user. If a credit card number needs to go through the blockchain mechanism of proofing where the user's consent for each transaction should be, then the black market of such data slowly loses its meaning and value. (Fig. 2.18.)
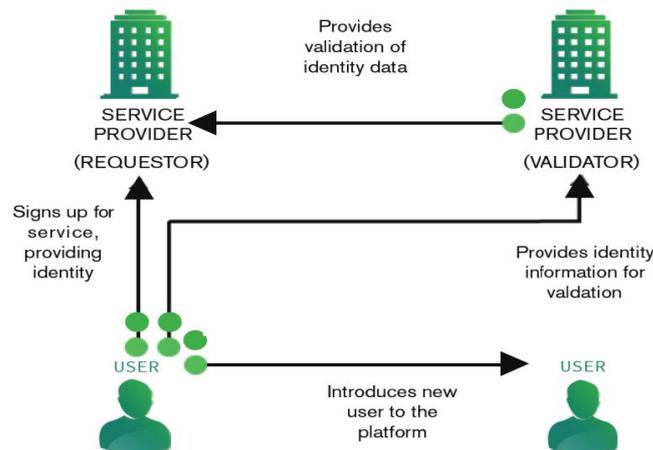
Figure 2.18. Civic concept

HYPR is a young company, founded in 2014. Their business model is based on merging biometric identification methods and blockchain technology. Biometric identification can replace a classic identification with a username and password, is faster and safer. Biometrics can recognize different parts of the human body such as palm geometry, fingerprint, eye iris, scent, face and many long physiological elements unique to the individual. Biometrics is a very good way of verifying an individual's identity because it is very difficult or impossible to forge it. HYPR therefore offers a password-free authentication platform with biometric encryption. The company does not deal with the development and production of identification devices but develops a distributed security system. As mentioned above, every digital data can be used to insert some of the cryptographic algorithms and get their hash.

This hash can be used to validate these digital data without the need for a validator to have a copy of that data. E.g., we read our finger on a fingerprint reader on a mobile phone, and a company that has access to the hash of our fingerprint in digital form can confirm our identity, without the possibility of being false as we do. Digital print is just a part of the offering that is offered. HYPR supports many types of biometric data, from simple authentication algorithms to face and speech algorithms to much more complex algorithms such as keyboard typing, rhythm writing on mobile devices, or the way we walk. With blockchain and data decentralization, authentication becomes much faster and simpler. Each user is responsible for their biometric data, such as on his mobile device. This avoids massive data theft, while individual theft may still be possible if the user is not careful enough to protect their data and devices. Such a system based on blockchain technology is resistant to Denial of Service (Denial of Service), which is a better centralized system. DoS attacks are attacks on some computer service in order to disable its use.

In this case, instead of attacking a single server used to authenticate data, DoS attackers should identify and attack all blockchain nodes in that system. The company emphasizes that protecting against DoS attacks is equally important and the interoperability of business processes. There is currently no possibility of authentication between two different corporate entities such as a bank and an insurance company. Each company has a different identity database, and they are not interoperable. Using

2021-1-SI01-KA220-VET-000034641

blockchain technology, we can have an interoperable distributed mainstream identity book between multiple entities without the need for complex and expensive infrastructure. Thus, the insurance company can prove our identity to the bank through biometric data.

The problem of proof of identity does not only appear in people. It may also be present in various products such as medicines, luxury products, diamonds, electronics, music, software, etc. These products are often counterfeit, causing damage to manufacturers in billions of dollars. People behind the Blockverify project want to reduce the number of counterfeit products on the market by preventing duplicate appearances. Different companies from different industries can register and track their products using Blockverify and blockchain technology. The company believes that improvement in counterfeit products can only be achieved by using decentralized, scalable, and safe solution attacks. Blockverify has its own private blockchain, but it also uses Bitcoin's blockchain to record important changes in its chain. Their chain is highly scalable and transparent so that each manufactured product can enter into it as an asset. After that, each of these assets will be added to the blockchain and assigned a unique hash. Anyone with that hash can access blockchain and check whether the product is valid or not. The primary goal of the company is to address the problem of counterfeit medicines, which is first on the scale of counterfeit products, but also one of the more dangerous counterfeit products because it directly affects people's health and causes millions of deaths per year. Another problem that a company wants to solve is the problem of verification of ownership. Thanks to blockchain technology, ownership changes can be easily recorded permanently. By this mode, individuals are prevented from making duplicate records and unauthorized changes.

Value

Blockchain is one of the disruptive technologies that are often called technology that will change the world and enable a new revolution. Blockchain represents a decentralized database that is publicly available to everyone via the Internet. Take for example databases and registers owned by the state and its institutions such as ministries, banks, mobile operators, etc., and all listed registers and data are published publicly, by placing them in the blockchain. It allows us to access all data concerning our own, with authorization, with access to the Internet. Likewise, we can present the same information to the other side as soon as we need to prove their identity, valid information, or some information.

System Architecture

If we talk about a classical identity of an individual and mention a personal identity card, a birth certificate, a patronage, a driving license, or a faculty diploma, then in terms of the digital identity of an individual we can talk about an e-personal ID card, e-birth certificate, e-homepages, e-driver's license, or e-diplomas. The e-name tag is electronic, meaning that these documents also have a digital component. This digital component can be, for example, an electronic data carrier (chip) that stores certain data or certificates that are readily uploaded to the computer by a reader if needed. The data being displayed are centralized and are guaranteed by and responsible to the institution issuing them and where the data is stored. (Fig. 2.19.)

59   2021-1-SI01-KA220-VET-000034641

Digital identity does not necessarily have to be a physical document or document. It also includes our email addresses as well as various user accounts and profiles on the internet, such as an eCitizen account, Facebook profile, email etc.
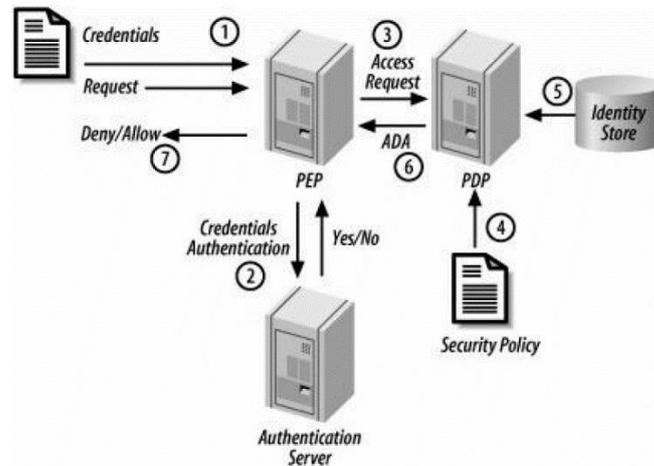

Figure 2.19. Authentication process

2.26 Digital twinning

Methodology

First, you must define the scope and objectives. The first step in creating a building automation digital twin is to define the scope and objectives of the project. This may include identifying the building systems to be modelled, the data sources to be integrated, and the performance metrics to be monitored.

Building automation systems generate a vast amount of data, which can be used to create a digital twin model. The data sources may include sensors, actuators, controllers, and other building automation devices. It is important to identify these data sources and establish a data collection and management strategy.

Once the data sources are identified, the next step is to develop the digital twin model. This involves creating a detailed representation of the building systems and their interactions, incorporating data from the identified data sources, and defining the performance metrics to be monitored.

Azure provides a range of services that can be used to build and deploy digital twins, including Azure IoT Hub, Azure Digital Twins, and Azure Stream Analytics. These services can be used to integrate data from various sources, process and analyse the data, and visualize the results.

Once the digital twin is developed, it can be deployed to the target environment and tested. This involves verifying that the model accurately reflects the behavior of the building systems, and that the performance metrics are being monitored correctly.

As the building automation digital twin is used, it may be necessary to optimize and refine the model based on feedback and new data. This involves analysing the data generated by the model, identifying areas for improvement, and making updates to the model, as necessary.

Value

Implementing a well-modelled digital twin will bring a lot of insight into how the current system is working in terms of efficacy. Depending on the sensors used, you can model the air quality in a given area at a given time, the temperature in a room, or the luminosity in a room over the course of the day. With this data you can manage these parameters and make sure they are at the optimal levels throughout the day, by increasing our decreasing the output of the various systems. This makes sure that no energy is wasted and creates a more pleasant environment.

System architecture

Physical Assets: HVAC systems, lighting, security systems, elevators, and other building systems.

Azure Digital Twins: Virtual representations of physical assets that can be used for simulation, monitoring, and analysis. In this system architecture, Azure Digital Twins is used to create digital twins of the physical assets.

Azure IoT Hub: Collects the data from sensors and other devices in the physical environment and send it to the digital twins for analysis.

Azure Stream Analytics: Used to process the real-time data coming from the physical assets and generate alerts or triggers based on predefined rules.

Azure Functions: Used to run custom code in response to triggers generated by Azure Stream Analytics. For example, if the temperature in a room exceeds a certain threshold, an Azure Function could be triggered to turn on the air conditioning.

Azure Time Series Insights: Used to store and analyse the data generated by the digital twins and the physical assets. This allows you to gain insights into the performance of the building systems and identify areas for improvement.

2.27 Disaster prevention platform

Methodology

Sensor data is generated **when a device detects and responds to some type of input from the physical environment**. Often coming together in a network, sensors generate mass quantities of sensor data that may or may not be immediately useful for decision-makers.

The sensor data and test data are collected **using the OPC client**, e.g., solution temperature, pH, ORP, zinc powder dosage, flow rate of feeding solution, etc. These real-time values of process variables are important for process monitoring and setting of manipulated variables.

Observability is the ability to assess an internal system's state based on the data it produces. An observability solution analyses output data, provides an assessment of the system's health, and offers actionable insights for addressing problems across your IT infrastructure.

Observability wouldn't be possible without monitoring. Monitoring is the collection and analysis of data pulled from IT systems.

The pillars of observability are the different kinds of data that a monitoring tool must collect and analyse to provide sufficient observability of a monitored system. Metrics, logs, and distributed traces are commonly referred to as the pillars of observability. Azure Monitor adds "changes" to these pillars.

When a system is observable, a user can identify the root cause of a performance problem by looking at the data it produces without additional testing or coding. Azure Monitor achieves observability by correlating data from multiple pillars and aggregating data across the entire set of monitored resources. Azure Monitor provides a common set of tools to correlate and analyse the data from multiple Azure subscriptions and tenants, in addition to data hosted for other services.

The development of a disaster prevention platform follows a systematic methodology that includes the following steps:
- Risk assessment: The first step is to identify potential hazards and assess the risks associated with them. This involves collecting data on historical disasters, as well as current and potential hazards in the area.
- Requirements gathering: The next step is to gather requirements from stakeholders, including government agencies, emergency responders, and citizens. The requirements should include user needs, system functionality, and technical requirements.
- System design: Based on the requirements gathered, the system design is created. This includes architecture design, data model design, and user interface design.
- System development: Once the design is finalized, the system is developed using appropriate software development methodologies, such as agile or waterfall.
- Testing and validation: After the development phase, the system is tested and validated to ensure that it meets the requirements specified. This includes functional testing, performance testing, and security testing.
- Deployment and implementation: Once testing is complete, the system is deployed and implemented in the production environment. This involves configuring the system, training users, and addressing any issues that arise during the deployment process.
- Monitoring and maintenance: The final step is to monitor and maintain the system to ensure that it continues to function correctly and meets user needs. This includes regular updates, bug fixes, and user support.

Overall, the methodology used in the development of a disaster prevention platform should be flexible, iterative, and responsive to changing user needs and requirements. The platform must also be continuously evaluated and updated to ensure its effectiveness in preventing or mitigating disasters.

Value

A disaster prevention platform has the potential to create significant value in multiple ways:

- Risk reduction: It helps in reducing the risk of disasters by identifying the potential hazards, analysing the vulnerabilities, and providing solutions to mitigate the risks.
- Early warning systems: It can provide early warning systems that alert the public and emergency responders to impending disasters, allowing them to take necessary actions to minimize damage and loss of life.
- Effective communication: It enables efficient communication between different stakeholders, such as emergency responders, government agencies, and citizens, helping them to work together and respond promptly to emergencies.
- Resource allocation: It helps in better allocation of resources by tracking the location of responders, supplies, and equipment necessary for disaster response.
- Preparedness and resilience: It enable people and organizations to prepare and become more resilient to disasters, thereby reducing the impact of such events.

Overall, a disaster prevention platform can save lives, reduce damage to property, and minimize the economic impacts of disasters. It is, therefore, a valuable investment for governments, organizations, and individuals to create and implement such a platform.

System Architecture

A disaster prevention platform architecture typically consists of the following layers:

- Data acquisition layer: This layer is responsible for collecting data from various sources such as sensors, satellites, social media, and other data sources that provide information about potential disaster events.
- Data storage and management layer: This layer stores and manages the collected data in a centralized repository. It includes data storage, data processing, and data analytics components.
- Analytics and decision support layer: This layer is responsible for using analytics and machine learning algorithms to analyse the data collected and provide insights into potential disaster events. It includes real-time analysis, predictive modelling, and decision-making capabilities.
- Alert and notification layer: This layer is responsible for generating alerts and notifications to the relevant stakeholders, such as emergency responders, government agencies, and citizens. It includes automated alerts and notifications, as well as customized notifications based on user preferences.
- Visualization and reporting layer: This layer is responsible for visualizing the data collected and analysed in the form of reports, dashboards, and maps. It includes visualization tools that help stakeholders to monitor and track disaster events and their developments.

- Integration layer: This layer is responsible for integrating the disaster prevention platform with various other systems, such as emergency response systems, communication systems, and logistics systems.

Overall, a disaster prevention platform architecture should be designed in a way that is scalable, flexible, and able to handle high volumes of data in real-time. The architecture should also be secure, with appropriate data access controls and authentication mechanisms to protect sensitive data.

2.28 Distribution of parcels in a geographical region with the help of autonomous drones

Value

The urban agglomeration as well as that on the roads and highways makes the distribution of postal parcels a difficult activity. In the not-too-distant future it may be possible to use drones in daily activities. It is possible that the first economic activities in which drones are used will be those in which parcels are transported.

This is less risky than with driverless cars because drones do not have to move along a well-defined lane like cars. Autonomous drones can become a reality in a short time and be used for the transport of small and medium-sized parcels.

Their movement between two points can be controlled by software that incorporates artificial intelligence and IoT technologies. Through the parcel transport system with autonomous drones, time is saved, land traffic is decongested. The coordinates of the destination points are given by the GPS system.

For parcel distribution, there are already automatic systems in big cities (in Romania) where parcels can be picked up. In the current system, an operator places the parcels in containers with drawers. The drawers can be accessed with a code. The owner of the parcel is informed that the parcel he is waiting for is located at the point located near his home. He is informed of the drawer code and the code with which he can open the drawer.

The person moves to the indicated place, which is very close to his home, and with the data he receives on his mobile phone, he can pick up the parcel. Currently, parcels are transported by local couriers by car and stored in containers.

If the drones will come directly to these containers and the operator is just waiting for them to unload the parcels and distribute them in the drawers, the parcel distribution activity will be easier and more efficient. Land traffic will be more fluid.

Application architecture

Although a good part of the infrastructure exists in the city of Cluj-Napoca in Romania, because there are smart containers with drawers. These containers are distributed over the entire surface of the municipality. For the proposed application, they must be adapted to receive the drone with the parcels. Now, the parcels are brought by the courier company's staff and put in the container drawers. Parcels are transported from a central warehouse in Cluj.

Within the application, each container has an address consisting of the container's GPS coordinates. The drone finds the container based on these coordinates. Directing the drone to land on the container is done by a beam of electromagnetic or ultrasonic radiation emitted locally.

The travel route from the central warehouse to each distribution point is established by the main application that is stored in the central parcel distribution point. So, the system has the following components:

The actual application that establishes the route of the drones. This application is located at the headquarters where the parcels are distributed.

A software component that is stored on the drone and that directs the drone according to the coordinates imposed by the central application, the structure of the buildings in the city and other considerations (e.g., it must avoid collision with other drones, etc.)

A component stored on each container that directs the drone to land and take off so that these operations can be done in complete safety. At the same time, the component stored on containers also performs other operations: it manages the activity of picking up parcels from the drawers and reports this to the head office. Check if all the drawers are full. If all the drawers are full, no more packages are accepted.

2.29 Document similarity detection and document information extraction system

Methodology

When we have exhausted all the modelling options, we look for a new direction. The method s presented here is an exploration of deep learning techniques to improve the information extraction results. To evaluate the techniques, a dataset with more than 25,000 documents has been compiled, anonymized, and published. It is already known that convolutions, graph convolutions, and self-attention can work together and exploit all the information present in a structured document. Here, we examine various approaches such as siamese networks, concepts of similarity, one-shot learning, and context/memory awareness as deep learning techniques. Information extraction tasks are not a new problem. Information extraction starts with a collection of texts, then transforms these into information that is more readily digested and analysed. It isolates relevant text fragments, extracts pertinent information from the fragments, and then pieces together the targeted information in a coherent

framework. The relevant collection of texts for this study is the content within business documents such as invoices, pro forma invoices, and debit notes. Example of an invoice and an extraction system together with its output. (Fig. 2.20.)
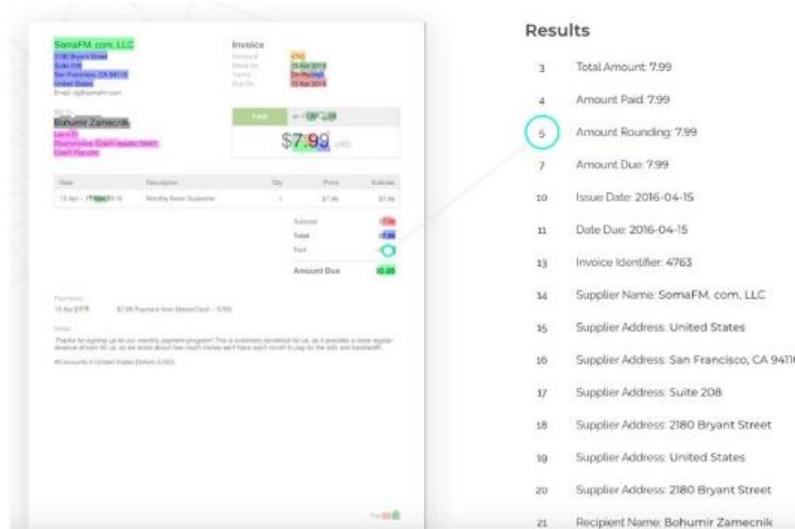

Figure 2.20. Invoice and an Extraction System together with its Output

The features of each word-box are either geometrical (to construct graphical CNN, reading order, and normalized (left, top, right, bottom) coordinates), textual (the count of all characters, numbers, the length of word, the count of first two and last two characters and trainable word features are one-hot encoded, deaccented, lowercase characters) or image (each word-box is cropped from the image).

The five inputs namely down sampled picture, feature of all word-boxes, 40 one-hot encoded characters for each word-box, neighbour-ids, position id by geometric ordering are concatenated to generate an embedding vector. The transformer approach is used to the position embedding vector. Image is stacked, max-pooled and morphologically dilated to generate 32 float features. Before attention, dense, or graph convolution layers are used, all the features are simply concatenated. Basic building block definition ends with each word-box embedded to a feature space of a specified dimension (being 640 unless said otherwise in a specific experiment). The following layer, for the "Simple data extraction model", is a sigmoidal layer with binary cross-entropy as the loss function. This is a standard setting since the output of this model is meant to solve a multi-class multi-label problem. The learning framework spans: 1) The system needs to keep a notion of already-known documents in a reasonably sized set. 2) When a "new" or "unknown" page is presented to the system, search for the most similar page (given any reasonable algorithm) from the known pages. 3) Allow the model to use all the information from both pages (and "learn from similarity") to make the prediction. Nearest neighbour definition: For one-shot learning to work on a new and unknown page (sometimes denoted reference), the system always needs to have a known (also denoted as similar or nearest) document with familiar annotations at its disposal. Embedding for nearest neighbour search is prepared by removing the latest layer and adding a simple pooling layer to the document classification model. This modified the model to output 4850 float features based only on image input. These features were then assigned to each

page as its embedding. These embeddings are held fixed during training and inference and computed only once in advance.

Baselines: We do not have any benchmark for comparison. Therefore, some models are prepared as baselines in the process. Simple data extraction model without any access to the nearest known page. Copy pastes, overlay the target classes from the nearest known page word-boxes. It provides counterpart to triplet loss and pairwise classification. Oracle always correctly predicts the nearest known page classes. Fully linear model without feature picture dana, provides a counterpart to the query and answer approach. Model architectures, every single one of the architectures is trained as a whole, no pre-training or transfer learning takes place, and every model is always implemented as a single computation graph in TensorFlow. Triplet Loss architecture, using Siamese networks canonically with triplet loss. Pairwise classification, using a trainable classifier pairwise over all combinations of word-box features from reference and nearest page. (Fig. 2.21.)
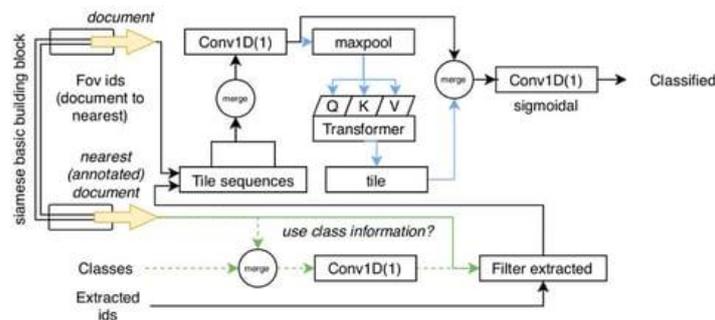

Figure 2.21. Query answer architecture

Query answer architecture (or "QA" for short), using the attention transformer as an answering machine to a question of which word-box class is the most similar.
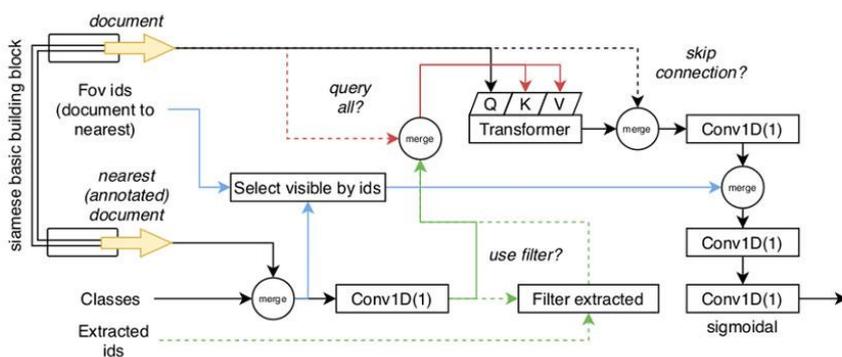

Figure 2.22. The filter mechanism

The filtering mechanism addresses only the annotated word-boxes from the nearest page. The tiling mechanism takes two sequences – first, the sequence of reference page word-boxes and secondly, the sequence of nearest page filtered word-boxes; and produces a bipartite matrix. The model selected in each experimental run was always the one that performed best on the validation set in terms of loss.

The basic building blocks present in every architecture were usually set to produce feature space of dimensionality 640. (Fig. 2.22.)

Value

We have verified that all possible parts of the architecture are needed in the training and prediction of the Query Answer model to achieve the highest score. What is the effect of the size of the datasets? By exploring the effect of the size of the training dataset and/or the search space for the nearest pages, we could ask if (and when) the model needs to be retrained and determine what a sample of a difficult-to-extract document looks like. How to improve the means of generalization? Currently, the method generalizes to unseen documents. In theory, we could desire a method to generalize to new classes of words, since this way the model needs to be retrained if a new class is desired to be detected and extracted. The model can fit into just one consumer-grade GPU and trains from scratch for at most four days using only one CPU process.

System Architecture

A classical heuristic way to generally improve a target metric is to provide more relevant information to the network. The idea of providing more information is fundamental – even for simpler templating techniques as problems cannot necessarily be solved using templates alone. The research question will focus on a similarity-based mechanism with various model implementations and whether they can improve an existing solution. In the background, we have assessed that none of these methods is well-suited for working with structured documents (like invoices), since they generally do not have any fixed layout, language, caption set, delimiters, or fonts. For example, invoices vary across countries, companies and departments and change over time. To retrieve any information from a structured document, it must be understood.

In one-shot learning, we are usually able to correctly identify classes by comparing them with already-known data. One-shot learning works well when the concept of similarity is utilized. For similarity to work, two types of data must be recognized – unknown and known. For the known data, target values are known to the method and/or the model. To classify any unknown input, the usual practice is to assign the same class to it as this is the most similar-known input. Siamese network is used for similarity in this type of work meaning the retrieval of similar documents that need to be compared. This is performed using the nearest neighbor search in the embedding space for this work. The loss used for similarity learning is called triplet loss because it is applied on a triplet of classes (R reference, P positive, N negative) for each data point:

$$L (R, P, N) = min (||f (A) - f (P)||2 - ||f(A) - f(N)||2 + \alpha, 0)$$

Where α is a margin between positive and negative classes and f is the model function mapping inputs to embedding space (with the Euclidean norm). The main unit of our scope is every individual word on every individual page of each document. For the scope of this work, we define a word as a text segment that is separated from the rest of the text by (at least) a white space, and we will not consider any other

text segmentation. Inputs and outputs: Conceptually, a document's entire page is considered the input to the whole system. Each word – together with its positional information (or word-box for short) – is to be classified into zero, one, or more target classes as the output.

We are dealing with a multi-label problem with 35 possible classes in total. The dataset and the metric: Overall, we have a dataset with 25,071 PDF documents, totaling 35,880 pages. The documents are of various vendors, layouts, and languages, and are split into a training, validation, and test set at random (80 % / 10 % / 10 %). A validation set is used for model selection and early stopping. The metric used is computed first by computing all the F1 scores of all the classes and aggregated by micrometric principle. (Fig. 2.23.)
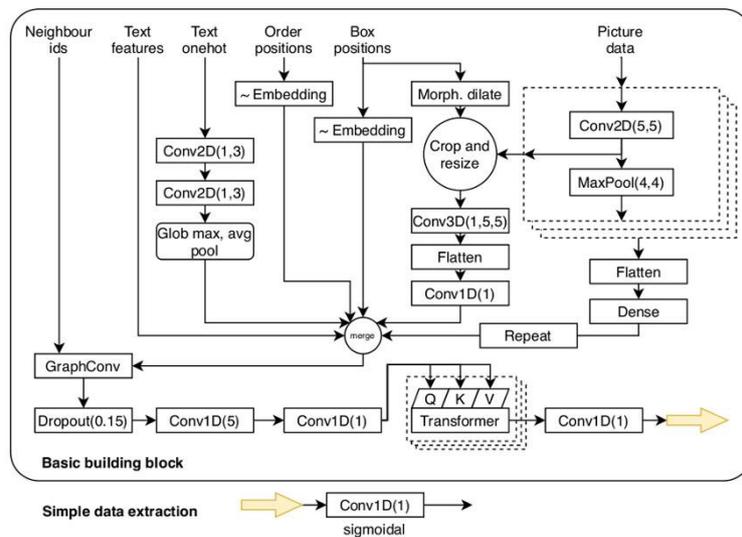
Figure 2.23. Micrometric principle

2.30 Document translation

Methodology

Create azure storage blobs to store the documents, and an optional storage blob for a glossary if needed.

One blob will be used as the Source URL, meaning the documents that are to be translated.

One blob will be used as the Target URL, meaning the location where the translated documents are to be stored.

The last optional blob is the storage where a glossary of terms and definitions are kept should it be necessary to use specific translations for words, such as industry jargon.

The Azure translator service needs certain accesses to the blobs. For the Source URL blob, it needs Read and List access. For the Target URL blob, the service needs Write and List access. For the optional Glossary it needs Read and List access.

These access parameters must be set on the storage blobs and then the translator service is granted a Token that allows it to interact with the storage blobs. (Fig. 2.24.)
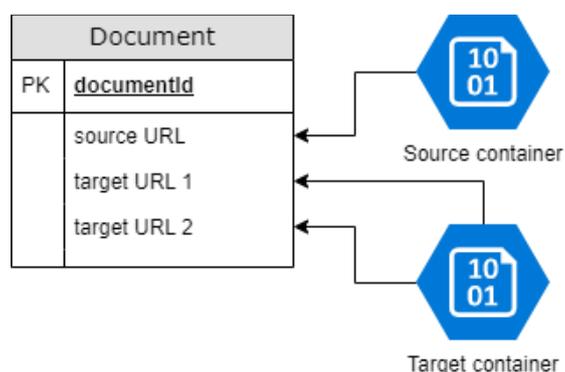

Figure 2.24. Source and Target document

Additionally, there exists a database that keeps references to the URL to the original document, and to the translated versions of the same document, making it easy to connect the various documents to each other.

Value

Being able to supply users with content in their language lowers the barrier for them interacting with your systems and websites. This can give you a competitive advantage over other businesses that are working in the same area or industry. By using automated translations, you can reduce the amount of work needed to be done by staff so they can concentrate on other tasks.

System architecture

Shown below is a diagram of the architecture and workflow from when a User uploads a document that is supposed to be translated, onto an azure storage location that has been set up to keep the documents. If this upload is successful, an Azure function initiates the cognitive services and translates the document. It uses a glossary of terms if applicable. It then writes the document it has translated onto an Azure storage location and after the upload is successful, a new entry in a database that has the references to the original document, and all the translated documents. (Fig. 2.25.)
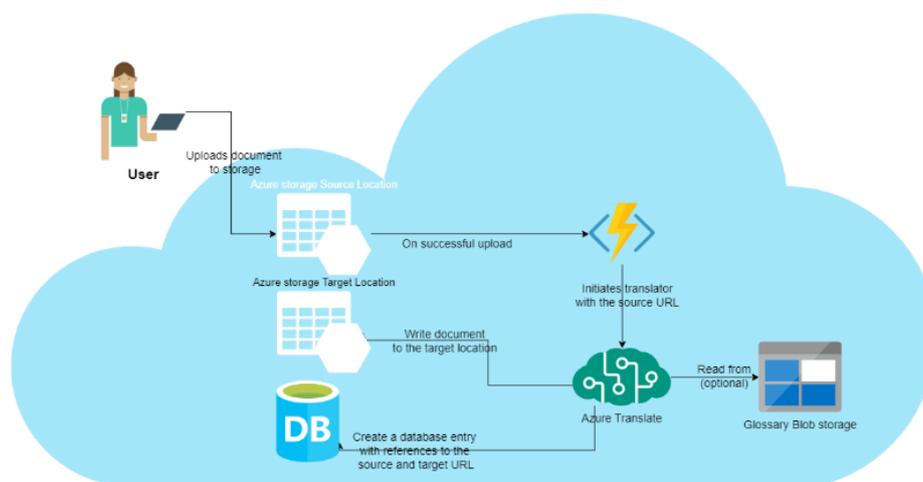
Figure 2.25. Diagram of architecture

2.31 Dynamic website hosting

Methodology

Static hosting can be compared to any bookshelf we have in our office. Considering that while setting up the bookshelf, we have included all the books that we wanted to put and no others are left in our collection so everyone who visits the bookshelf, will see the same content every time. That's analogous to static websites. A static website contains web pages with fixed content and publishing such websites to a web server is called static hosting. Simply put, static web hosting supports fixed-content, HTML-based websites that display the same information to all visitors. When a user's web browser retrieves a static website from a static web hosting server, the entire page is already constructed in HTML files (along with possibly CSS and JavaScript). Github pages, Netlify, Firebase, etc are some of the free static web hosts.

Dynamic hosting can be described as our bookshelf having a provision that we can update the books when someone visits it inclined with the requirements so the content will not be the same for every visitor. This is analogous to dynamic websites. A dynamic website is a collection of dynamic web pages whose content changes in real time. Hosting such web pages to a web server is dynamic hosting. It accesses content from a database or Content Management System (CMS). Therefore, when you alter or update the content of the database, the application code generates runtime files and the content of the website is altered or updated dynamically. Local hosting refers to the hosting on local computer that a program is running on. For example, if we are running a Web browser on our computer, our computer is considered to be the "localhost." Whenever we call the localhost, internal communication happens within your system to make an application run on our computer.

Value

The server-side code used to build a dynamic website can generate real-time HTML pages constructed to fulfil individual user requests. While static websites tend to be informational, dynamic websites contain interactive, continually changing elements. As a result, web developers typically use a

combination of client-side and server-side programming to create a truly interactive website experience for visitors. Dynamic websites generate and display content based on the actions taken by a user. The level of change that takes place depends on the developer's skill and how intricate they make the interactive elements of a dynamic website. Let's think about a user profile you may have set up on a website like Amazon or Walmart. Every time you visit the page, you see recommendations chosen for you based on past purchases. You can also pull up information about your account or past orders. The site generates a unique experience for you based on our past actions.

Dynamic websites provide more website functionality and enable user interaction, let us request and store information in an organized way, display content based on the user's needs. These types of websites enable additional website flexibility by allowing connection to a CMS, including ability from multiple users to adjust the content. In such way, it is less costly to adjust and changes versus a static website. Dynamic sites are more likely to attract recurring customers and visitors.

System Architecture

System architecture is presented in the Figure 2.26. below.
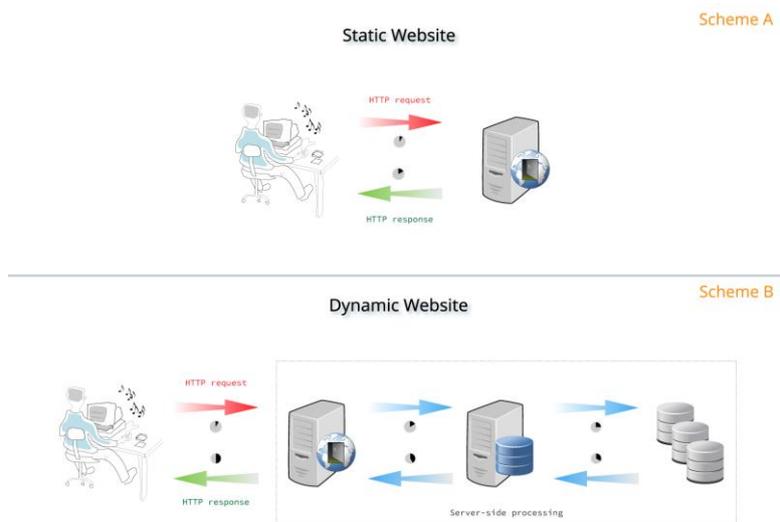
Figure 2.26. System architecture

2.32 Dynamic web site with data storage in a database

Value

Websites are very necessary not only for advertising but also for accessing information through internet browsing. Even small companies have websites through which they advertise. These sites are very useful for informing potential collaborators about what the company can offer at a given time.

For example, a pizzeria or any restaurant can inform potential customers through its own website about the daily menu it offers. This can be changed. For a website with static content, it is very difficult to

make these changes daily. A person must deal daily or almost daily with updating the website.

In addition, this person must have specialized knowledge. This person must consult the management of the company to obtain the data that must be modified. It is much easier if the website has content that can be dynamically modified by someone in the company who has no knowledge of how to create or modify a website.

A site of this type saves money and time and is efficient for the company. The content can be changed as often as desired without the involvement of a specialized person.

This type of site can also ask for the opinion of customers about the quality of the services offered. Knowing the opinion of the customers, the company can modify its internal policy in accordance with the demands of the customers.

This application can be used for any company that offers services to customers similar to those of a pizzeria. For example, it can be used in restaurants so that the customer knows if the dish he wants is on the menu that day. Florists can also use this type of application due to the fact that they can offer customers personalized bouquets. Customers can also express their opinions and suggest ways to improve the services offered by the company to the management of the companies.

Application architecture

The architecture of the application consists of HTML pages that display the information that the company wants to send to the public space. The information in the HTML page is modified using Java script functions written on separate pages. The information to be modified is stored in a database. The communication of the HTML page with the database is done with the help of the PHP language through Java Script.

An authorized person within the company can access the database through an HTML page of the site (Fig. 2.27.). After authentication, the person in question can access the HTML pages that need to be modified and choose from the database the information to be modified. This information can be displayed in the form of text or in the form of images. The information needed to modify the page is stored in the database.
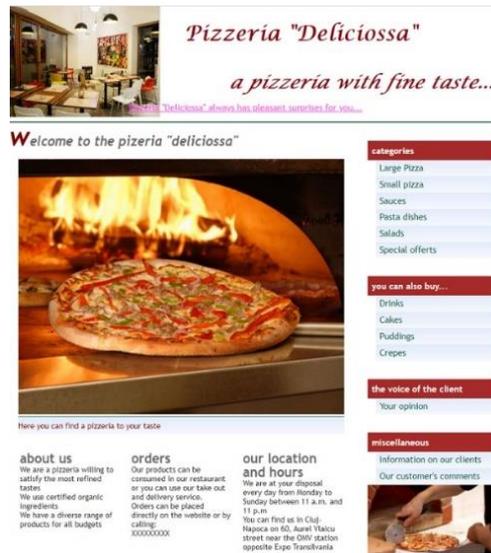
Figure 2.27. Page HTML. Adopted from Pizza Lipa, 2023

The website also contains pages through which it relates to customers. He can ask for their opinion or build customer loyalty. The authorized person can read customer opinions in a web page to which only he has access. As a result, some web pages can be accessed publicly and some pages from the same site are accessed based on authentication using the password and user account only by the company's staff.

The stages of making the application are:

Establishing the content of web pages and the appearance of these pages
In this stage, the content of each page and the way in which the images and texts are displayed are established. The color scheme of each page is also established by mutual agreement.

Creating the database that the application will access
In this stage, the types of data that will be stored in the application, the format of these data are determined. Depending on these elements, the tables contained in the database and the relationship between them are established.

Writing the application
In this stage, the codes for creating the HTML pages are written, using the HTML, CSS, Java Script and PHP languages. The HTML and CSS languages determine the appearance of the pages and the structuring of the information on the page. The functions that modify the images and text in the HTML pages are written in the Java Script language. The PHP language ensures the relationship between the application and the database through the functions that are written under this language.

Testing and adjusting the application
After writing the codes and verifying the application by the executor, the web application must be publicly accessible. In this sense, a domain is rented by the owner of the application and tested. The tests consist of checking how the application is seen by different browsers. Check if the information can be easily changed using the specific pages created in the application. It also checks how the application

2021-1-SI01-KA220-VET-000034641

relates to the database.

Special attention must be paid to the way in which the information is entered in the database by the application. It is important that the information entered in the database tables is correct and not truncated or entered in other columns due to errors in writing the application code.

2.33 E-commerce Application

Methodology

The methodology for developing an e-commerce application involves the following steps:
- Requirements Gathering: The first step is to gather requirements from stakeholders, including businesses, customers, and other end-users. This includes identifying the key features and functionalities needed for the e-commerce application, such as product listings, shopping cart, checkout process, payment gateway integration, and inventory management.
- Planning and Design: Based on the requirements gathered, the next step is to plan and design the e-commerce application. This involves creating wireframes, prototypes, and mock-ups to illustrate the user interface and the user flow. Additionally, the technical architecture of the system is defined, and the development team establishes the technology stack, development methodology, and deployment strategy.
- Front-end Development: Once the design is complete, the front-end development team starts developing the user interface and the user experience components. This includes designing the website's look and feel, UI/UX design, and creating the functionality components.
- Back-end Development: After the front-end development is complete, the back-end servers and APIs are developed, following the technical architecture defined in Step 2. This includes integrating with payment gateways, inventory management systems, and other third-party applications and services.
- Quality Assurance: After the development is complete, the e-commerce application is tested thoroughly to ensure that it meets the requirements specified earlier. This includes functional testing, performance testing, security testing, and user testing to ensure that the application is user-friendly and error-free.
- Deployment and Launch: After the application is developed, tested, and approved, it is deployed to the production environment, and the launch process is initiated. During this phase, the application is monitored carefully to ensure that everything is working fine and bug-free.
- Maintenance and Updates: As part of the e-commerce application methodology, new features are added, bugs are fixed, and the application is maintained on an ongoing basis.

Overall, the methodology for developing an e-commerce application requires close collaboration between business owners, developers, designers, testers, and other stakeholders. It should be iterative, allowing for feedback, updates, and modifications throughout the development process, with a goal to deliver a user-friendly, secure, and seamless shopping experience for customers.

Value

An e-commerce application can add significant value to businesses by providing a seamless and convenient platform for customers to buy products and services online. Here are some of the ways that an e-commerce application adds value:

- Increased sales revenue: An e-commerce application opens up new sales channels allowing businesses to reach customers who prefer shopping online. This can increase their sales revenue and help businesses to reach a wider audience.
- Reduced costs: E-commerce applications can reduce many of the costs associated with traditional sales channels, such as rent costs, advertising, and payment processing. This results in increased profits and a more cost-effective way to sell products and services.
- Improved customer experience: Customers can take advantage of the convenience of online shopping, browsing through products, comparing prices, sharing ratings and reviews, and adding items to their online cart, all in the comfort of their own homes. It provides a better shopping experience, leading to higher customer satisfaction.
- Increased customer engagement: An e-commerce application can help businesses to keep customers engaged, establish a loyal customer base and generate repeat customers through personalized suggestions, rewards, and loyalty programs.
- Insights and Analytics: E-commerce applications provide businesses with comprehensive data regarding customer behavior, buying patterns, and preferences, which can reveal opportunities to improve the online shopping experience and drive further sales.
- Scalability: E-commerce applications can scale with the business's growth, making it simple for businesses to add new products or services, add new features, and manage a growing customer base.

Overall, by providing an easy and convenient platform for customers to shop online and providing businesses with a more cost-effective way to sell products and services, e-commerce applications can significantly enhance productivity, sales, revenue growth, and customer relationships.

System Architecture

An e-commerce application system architecture comprises several components:

- Client-side architecture: The client-side architecture comprises the user interface components which the customers interact with. This includes the web browser and the presentation layer, which renders the application interface.
- Application server: The application server comprises the server-side components that interact with the data management layer, security, and integration services. This layer includes the business logic, which manages data transactions and process flows.
- Database layer: The database layer is responsible for storing and managing application data. It includes the actual data storage and management systems, such as file systems, relational databases, or other data storage infrastructures.

- Payment gateway: Payment gateway is a critical component of an e-commerce application that securely processes financial transactions, including credit card and other digital payment methods.

2.34 Electronic catalogue with students' school results

Value

The results of the students in the written and oral evaluations are recorded in most cases in paper catalogues. In these catalogues, the grade, the date it was obtained and the subject for which the assessment was made are recorded. The analysis of the situations of each student is usually done periodically by doing the calculations manually.

A modern and more efficient system is to record the marks obtained by students following the evaluations in a database. The advantage over the classic system consists in the fact that the student's situation can be analysed at any time, and upon obtaining a low grade that places the student at the limit of possibility, the application issues a warning message.

The student can access the data from the electronic catalogue, but he only has access to his own grades and not to the grades of his colleagues.

The application creates different statistics that can be useful to the student. An example is the ranking of students in each study subject based on the grades obtained in a certain test or based on the final average in the respective subject.

Application architecture

The application is made up of a database organized on tables in which the data of each student and the results obtained are stored.

The application itself is of the website type, to which each teacher has access through a password. The teacher has access only to the tables in the database in which the results obtained by the student in the subject taught by the respective teacher are recorded.

The application automatically calculates the average obtained by each student in each subject at that time. There are situations in which the student has a situation below the promotion limit and in this case the application issues a warning message that can be sent by email or mobile phone.

The database is located on a server, as is the website application. The language in which the application is written is HTML, JavaScript, and PHP. Java script makes the connection between the functions written in PHP and the page written in HTML.

2.35 Facilities Access Control

Explanation

In physical security and information security, access control (AC) is the selective restriction of access to a place or other resource, while access management describes the process. The act of accessing may mean consuming, entering, or using. Permission to access a resource is called authorization.

The management of access to facilities over an undefined area, with different facilities, providing real time information to the management team of entrance/exit data and occupancy

Electronic Access Control

Electronic access control (EAC) uses computers to solve the limitations of mechanical locks and keys. A wide range of credentials can be used to replace mechanical keys. The electronic access control system grants access based on the credential presented. When access is granted, the door is unlocked for a predetermined time and the transaction is recorded. When access is refused, the door remains locked, and the attempted access is recorded. The system will also monitor the door and alarm if the door is forced open or held open too long after being unlocked.

When a credential is presented to a reader, the reader sends the credential's information, usually a number, to a control panel, a highly reliable processor. The control panel compares the credential's number to an access control list, grants or denies the presented request, and sends a transaction log to a database. When access is denied based on the access control list, the door remains locked. If there is a match between the credential and the access control list, the control panel operates a relay that in turn unlocks the door. The control panel also ignores a door open signal to prevent an alarm. Often the reader provides feedback, such as a flashing red LED for an access denied and a flashing green LED for an access granted.

The above description illustrates a single factor transaction. Credentials can be passed around, thus subverting the access control list. For example, Alice has access rights to the server room, but Bob does not. Alice either gives Bob her credential, or Bob takes it; he now has access to the server room. To prevent this, two-factor authentication can be used. In a two-factor transaction, the presented credential and a second factor are needed for access to be granted; another factor can be a PIN, a second credential, operator intervention, or a biometric input.

There are three types (factors) of authenticating information:
- something the user knows, e.g., a password, passphrase or PIN.
- something the user has, such as smart card or a key fob.
- something the user is, such as the user's fingerprint, verified by biometric measurement.

Passwords are a common means of verifying a user's identity before access is given to information systems. In addition, a fourth factor of authentication is now recognized: someone you know, whereby

another person who knows you can provide a human element of authentication in situations where systems have been set up to allow for such scenarios. For example, a user may have their password, but have forgotten their smart card. In such a scenario, if the user is known to designated cohorts, the cohorts may provide their smart card and password, in combination with the extant factor of the user in question, and thus provide two factors for the user with the missing credential, giving three factors overall to allow access observing the "future needs" of a platform for advanced analysis of video content using machine and deep learning in the field of surveillance/security and traffic/transport is the basis of the differentiation of this project and consists of three main starting assumptions:

Computer Access Control

In computer security, general access control includes authentication, authorization, and audit. A narrower definition of access control would cover only access approval, whereby the system makes a decision to grant or reject an access request from an already authenticated subject, based on what the subject is authorized to access. Authentication and access control are often combined into a single operation, so that access is approved based on successful authentication, or based on an anonymous access token. Authentication methods and tokens include passwords, biometric analysis, physical keys, electronic keys and devices, hidden paths, social barriers, and monitoring by humans and automated systems.

In any access-control model, the entities that can perform actions on the system are called subjects, and the entities representing resources to which access may need to be controlled are called objects. Subjects and objects should both be considered as software entities, rather than as human users: any human users can only have an effect on the system via the software entities that they control.

Although some systems equate subjects with user IDs, so that all processes started by a user by default have the same authority, this level of control is not fine-grained enough to satisfy the principle of least privilege, and arguably is responsible for the prevalence of malware in such systems.

More to Explore: Poindev (2022)

2.36 Facilities Access Control

Methodology

The methodology for implementing Facilities Access Control involves several stages, as follows:
- Risk Assessment: Identify the areas of the facility that require access control measures based on risk assessment and analysis. This may include sensitive areas that contain valuable equipment, information, or facilities.
- Design: Once the risk assessment is complete, a design for the access control system is prepared. This includes determining the types of access control measures required for each area and the method of authorization (e.g., credentials, biometric data, or dual authentication).

- Installation: The installation stage involves installing the necessary hardware, software, and electronic access control devices. This may include installing card readers, electronic locks, and control panels.
- Configuration: Once the system is installed, it needs to be configured based on the design. This includes setting up user accounts, configuring access control permissions, and testing the system.
- Testing: The access control system is tested to ensure that it meets the design requirements. This typically involves testing the system for reliability, accuracy, and security.
- Training: Once the system is fully configured and tested, training is provided to end-users on how to use the system. This includes instruction on how to properly access protected areas, how to operate the access control devices, and how to respond in case of emergency situations.
- Maintenance: The maintenance stage involves ongoing management of the access control system. This includes monitoring the system performance, updating, and patching software, replacing, or repairing equipment where required, and ensuring that the system remains aligned with the organization's risk management policies and procedures.

Overall, implementing Facilities Access Control requires careful planning, design, installation, configuration, testing, training, and maintenance. By deploying an access control system, organizations can ensure the security of their facilities and assets and provide a safe and secure environment for their employees and visitors.

Value

Facilities Access Control offers substantial value to organizations in different ways. The following are some of the ways organizations can benefit from Facilities Access Control:

- Improved Security: Facilities Access Control can improve the security of the organization by reducing the risk of theft, unauthorized access, and vandalism. This can help protect the organization's assets, information, and reputation.
- Controlled Access: Facilities Access Control helps restrict access to sensitive areas of the facility to authorized persons only. This can provide confidence to employees that their work environment is secure from trespassers, ensuring their personal safety.
- Enhanced Regulatory Compliance: Most industry standards have security and access control measures as part of their regulations, Facilities Access Control can improve compliance rates and make compliance verification easier.
- Increased Efficiency: Facilities Access Control can enhance efficiency by automating access checks, which can help reduce administrative overhead, and minimize discrepancies or errors when checking for authorized access.
- Audit Trail Creation: Electronic Facilities Access Control system will keep a log of all access requests and activities which can provide management with visibility on who accessed what area and when an event occurred. This can help management identify potential security risks.

Overall, Facilities Access Control provides significant value to an organization by enhancing security, increasing regulatory compliance, creating audit trails and, in turn, improving efficiency, protecting assets, and ensuring employee safety.

2021-1-SI01-KA220-VET-000034641

System Architecture

The system architecture of Facilities Access Control can vary depending on the requirements and size of the facility. However, the following are the components that typically make up an Access Control system:

- Electronic Access Control Devices: Access control devices are the hardware components that restrict access to restricted areas within a facility by verifying user credentials. The devices can range from card readers, biometric scanners, keypads, or touchless technologies.
- Electric Locks: Electric locks are powered locks that control the physical entry or exit of a passage or door within a facility. The locks can be triggered remotely once the user's credential is authenticated by the electronic access control devices.
- Access Control Software: The software is responsible for managing the user credentials, determining access rights and permissions, logging and tracking incidents and activities, and generating reports for management review.
- Databases: Databases serve as the repository for the user's profile and information. The databases store a list of users with their respective qualification levels and access rights to properties and facilities.
- Network Connectivity: Network Connectivity utilizes networks to enable control over access entries into restricted spaces.
- Monitoring and Alert Systems: These systems send notifications to management when there are suspicious.

2.37 Facilities Management

Methodology

The following are some of the common steps involved in the facilities management methodology:

- Assessment and Analysis: The first step in facilities management methodology is to conduct a thorough analysis of the facilities to identify its strengths, weaknesses, and areas for improvement. This includes analysing the building systems, equipment, occupancy data, maintenance logs, and other relevant data.
- Strategic Planning: Based on the analysis, the facilities management team develops a long-term strategic plan to address the issues identified and develop a roadmap for improving the facility's functionality, efficiency, and sustainability.
- Budgeting and Resource Allocation: Facilities management teams develop budgets to support their strategic plans. The budgeting process allocates resources among maintenance, repair, and building improvement projects.
- Implementation: Facilities management teams implement the strategic plan, which includes a range of activities such as preventative maintenance, repairs, and modernization projects.
- Performance Measurement: Facilities management teams track and measure facility performance through key performance indicators and metrics. This can include tracking energy consumption, space utilization rates, and maintenance costs.

- Continuous Improvement: Based on the performance measurement data, the facilities management team makes improvements to the facility management strategy and implements new initiatives to further advance the facility's functionality and efficiency.

Overall, implementing an effective facilities management methodology requires a combination of technical knowledge, analytical skills, and a proactive approach to managing assets' lifecycle. The facilities management methodology's success is largely dependent on integrating personnel and technology such as IoT sensors, data processing and analysis tools, and machine learning to optimize the facility's operation's efficiency and sustainability.

Value

Facilities Management (FM) can deliver significant value to an organization by ensuring that the built environment supports the efficient functioning of its core activities, providing safe, functional, and comfortable facilities. Here are some specific benefits of Facilities Management:

- Reduced operational costs: Facilities Management can help to reduce operational costs by identifying and eliminating waste and inefficiencies, optimizing space utilization, and improving energy efficiency.
- Improved asset management: FM can help an organization to manage the physical assets of its facilities, optimizing their use and extending their useful life, which can deliver cost savings in the long run.
- Enhanced safety and security: FM can help to ensure that facilities are safe and secure for occupants, minimizing the risk of accidents and security breaches.
- Increased productivity: By providing a safe, comfortable, and functional environment, FM can improve occupant satisfaction, boost morale, and increase productivity.
- Regulatory compliance: FM can help an organization to comply with various regulations and standards related to the built environment, such as safety codes, environmental regulations, and accessibility requirements, avoiding costly legal penalties.
- Improved sustainability: FM can help an organization to implement sustainable practices, reduce environmental impact, and promote responsible resource use, demonstrating social responsibility and enhancing brand reputation.

Overall, Facilities Management can deliver significant value to an organization by improving the efficiency of its operations, enhancing the quality of its facilities, and promoting sustainability and social responsibility.

System Architecture

The architecture of Facilities Management (FM) solution depends on the specific needs and goals of the organization. The following are some common elements that may be included in an FM architecture:

- Facility data management: The FM solution may collect and manage a range of data about facilities, such as building systems, occupancy data, and energy usage patterns. This data is usually stored in a database or data warehouse.

- Data analytics: The FM solution may use data analytics tools and algorithms to analyse facility data. Analytics can help identify trends, patterns, and anomalies, enabling predictive maintenance and optimized decision-making.
- Work Order Management: The FM solution can include a work order management system to create and manage work orders, send them to technicians, track progress and manage budgets and other resources.
- Asset Management: The FM solution can include an asset management module, Asset tracking helps FM teams capture information such as the asset description, specification, its location, date of purchase and lifetime values.
- Resource scheduling: The FM solution may include features for scheduling and allocating resources such as labor, materials, equipment, and space. An intelligent scheduling system can quickly identify, and fill slots left open due to technician unavailability, delays in material delivery, or even weather-related absences.
- Reporting & Dashboards: The FM solution may include customizable dashboards and reports that provide a clear view of the status of facility operations, maintenance, asset performance, and work order progress. Reports and dashboards help management teams analyse information to make data-driven decisions about how to optimize FM processes.

The Facilities Management architecture is intended to provide an integrated and holistic view of facility management processes, enabling efficient and optimized operations and maintenance. The architecture employs solutions that cover all the needs of facilities management, identify trends, and anomalies, allocate resources and track spending, and report performance.

2.38 Facilities Occupancy Data

Methodology

Occupancy data is derived from occupancy analytics. Occupancy data is collected by IoT sensors that monitor and record the movement of people throughout a building. Occupancy analytics do not track individuals but assess the number of people using a space over time. The places we visit, either for work, learning, health, or leisure, get transformed to comply with social distancing and government guidelines, many new and different decisions will need to be made. Spaces will continuously need to adapt with the changing guidance bringing new challenges for building owners and occupiers on how to best maximize the efficiency of their space without compromising on safety and impacting heavily on people's wellbeing. These important decisions could be made using best judgement with perhaps a degree of trial and error but what if they could be made using data instead? And, what if the systems in these spaces could adapt to the changes by themselves and provide valuable insights to ensure the right choices are made?

By using data from lighting occupancy sensors, spaces can be mapped out showing busy or quiet periods and highlighting high traffic and bottlenecks areas. The data can be visualized on a floorplan of a building with the ability go back hours, days or longer to help assess how effective measures have been. This feature is particularly useful in managing social distancing as layouts can be altered, for example, if it is

known that a certain area has a high cluster of people, then the layout can be changed to prevent clustering.

Value

Access to occupancy data also brings further benefits. By choosing to close off certain areas, significant savings on energy, maintenance and cleaning costs could be achieved. The data could also help to decide exactly how much space should be closed off helping to achieve the right balance of optimization, safety, and wellbeing.

System Architecture

Facilities occupancy data architecture involves several components, including the following:
- Occupancy sensors: These sensors detect the presence of people in a given room or space using different technologies, including motion, heat, and optical.
- Data communication network: Occupancy sensors transmit information to a central communication network that aggregates, stores, and analyses data.
- Data storage: Facilities occupancy data need to be stored for future analysis and decision-making. Data storage for occupancy data can be on-premises or in the cloud.
- Data analysis tools: Occupancy data analysis tools provide the ability to identify and visualize trends, patterns, and anomalies.
- Dashboards and reporting: Facilities occupancy data dashboards offer a comprehensive view of occupancy metrics. These dashboards provide real-time information about spaces' use and help identify areas for optimization in facilities.
- Integration with other systems: Integration with other building systems such as HVAC, lighting, and access control provides further insight into occupancy data.
- Artificial intelligence and machine learning: Facilities occupancy data can be leveraged for machine learning models to predict occupancy and utilization patterns.

Overall, the architecture of facilities occupancy data systems is designed to create an efficient and sustainable building. Data monitoring and analysis tools enhance building management, reduce energy consumption and costs while optimizing occupant comfort. Good occupancy data architecture must ensure that data is accurate and reliable and facilitate effective decision-making.

2.39 File Comparison

Methodology

There are different methodologies for file comparison, but here are some common steps that are typically followed:
- Identify the files to be compared: This involves selecting the files to be compared and ensuring that they are in compatible formats.

- Choose a file comparison tool: There are a variety of file comparison tools available, ranging from simple text editors with built-in comparison features to more advanced software programs that can compare files in different formats.
- Configure the comparison settings: Depending on the tool being used, there may be different settings that can be configured, such as the type of comparison to be performed (line-based, character-based, binary), which parts of the files to compare, and whether to ignore certain types of changes (e.g., whitespace differences).
- Perform the comparison: Once the comparison settings have been configured, the tool is run to perform the comparison. The tool will highlight any differences between the files and may provide the option to perform specific actions based on the identified differences, such as merging changes.
- Review and validate the results: The results of the comparison are reviewed to ensure that the identified differences are correct and to determine any actions that need to be taken in response to the differences.
- Document the comparison and results: The details of the comparison and its results should be documented for future reference, such as in a report or log file.

Overall, the file comparison methodology involves selecting the files, choosing a tool, configuring the comparison settings, performing the comparison, reviewing, and validating the results, and documenting the process and outcomes.

System Architecture

The architecture of a file comparison system depends on the specific tool or software being used. Here are some common elements that may be included:

- Input files: The input files are the files that need to be compared. They can be in different formats, such as text, binary, or document files.
- Comparison engine: The comparison engine is responsible for performing the file comparison. It analyses the input files, identifies differences between the files, and generates a report summarizing the comparison results.
- Configuration settings: The configuration settings allow the user to customize the comparison process. Depending on the tool being used, the settings may include options such as which parts of the file to compare, the type of comparison to perform, and whether to ignore certain types of changes.
- User interface: The user interface allows the user to interact with the comparison tool. It may provide options for selecting input files, configuring settings, and reviewing comparison results.
- Report generation: Once the comparison is complete, the system generates a report summarizing the results of the comparison. The report may include details such as the number of differences found, the type of differences, and the location of the differences in the files.

The file comparison architecture is designed to make it easy for users to input files, configure settings, run the comparison, and review the results. Some file comparison systems are standalone tools, while others are integrated into other software applications, such as code editors or document management

systems. The architecture is typically optimized for speed and accuracy to provide users with reliable and timely comparison results.

2.40 File storage system using hybrid cryptography cloud computing

Methodology

The encoding method converts real data into an unintelligible form. There are two types of encryption methods: symmetric code decryption and shared key encryption. One technique employs variables to transform the data into an unintelligible structure. Consequently, the authorized individual has access to the information stored on the internet platform. A document was visible to everyone. International Data Encryption Algorithm (IDEA), Triple Data Encryption Standard (3DES), AES, blowfish, BRA and Data Encryption Standard (DES) were symmetrical key cryptography methods. The biggest problem was having the secret for the user in a multi- processor program. The technique provides a low decoding time of information and encryption have a low level of protection. RSA and ECC algorithms are used for public key cryptography. In symmetrical encryption techniques, the public and private keys were combined.

These methods provided a large amount of security, but increased the time required to encode and decode the information. Steganography conceals the appearance of secretive information in a package. The availability of information is not obvious for all with this method. Information was only available to the intended recipient. Textual Steganography technology should be used to ensure data safety. The client's private data is hidden in a message cover picture. When you add text to a message cover image, it looks like a normal text document. If an unauthorized individual discovers a Word document, sensitive information cannot be accessed. If an unauthorized person tries to restore actual information, considerable time was needed. Word has been encrypted and decoded using DES algorithm. Text stenography has the advantage of allowing word privacy. In contrast with image Steganography, word Steganography requires the smallest amount of space. For imaging Steganography, the 3-bit LSB approach has been used. Sensitive customer information was hidden behind the coverage image. It hides a large amount of information in an image using LSB cryptographic technology. The encryption algorithm was executed in a high bandwidth design. AES employs symmetric keys to encrypt information. It supports three major kinds of secrets. It takes 12 turns a 192-bit key, 14 turns at a 256-bit key, and 10 turns, a 128-bit key. Cryptography or decoding timing was less than the enhanced AES method. The modified AES system was able to achieve a significant reduction in time. It uses a specific key for decoding and encrypting messages. The key was 128 bits long. A methodology, many steps were performed arbitrarily so that an unauthorized consumer could even assume the steps of a method [11].

One of the benefits of symmetric key cryptography techniques was their broadband. The enhanced DES algorithm uses a key size of 112 bits to encode and decode data. Two keys are used for data encryption. The 128-bit result of the DES method was divided into two halves. For encrypting and decrypting, the 3DES method requires a considerable amount of time. In comparison to DES and 3DES, the enhanced DES algorithm could provide superior efficiency. One byte at a time was used for name-dependent cryptographic algorithms. It uses a personal key to encrypt and decode data. The method of producing

symmetric keys was used for authentication performed. It protects the privacy of the data. As it runs on a single byte per duration, this technique has a significant maximum duration to turn the information into an encrypted message. To address data processing and security problems, the author has developed a special security framework. Private and public cloud storage containers have been used in this strategy to enhance data security.

Value

Cryptography and Steganography should be employed to address cloud infrastructure issues. Blowfish, RC6, BRA and AES methodologies are used to protect the data brick. LSB has been used to protect sensitive information. To ensure the integrity of the data, the SHA1 hashing algorithm is used. To find the lowest delay score, the multiple processors must be used. The proposed security protocol meets data protection, high security, minimum latency, and identification and privacy requirements. The proposal was correlated with the AES method; text document encryption used 17-20% shorter duration. In comparison with the proposed technique, decrypting AES messages takes 15 to 17 per cent more. Compared to the recommended hybrid model, Blowfish requires 12 to 15% more to encrypt information. Compared to Blowfish's algorithm, decrypting text files using a hybrid approach takes 10% to 12% less time. In the future, combine public key cryptographic techniques to reach a tremendous level of security.

System Architecture

As shown below, the cloud proprietor and cloud user are part of the network infrastructure. The information is updated on the internet platform through the internet operator. A document was separate to octets. the multiprocessing method, each component to the document was encrypted at the same duration. On a cloud platform, the encoded file is saved. The cryptographic keys were kept in the cover image. The multiprocessing scenario was proposed as an internet application. Allows several clients to view files stored on the cloud platform. On document demand, the client also receives a Steganography image through email, which contains crucial data. The document was decoded using the inverse procedure. (Fig. 2.28.)

Figure 2.28. 3DES

87   2021-1-SI01-KA220-VET-000034641
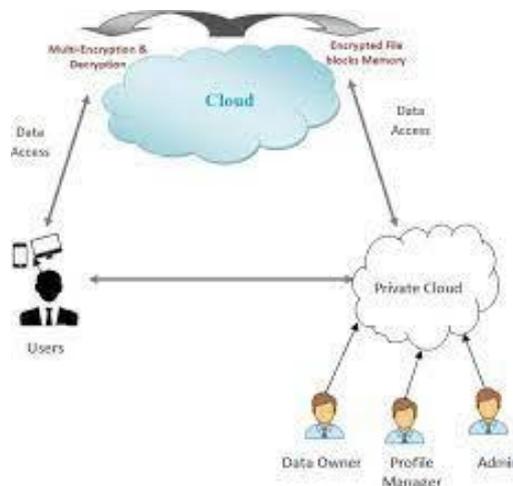
3DES is a significantly improved version of DES in cryptography. DES was being used three times in the 3DES algorithm to boost secure communication. As a result, 3DES would continue to be a flexible cryptography standard in the future is shown below (Fig. 2.29.).
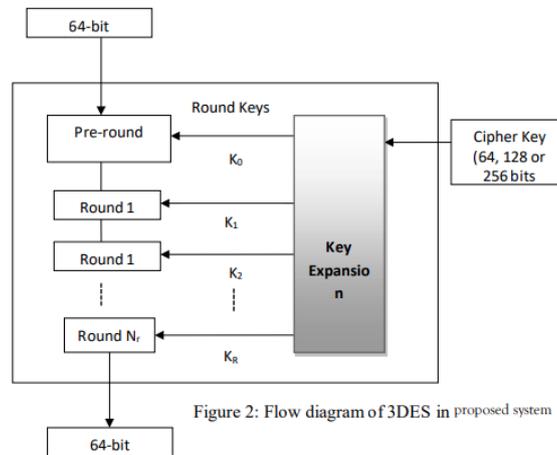
Figure 2: Flow diagram of 3DES in proposed system

Figure 2.29. 3DES flexible cryptography

2.41 Handling traffic spikes

Methodology

- Planning: This involves analyzing past traffic patterns, identifying potential causes of spikes, and determining the resources required to handle spikes.
- Scaling: This involves increasing the capacity of the infrastructure to handle traffic spikes, either vertically by adding more resources to existing servers or horizontally by adding more servers.
- Load testing: This involves simulating traffic spikes and testing the infrastructure's ability to handle them.
- Monitoring: This involves monitoring the infrastructure for signs of traffic spikes and taking corrective action as necessary.
- Optimization: This involves fine-tuning the infrastructure to improve performance, reduce costs, and ensure that it meets the organization's changing needs.

Value

Handling traffic spikes provides several benefits, including:
- Improved performance: By handling traffic spikes, organizations can ensure that their website or application continues to perform optimally, even during periods of high traffic.
- Increased reliability: By preventing crashes and slowdowns, organizations can improve the reliability of their website or application, which can increase user satisfaction and loyalty.
- Reduced costs: By scaling the infrastructure only, when necessary, organizations can reduce costs associated with maintaining excess capacity.

- Improved scalability: By designing the infrastructure to handle traffic spikes, organizations can improve its scalability, which can provide greater flexibility and agility.

System architecture

The system architecture for handling traffic spikes typically involves the use of load balancers, which distribute traffic across multiple servers to prevent overloading any one server. Additional servers can be added to the infrastructure to handle increases in traffic, either manually or automatically through auto-scaling. Content delivery networks (CDNs) can also be used to reduce latency and improve performance by caching content closer to users. Monitoring and optimization tools are used to ensure that the infrastructure is performing optimally and meeting the organization's needs. (Fig. 2.30.)
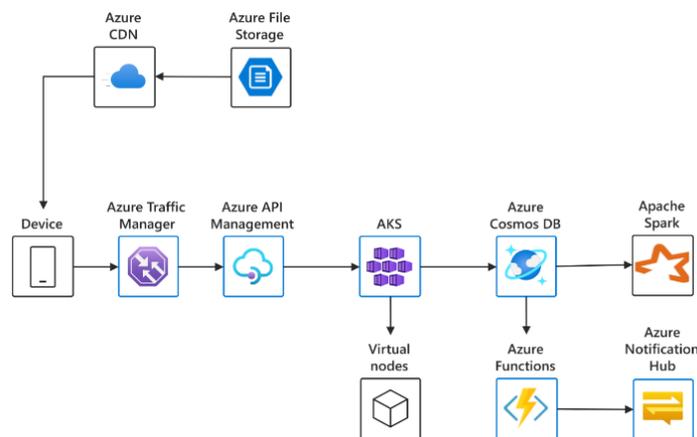
Figure 2.30. Content delivery network. Adapted from Azure Microsoft, 2023.

2.42 Host a static website using AWS (or other clouds)

Methodology

In this part main architectural components of the system are explained. The first mandatory service required is S3. It is a secure and durable object storage service that enables you to store your files in cloud-based storage units called buckets. Every file inside a bucket has a unique URL associated with it that can be used to access it if the user has the privileges. As a first step, you would need to upload the complete contents of a static website into an S3 bucket. Enabling web hosting and making the bucket public will enable it to serve the content of our website. When we have hosted your website on S3, we will have an AWS region-based default website endpoint which we can directly use to access the website. We also have the option to map your custom domain to point to our static website default endpoint, and this is where Route53 comes into play. To provide an authentication mechanism for our website, we can use AWS Cognito. We can create a pool of users and provide a login to the users in that pool. Cognito user pools offer functionality for federated identity providers (for example, login via Facebook and Google), password recovery, and user authorization security in the cloud.

Microservices development (Using Lambda, API Gateway, and RDS) can be delivered using a combination of API Gateway and AWS Lambda. These microservices can be used directly by our website to GET and POST data from a data source. Hence, our static website can also have the functionality of a

dynamic webpage. For example, on an event (e.g., click of a button, form submission, etc.) we can use Jquery Ajax to call the microservice to load data and display on the web page. We can create microservices for complete CRUD operations and more with the help of these services within our website to make your site act fully dynamic. Similarly, we can create microservices which interact with other AWS services like SES in case you want to send email notifications through the website. Basically, we use the AWS SDK to write Lambda functions that provide functionality for your microservice. Through Lambda, we can communicate with any AWS service. API Gateway integrates with Lambda to provide us with an API endpoint.

Value

If our website is mainly informational, and we do not expect too many changes, and the goal is to achieve easy management and low cost, then S3 should be our first choice. Also, using a service-oriented architecture, as shown above, can dynamically display data and post data back to any data source, which gives you the power to make our website achieve many functionalities with low cost and easy maintenance. (Phil Windley, Digital identity, p.37, 2018) (Gupta M., Blockchain for dummies, 2nd IBM limited edition, p.14, 2018)

System Architecture

System Architecture is as seen on the picture 31 below.
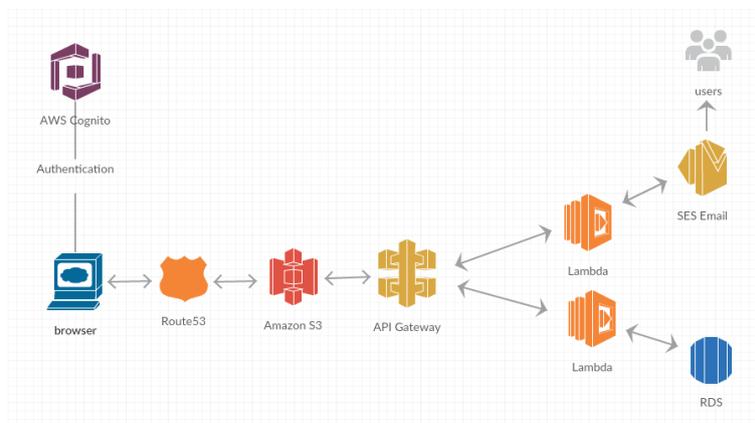
Figure 2.31. System architecture

2.43 Instant Messaging applications

Methodology

Instant messaging is used for real-time communication among users on the internet. Enterprise and consumer users find it an immediate, convenient, and flexible alternative to email. IM'ing is faster than email and more direct than other asynchronous forms of communication.

Instant messaging systems tend to facilitate connections between specified known users (often using a contact list also known as a "buddy list" or "friend list"), and can be standalone applications or integrated into e.g., a wider social media platform, or a website where it can for instance be used for conversational commerce. IM can also consist of conversations in "chat rooms". Depending on the IM protocol, the technical architecture can be peer-to-peer (direct point-to-point transmission) or client–server (an IM service center retransmits messages from the sender to the communication device). It is usually distinguished from text messaging which is typically simpler and normally uses cellular phone networks. Instant messaging applications can store messages with either local-based device storage (e.g., WhatsApp, Viber, Line, WeChat, Signal etc.) or cloud-based server storage (e.g., Telegram, Skype, Facebook Messenger, Google Meet/Chat, Discord, Slack etc.).

Value

Instant messaging applications provide various types of value to both individual users and organizations. Some of the key values that instant messaging applications offer include:

- Improved communication: Instant messaging applications provide users with a platform to communicate in real-time, which can improve communication by providing instant feedback and reducing response times.
- Increased connectivity: Instant messaging applications allow users to connect with others regardless of their location, providing greater connectivity and easier access to networks of people.
- Effective collaboration: Instant messaging applications make it easier for teams to collaborate and share knowledge in real-time, enhancing productivity and enabling better decision-making.
- Cost savings: Instant messaging applications can replace more expensive communication channels such as phone calls and email, resulting in cost savings for individuals and organizations.
- Enhanced customer service: Organizations can use instant messaging applications to provide better customer service, responding to inquiries and issues in real-time, and improving customer satisfaction.
- Personalization: Instant messaging applications often enable users to personalize their messaging experience with custom themes, emojis, and other features, which can enhance user satisfaction and engagement.

Overall, instant messaging applications provide a range of benefits to both individual users and organizations, including improved communication, increased connectivity, effective collaboration, cost savings, customer service, and personalization. As the use of these applications continues to grow, new ways of leveraging their value will emerge, continuing to drive innovation and change in how we communicate and collaborate.

System Architecture

Software developers usually concentrate on the features that attract end-users rather than security-related issues when designing and implementing their software products. The software industry is plagued by such practices which treat security as an afterthought when planning and producing a new product. With the prevalence of IM and the millions of users benefiting from its services, it is very likely

that attackers and creators of malicious programs will take advantage of the situation and exploit vulnerabilities in IM systems to infect a large sector of Internet community. In this section, we discuss some of the security features in IM systems and shed light on the threats surrounding them.

2.44 Manage virtual network

Methodology

The methodology for managing virtual networks typically involves the following steps:

* Planning: This involves defining the requirements for the virtual network, including the number and types of VMs, the network topology, and the security requirements.
* Design: This involves creating a network design that meets the requirements, including configuring subnets, creating virtual network gateways, and defining routing tables.
* Implementation: This involves deploying the virtual network, including creating and configuring virtual machines and other resources, and connecting them to the virtual network.
* Monitoring: This involves monitoring the virtual network for performance, security, and compliance issues, and taking corrective action as necessary.
* Optimization: This involves fine-tuning the virtual network to improve performance, reduce costs, and ensure that it meets the organization's changing needs.

Value

* Improved scalability: Virtual networks can quickly scale up or down as needed, providing greater flexibility and agility than traditional network infrastructure.
* Reduced costs: Virtual networks eliminate the need for on-premises network hardware and associated maintenance costs.
* Enhanced security: Virtual networks can provide more robust security than traditional networks, with features like network security groups, virtual private networks (VPNs), and distributed denial of service (DDoS) protection.
* Improved performance: Virtual networks can be optimized for performance, with features like load balancing and traffic routing that can help ensure high availability and low latency.

System architecture

The system architecture for a virtual network typically involves the use of software-defined networking (SDN) technologies, which separate the network control plane from the data plane. Virtual switches and routers are used to connect virtual machines and other resources, and virtual network gateways are used to connect the virtual network to on-premises networks or the internet. Monitoring and optimization tools are also used to ensure that the virtual network is performing optimally and meeting the organization's needs. (Fig. 2.32.)
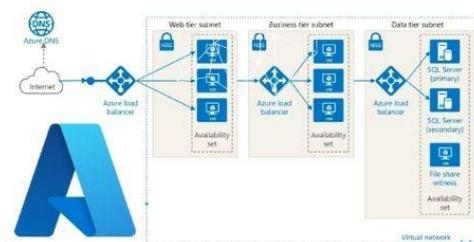
Figure 2.32. Software defined networking

2.45 Migrate to cloud

Methodology

The migration process involves a series of steps, which include:

- Assessment: This involves analyzing your current IT infrastructure, identifying potential areas for improvement, and determining which applications are suitable for migration.
- Planning: This involves creating a roadmap for the migration process, identifying the necessary resources, setting timelines, and defining migration strategies.
- Execution: This involves implementing the migration plan, including transferring data, configuring applications, and deploying the infrastructure.
- Validation: This involves testing the migrated applications and infrastructure to ensure they function as expected.
- Optimization: This involves fine-tuning the migrated systems to improve performance and ensure they are fully optimized for the cloud environment.

Value

Migrating to the cloud provides several benefits, including:

- Improved agility and scalability: Cloud-based infrastructure allows organizations to quickly scale their resources up or down as needed, providing greater flexibility and agility.
- Reduced operational costs: Cloud-based infrastructure eliminates the need for on-premises servers and associated hardware and maintenance costs.
- Enhanced security: Cloud providers typically have more robust security measures in place than on-premises servers, providing improved data protection.
- Improved accessibility: Cloud-based applications and data can be accessed from anywhere, enabling remote work and collaboration.

System architecture

The system architecture for a cloud-based infrastructure typically involves the use of virtual machines, containers, and microservices to create a highly scalable and flexible environment. Cloud-based infrastructure also typically leverages automation tools to streamline deployment, management, and maintenance.

2.46 Monitoring the activities carried out by agricultural machinery on a given surface

Value

The rather difficult conditions in which agricultural works are carried out demand the people who drive the machines involved in these works. The quality of the work also depends on the level of fatigue accumulated by the people who work on these machines. The technological developments that have made it possible to build self-driving cars are also applicable in agriculture. Machines can be built that work without being directly controlled by humans. These machines incorporate artificial intelligence and robotics and IoT technologies.

These machines can be supervised from a fixed point located somewhere on the perimeter of the surface on which the works are carried out. In this supervision point, optimal working conditions can be created for the person supervising the works. In addition, the machines can be equipped with sensors and cameras so that the quality of the work in progress can be checked in real time. In addition, several machines working simultaneously can be supervised from that surveillance point.

Since the person supervising the works is relieved of the obligation to drive the machine, he can closely check the quality of the work carried out thanks to the information received from the sensors and video cameras that are mounted on the machines. The time for carrying out the works decreases significantly because the work rate of the machines remains constant and does not depend on the fatigue of the machine driver (in the situation where he would have driven the machine directly.

Moving the equipment from the garage to the work point in the field can only be done with one person driving the vehicle in front and the other equipment following the driving vehicle. The driving vehicle can be of the type of a minibus modified so that the equipment for monitoring agricultural machinery is placed.

Application architecture

The system consists of several applications. A central application located in the activity monitoring center. Local applications implemented on agricultural machines through which the machine communicates with the central application. The device communicates its position in relation to the monitoring point.

The central application communicates to the machine the commands necessary for the machine to move correctly on the surface it is processing. This is very important because an incorrect movement leaves portions of land improperly worked. The machines can also transmit certain information about the operations that are performed by the agricultural machine. This information can be stored in a database so that it can be analysed later to improve the agricultural technology that is used on the respective land.

In conclusion, the system consists of:
- The central application implemented in the monitoring centre.
- Local applications implemented on agricultural machinery.
- The database where the data collected from the field is stored.

Additional Resources: (Executivegov,2022) (D1 awsstatic, 2023) (Blog DTL,2018)

2.47 Monitoring the physiological parameters of athletes during training

Value

During training, athletes subject their bodies to high demands that change the values of physiological parameters. Such parameters can be blood pressure, pulse, reaction times to certain stimuli.

In order to avoid the occurrence of accidents during training and to appropriately dose the athlete's effort according to his physical and mental state, the monitoring of physiological parameters is welcome.

System architecture

The monitoring system is composed of a number of sensors that are attached to the athlete's body and an application implemented on a computer or mobile phone that captures the signals transmitted by the existing sensors on the athlete's body.

The application processes these signals and provides data about the physical and mental state of the athlete who is training. In this way, the person leading the training receives information in real time about how the athlete's body responds to the training he is doing.

In this way, the coach can dose the training so that the athlete's body is not overloaded.

System realization stages
- Choice of monitored physiological parameters.
- In addition to blood pressure, pulse value and other specific parameters can be monitored.
- Establishing the type of sensors used for monitoring.
- In order to reduce the inconveniences related to wearing the sensors, miniature sensors must be

chosen that are attached to the athlete's body.
- Creation of a database in which the parameters collected from the athlete's sensors are stored.
- The database stores the values collected by the sensors over a period.
- Writing the application for the analysis of the athlete's condition based on the values collected from the sensors and the existing values in the database.
- Testing and adjusting the application

As with any software application in operation, situations may arise that were not foreseen when the application was written. These situations lead to blocking the operation of the application. Following the detection of these anomalies in operation, the application code can be adjusted so that the application works.

2.48 Operate several projects simultaneously

Methodology

The methodology for operating several projects simultaneously using Google Cloud Platform typically involves the following steps:
- Planning: This involves defining the goals, objectives, and scope of each project, identifying the resources required for each project, and creating a project plan for each project.
- Resource allocation: This involves allocating resources, such as virtual machines, storage, and networking, to each project, based on its requirements and priorities.
- Security and access management: This involves setting up security policies and access controls to ensure that each project is secure and accessible only to authorized users.
- Coordination: This involves coordinating the activities of different projects to ensure that they do not conflict with each other, and that resources are allocated efficiently.
- Monitoring: This involves monitoring each project for progress, risks, and issues, and taking corrective action as necessary.

Value

Operating several projects simultaneously using Google Cloud Platform provides several benefits, including:
- Scalability: Google Cloud Platform provides scalable computing resources, such as virtual machines, storage, and networking, allowing organizations to scale resources up or down as needed.
- Security: Google Cloud Platform provides a range of security features, such as firewalls, encryption, and access controls, ensuring that each project is secure and accessible only to authorized users.
- Cost-effectiveness: Google Cloud Platform provides cost-effective computing resources, allowing organizations to pay only for the resources they use, and avoid upfront capital expenditures.
- Collaboration: Google Cloud Platform provides collaboration tools, such as Google Drive and Google Docs, allowing team members to collaborate on projects in real-time, from anywhere.

System architecture

The system architecture for operating several projects simultaneously using Google Cloud Platform typically involves the use of several Google Cloud Platform services, including:

- Compute Engine: This service provides virtual machines for running applications and services.
- Cloud Storage: This service provides scalable, durable storage for data.
- Cloud Networking: This service provides virtual networking for connecting resources across different projects.
- IAM: This service provides access management for controlling access to resources.
- Stackdriver: This service provides monitoring and logging for applications and services.

- In addition, organizations may use other Google Cloud Platform services, such as Cloud SQL, Cloud Bigtable, and Cloud Pub/Sub, depending on their specific requirements.

2.49 SAP Build

Methodology

The methodology for SAP Build involves the following steps:

- User Research: The first step is to conduct user research to understand user needs, preferences, and requirements. This can be done through surveys, interviews, focus groups, or other research methods.
- Ideation: The next step is to brainstorm and generate ideas for the design of the user interface. This can be done through workshops, design sprints, or other ideation sessions.
- Prototyping: After identifying the best ideas, the next step is to create an initial prototype using SAP Build. This involves selecting design elements, functionality, and creating wireframes and mockups.
- Testing and Evaluation: Once the prototype is ready, it should be tested and evaluated by users to ensure it meets their needs and requirements. User feedback is critical at this stage as it helps identify areas where improvements can be made.
- Design Updates: Based on user feedback and testing results, the design can be updated and improved. This may involve making changes to the layout, functionality, or overall design principles.
- Development: Once the design is finalized, the next step is to develop the application. Many development tools can be used to develop the application, but SAP Build is integrated with SAP's broader development and deployment ecosystem, allowing for seamless integration.
- Deployment: Once development is complete, the application can be deployed to the production environment.
- Ongoing support: After deployment, ongoing support and maintenance of the application will be required to ensure its continued success and user satisfaction.

Overall, the SAP Build methodology focuses on the design and development of user-friendly applications that meet the needs and preferences of SAP users. It emphasizes user research, iteration, and validation to ensure that the final application is both effective and intuitive to use.

Value

SAP Build provides significant value to organizations by enhancing the user interface design and development process, leading to improved user satisfaction and productivity. Here are some of the ways that SAP Build adds value:

- Faster design and development: SAP Build's drag-and-drop interface and pre-built design elements enable fast and efficient user interface design. This reduces the time and effort required for designing and developing user interfaces, allowing organizations to develop applications in less time.
- Improved collaboration: The collaborative environment provided by SAP Build allows business teams and technical developers to work together more effectively, resulting in quicker development and delivery of applications.
- Enhanced user experience: SAP Build provides ready-to-use templates, design guidelines, and components that help design intuitive and user-friendly interfaces. The resulting applications provide a better user experience, leading to higher user satisfaction and engagement.
- Better visualization: SAP Build provides visualization capabilities that allow stakeholders to visualize and test concepts before committing code. This helps identify potential design flaws early on to reduce development delays.
- Faster feedback and iteration: Users can provide feedback and iterate on designs much faster with SAP Build's prototyping capabilities. This reduces the time and costs associated with changes to an application design.
- Reduced development costs: By reducing the time and effort required to design and develop user interfaces, SAP Build can help reduce development costs, allowing organizations to allocate resources to other business-critical areas.

Overall, SAP Build adds value to organizations by streamlining the application design and development process, improving user experience, and reducing the costs associated with development. With SAP Build, organizations can create better applications that meet user needs and requirements, leading to increased user satisfaction and productivity.

System Architecture

The SAP Build system architecture consists of several components, including:

- Front-end components: The front-end components of the SAP Build system architecture include the browser-based user interface, which allows business users and developers to design and build user interfaces. It also includes the design elements, templates, and design guidelines that are integrated into the front-end tools that provide a drag-and-drop interface.
- Back-end components: The back-end components of the SAP Build system architecture consist of the application server, which stores the data generated by the users, and the server-side software that processes the data. This layer also includes the integration tools and APIs to integrate with other systems, such as SAP Fiori and SAP Cloud Platform.

- Database: The database is an integral part of the SAP Build system architecture, which stores the data generated by the users, such as user interface designs, user feedback, and other application data.
- Authentication and Authorization: SAP Build system architecture includes authentication and authorization services that ensure secure access to the system. It provides role-based access control to specific functions, which ensures that only authorized users can access specific parts of the system.
- Monitoring and Analytics: The SAP Build system architecture also includes monitoring and analytics capabilities, providing insights into usage patterns, design themes, and app performance issues. These capabilities enable organizations to identify areas for improvement and optimize the user experience.

Overall, the SAP Build system architecture is designed to be flexible and scalable, allowing the platform to be used to develop applications for a range of use cases and industries. The architecture is designed to enhance collaboration between business users and developers to create user interfaces that meet user needs and preferences while adhering to SAP's design principles.

2.50 Reconfiguration of public transport routes in a city

Value

The means of public transport circulate on well-established routes according to a precise schedule established by the management of the company in agreement with the municipality. During peak hours, the means of transport are in great demand, circulating loaded close to the maximum value imposed by the construction.

There are sometimes situations in which on certain routes the maximum load level of the means of public transport is exceeded in certain periods of the day. During peak hours, the municipality introduces races at very short time intervals to reduce the discomfort of travelers.

During normal traffic hours, the means of public transport circulate loaded far below the number of passengers allowed by the manufacturer of the bus, tram, etc. In order to make costs more efficient, the time intervals at which the means of transport circulate are longer so that they can transport a larger number of passengers.

On some routes, means of transport circulate at relatively long-time intervals, which causes major discomfort to the citizens who live in the respective area.

The application reconfigures the route of public transport depending on the number of applicants at each station and the destination they want to reach.

In this way, buses can be put into circulation that cover a route configured in such a way as to reduce the waiting time in stations for travellers, to offer citizens the opportunity to reach their desired

destination faster and to make transportation costs more efficient.

In this way, the quality of transport increases. In addition, the application can also determine the number of means of transport needed during peak hours so that travellers travel in conditions of increased comfort and transport costs are low.

Application architecture

The application requires smart panels in each station. Each traveller must specify the place he wants to reach by touching a touch screen on which a map of the city is displayed. The application centralizes the number of passengers from each station and the destination where the passengers want to go.

Following an analysis based on algorithms used by AI, the application reconfigures the route of public transport in traffic or public transport can be put into circulation to meet the requirements of citizens traveling in the city and to make costs more efficient transport.

The realization of the application involves the introduction of touch screen panels in each station and the creation of stations for the means of public transport on the territory of the city so as to effectively cover the transport throughout the city.

The smart panels can also display additional data such as the waiting time for the bus traveling on a certain route and the configuration of the bus. And on the bus, there are panels that display the route the bus travels on and announce the following stations.

The realization of the application involves the following stages:
- Analysis of the city map and the establishment of all variants of routes on which the buses will run. These routes depend on several possible scenarios.
- Choosing the type of panels that will be installed in the stations.
- Installation of panels in stations
- Writing the application
- The language used to write the application can be Python. In writing the application, specific artificial intelligence algorithms are used, based on which to reconfigure the bus route. The bus drivers receive the information from the application and follow the route established by it.
- The application has a pilot stage in which data is collected regarding the preferences of citizens and the places where stations should be located on the territory of the city are determined so that the transportation needs of the citizens can be met.
- Testing and adjustment stage
- In this stage, it is checked whether the application corresponds to the purpose for which it was created, and the codes are adjusted if necessary.

This application is especially usable in the perspective in which artificial intelligence allows there to be vehicles that drive autonomously without a driver. In this case, the buses will be guided directly by the application. This will establish the route of the buses depending on the number of passengers waiting

in the stations in the city and the destination they want to reach.

There are intelligent terminals in the buses that also display data regarding the estimated time interval in which the bus will arrive at the respective station depending on the route the bus is going to take after configuring the route.

2.51 Remote-controlled smart devices in smart home/office

Methodology

IoT devices and low-power wireless sensor network has many potential wide-ranging applications in the retail industry. They can be used in relation to any of the retail process steps, from supply chain management to customer experience. In this study we emphasize the relations that indirectly connect environmental data to widely used KPIs in existing ERP, DWH and BI systems. Focus is on real time measurements and decision making based on ambient conditions in the stores, especially when the IoT real-time monitoring system is interoperable with the main building air-conditioning system. Such systems can then be used to improve customer experience by optimizing customer behavior through changing environmental conditions in a retail store.

The industry survey shows that 37% of food and grocery companies already experiment with low power IoT technology (e.g., Bluetooth low energy, Zigbee…) or have successfully initiated IoT services or products and further 58% are planning to expand their utilization of the technology implementation and exploitation. IoT engineering helps to improve both internal operations and customer facing processes. Researchers identifies several operational benefits including personalization, dynamic pricing, inventory tracking and monitoring, and recommendations. Energy efficient smart thermostats and lighting are also mentioned. Sensors provide real-time stock information which is used to improve demand forecasts and optimize inventories. IoT improves monitoring and control by coding and tracking objects. That allows companies to become more efficient, accelerate processes, decrease errors, and avoid theft. The real-time data provided by low-power wireless sensor network allows stakeholders to make better operational decisions. IoT-connected smart labels provide means for identification of products as well as for providing additional information. This information can be combined with personalization and recommendation services to enrich shopping experience with pervasive displays and smart things. IoT engineering provides means to engage customer throughout the product life cycle. For example, smart textiles can communicate with smartphones to process biometric information. The life-cycle support requires an adequate NB IoT architecture ensuring efficient and secure data processing.

Value

Implementation of the IoT data analysis platform includes empirical data analysis of relation among the environmental conditions and customer behavior as well as sales performance has been conducted. The results of the mathematical analysis should be used to configure the platform for enactment of improvements of the environmental conditions. The statistical analysis should be used to confirm that the sales performance is significantly affected by the air quality and humidity. The static analysis of

historically accumulated data should help dynamic adjustment of the data analytical models, and also integration of real-time point-of-sales data for dynamic pricing and personalized recommendations. This study does not discuss applied controls which alter the environmental conditions, but one should be aware that technological implementation of such controls is necessary to check actual impact on customer behavior and sales performance.

System Architecture

System architectures is shown in the Figure 2.33. below.
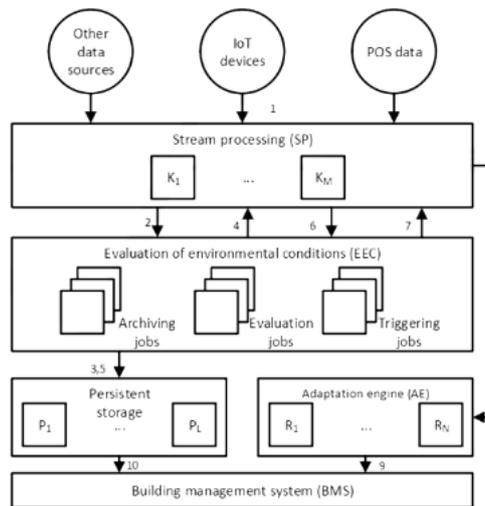
Figure 2.33. Components of ioT data analytics platform

2.52 Resource and application access management

Methodology

Define IAM requirements: Before setting up Azure IAM infrastructure, it is important to define the IAM requirements for your organization. This may include identifying the users and groups that require access to Azure resources, defining roles and permissions, and determining authentication and authorization requirements.

Create an Azure Active Directory (AD) tenant: Azure AD is the identity and access management service that provides centralized identity management and authentication for Azure resources. To create an Azure AD tenant, you can follow the steps provided in the Azure portal.

Configure users and groups: Once the Azure AD tenant is created, you can configure users and groups by adding them to Azure AD. This can be done through the Azure portal or by using Azure AD PowerShell commands.

Create custom roles: In Azure, custom roles can be created to define granular permissions for users and groups. This involves defining the permissions required for a particular role and assigning the role to specific users or groups.

Implement multi-factor authentication (MFA): MFA can be used to provide an additional layer of security for Azure resources. This involves requiring users to provide a second form of authentication, such as a code sent to a mobile device, in addition to a password.

Configure access policies: Access policies can be configured to control access to specific Azure resources. This involves defining the conditions under which a user or group is granted access to a resource, such as based on IP address or device type.

Monitor and manage IAM infrastructure: Once the IAM infrastructure is set up, it is important to monitor and manage it to ensure that security and compliance requirements are being met. This can be done through the Azure portal or by using Azure AD PowerShell commands.

Value

Properly implemented Access Management can be crucial for any business, and the value gained can be immeasurable. Lack of, or improperly configured access can leave the organization vulnerable to unauthorized and malicious access that can both lead to theft of resources or data, and a potentially huge blow to the organization's reputation.

It is also a great tool to minimize the risk for more accidental issues, such as an otherwise trusted employee making a mistake and deletes resources that is needed by others, or if they forget to deprovision some resource resulting in a huge billing cost.

System architecture

System architecture is shown below on Figure 2.34.
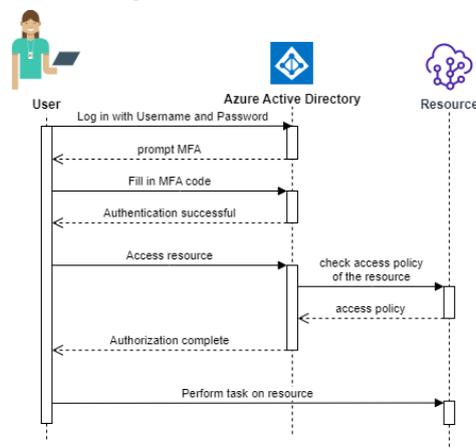
Figure 2.34. Azure AD

When Azure AD checks the policies on the resource it confirms the User has access based on criteria that has been defined for the policies in the conditional access. E.g., the user's groups, roles, location, and device. (Fig. 2.35.)
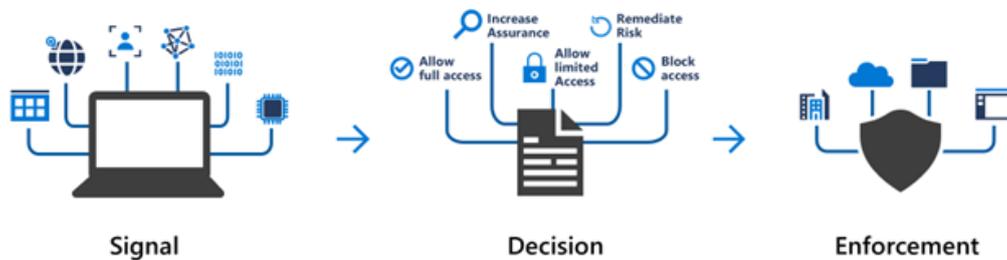
Figure 2.35. Conditional access flow. Adopted from "What is Conditional Access in Azure Active Directory?" by Microsoft learn.

While the user is performing these actions, they are also logged. The user credentials used to log in, the resource they are trying to access, the operations they perform, and the time signature for when it happened are stored and can be used to create a report for the resource, or the user. This can then be used to detect unwanted access, malicious or otherwise, and makes troubleshooting much easier when situations arise.

2.53 Rule-based phishing website classification

Methodology

For the purpose of collecting data, we built a system called Lino that is used to record the available server and client variables, with the corresponding website used as a graphical interface. Lino consists of news and articles taken from other sources where the news is usually attractive and current to attract a larger number of potential users as well as robots. The system collected data for three months, from September until December. Lino has no specific functionality; instead, it is being used to display the content, text fields, and other website elements that are used for the collection of user actions. In other words, from every user's request Lino tries to elicit useful information that will be used in next steps. Since we are interested in malicious clients, we decided to hide in the page specific keywords, so called Google Dorks, used by attackers to find victims of a specific attack.

Example of a keyword listed as Google Dork:
inurl: »wp-download.php?dl_id="
Keywords are intended for the Google search engine queries, as in the example above where the attacker seeks vulnerable parts of the popular blog engine Wordpress. Also, on the site we hide popular spammer addresses, to attract more email harvesters. To reduce the risk that we do not collect enough data in each time interval we made multiple instances of the system. In total we used three instances that contained links to other instances.

For each user who accessed the system Lino generated a session identifier which had a duration of 60 min. From every user request Lino captured the following data:

- Request timestamp.
- Client IP address.
- Client port used for the connection (web browser takes a random port).
- Requested path to the Web server.
- Query parameters used in the request.
- Request method used (GET, POST, HEAD ...).
- Type of protocol that is used.
- If the client sends data using the POST method, the content is written.
- Previous references from which the user came to the currently requested page (HTTP Referrer field).
- User Agent.
- Session identifier.

Before labelling a. dataset, it was necessary to reduce and generalize the data. We decided to split into sessions all requests to the application Lino (over 12,500 of them). The session can be viewed as a unique behavioural profile of a client who joined the system in each time interval (in this paper we used a period of 4 hours). Session identifies requests that were made by a unique client. The client's session grouping is necessary to select a specific time range in which the user has been active on the website.

Following in Table 2.1. are the criteria by which we add sessions to clients.

Table 2.1. Time ranges of the data collected

Feature description	Feature name and type
Requests made in specific session	nurn_requests (numeric)
Requests with HTTP HEAD parameter it was often the case that the robot checks the availability of website using a HEAD method	head (boolean)
Since we had hidden links in the application, we calculated the percentage of accesses to those links	percJiiddenJ/inks (numeric)
Standard deviation in the intervals between queries in a single session	std_dev-timedelta (numeric)
Duration of the session	session-duration (numeric)
The user receives a session identifier which is valid for 60 minutes. Users accessing without a browser or with disabled JavaScript change session identifiers on every request	session-change (boolean)
Sent data via a web form	post-data (boolean)
Accessing a robots.txt file, which contains the rules of behaviour for robots	robots (boolean)

All requests with the same IP address and user agent in the time range of four hours (counting from the first request) are considered as a new session.

When creating each session, we also store the features presented in Table. Based on the above-mentioned criteria, we got about 3,500 sessions, which were used to build a learning set, also storing additional attributes which included: reverse DNS queries for client IP address, client country of origin,

client's AS number of the service provider. We automated the labelling procedure using publicly available data from MaxMind's GeoIP database. After the automatised labelling procedure we manually skipped through the records to correct the wrongly classified instances, for examples unknown robots. We used aforementioned client based and session-based data as features to identify if a new session is a robot or a human. Unfortunately, due to limited number of robots collected we could further segment the dataset on malicious and normal robots. Lino fulfilled its primary purpose and collected mostly robots' sessions. Session's distribution by user types is the following: Human 9 %, Robot 90 % and Unknown 1 %.

Value

The main drawback with selected features is the detection of human visitors which as consequence gives a high false positive rate for robot detection. The root cause is probably the asymmetric dataset where there are many more robots than human visitors. A recommendation would be to expand the number of human visitors to get a more balanced dataset[4]. The reason for these shortcomings is too short time used for data collection and perhaps ambiguous texts that are found in Lino. The system should be set to parse a more popular domain and adapt these texts to target a smaller, more specialized, audience. Also, we should optimize vectors of attacks (dorks) incorporated in Lino.

In addition to the above-mentioned shortcomings this study consists of several contributions, keep in my mind that we did not want to present a perfect classifier. The first contribution is the way in which the data is collected, in most available literature datasets are used from activity logs of web servers we use a trap system for fooling and catching robots. There is no doubt that Lino should collect more client features, and this is planned for future implementations and development. Also, features used in Lino are slightly different from previously used features. Researchers often look at error ratio, percentage of images in data and other similar features. We do not analyse these features because it is not relevant to Lino nor is it possible that we produce a query that will cause an error in Lino. Lino allows us to collect more useful data related to user behavior. For example, whether the user has posted something in a non-existing form, if he clicked on hidden links or if he constantly changed identifier sessions. In relevant literature no one uses variables that are related to the time interval between queries. We use the standard deviation of the time elapsed between queries, which according to our feature selection does not contribute to the classification. Although it proved to be insignificant with the current dataset, automatized procedures should have lower standard deviations and a regular difference between queries.

System Architecture

System Architecture is shown below in the Figure 2.36.
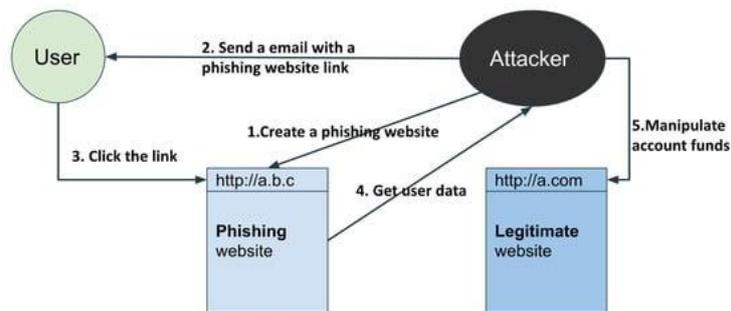
Figure 2.36. System Architecture

2.54 Set up load balancers

Methodology

The Azure Load balancer works at the transport layer of OSI model, where it is the single point of contact between the user and the server layer. It distributes the incoming traffic between the available servers based on the load balancing rules that have been configured in Azure, and the current health of the servers using health probes.

Load balancers are set up with either/or a public load balancer and an internal load balancer.
Public load balancers provide the entry-point for the wider internet, where it takes the incoming public IPs and translates them into a private IP suitable for use within the virtual network. The public load balancer also does the reverse, where it will receive data from within the virtual network (most probably some form of response from the server) and then translate the private IP into a public IP before returning it to the client.

The internal load balancer can only be used within the virtual network and will only accept requests from private IPs. Meaning they cannot be accessed by the internet. This type of load balancer will usually be used to distribute the workload between servers or virtual machines running the business logic of a system.

Health probes are used to determine the current health status of the servers in the backend pool. The health probe for a load balancer is created and configured when the load balancer itself is created. During the configuration you can set the threshold for what is considered an unhealthy server, and when a probe does not respond the balancer will not utilize the server the probe is connected to.

Value

The generated value from implementing load balancing to a server or virtual machine infrastructure can be quiet.

Improved performance: By distributing traffic across multiple servers or virtual machines, load balancers can improve the responsiveness of applications and reduce the risk of downtime due to overloading.

Increased availability: Load balancers can also help ensure that applications remain available even in the event of a server or virtual machine failure, by automatically redirecting traffic to healthy instances.

Scalability: Load balancers can be used to scale applications horizontally by adding additional instances as needed, without impacting performance or availability.

Simplified management: Azure Load Balancer integrates with Azure Virtual Network, allowing for easy management and configuration through the Azure portal, PowerShell, or the Azure CLI.

System architecture
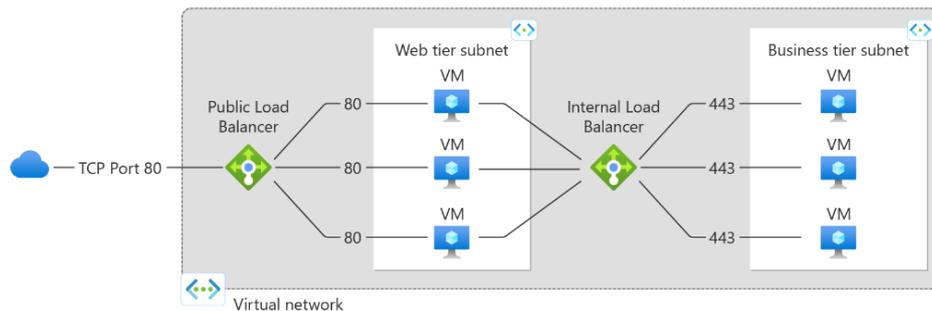
Figure 2.37. Load balancer in Azure

System architecture of a load balancer in Azure where the incoming traffic from the public internet is being distributed amongst available Virtual Machines handling the public traffic and then an internal or private balancer distributing the requests from the client-facing applications to the Virtual machines running the business logic (Fig. 2.37).

2.55 Smart traffic management

Methodology

Examples of the use of such a platform can be presented through scenarios, of which we cite some typical examples.

The detection of vehicles in traffic is one of the key examples of the use of video surveillance for traffic monitoring. In such a scenario, video processing can be applied to observe traffic flow rates on highways that could be used to predict travel times, dynamically calculate tolls, etc. A sequence of video images captured by video cameras is analysed using behavioral detection and monitoring. The number of vehicles detected from each video frame is communicated externally as an output to the surveillance/security and traffic/transport platform.

Detection of special events (incidents) is another example of using video surveillance to monitor traffic. Videos obtained from video cameras can be used to detect traffic accidents, vehicle breakdowns, poor

2021-1-SI01-KA220-VET-000034641

road conditions, etc. Level gap curves are obtained for the traffic flow in each lane, and a change in the curves at a high level indicates an accident.

Pedestrian/cyclist monitoring is another example of transportation-based video surveillance. Information such as direction of travel, density of pedestrians/cyclists, average speed, etc. Can be used to improve safety measures, design of city intersections and signalized crossings, road advertising, heating and ventilation in public spaces, etc. For example, in such change scenarios dimension occurring in the bounding box of each pedestrian/cyclist in the video source can be used to estimate their direction of movement.

In the field of monitoring public areas, the platform enables the identification of security threats through the analysis of the behavior of objects, the identification of persons who are considered high-risk, that is, through the monitoring of unusual or inappropriate situations that can be detected.

Observing the "future needs" of a platform for advanced analysis of video content using machine and deep learning in the field of surveillance/security and traffic/transport is the basis of the differentiation of this project and consists of three main starting assumptions:

Assumption 1: Focus on data or video content as the basis of solutions in the field of surveillance/security and traffic/transportation
Thinking about what kind of data we need to be able to place it on the platform and use it for solutions in the field of surveillance/security and traffic/transport and aligning these conclusions with visions of needs and aligning with future strategies in the field of surveillance/security and traffic/transport.

In addition to the type of data, it is also important how we plan to process this data (for example, video processing is very different from text processing, so we need to think about what tools, technologies, and techniques we use to incorporate this data into our data analysis platform).

Assumption 2: Emphasis on value creation in surveillance/security and traffic/transportation
The question that follows is how we get business value from the previously mentioned data in the field of surveillance/security and traffic/transportation. This question needs to be broken down according to the project's proposed user group, which is a stakeholder. Sometimes it is enough to start a dashboard or run a report to solve problems in the area of monitoring/security and traffic/transportation. However, sometimes it is necessary to make an ad hoc insight or go deeper into data research, i.e., progress towards machine learning methods and artificial intelligence. By emphasizing value creation in surveillance/security and traffic/transportation, we are moving from simply being responsive to a stimulus into a truly predictive realm, and ultimately to the point where we want to be able to predict.

Assumption 3: Provide flexible IT support in the area of surveillance/security and traffic/transportation.
Today, we operate in a hybrid environment that combines on-premises and cloud solutions in the areas of surveillance/security and traffic/transportation. Sometimes it's because the company has different data needs at different times, sometimes it's simply because the company can't get the information and confidence to incorporate the new technology into the business. The project entails control over our

architecture specifically for surveillance/security and traffic/transportation applications, which actually allows us to stream data directly from the cameras in real time and allows us to inject some machine learning models that we will train in the cloud and actually inserted into the flow of data so that they can become very useful in terms of supporting actions and decisions to be initiated or made in the field of surveillance/security and traffic/transportation.

Value

Project is based on the breakdown of video content into basic parts/functionality in the field of surveillance/security and traffic/transport, namely:

Functionality 1: Object detection in the field of surveillance/security and traffic/transportation
The human eye can distinguish a pedestrian, a vehicle, or a suitcase at a glance. It is a more difficult task to distinguish an object that represents a risk. This is precisely what machines can be trained to do using object detection models in the surveillance/security and traffic/transport fields. The learned model can be used to detect an object in a video frame. One model has the ability to detect objects from multiple types and categories, or simultaneously detect multiple objects in the frame. This can be mentioned as the foundation of computer vision because for all other cases, the object detection output will be used as an input parameter for use in the field of surveillance/security and traffic/transportation.

Functionality 2: Object recognition in the field of surveillance/security and traffic/transportation
Object recognition is an extended version of object detection that uses object detection in the initial phase and then maps the detected image to a known related pattern dataset to match features and try to recognize a unique object in the surveillance/security and traffic/transportation fields. Object recognition is widely used for face recognition, license plate recognition, recognition of high-risk situations such as crowds at concerts, recognition of people who are prohibited from entering certain facilities, and the like.

Functionality 3: Behavior of objects in the field of surveillance/security and traffic/transportation
Object behavior is a more advanced scenario and can be used to monitor the behavior of objects. In behavioral tracking and analysis, object detection is used to build an initial input and the detected object will be tracked throughout the video in the surveillance/security and traffic/transportation areas. Object tracking can be considered a complex process for several reasons, for example if it is considered a person, it is very challenging due to perceptual details and interferences, such as poses, illumination and lighting conditions. For training such systems in the field of surveillance/secure sti and traffic/transport, there are multiple objects tracking algorithms and methods that can be used to track a detected object in order to create value from that data.

By combining these functionalities, we are able to solve complex scenarios, such as "detection of abandoned luggage at the airport".

System Architecture

The system we want to run that uses image processing in the field of surveillance/security and traffic/transportation, as opposed to a surveillance system, will need to provide much more functionality than just the chin of the image. Such a system should collect data from video sources, apply video processing, apply business logic in the area of surveillance/security and traffic/transport, manage business logic in the area of surveillance/security and traffic/transport, allow the user to see interesting data via a dashboard created according to the needs of the surveillance/security and traffic/transport areas, generate and manage alerts, and operate equipment that provides video summaries. (Fig. 2.38., 2.39.)
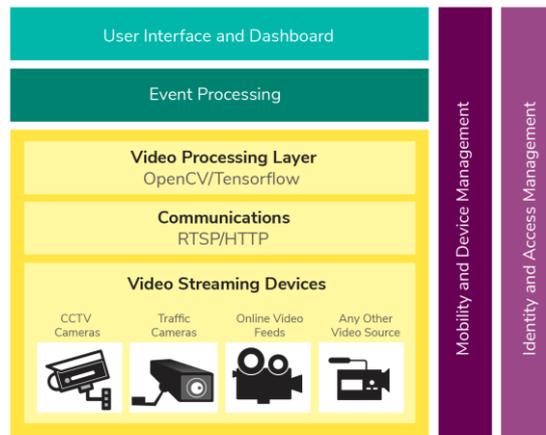
Figure 2.38. Reference architecture One. Adopted from Medium, 2022
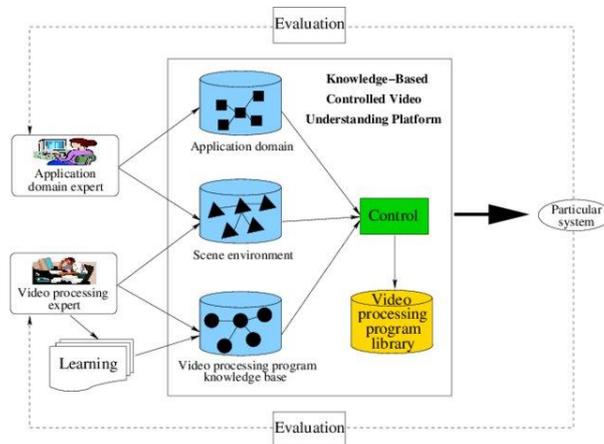
Figure 2.39. Reference architecture Two

The Figure 2.40. below shows a reference architecture (basic and conceptual schematics).
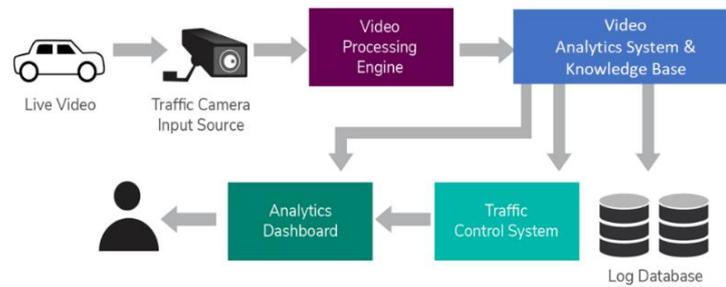
Figure 2.40. Reference architecture Three. Adopted from Medium, 2022

The reference architecture mentioned above can be used in many different use cases in the field of surveillance/security and traffic/transportation: vehicle detection, traffic control by monitoring vehicle density, vehicle tracking by reading vehicle license plates, pedestrian behavior monitoring at the crosswalk, phone usage monitoring while driving or for monitoring fast vehicles. Video feeds from traffic cameras can be used as inputs to a surveillance/security and traffic/transportation platform along with other available data, creating the basis for insight modelling that combines all available information, records events in a knowledge base, and learns from rules that prove useful.

The figure above explains how video processing can be implemented in a control system in the field of surveillance/security and traffic/transportation. Live video feeds taken from available traffic cameras can be used as input to the system. The video processing engine then processes the video, and the extracted data is sent to the database for processing. According to the processed data, the traffic management system will control the devices connected to it. If the flow processor detects an over speeding driver, it will report it directly to the appropriate monitor using the analytics dashboard server capabilities. The rules are further recorded in the knowledge base and forwarded to the end user, after which the violation is reported in a separate database.

2.56 Supply real-time sales data

Methodology

Define the schema for the data to be collected and stored in both the Cloud SQL and the Datastore NoSQL and make sure that the schema aligns with the data that will be collected.

Collect real-time data using Cloud Functions, Cloud Pub/Sub, and Dataflow from the various points of sales and send the data to the databases, making sure it conforms to the schema.

With the data transformed into a suitable format in the databases, use BigQuery to analyse the data in real-time by creating queries that can aggregate, transform, and visualize data from both Cloud SQL and Datastore NoSQL.

With the queries built to extract the data needed, set up real-time data pipelines in BigQuery to monitor and analyse the data streams as they come in.

Once the data has been analysed, use a visualization tool to create custom dashboards and reports on the data and present it to the users so that they can make decisions based on the insights provided.

Value

Processing data in real-time enables all parts of the value chain to see the status of operations without delay and make better-informed decisions that help avoid these problems.

Seeing in real-time the effects of promotions and exclusive offers lets you take action based on the results. Trying out a promotion in a subset of locations first and then act according to the results, either stopping the promotion or rolling it out to all other locations as well.

Getting an up-to-date picture of what products are in demand means you can better anticipate the need for restocking in time and not run the risk of running out of product or having to order unexpected shipments.

System architecture

Google Cloud SQL database to store all the structured and relational data gathered.

Datastore NoSQL database to store all the unstructured data constantly coming in from the data streams provided by the points of sales.

BigQuery for aggregating all the data that is being sent to the databases and transformed so that it can perform real-time analysis on the data, outputting formatted data that can be used to present visual data to the end user.

2.57 The graphic interface for programming at a car service combined with a website

Value

For diagnosis appointments, companies usually use a call center where car owners who want to make a car diagnosis must call and get an appointment.

This has some drawbacks:
- the phone number of the call center is often busy because all the operators are engaged in other conversations, and they have to either wait or return with the phone.
- in the dialogue with the operator from the call center there are some misunderstandings regarding the nature of the problem that needs to be solved.

By creating a website with a friendly graphic interface and a desktop application, the problem can be solved in a way that makes programming for diagnostics easier.

The website helps the client to schedule online, and the desktop application offers the client the possibility to schedule at the service headquarters. The graphic interfaces of the two applications have a page where the customer creates a customer account secured by a password.

After creating the customer account, the person interested in diagnosing the car can access the service scheduling page. On this page, the company displays the departments to which the client can go for diagnosis depending on the problem the car has.

Each department has a number of employees who are specialized in solving certain problems. The page provides the names of the employees, including their picture and details about their specialization.

The customer can opt for one or another specialist from the department who deals with fixing the problem the customer's car has. The respective page also shows the times when the client can be scheduled for the diagnosis.

The web application can be accessed at any time and the client can decide for himself the department and the specialist he wants to make an appointment with, based on the information provided by the application.

After creating the appointment, the application sends the client an email informing him about the details of the appointment. This avoids unpleasant situations in which the client schedules, but the schedule is erroneous for various reasons.

An information message is sent to the client with a certain time interval before the scheduled date to remind the client of the scheduled date and time. The desktop application is implemented on a local computer within the company and has the same programming pages as the online application.

This application offers the possibility to the client to make a new appointment if, after the diagnosis, it is found that the problem for which he came is more complex and he needs to call another department for the complete solution of the defect.

System architecture

The application has four elements: the website, the desktop application for the client, the desktop application for the application administrator, and the database that is accessed by the three applications.

The website also displays general data about the company, through buttons or links it directs the visitor to create a customer account and schedule. Also, the site contains a page through which the client has the opportunity to convey his appreciations or criticisms regarding the quality of the company's management performance.

The desktop application located on a local computer at the company headquarters can offer more data to the client because it is located inside the company and can be secured better.

The desktop application on the local computer can display data such as the bonuses granted by the company, the prices for the components used in the repair of defects, the prices of repairs, the stages of repair of each defect, etc. data that must not reach competing companies.

The desktop application reserved for the management of the unit contains dedicated pages through which the management can change the data included in the web and desktop applications reserved for customers. If a specialist is on vacation or is sick during a certain period, he will not appear in the client applications.

The timetable can be changed, or certain extreme situations can occur that cancel the appointment. This information is changed in the application through the desktop application reserved for the management of the unit.

Also, the management of the unit can access the assessments made by customers and in this way find out the level of customer satisfaction. The data accessed by clients are stored in a database located on a local server or a server in the cloud. The advantages of using a cloud server are that in case of unpleasant events at the company headquarters, the data on the cloud server is not affected.

Also, the company does not have to make the investment for the local server or hire staff to manage the server.

In principle, the customer account creation page on the two applications contains an electronic format in which the customer fills in the data and sets the password. The customer account can be the email address or a nickname that he chooses.

After creating the customer account, pressing a button or a link directs the customer to the page where he can program. Programming is done by choosing from several drop-down lists the compartment, the name of the specialist preferred by the client.

On this page there is a button through which the customer can access additional information that can help him make the correct schedule.

2.58 Video conference system

Methodology

The video conferencing process can be split into two steps: compression and transfer.

During compression, the camera and microphone capture analog audiovisual (AV) input. The data collected is in the form of continuous waves of frequencies and amplitudes. These represent the

captured sounds, colors, brightness, depth, and shades. Once captured, codecs convert data into digital packets, typically with compression to minimize bandwidth usage.

During the transfer phase, packets are sent over the network, typically to the cloud service provider, which then transmits them to other conference participants (and combines voice and video from multiple participants).

Once packets reach the endpoint, the codecs decompress the data. The codecs convert it back into analog audio and video. This enables the receiving screen and speakers to correctly view and hear the AV data.

Components of video conferencing systems

The components of a video conferencing system include the following:

- A network for data transfer, such as wired/wireless local area network, wide area network, cellular wireless and residential broadband.
- Two or more video cameras or webcams that provide video input.
- Two or more microphones -- either an external microphone or one built into the accessing device.
- A computer screen, monitor, TV, or projector that can broadcast video output.
- Headphones, laptop speakers or external speakers that can be used for audio output.
- Codecs, which can be hardware- or software-based, to reduce bandwidth by compressing and decompressing AV data. They typically include acoustic echo cancellation capabilities, which reduce audio delays to support real-time communication. Codecs may also include features like noise cancellation and acoustic fencing to minimize background noise during conferences.

Value

Video conference systems can deliver significant value to an organization by fostering communication, collaboration, and productivity across geographically dispersed teams. Here are some specific benefits of a video conference system:

- Increased productivity: Video conferencing systems can help to increase productivity by reducing travel time and allowing for more efficient meetings, leading to more time spent on core activities.
- Improved collaboration: Video conferencing systems provide a platform for real-time interaction and collaboration, regardless of physical location, improving teamwork and decision-making.
- Enhanced decision-making: Video conferencing systems can lead to better decision-making by facilitating faster and more effective communication among remote teams.
- Reduced costs: Video conferencing systems can enable significant cost savings by reducing travel and transportation expenses, including flights, accommodation, and transportation.

- Increased flexibility: Video conferencing systems provide greater flexibility and convenience, allowing participants to attend meetings from anywhere with an internet connection, even from home, increasing employee satisfaction.
- Environmental benefits: Video conferencing systems promote environmentally responsible business practices by reducing the carbon footprint associated with business travel.

Overall, video conferencing systems can deliver significant value to an organization by increasing productivity, enhancing collaboration and decision-making, reducing costs, increasing flexibility, and promoting environmental responsibility.

System Architecture

The architecture of a video conference system depends on the specific system being used. However, here are some common elements that may be included:
- User Interface: The video conference system usually includes a user interface that enables users to start and join meetings, share screens and documents, and manage participants.
- Video and audio transmission: The video conference system uses audio and video codecs to encode and decode audio and video signals for transmission between participants. The system may include gateways and firewalls to manage different network protocols, jitter and packet loss.
- Data sharing: The video conference system may include features for sharing files and documents during the meeting, such as screen sharing, virtual whiteboards, and co-authoring tools.
- Recording and Playback: The video conference system may include recording and playback.

2.59 VoD offering

Methodology

Set up the storage: AMS requires a storage account where you can store your video content. Make sure you configure the storage account to work with AMS so that it has the correct privileges to both read, list, and write data to and from the storage.

Upload your video content: Once your storage account is set up, you can upload your video content. You can do this through the AMS portal or through the REST API. Make sure you understand the supported video formats and codecs, as well as the best practices for encoding your video content.

Create an asset: In AMS, an asset represents a single video file or a collection of related video files. Create an asset for each video you want to publish. You can do this through the AMS portal or through the REST API.

Encode your video content: AMS uses encoding to convert your video files into different formats and bitrates. You can use Azure Media Encoder, Azure Media Encoder Premium, or third-party encoding

tools to do this. Make sure you configure the encoding settings based on your video content and the devices you want to support.

Publish your video content: After encoding, you need to publish your video content. You can do this through the AMS portal or through the REST API. When you publish your video content, AMS generates streaming URLs that you can use to play your video content.

Embed your video content: Finally, you can embed your video content in your website or application using the AMS player. The player supports a variety of features, including adaptive streaming, closed captioning, and multiple audio tracks. You can customize the player's look and feel to match your brand.

Value

The business can provide videos on their own website and lets them create a unique media player that has their own personal design using their own branding. They can provide videos such as product presentations or reviews, or instructional videos, either for internal use or for external users.

System architecture

The architecture for a VoD system is comprised of several Azure cloud services.
Azure Media Services is the core component of the system and is responsible for ingesting, processing, and delivering your video content. It includes numerous services, such as Encoding, Streaming, and Indexing, that work together to provide a comprehensive VoD solution.

AMS uses Azure Storage to store your video content and metadata. You can choose between various storage options, including Blob storage and Azure Files, depending on your requirements.

Azure Content Delivery Network (CDN) is used to distribute your video content to users worldwide. It helps reduce latency and improve the user experience by caching content closer to end-users.

AMS can be integrated with Azure Active Directory (AAD) to enable secure authentication and authorization of users. AAD provides robust identity and access management capabilities that can be used to control access to your video content.

The web application provides a user interface for your VoD solution. It can be used by administrators to manage video content, as well as by end-users to search, browse, and view videos. The web application can be hosted on Azure App Service or any other web hosting service.

The AMS Player is a client-side component that allows users to view your video content. It provides a variety of features, including adaptive streaming, closed captioning, and multiple audio tracks. The player can be embedded in your web application or used as a standalone player.

2.60 Water supply management using distance readers in water supply networks

Methodology

Water supply as an industry in general sees an increasing need for advanced analysis, controls, and automation to better manage various and numerous facilities operations. Advanced processes/programs can deploy and upgrade systems automation on various objects in different time intervals and especially when available financing that can lead to multiple vendor systems. In addition, business information technology systems for asset management, finance, geospatial data, and laboratory information are also necessary for overall management business environment in water supply.

By adding new, smart, measurement systems in water supply, the complexity also increases of such systems where, potentially, millions of data streams must be monitored and managed to increase billing efficiency, control water losses, and relieve pressure to the management of such systems. With a multitude of business and engineering solutions, quickly accessing, integrating, visualizing, and using the mentioned data in a meaningful way to make better business decisions, it can be very complex, difficult, and expensive. Having fast and simple, instant ("real time") access to current data and of active access to historical time series of data, enables providers decision making quality decisions based on factual data. In general, the industry usually calls the systems that enable insight into operational data kind of "warehouse", but modern tools have enabled developments in the last few years interactive solutions and visualizations and without a classic, structured data warehouse.

You tools use advanced control algorithms built into visualization tools. The most advanced of these systems are scalable and can manage and integrate operational and business data into one unit. On the other hand, "cloud" based services and infrastructure it enabled low-cost collection, availability, and scalability/capacity ready for various analyses. Practice shows that it is efficient to monitor data in two groups. Primary data refer to collection of raw data, usual measurement of water quality or water quantity. It may be measurements of currents or pipes collected by technicians or sensors (automated sampling technologies). Water quality refers to chemical, physical, and biological properties water. It is a measure of the state of the water, usually in relation to the requirements of some ecological process or anthropogenic purposes. Water quality is not a single measurement but a latent factor that relates on hundreds of water characteristics. On the other hand, the amount of water refers to the amount of water or to the speed of movement of the amount of water. Information on the amount of water is often linked to other aspects of water management such as water or fiscal value of water. Historical quantity records of water are more consistent than historical records of water quality.

Qualitative and quantitative measurements they evolve from time-intensive sampling methodologies that require technicians' special patterns to automated, more common processes performed by sensors. Today, for example, the Us the Geological Survey4 (USGS) has 1,908 sample locations across the country it measures water quality and transmits data at fixed intervals of 15 to 60 minutes with use automated recording equipment. Sensor measurements are useful because they increase significantly amount of data and are also available for water quality. The primary objective of quality and quantity

monitoring water is to mark variability. It is therefore important that individual measurements can be made appropriately compare time and space. This requires standardization of measurement procedures. Although they are each measurement something unique, specialized techniques and technologies are designed how would reduce variability and Favor comparability. Water is measured through many different methods and recorded in many different formats. For example, meter gauges are used to measure river depths, sensors are used to measure contents nitrogen, satellites are used to measure the volume of underground water.

Depending on water characteristics and resource availability, water is measured differently between regions. The central dogma is that the more data that can be collected, the more data can be collected quality of data-based secondary processes. Secondary data is information obtained by direct hydraulic measurements or sensors.

In contrast to the standardized measurements of primary data, secondary data are more adapt to specific circumstances and wishes. For example, models, receiving methods of primary data from various sources and compiling them to estimate another variable can use historical rainfall and stream flow gauges to predict availability of water for a specific municipality. Other models determined how the release of water from tank to provide habitat for fish. Models are often associated with regulatory and institutional framework, such as water rights, stream flow requirements and permitting for Environment. There is a much greater diversity among derived water data than primary data on leads. Some are small, effective for the local area. Others are larger, applicable to regions like the catchment area of the river. Attributes of secondary data depend on the sponsoring entity of the application, on the mission and priorities of the budget. Therefore, it is often easier to share primary data from secondary data.

Value

Today, in home-made construction with modest financial expenses, a prototype device with a function almost unimaginable can be made only a decade ago. If the price of the central platform, which can be performed with any standard computer, is neglected, less than EUR 40,00 was spent for the entire platform. Of course, the knowledge required to build the system and the time spent on development are incomparably greater.

During operation, most attention and time was spent on developing the sensor platform as the most critical part of the system. Ultimately, a completely autonomous and reliable sensor platform was successfully designed and built, which, in addition to its basic function, had to serve as an intermediary for adjusting the LoRa module and as an instrument for measuring range. To achieve the longest possible autonomy of the sensor platform, studying the current through the sensor platform, the consumption of microcontrollers and radio modules was gradually reduced to almost one fifth of the original consumption while waiting.

This is completely done by software shutting down individual components of the system at times when we do not need them and quickly turning them on when the need arises. The operation of the entire

system was finally tested using a signal generator, thus confirming the correct operation of all three parts of the system. To obtain confirmation that the developed system can meet the requirement for radio range, the distance was measured at 92 different points and the measurement covered an area of 21.5 km2. The radio reach distances obtained by measuring in this paper fully confirmed that LoRa is a more than acceptable solution for water flow control in houses where the owners do not live most of the year and are located within or on the wider periphery of the settlement. The obtained results show that such a prototype could be applied in practice even now, without major changes in implementation, only with the connection to the solder plate and placement in a suitable housing. Also, by using other types of sensors, the prototype can serve as a basis for collecting various other information from less frequently visited locations.

System Architecture

One of the common disasters that can befall an object that is rarely inhabited is the rupture of a water pipe. This phenomenon is mainly caused by the freezing of residual water in the elbows of pipes and valves, during the winter. As the temperature rises, if the main valve was not closed or it was damaged, water will leak. The aim of this research is to propose a system that will inform the owner or the person caring for such a facility that an adverse event has occurred. It must be considered that the electricity is mostly turned off at such facilities, and therefore the internet is not available either.

The system that will perform such a task must therefore have its own power supply with the longest possible life of autonomous operation and must not rely on becoming connected to the Internet from the facility itself. As a suitable solution for reporting unwanted water flow, the system is proposed in this paper. This system consists of three parts: (i) water flow sensor; (ii) LPWAN central transceiver; (iii) background system (backend). The water flow sensor is located on the building itself. Its role is to measure the flow of water through a pipe. If a flow occurs, then the sensor must report its size to the central transceiver. It must also report the moment when the water stopped flowing through the pipe. The sensor itself consists of a water flow meter, a microcontroller and a LoRa LPWAN module.

Depending on the type of water flow meter, it can be placed immediately behind the water meter or as in the case of this work in which a water meter of section R 1/2 '' was used, in a place like a garden tap. The microcontroller and the LoRa LPWAN module can be a few meters away from the measuring point, depending on the voltage drop on the connection cable between the microcontroller and the water flow meter. This circuit is completely autonomous. It can run on battery power for a long time and does not need a commercial network connection. The central transceiver, as well as the sensor platform, uses a LoRa LPWAN module that has the same communication parameters set as on the sensor platform. The receiver also consists of a microcontroller connected to an LPWAN transceiver. In this case, we use a type of microcontroller that has a built-in Wi-Fi 802.11n module. (Fig. 2.41.)

2021-1-SI01-KA220-VET-000034641
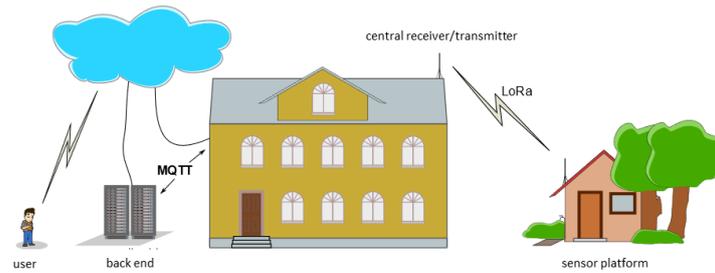
Figure 2.41. System architecture

This dual connection allows it to forward the messages it receives from the sensor via the LoRa LPWAN receiver to the background system via a local or public IP network. Like the sensor platform, the central transceiver can be powered by a battery, but it is recommended to connect it to a mains socket. The position of the central transceiver is not tied to the position of the background system, but it is important that the LoRa module on it and the LoRa module on the sensor platform are within radio range and that it is possible to connect to an IP router equipped with Wi-Fi receiver/Wi-Fi router. Using the MQTT protocol, the central transceiver will notify the background system of the occurrence or cessation of water flow at the sensor location. In addition to these two messages, the sensor also sends periodic messages.

They are important to confirm that everything is OK with the sensor. In the system proposed in this paper, periodic normal state messages are sent every 12 hours, while in the case of an alarm, messages are sent every 15 minutes. As a background system, one Raspberry Pi microcomputer was used. Of course, depending on the needs, it is possible to use any other Intel or Arm based computer, powered by Linux, Windows or MacOS operating system. The background system receives messages from the central transceiver via the MQTT intermediary. An intermediary program (middleware) is subscribed to the MQTT messages related to the described alarm system, which prepares the received messages in a suitable format and saves them in a time-oriented database (TSDB). As part of the background system, a web-based user interface has been added that allows you to see if the water flow sensor is in an alarm state and when it last responded to the central transceiver.

The same application can report a message to the end user via email or via the instant messaging system. For an example of this paper, the Telegram instant messaging system was used. The whole system is designed as a demonstration and only one water flow sensor and one central transceiver are used. By introducing the LoRaWAN communication protocol, which is a software upgrade of the existing system, and using a multi-channel central transceiver, the system can be expanded to several sensors that send data to multiple central transceivers. Also, the background system does not have to be on a single computer and can be deployed to multiple servers as needed. In this way, it is possible to build a very robust and flexible alert network that covers more widespread areas. (Fig. 2.42., 2.43.)

2021-1-SI01-KA220-VET-000034641

Figure 2.42. Central transceiver antenna position and measuring ranges
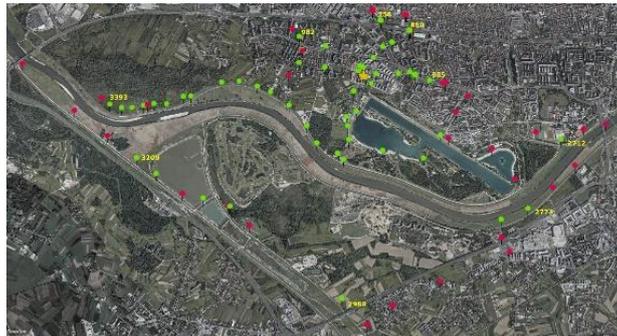
Figure 2.43. Central transceiver antenna position and measuring ranges

2.61 Webstore

Methodology

Decide on what kind of database is most suited for the use-case, some use-cases benefit more from using a relational database while in others, a document database would be a more appropriate.

Server technology must also be decided upon, whether one should use a Node – Express JavaScript stack or if PHP, Python, Java, or .NET would be best suited. This would depend a lot on the skillset of the developers building and maintaining the server.

Design data model, ensuring all relationships between inventory, orders and customers are considered.

Create a server to handle incoming requests from a front-end by building an API (Application Programming Interface) that will receive requests and then pass requests onto the Database, making any necessary posts, deletes or reads needed.

Value

Being able to keep track of inventory and a record of any orders in progress and previous orders are many. Having access to a concurrent and reliable inventory count can be used by inventory managers to minimize the risk of running out of products which can result in loss of sales, order backlogs and

dissatisfied customers. With accurate data about stock levels the managers can order in new stock as it gets purchased, and with historical data about previous sales, they can also anticipate peaks in the sales of certain items. For example, grocery stores stock up on certain items during the holidays because of expected peaks of sales for those items that normally have low sales numbers. E.g., turkeys.

Having an up-to-date customer database and a record of their order history lets you keep track of which items are popular with which customers. This can let you create personalized offers to increase their satisfaction with your business and increase customer retention. Having records of all sales to each customer can also help with customer support should anything go wrong, and customer support staff can easily see which items were ordered at a specific date.

System architecture

A possible (and simplistic) architecture for this webstore could look like this, with the Entity model diagram of the database (Fig. 2.44.).
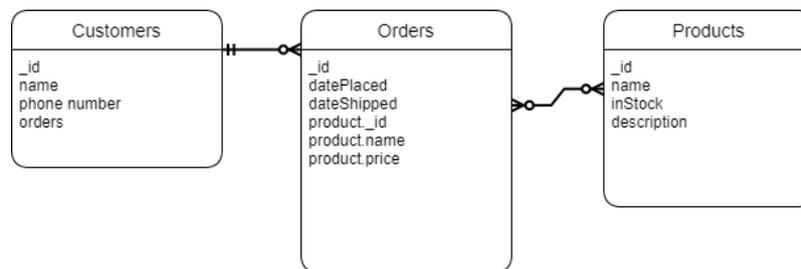
Figure 2.44. Entity model

And the sequence diagram of a user placing an order via an (already existing in this case) storefront (Fig. 2.45.).
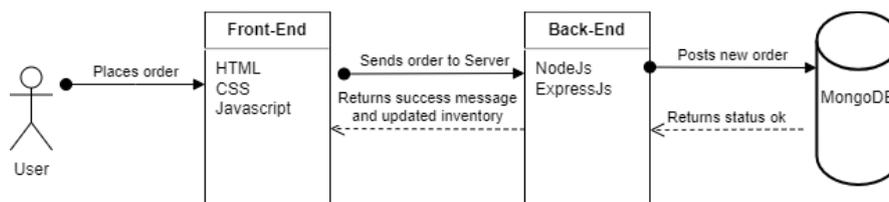
Figure 2.45. Sequence diagram of a user placing an order.

2.62 Web application for the online completion of a company's staff timesheet

Value

The daily monitoring of the activity of a work team is necessary in order to know the rhythm of the works in which the team is involved. For companies that carry out activities in the field of construction, it is specific that teams are deployed in several points in a geographical area.

Among these data is the daily duration of the work schedule for each member of the team. The classic option is to complete a word or excel file by the team leader and send it by email to the company headquarters. It is a variant that requires a high volume of work for the personnel who centralize the data.

The application gives the possibility to directly fill in an electronic form from the web page of the daily attendance for each person in the team. The data from this form are retrieved in a database. The application has the advantage that it allows the creation for each member of the team of a table in the database containing the data regarding the activity carried out over a period. This table can be correlated with other data.

A desktop application allows access to the personnel database that must centralize the data for the entire team in order to calculate the salary for each team member.

The application also allows the creation of reports requested by the company's management in order to analyse the efficiency of the team's activity. In addition, errors that occur when transcribing data from files received by email into the central file are avoided and time is saved.

Application architecture

The application is written in HTML, Java script and PHP. It contains a password-based access page that the team leader (teams) accesses.

In an electronic form, the names of the team members and the date for which the attendance is completed are displayed.

The data entered in the electronic form completed by the team leader are sent to a database.
The application has several pages that are accessed by personnel from the HR department and that allow a centralization of the data entered in the database and the creation of the various reports necessary for the analysis of the activity of the team (teams).

As with any software application, the four stages are completed.
* Analysis of the application that involves establishing the table columns from the electronic form that appears on the page accessed by the team leader.
* Establishing the tables that make up the database and creating the database.
* The data belonging to the team members are entered in the database.
* Writing the application
* HTML, JavaScript and PHP languages are used in writing the application. The PHP language manages the functions that access the database, and the Java script language makes the connection between HTML and the PHP language.
* Application testing
* It is the final stage in which it is checked if there were no errors in writing the code. At this stage, the code can be modified if it is found that the application is not working correctly.

2.63 Website hosting with static content

Methodology

The five stages of the Web development life cycle:

1. Plan
2. Design
3. Develop
4. Test
5. Deploy

Figure 2.46. 5 stages of web application development

Note that in the infographics they have used the term Research and Analysis instead of planning.
In addition, some sources add two more steps, Research before planning, and Maintenance after Deployment. (Fig. 2.46.)

Plan

This stage is used to decide between a lot of the existing possibilities for what the website should be and how to develop it. These decisions should include, but not be limited to, things such as what technology stack should be utilized, (e.g., a **MERN** (MongoDB, Express, React, Node) or **LAMP** (Linux, Apache, MySQL, PHP/Perl/Python) stack), creating wireframes and initial design sketches, the structure of the content and estimating the timelines for the various stages in the project.

Design

Creating and laying out the design of the website. This includes designing the UI/UX elements as well as the visual aesthetic by deciding the typography, color pallet and the design of presentation of the content itself such as the menus, and buttons used on the site.

2021-1-SI01-KA220-VET-000034641

Co-funded by
the European Union

Development

The development stage is where the website is built from the decisions made in the design stage. Here the development of the front-end and back-end parts of the website will be created, allowing users to interact with the website and the business to create, store and provide information and functionality to the users.

The two parts can be roughly separated into the user-facing website that the user sees when they access the website (The front-end) and the 'under-the-hood' parts with the communication, between, for example, a server and a database (The back end)

Test

During the testing stage the website and all its functionality is tested to make sure it is ready to be used by the end-users.

This is done by performing several types of tests:

Unit testing, making sure all parts of the website work as intended.

Stress testing, running the system over capacity to determine the absolute maximum stress the servers can handle.

Integration testing, making sure all parts of the website can communicate with each other correctly and providing the expected results.

Load testing, testing the system performance when it is put under the kind of load is expected at any given point. E.g., 200 users accessing the website at the same time.

Deploy

After testing has been completed has been completed and all tests have passed it is time to deploy the website. The code for the website is uploaded onto the chosen hosting and app platform provider.

After the website has been deployed it is necessary to keep maintaining the site, making changes based on user feedback and fixing bugs that managed to go unnoticed during the previous stages. This is also why many people add Maintenance as a final stage for the lifecycle.

Value

The value of having a website for a business, whether this business is for a one-person company or a large organization, is manifold. The most immediate and obvious is allowing your business to have a constant presence in the world, allowing users and customers to have 24/7, global access to your online

space, driving engagement and increasing the reach of your products or services, which in turn drives sales. Costs can also be reduced as you can provide information on your website about commonly asked questions such as, an address or phone number to a store location or information about product or service offerings which might have had to be handled by someone in the business taking away from performing other duties.

In addition, being without a website puts you at a significant disadvantage to competitors who will most likely have their own online presence.

Beyond these immediate benefits, there are also other values that can be gathered over a longer period by monitoring the data traffic to your website. With this data you can get insights into the geographic locations of your customer bases, what kind of products they are interested in and the demographic you are hitting. With this information you can more accurately tailor your experiences to the needs and wants of your customers. For example, by offering personalized discounts and offers based on their engagement on your website.

System architecture

There are a lot of options for how one could host a static website. All the big cloud providers have their own solutions offering different solutions to achieve the same goal. In addition to the "big three" there are also other providers who specialize more in web hosting such as Netlify, Digital Ocean or Vercel (Fig. 2.47.).
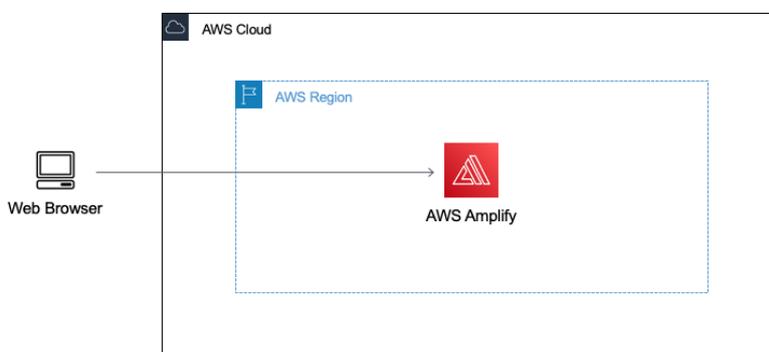
Figure 2.47. Static website source code

- The static website's source code is stored in a GitHub repository.
- AWS Amplify connects to the GitHub repository and automatically builds and deploys the website when changes are made to the repository.
- The built website is stored in an Amazon S3 bucket set up by Amplify.
- Amplify also uses Amazon CloudFront to cache the website's static content and deliver it to users from the edge location closest to them.
- When a user visits the website, the request goes through CloudFront to the S3 bucket, which retrieves the website's static files and sends them back to the user's browser.

REFERENCE

1. Blog DTL. (2018). AWS Insight: The Role of the Cloud in Powering U.S. Elections. Retrieved from: https://www.dlt.com/blog/2018/10/31/aws-insight-role-cloud-powering-u-s-elections

2. D1 awsstatic. (2023). Digital Transformation and IT Modernization for Elections in AWS. Retrieved from: https://d1.awsstatic.com/whitepapers/digital-transformation-and-it-modernization-for-elections-in-aws.pdf

3. Executivegov. (2022). AWS, Nonprofit Partner to Offer Midterm Campaign Cybersecurity Services. Retrieved from: https://executivegov.com/2022/09/aws-nonprofit-partner-to-offer-midterm-campaign-cybersecurity-services/

4. Gupta M. (2018). Blockchain for dummies. 2nd IBM limited edition. Retrieved from: https://www.ibm.com/downloads/cas/36KBMBOG

5. Medium. (2023). Use AWS to compare Inauguration speeches of Obama and Trump. Retrieved from: https://medium.com/@szekelygergoo/use-aws-to-compare-inauguration-speeches-of-obama-and-trump-670068ea39d5

6. Mostafa, S. (2023). Host a Dynamic Website on AWS. Retrieved from: https://www.linkedin.com/pulse/host-dynamic-website-aws-sara-mostafa/

7. Pointdev. (2022). Windows System Management Software. Retrieved from: https://www.pointdev.com/en/ideal-administration/index.php?gclid=Cj0KCQjw27mhBhC9ARIsAIFsETGJ_Xx6pog0rUrsSiKl3AQ-cqGUI0DZWC9Err8SnsR9oyZvWvUtFGwaAl-wEALw_wcB

8. Windley, P. (2018). Defining digital identity. Retrieved from: https://www.windley.com/archives/2023/01/defining_digital_identity.shtml

2021-1-SI01-KA220-VET-000034641

APPENDIX

Appendix 1: Code Snippets

Usecase: Chatbot for students in EDU institution

Importance of Natural Language Understanding (NLU) cannot be highlighted enough but that is the main reason why this thesis is even considering to be enrolled. From technology perspective Microsoft has really great service to offer. Language Understanding Service (LUIS) is one of the best NLU solution on the market. However, every Microsoft service that is somehow in relation with NLU is coupled to LUIS in the background. With LUIS it is easy to add language understanding to any app. It is designed to identify valuable information in conversations, LUIS interprets user goals (intents) and distils valuable information from sentences (entities), for a high quality, nuanced language model. LUIS integrates seamlessly with the Azure Bot Service, making it easy to create a sophisticated bot (Fig. 3.1.).

```json
{
    "query": "Book me a flight to Cairo",
    "topScoringIntent": {
        "intent": "BookFlight",
        "score": 0.9887482
    },
    "intents": [
        {
            "intent": "BookFlight",
            "score": 0.9887482
        },
        {
            "intent": "None",
            "score": 0.04272597
        },
    ],
    "entities": [
        {
            "entity": "cairo",
            "type": "Location",
            "startIndex": 20,
            "endIndex": 24,
            "score": 0.956781447
        }
    ]
}
```

Figure 0.1. LUIS in Action

For example, for query like "Book me a flight to Cario", LUIS is capable to turn results in a JSON form, where could be found valuable information like BookFlight as an Intent with 98 % of accuracy and entities like Cairo as a Location entity with 95 % of accuracy. Even if Bots and NLU are pretty mature technologies there are still possibilities to some Students questions stays unanswered or misunderstood. These situations should be well treated, and students should have another possible option to fulfil their request. One of common approaches for that situation are quick replies. Quick replies are small buttons or menus which have already prepared and predicted possible questions which can be written but also chosen by pressing the right predicted question (Fig. 3.2).
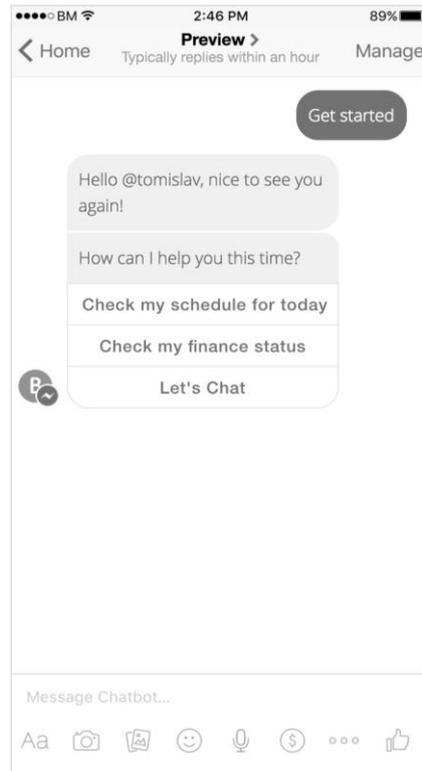
Figure 0.2. Quick in replies

Other possible solution is to offer to chat or call Student Office Desk Staff directly but that should be only in rear cases. The main idea of Student Service Support Chatbot is to decrease amount of student calls to minimum.

Usecase: Digital asset certification using distributed ledger/blockchain.

Application Modules

This type of application is intended for private blockchain. This means that each educational institution should have its own stream that only the people in the institution have the authority to store the diploma. All streams are stored in the main book that is distributed to all nodes, ie educational institutions in this example. The more nodes in the chain, the better, because the chain becomes ever stronger and safer.

The application consists of three modules:
1.   Module for entering a diploma.
2.   Diploma check module.
3.   Diploma printing module.

The first module is for entering a diploma. It switches the entered data into a hexadecimal form and stores them in the chain and returns the transaction ID (txid) back. Transaction ID is a private key that is awarded to a graduate student because it can be used to check the diploma data in the chain. The Diploma Check Module, combined with the OIB and Transaction ID, sends a query to the chain and

verifies whether there is a record in the chain. Thereafter, it gives a positive or negative answer, depending on whether there is a really required degree in the chain and whether it complies with the OIB entered. The diploma printing module prints a diploma on the screen in PDF format. All of the modules listed in this example are displayed in the command-line text interface, i.e., in the Ubuntu Operating System Terminal. They can also be programmed into a web application and used in WEB browsers.

User Roles

After the student successfully completes the faculty and defends his graduate thesis, the faculty system reports that the student has graduated. With this application and the module for entering the diploma, an authorized person at the university will enter the name, last name and OIB graduate student and this information will be stored in the chain. As feedback, he receives a Transaction ID which gives the student and enrolls on the original print diploma. It can also be printed in the form of a bar code whose scan is the value of the Transaction ID (Fig. 3.3.).

Figure 0.3. Showing the module for entering the diploma

The student gets his deserved diploma and his private key diploma, which in this case is 80bbfd9b068259c1f02a72b7196417c5464c54a4b68cfaf6e824777e268ff747. He then reports for a job and after a call from the employer goes to the job interview. The employer asks for a degree to check his qualifications. The procedure is currently being conducted so that the employer contacts the educational institution to verify the validity of the diploma, most often in writing. This process is long-lasting and consumes a lot of resources. But in this case, the employer gets a diploma with a private key. The employer then appoints the OIB person applying for a job and the public key in the application. This way in a fraction of a second returns the information on the validity of the diploma (Fig. 3.4.).

```
Molim Vas odaberite opciju:
1. Unos novog dokumenta
2. Provjera postojanja dokumenta
Odaberite broj [1,2]? 2
OIB:
68816974905
Ključ:
80bbfd9b068259c1f02a72b7196417c5464c54a4b68cfaf6e824777e268ff747
Goran Fijačko diplomirao na Visokom učilištu Algebra, smjer Multimedija, 15.10.2018. u Zagrebu.
Prikazati diplomu?
1. Da
2. Ne
Odaberite broj [1,2]? 1
Diploma će se prikazati u PDF-u!
```

Figure 0.4. Show the diploma verification module

After the application's confirmation is answered, the screen prints. The name and surname of the student, educational institution, orientation, date, and place of graduation are written in the print. The employer eventually has the option of printing a copy of the diploma for his own archive. If you choose a print option, the diploma will be generated and opened in PDF format.

For ease of use of the application after release to production, it is a better choice to use it as a WEB application. This means that everything shown will be moved to a web server and the application will access the https protocol (e.g., via URL https://www.diplome.hr) in web browsers. This means that users only need an internet connection and an account in the application to check the validity of the diploma quickly and securely.

Usecase: Remote-controlled smart devices in smart home

In order to interpret the effect of the ambient conditions of the stores to customer behavior we can use IoT sensors to measure brightness, temperature and humidity and determine/control their influence on the customer basket. This involves determining thresholds for unfavorable brightness, unpleasant temperature and inadequate humidity levels. The technological solution should be deployed in the form of a decision support system that can analyse the mutual relationships between IoT collected data, specific product groups and overall transactions in the store. Part of the decision support system should be able to control technical conditions in automated manner through interoperable interface embedded into existing air-conditioning systems. Since ambiental conditions are usually not equal in the entire store because some products can require different conditions (e.g., frozen food has different acceptable ambiental temperature range than other food), we should include into analytical data sets the store zone that identify particular area requiring specific environmental conditions.

The proposed data points are divided into two granularity levels: shop visit and product bought. The data points are collected from existing transactional databases and the data store containing IoT sensors real-time data.

Transactional data source tables are as follows in Figure from 3.5 to 3.10.

Environment

Field Name	Data Type	Description (Optional)
Time	AutoNumber	Time stamp (granulation level is arbitrary)
StoreAreaID	Number	The area of the store
Temperature	Number	Mean temperature of the store area
Brightness	Number	Mean brightness of the store area
Humidity	Number	Mean humidity of the store area

Figure 0.5. Environment

Transactions

Field Name	Data Type	Description (Optional)
VisitID	AutoNumber	
ProductID	Short Text	Bought product during the visit
Quantity	Number	Quantity of the bought product

Figure 0.6. Transactions

StoreAreas

Field Name	Data Type	Description (Optional)
StoreAreaID	AutoNumber	
StoreAreaName	Short Text	The name of the area of the store

Figure 0.7. StoreAreas

Products

Field Name	Data Type	Description (Optional)
ProductID	AutoNumber	
ProductName	Short Text	The name of the product
ProductCategory	Short Text	The category of the product
ProductSubcategory	Short Text	The subcategory of the product
ProductWeight	Number	The weight of the one unit of the product
StoreAreaID	Short Text	The area in which the product is located in the store
Price	Number	Price of one unit of the product

Figure 0.8. Products

Visits

Field Name	Data Type	Description (Optional)
VisitID	AutoNumber	
EnteringTime	Date/Time	Time when customer entered the store
ExitingTime	Date/Time	Time when customer arrived to the cash register

Figure 0.9. Visits

ETL relations
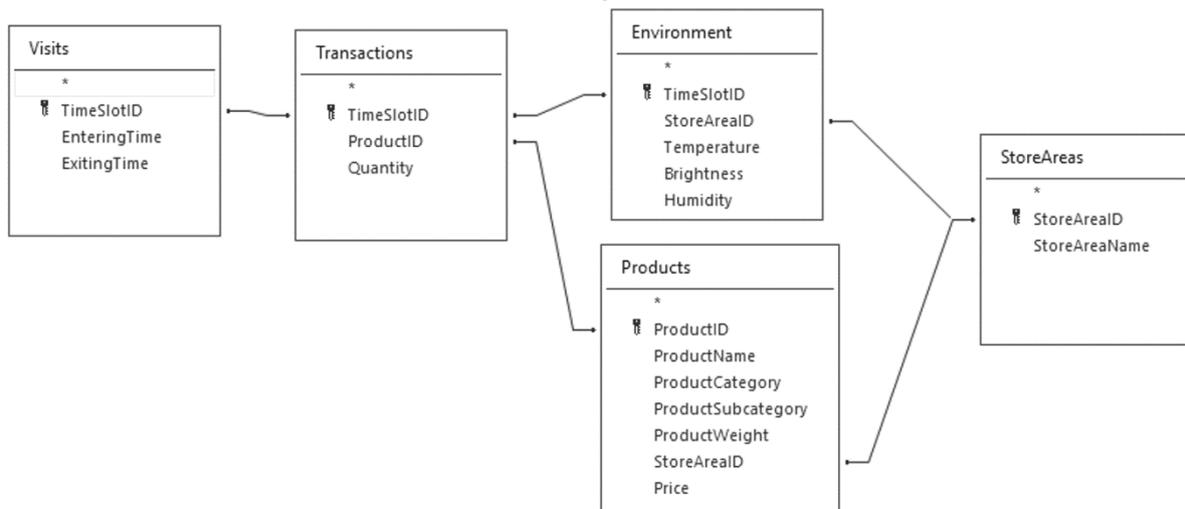
Figure 0.10. ETL relations

The variables available to analyse shop visits after applying ETL procedure are (Fig. 3.11.):

Field Name	Data Type	Description (Optional)
TimeSlotID	AutoNumber	
EnteringTime	Date/Time	
ExitingTime	Date/Time	
ProductID	Number	
Quantity	Number	
Environment_StoreAreaID	Number	
Temperature	Number	
Brightness	Number	
Humidity	Number	
ProductName	Short Text	
ProductCategory	Short Text	
ProductSubcategory	Short Text	
ProductWeight	Number	
Products_StoreAreaID	Number	
Price	Number	
StoreAreaName	Short Text	

Figure 0.11. The variables available to analyse shop visits after applying ETL

As target variables for machine learning we can now derive:

- Number of items (N)– number of different products purchased by a customer in one store visit (i.e., number of items in shopping basket).
- Weight of purchases (W) – weight of all products purchased by a customer in one store visit.
- Quantity of items (Q) – quantity of items of all products (summed across all types of products) purchased by a customer in one store visit.

Other group of possible target variables – retail business indicators - are described separately in the next section.

Usecase: Automation of tasks using cloud based services

To demonstrate how to carry out an MBA programming language R was used, and the arules package, along with some code included as a proof-of-concept. The example used is available at arulesViz Vignette and use a data set of grocery sales that contains 9,835 individual transactions with 169 items. First step was to look at the items in the transactions and, in particular, plot the relative frequency of the 25 most frequent items. This is equivalent to the support of these items where each itemset contains

2021-1-SI01-KA220-VET-000034641

only the single item. The bar plot illustrates the groceries that are frequently bought at this store, and it is notable that the support of even the most frequent items is relatively low (for example, the most frequent item occurs in only around 2.5% of transactions). These insights were used to inform the minimum threshold when running the Apriori algorithm; for example, we know that in order for the algorithm to return a reasonable number of rules we'll need to set the support threshold at well below 0.025. (Fig. 3.12.)
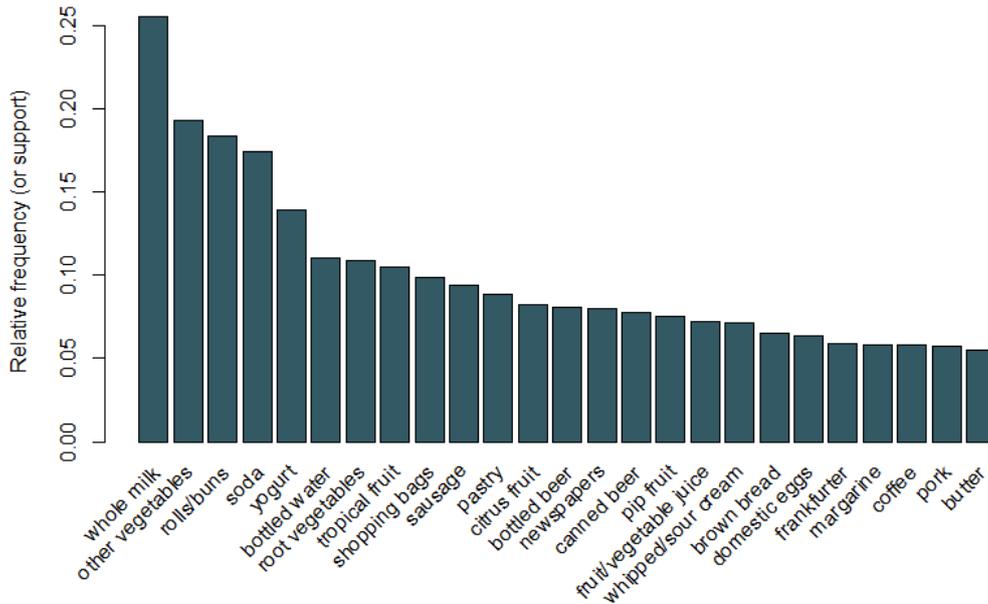
Figure 0.12. Bar plot of the support of the 25 most frequent items bought

By setting a support threshold of 0.001 and confidence of 0.5, we can run the Apriori algorithm and obtain a set of 5,668 results. These threshold values are chosen so that the number of rules returned is high, but this number would reduce if we increased either threshold or support. Experimenting is recommended with these thresholds to obtain the most appropriate values. Whilst there are too many rules to be able to look at them all individually, we can look at the five rules with the largest lift (Table 2).

Table 0.1. Five rules with the largest lift

Rule	Support	Confidence	Lift
{instant food products, soda}=>{hamburger meat}	0.001	0.632	19.00
{soda, popcorn}=>{salty snacks}	0.001	0.632	16.70
{flour, baking powder}=>{sugar}	0.001	0.556	16.41
{ham, processed cheese}=>{white bread}	0.002	0.633	15.05
{whole milk, instant food products}=>{hamburger meat}	0.002	0.500	15.04

2021-1-SI01-KA220-VET-000034641

These rules seem to make intuitive sense. For example, the first rule might represent the sort of items purchased for a BBQ, the second for a movie night and the third for baking. Rather than using the thresholds to reduce the rules down to a smaller set, it is usual for a larger set of rules to be returned so that there is a greater chance of generating relevant rules. Alternatively, we can use visualization techniques to inspect the set of rules returned and identify those that are likely to be useful. Using the **arulesViz** package, rules by confidence, support and lift are plotted. This plot illustrates the relationship between the different metrics. The optimal rules are those that lie on what's known as "support-confidence boundary". Essentially, they lie on the right-hand border of the plot where either support, confidence or both are maximized. The plot function in the *arulesViz* package has a useful interactive function that allows you to select individual rules (by clicking on the associated data point), which means the rules on the border can be easily identified (Fig. 3.13.).
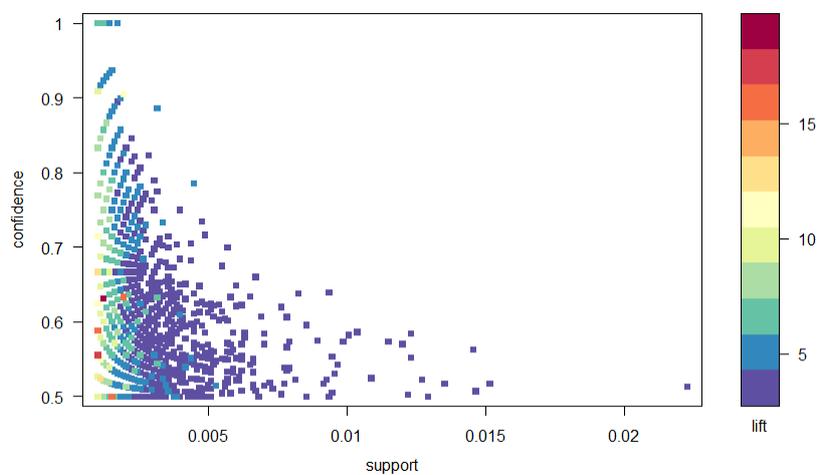
Figure 0.13. A scatter plot of the confidence, support, and lift metrics

There are lots of other plots available to visualize the rules, but one other figure that we would recommend exploring is the graph-based visualization of the top ten rules in terms of lift (more than ten rules can be included, but these types of graphs can easily get cluttered). In this graph the items grouped around a circle represent an itemset and the arrows indicate the relationship in rules. For example, the purchase of sugar is associated with purchases of flour and baking powder. The size of the circle represents the level of confidence associated with the rule and the color the level of lift (the larger the circle and the darker the grey the better).
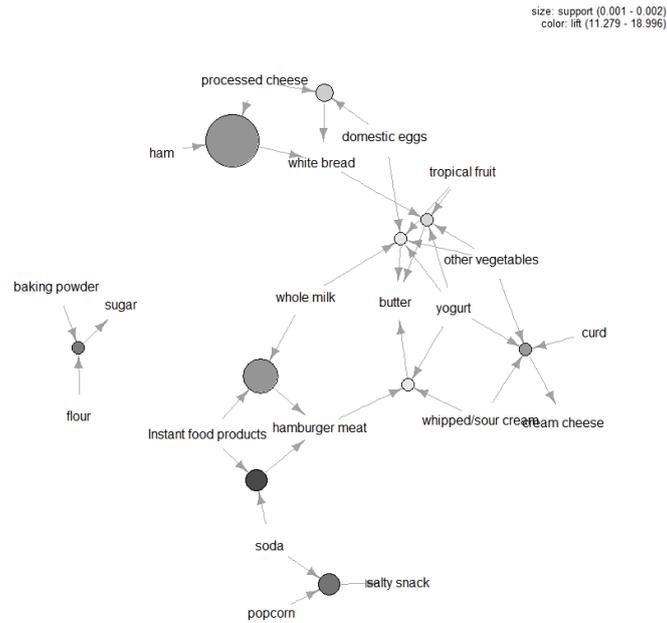
Figure 0.14. Graph-based visualisation of the top ten rules in terms of lift

Market Basket Analysis is a useful tool for retailers who want to better understand the relationships between the products that people buy. There are many tools that can be applied when carrying out MBA and the trickiest aspects to the analysis are setting the confidence and support thresholds in the Apriori algorithm and identifying which rules are worth pursuing. Typically, the latter is done by measuring the rules in terms of metrics that summarize how interesting they are, using visualization techniques and also more formal multivariate statistics. Ultimately the key to MBA is to extract value from your transaction data by building up an understanding of the needs of your consumers. This type of information is invaluable if you are interested in marketing activities such as cross-selling or targeted campaigns.

R Code

```
library("arules")
library("arulesViz")
 #Load data set:
data("Groceries")
summary (Groceries)
#Look at data:
inspect (Groceries [1])
LIST(Groceries) [1]
 #Calculate rules using apriori algorithm and specifying support and confidence thresholds:
rules = apriori (Groceries, parameter=list (support=0.001, confidence=0.5))
#Inspect the top 5 rules in terms of lift:
inspect (head(sort(rules, by ="lift"),5))
#Plot a frequency plot:
itemFrequencyPlot(Groceries, topN = 25)
 #Scatter plot of rules:
```

```
library("RColorBrewer")
plot(rules,control=list(col=brewer.pal(11,"Spectral")),main="")
#Rules with high lift typically have low support.
#The most interesting rules reside on the support/confidence border which can be clearly seen in this plot.
 #Plot graph-based visualisation:
subrules2 <- head(sort(rules, by="lift"), 10)
plot(subrules2, method="graph",control=list(type="items",main=""))
```

Usecase: Water supply management using distance readers in water supply networks

The LoRa protocol is a modulation of wireless data transmission based on existing Chirp Spread Spectrum (CSS) technology. With its characteristics, it belongs to the group of low power consumption and large coverage area (LPWAN) protocols. Looking at the OSI model, it belongs to the first, physical layer. The history of the LoRa protocol begins with the French company Cycleo, whose founders created a new physical layer of radio transmission based on the existing CSS modulation. Their goal was to provide wireless data exchange for water meters, electricity, and gas meters. In 2012, Semtech acquired Cycleo and developed chips for client and access devices. Although CSS modulation had hitherto been applied to military radars and satellite communications, LoRa had simplified its application, eliminating the need for precise synchronization, with the introduction of a very simple way of encoding and decoding signals. In this way, the price of chips became acceptable for widespread use. LoRa uses unlicensed frequency spectrum for its work, which means that its use does not require the approval or lease of a concession from the regulator. These two factors, low cost and free use, have made this protocol extremely popular in a short period of time.

The EBYTE E32 (868T20D) module was used to create the project. The module is based on the Semtech SX1276 chip. The maximum output power of the module is 100 mW, and the manufacturer has declared a range of up to 3 km using a 5dBi antenna without obstacles, at a transfer rate of 2.4 kbps. This module does not have an integrated LoRaWAN protocol but is designed for direct communication (P2P). If it is to be used for LoRaWAN, then the protocol needs to be implemented on a microcontroller. Communication between the module and the microcontroller is realized through the UART interface (serial port) and two control terminals which are used to determine the state of operation of the module. The module will return feedback via the AUX statement.

LoRaWAN is a software protocol based on the LoRa protocol. Unlike the patent bound LoRa transmission protocol, LoRaWAN is an open industry standard operated by the nonprofit LoRa Alliance. The protocol uses an unlicensed ISM area (Industry, Science and Medicine) for its work. In Europe, LoRaWAN uses the ISM part of the spectrum that covers the range between 863 - 870 MHz [4]. This range is divided into 15 channels of different widths. For a device to be LoRaWAN compatible, it must be able to use at least the first five channels of 125 kHz and support transmission speeds of 0.3 to 5 kbps. Due to the protection against frequency congestion, the operating cycle of the LoRaWAN device is very low and the transmission time must not exceed 1% of the total operation of the device.

In addition to defining the type of devices and the way they communicate via messages, the LoRaWAN protocol also defines the appearance of the network itself [5]. It consists of end devices, usually various types of sensors in combination with LoRaWAN devices. The sensors appear to central transceivers or concentrators. One sensor can respond to multiple hubs which improves the resilience and range of the network. Hubs are networked to servers that process incoming messages. One of the tasks of the server is to recognize multiple received messages and remove them. Central transceivers must be able to receive many messages using multi-channel radio transceivers and adaptive mode, adapting to the capabilities of the end device. The security of the LoRaWAN network is ensured by authorizing the sensor to the central transceiver, and messages can be encrypted between the sensor and the application server via AES encryption.

MQTT is a simple messaging protocol. It is in the application layer of the TCP / IP model (5-7 OSI models). It was originally designed for messaging in M2M systems (direct messaging between machines). Its main advantage is the small need for network and computer resources. For these reasons, it has become one of the primary protocols in the IoT world. This protocol is based on the principle of subscriptions to messages and their publication through intermediaries. An intermediary, commonly called a broker, is a server that receives and distributes messages to clients who may be publishers of messages or may be subscribed to them to receive them. The two clients will never communicate with each other.

The most important segment of the sensor platform is its reliability. To make sure that an accident occurs in time, we must first ensure the reliability of the platform. Precisely for this reason, in the solution proposed in this paper, periodic reporting from the sensor platform to the system is set. The device will report periodically every 12 hours, and this is taken care of by the alarm system on the microcontroller. Namely, STM32F411 is equipped with a clock that monitors real time (RTC) and offers the ability to set two independent alarms. In this case, one of them oversees waking up the process that sends periodic messages with the current state of the measured water flow through the meter.

Before the software implementation of the measurement, it should be noted that the pulse given by the sensor at its output voltage is 5 V. Although the used microcontroller will tolerate this voltage at its input, it is better to lower it to the declared input value of 3.3 V. Such voltage is obtained by two resistors, one with a value of 10 kΩ and the other of 22 kΩ, connected in a simple voltage divider [9]. The connection method is clearly shown in the diagram. The flow volume measurement itself is done by monitoring the number of pulses sent by the water sensor via a standard time counter. Each pulse will be registered by the microcontroller as an interrupt. When pulses appear, it is possible to measure the flow and report it via LoRa radio transmission.

The frequency of the timer is set to 1 MHz via a divider. By comparing the number of clock cycles between the two interrupts, one can very easily obtain the pulse frequency given by the water flow sensor. Knowing the pulse frequency and pulse characteristic, the water flow can be calculated using pre-defined procedure.

The first measured flow value greater than zero sets the sensor platform to an alarm state. As long as there is a flow, periodic advertising will take place every 15 minutes instead of every 12 hours. Five

minutes after the flow stops, the device will sound the end of the alarm, and the next call will be made regularly after 12 hours or earlier in the event of a new alarm. The alarm system works internally in such a way that the last measured value of the water flow is read every 5 seconds. This value, together with the current counter time, is continuously stored by the measurement process in the form of a time and flow structure. The read value is stored in a field the size of three elements. If after three readings all three elements in the field are equal, it can be determined that there was no flow in the last 15 seconds and the device exits the alarm state. The system waits another five minutes before announcing the end of the alarm over the LoRa connection. If the flow occurs again within these five minutes, the system will act as if the alarm has not stopped, that is, it will send a flow message after 15 minutes. (Fig. 3.15.)
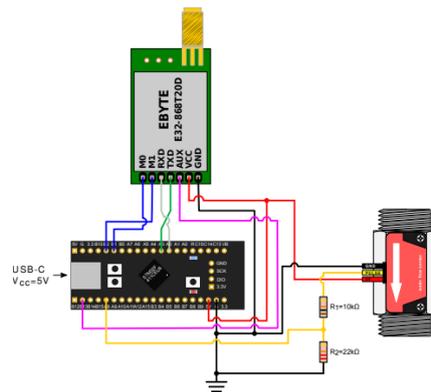
Figure 0.15. Water flow sensor connection diagram

LoRa notifications are intentionally delayed so that in the event of a constant occurrence and interruption of the flow, they would not often send radio messages.

Real life experience

During the measurement, the circuit is supplied with 5 V DC. This is the recommended operating voltage for the LoRa module and water flow sensor used, while the microcontroller can be powered by 5 V or 3.3 V. In this measurement, the first goal is to show that the peak current value will not reach a value greater than 300 mA, which is the maximum that the microcontroller circuit can withstand. This data allows us to power the entire circuit through the microcontroller using the built-in USB port and thus simplify the appearance of the entire sensor. The second goal is to reduce power consumption to prolong the autonomy of the sensor operation as much as possible. As an external power supply, a laboratory power supply R-SPS3010 from Nice-power was used, which can provide a stable operating voltage from 0 to 30 V with a current of up to 10 A. The universal measuring instrument UT139B from UNI-T is connected in series. It is set to measure milliamperes during the measurement, keeping the maximum measured value on the screen.

Range measurement

The range was measured from the Zagreb settlement of Vrbani 3, which is located next to Lake Jarun. This location gives us an insight into what range can be expected in urban and what in rural conditions. Namely, from the central transceiver to the north there is a very urban part with many residential buildings and dense traffic infrastructure, while on the south side is Lake Jarun and the Sava River, which are mostly green areas, smaller forests, and only a few lower buildings. The limiting factor is the position

of the antenna of the central transceiver, which was located on the first floor of a residential building, approximately 4 m above ground level and surrounded by buildings. When measuring on the side of the central transceiver, an omnidirectional antenna with a gain of 3.5 dBi was used, which is stationary placed on the outside of the window of a residential building. On the sensor side, for mobility, a smaller antenna with 2 dBi gain was used. The signal was sent in the open "out of hand". The position of each measurement was recorded via a GPS device on a mobile device and later transferred to Google Earth.

In Google Earth, it is possible to import recorded measuring points and measure the distance between them and the antenna of the central transceiver. According to the manufacturer's specification, the maximum range that can be expected from these modules is 3 km in almost ideal conditions with a 5 dBi antenna. To somehow approach this distance despite the unfavourable measurement position, the data transfer rate was reduced from the standard module settings from 2.4 kbps to 300 bps. Due to the small amount of data that needs to be transmitted, this is not a limiting factor in practice, and due to the low transmission speed, a smaller number of errors was obtained when recognizing the received signal and increased success in receiving messages over long distances. In figure below the measured range of the fabricated LoRa system is shown. The position of the central transceiver is shown with an asterisk, while the points from which the signal from the sensor managed to reach it are shown in green. Red dots indicate places where it was not possible to communicate between the sensor and the central transceiver. As expected, the largest range of 3393 m was achieved to the southeast, where apart from a couple of residential buildings near the antenna, there were no additional obstacles. Towards the southwest, the obtained result was 2773 m. However, according to the urban part of the city, the maximum achieved range was 982 m to the east, and to the north it was only 860 m. (Fig. 3.16. and 3.17.)

Figure 0.16. Central transceiver antenna position and measuring range
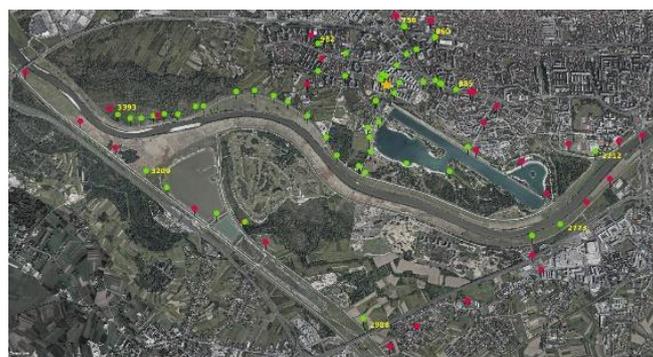
Figure 0.17. Central transceiver antenna position and measuring range

According to the specification, the maximum consumption of the used module is 130 mA. The measured consumption of the water flow sensor is 4 mA. The maximum current that can be conducted through the sensor board development board is 300 mA, and the circuit on the development platform used is designed so that the Vbus USB terminal and the 5 V terminals of the circuit are on the same bus. From this we can conclude that the entire interface with the sensor and the LoRa module can be powered by the USB interface. However, it is necessary to optimize the consumption so that the circuit can run on a commercially available battery for as long as possible. Table 3 shows the current measurements during the operation of the microcontroller. Here, the microcontroller operated with a maximum operating clock of 96 MHz and without any power optimization. Data are given separately for each element to make it easier to track optimization.

Table 0.2. Circuit current without optimization

Connected system components	Current [mA]	State
Microcontroller	26.65	Wait
Microcontroller	26.88	Event stop
Microcontroller + LoRa Module	39.16	Wait
Microcontroller + LoRa Module	121.5	Signal send
Microcontroller + LoRa Module + Sensor	42.51	Wait
Microcontroller + LoRa Module + Sensor	125.7	Signal send

As the flow sensor does not have the possibility of optimization, in Table the values of the current flowing through it are singled out and at the end of each step they will only be added to the obtained results. Table 4 shows that by reducing the operating clock, the current decreased by 11 mA, which is a decrease of slightly more than 40% in the consumption of microprocessors.

Table 0.3. Current through the water sensor

Current [mA]	State
3.35	Idle
4.03	Flow

The first step of optimization is to lower the processor clock to 48 MHz (Table 5).

Table 0.4. Current with reduced microprocessor clock speed

Connected system components	Current [mA]	State
Microcontroller	15.50	Wait
Microcontroller	15.91	Event stop
Microcontroller + LoRa Module	28.15	Wait

As the LoRa module on the sensor platform is not used for receiving messages, there is no need to keep it constantly active. Fortunately, this module has a mode in which it shuts down its radio transceiver. By changing the code on the microcontroller, an operating mode was introduced where the radio transceiver is turned on only when necessary. With this procedure, the total current through the microcontroller and the LoRa module dropped to 17.7 mA in standby mode. The STM32F411

microcontroller has various energy saving functions. One of them is a sleep state in which we stop the processor clock completely and listen only to interruptions coming from external devices or clocks.

As FreeRTOS was used in the paper, instead of directly sending the microprocessor to sleep, FreeRTOS tickles mode was used. In it, FreeRTOS stops working and puts the microprocessor to sleep. This lowers the current through the circuit consisting of the microcontroller and the LoRa module to 5.87 mA in standby mode, with the total current through the entire circuit now being only 9.22 mA in standby mode. Measuring the current strength has successfully shown how it is possible to use a USB port to power the entire circuit. Also, in several interventions on the program code of the microprocessor, it was possible to lower the current from 42.51 mA to 9.22 mA, which is a difference of 78%. This is very important because waiting time is the state in which the circuit is located almost all the time. Using a portable USB charger (power bank) with a capacity of 10000 mAh (the most common value at the time of writing), with such consumption can be counted on approximately 40 days of autonomous operation of the sensor. Radio signal acquisition showed very good results considering the power and position of the antenna.

This measurement is an indication of how even without a great search for the ideal antenna position, a quite decent range can be achieved with a device that has the output power of an average home Wi-Fi system. The maximum measured distance was 3393 m in terms of measurements from ground level and without optical visibility. There is also a large difference in the behavior of LoRa radio protocols between urban and rural areas. While in an uninhabited area the range exceeded the manufacturer's specifications, in places with several residential buildings, the range dropped sharply. It can be concluded that for the purpose of reporting adverse events in rural and remote areas, LoRa LPWAN is an excellent solution. Smaller range in the urban area is very easy to compensate with more densely placed central transceivers. (Fig. 3.18. and 3.19.)
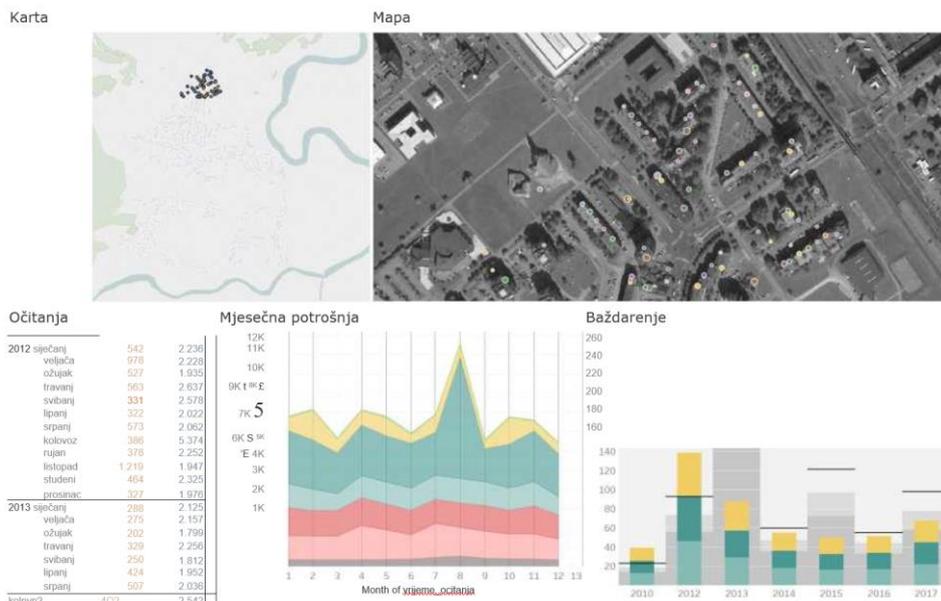
Figure 0.18. Signals

Figure 0.19. Signals

Use case: Rule-based phishing website classification

Before training the classification model it was necessary to choose the features that are relevant and useful for the classification process. To evaluate the features, we used the ranking of features based on the following methods:

- Information gain that ranks features based on the calculated information gain relative to the classification class, numerical features are first discretized.
- Gain ratio ranks features based on the calculated gain ratio. Gain ratio is calculated as Information gain divided by the entropy of the feature for which the ratio is computed.
- Symmetrical uncertainty is a measure that eliminates redundant and meaningless features, which have no interconnectivity with other features.
- Relief method was proposed by Kira and Rendell and is used for the selection of statistically relevant features, it is resistant to noise in the dana and the interdependence of features.

- Features are evaluated in a way that is randomly sampled from a given set of instances and take nearest neighbors that belong to the class. If the neighbors are aligned with instances the weighting factor increases, in contrast if the closest neighbors are different the weighting factor decreases. (Fig. 3.20.)
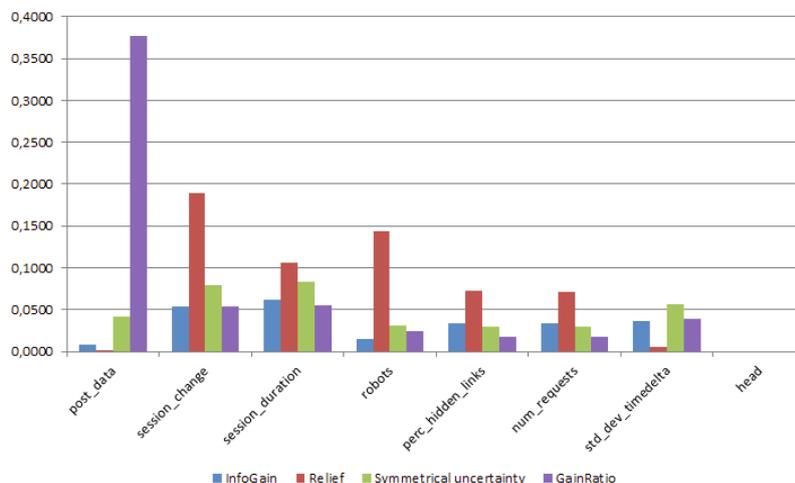
Figure 0.20. Comparison of different methods for feature selection

145   2021-1-SI01-KA220-VET-000034641

If we look at the ranked features, we see that the features that dominate the dataset are:

- post data, which shows us whether the client has filled/not filled the fake form in the Lino system.
- session change, which shows us if user, during the session, has changed the
- session identifier or not
- session duration, duration of the session in seconds
- robots, which shows us whether the user accessed/not accessed robots.txt. file, which defines the rules of robot conduct.

Aforementioned features were selected manually, we ranked all features according to the score of the feature selection method. We selected most significant features for our classification models, in our case the top five features.

Classification Model Selection for Bot from Human Differentiation

A prerequisite for using supervised learning methods and selecting the optimal subset of features is a labelled dataset. Selected features should contribute to the generalization of some classes, ie. for each class they should be able to make a unique behavioral profile. To evaluate the performance of the classification method we used the K-fold cross validation method. For our purposes we used k = 10 parts - the relevant literature states that k = 10 parts are a optimum number for estimating errors.

Decision Tree C 4.5

Firstly, for classification purposes, we evaluated a decision tree algorithm C 4.5, which is an upgrade of the classic algorithm ID3. Both algorithms are the result of research made by Ross Quinlan. C 4.5 uses a dataset for learning to create a redundant tree. In the case of using similar data, in the learning and validation set, classifier has good results but when we use an independent validation set the classifier usually produces bad results.
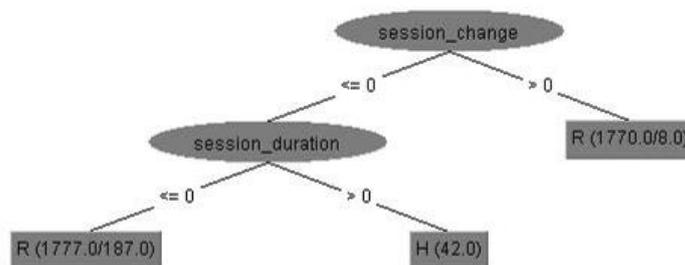
Figure 0.21. Pruned tree, using the full set of features

After building a redundant tree, the tree is converted to the IF/THEN rules and the algorithm calculates the best conditions for classification accuracy we remove the IF conditions if they do not reduce the classification accuracy. Pruning is done from the leaves to the root of the tree and is based on the pessimistic estimation of errors; errors are related to the percentage of incorrectly classified cases in the training dataset. Based on the difference in accuracy of rules and standard deviation taken from the binomial distribution we define a certain upper limit of confidence that is usually 0.25, based on which

the trees are pruned. For building our models with C 4.5 we set the confidence threshold for pruning to 0.25 and the minimum number of instances per leaf is 2.

	Class	TP Rate	FP Rate	F-measure	AUC
C 4.5 Experiment #1	Human	0.177	0	0.301	0.773
	Robot	1	0.823	0.972	0.773
C 4.5 Experiment #2	Human	0.793	0.002	0.872	0.985
	Robot	0.998	0.207	0.992	0.985
SVM Experiment #1	Human	0.265	0	0.419	0.801
	Robot	1	0.735	0.979	0.801
SVM Experiment #2	Human	0.962	0.006	0.942	0.976
	Robot	0.998	0.042	0.997	0.978

Figure 0.22. Classification results for C 4.5 and SVM, experiment 1 uses only selected features. Experiment 2 uses selected features plus Country and ASN of client.

Prior to the classification we removed the class instance of unknown visitors because they represented human attempts to attack with manually entered vales or using non-existing browsers. Method C 4.5 has resulted in the pruned tree showed, which is the same with the optimal selection of features and using the full set of features. It is important to note that the algorithm C 4.5 is very good at choosing features by using heuristics in creating and deleting subtrees.

If we look at the results in Figure above, we can see that for the given features (Experiment 1) we have a classification accuracy of 94.5% and a perfect rate of correct positives for robots (TP rate). Classifier badly classifies human visitors (TPR = 0.177) and degrades the classification ability of the robot where the rate is false positive rate is high - 0.823. Looking at F-measure we can say that a good classifier correctly detects robots while wrongly classifies human visitors and commonly (> 80%) declares them robots. We tested the C 4.5 classier with two additional features - client's country and ASN of service provider. These features were resolved from the IP address using the aforementioned GeoIP database. This subset (C 4.5 Experiment 2) is shown in Figure above. We reduced the number of false positives for the class Robot to 0.207, thus the result of classification for class Human was better 0.793.

Support Vector Machine

SVM is an algorithm that finds the maximum margin of separation between classes, while defining the margin as distance between the critical points which are closest to the surface of separation. Points nearest to the surface are called support vectors, the margin M can be seen as the width of the separation between the surfaces. Calculating the support vector is an optimization problem which can be solved using different optimization algorithms. The trick used in calculating SVM is using different kernel functions, which move unsolvable or inadequate problems to a higher dimension, where it can be solved. In our experiments we trained our SVM models with the sequential minimal optimization algorithm using a linear kernel $K(x, y) = < x, y >$ where _ = 1.0–12 and tolerance is set to 0.001, previously training data was normalized. For Experiment 1 features SVM performs better than C 4.5, the precision was 95.8%. Human visitors are still a problem, although SVM has much higher rate of true positives

(26.5 %). Increased detection rate of human visitor gives a lower rate of wrong detection of robots (73.5 %). F measure is very good for robots and much better for human visitors (even better than method C 4.5) - but still too low to be used - (0.419). With additional features Country and ASN (Experiment 2) we obtained a false rate for both classes under 5 %. The true positive rate was also high for class Human 0.962 and for class Robot 0.998. We can conclude that with this subset of features and with regular retraining to avoid the concept drift, this model is feasible for everyday use.

Usecase: Data loss prevention cloud-based system

Comparison of DLP solutions available on the market

Comparison of DLP solutions available on the market, based on the Gartner® Releases 2022 Market Guide for Data Loss Prevention: Key Takeaways.

Symantec

Based in Mountain View, California, Symantec has been on the DLP market since the acquisition of Vontu in 2007. Symantec has recently released Symantec Data Loss Prevention 15.0 and has component products for DLP Enforce, DLP IT Analytics, cloud storage (supports more than 65 cloud applications), Cloud Prevent for Microsoft Office 365, DLP for endpoint, DLP for network and DLP storage, as well as third-party security technology DLP API support, such as content retrieval, reporting, and FlexResponse for encrypting content or DRM application. Symantec continues to invest in DLP technology and improves its data protection business unit. In 2016, Symantec has made the acquisition of Blue Coat, which gives it the option of purchasing Elastica and Perspecsy for the Blue Coat, for which there are integration of DLP policies through the two-way REST API between Elastica and Symantec DLP. Symantec is a convenient choice for organizations that require advanced detection techniques and integration with CASB for a unique data protection policy.

Advantages
Symantec offers the most advanced detection techniques in the market with advanced functionality such as form recognition, image analysis, and handwriting recognition that can cover a wide range of data loss scenarios. Symantec supports a hybrid deployment model for several of its DLP products where Detection Servers installed on AWS, Azure, or Rackspace connect to a local DLP Enforce platform. Symantec's Smart Response system offers a wide range of administrative flexibility based on content actions that conform to the DLP rule. Its Vector Machine Learning (VML) DLP enables users to learn the DLP system by providing both positive and negative sample content. This could be useful if traditional matchmaking methods are not sufficient to match content correctly.

Weaknesses
Symantec clients expressed frustration when purchasing or updating Data Insight plugins for Symantec DLP, which is now owned by Veritas. Make sure your Symantec DLP vendor can also sell Veritas Data Insight if you are interested in this add-on. Monitoring and detecting sensitive data in cloud applications requires DLP endpoint detection and the required Symantec CASB connectors to achieve full

functionality. Clients express concern over the overall cost of implementing Symantec DLP, compared to competing products.

Digital Guardian

Established in 2002, Digital Guardian (formerly Verdasys) is headquartered in Waltham, Massachusetts. Access to Digital Guardian DLP is primarily via the DLP endpoint, with strong partnerships for DLP product network integration and DLP detection by October 2015, when Code Green Networks (CGN) was acquired through acquisition. Since then, it has launched it as a line of Digital Guardian Network DLP products. The Digital Guardian endpoint covers DLP, advanced threat protection, and endpoint detection and response (EDR) in a single agent installed on desktop computers, laptops and servers running on Windows, Linux, and Mac OS X, as well as support for VDI environments. The Digital Guardian Network DLP and the Digital Guardian Discovery product cover DLP networks, cloud data protection, and data discovery, and are offered as hardware, software applications, and / or virtual apps. During 2016, Digital Guardian worked on simplifying and integrating management capabilities between its DLP endpoints and assets from CGN acquisitions. Digital Guardian also has an existing partnership with Fidelis Cybersecurity Network DLP. Several Gartner clients recently talked about this partnership, and Gartner believes that, in addition to existing joint customers, the partnership will continue to reduce and eventually stop. Digital Guardian is a suitable choice for organizations with strong concerns about the legislation, particularly in the health sector and financial services, as well as the organization with the requirements of AD protection of intellectual property. Digital Guardian is also a good choice for organizations that require uniformity of DLP rules to work equally well in all Windows, Mac OS X and Linux operating systems.

Advantages
Clients report faster implementation times and successful projects when using the Digital Guardian product in combination with managed digital guardian services. Digital Guardian has integration with wider security products, including threat intelligence, network sandbox, User and Entity Analysis (UEBA), Cloud Data Protection, and Security Event Management (SIEM, including IBM QRadar and Splunk applications). Customers like the option of modular licensing for the DLP endpoint, with support for Windows, Mac OS X and Linux, and feature endpoints that can be licensed in any combination of device visibility and control, DLP and advanced threat protection. Digital Guardian's vision shows a strong understanding of technology, security, threats, and trends in the industry that will shape their bids.

Weaknesses
Digital Guardian does not have a common policy for end points and network products. The Digital Guardian Agent can not differentiate between personal and business accounts for Microsoft OneDrive. However, it can prevent the use of personal Microsoft OneDrive applications. Customers expressed concern about the speed of integration of the acquired CGN. Structured data indexing is not supported by the Digital Guardian endpoint agent, but this feature is available through CGN agent.

Forcepoint

In 2015, Raytheon and Vista Equity Partners completed a joint venture that combines Websense, a portfolio company Vista Equity and Raytheon Cyber Products. In 2016, the company gained two lines of Intel Security - Stonesoft and Sidewinder fireworks through acquisition - and restarted the combined company as Forcepoint. Raytheon has already ht municipal share Forcepoint's, and Vista Equity Partners holds a minority stake. Headquartered in Austin, Texas, Forcepoint has been a leader in the DLP product market, previously known as Raytheon-Websense, for several years. The Forcepoint DLP product line includes Forcepoint DLP Discover, Forcepoint DLP Gateway, Forcepoint Cloud Applications, and Forcepoint DLP Endpoint. During the years of delivery of DLP and integrated DLP modules for its secure web and e-mail gateway products, Forcepoint has created an outstanding DLP packet for network coverage, endpoints, and data discovery (both client and cloud), with special attention to the protection of intellectual property and the implementation of the compliance policy with the regulations. Forcepoint is a suitable choice for organizations with requirements for legal compliance and intellectual property protection or organizations that want to implement DLP virtual devices in the Azure public cloud infrastructure.

Advantages

Forcepoint DLP Endpoint can automatically encrypt / decrypt files via Microsoft RMS without removing RMS protection based on end-to-end data, motion data, and discovery rules. Forcepoint provides over 350 predefined rules and embedded component UEBA for additional security analytical features that perform incident risk rating, identify threats from internal users, point out endangered endpoints, and calculate data theft risk indicators to identify the most vulnerable users and activities. Indexing structured data, especially data-indexing support in Salesforce, cites clients as the key differentiator factor.

Weaknesses

Clients reported problems with technical support for indexing structured data. If you need to index structured data in the database, make sure you thoroughly test it on live data in your specific database environment. Raytheon's involvement in the defense market will help strengthen Forcepoint with additional intelligence and products. However, there is no success of security vendors owned by defense structures that have succeeded in commercial markets. Forcepoint's relevance in some geographic areas can be problematic because of Raytheon's strong American loyalty. Some Gartner clients have noted this complaint and see if this is causing your concern in your organization.

Intel Security (today: McAfee)

Over the past few years, Intel has changed its investment in and from various product lines several times and has not sufficiently considered these changes inside and outside the company. This has caused exhaustion of employees at alarming rates, many of which have been launched by new security companies or are employed by competitive security vendors. Historically, in many of Intel's security products, there has been a chronic lack of investment.

Intel's security approach was to integrate acquisitions with McAfee's ePolicy Orchestrator (McAfee ePO) policy management system, alert monitoring, and link security events between DLP Event Ends, Network Transfers, and Restricted Data on Storage Data in the Organization. The DLP 10.0 release has brought further improvements to DLP and updates to DLP online products in 2016 highlighted McAfee's renewed focus on data protection. Intel Security is a good choice for organizations that have significant resources invested in McAfee ePO and want a unique supplier that can provide DLP, device control and encryption.

Advantages

DLP integration within McAfee Web Gateway proxy supports decrypting and re-encrypting site traffic site traffic, including email service providers and cloud storage products. The capture database can index and store all visible network and endpoint components. Clients reported this useful to test new rules, forensic analysis of events that occurred prior to policy making and after-event investigation. It also supports e-discovery and legacy retention, as well as integration directly with the software Guidance Software and Access Data. McAfee DLP includes the basic level of data classification on the DLP 10 endpoint for Windows and Mac OS X and can still be firmly integrated with Titus and Bold James for various data classification options. DLP endpoint rules are aware of locations and may have different responses and content remedies when they are online when they are offline. The Federation of Security Innovations (SIA) is still robust and is a good way for Intel Security customers to maximize their DLP investment due to proven and tested integration of data product classifications, DRIF and UEBA suppliers.

Weaknesses

McAfee DLP supports native API integration with Cloud Data Box but support for other cloud applications and cloud storage support are missing. Intel Security has made some improvements to DLP Agent 10 on Mac OS X, but still lacks support for email, web, and cloud. Linux is not supported. Customers report that the configuration of DLP rules can be complex and disadvantageous compared to other DLP products. The future success of Intel Security in the DLP market will depend on their performance while acting as a company and whether the focus can be on data security tasks over a longer period of time.

Usecase: Dynamic web site hosting

How to deploy a dynamic website with AWS by uploading your website content into S3 bucket, create an EC2 instance to host web app on it as in this scenario EC2 acts like a public server all people from the world can visit this server.

Amazon S3 (Simple Storage Service) is a service offered by AWS for object storage through a web service interface. It can be used to store or retrieve any amount of data such as documents, images, videos, etc. S3 bucket is a resource in Amazon S3. It is a container where files and folders can be uploaded.

Amazon EC2 (Elastic Compute Cloud) is a service offered by AWS. It is considered as a virtual server.

2021-1-SI01-KA220-VET-000034641

IAM (Identity and access management) Role is used to give permission to service to do something on another service.

LAMP web server can be used to host a static website or deploy a dynamic PHP application that reads and writes information to a database.

Steps

Step 1: Create S3 Bucket

You will need to create an S3 bucket to put your website's files and folders. To do this, login into your AWS management console and click on Services on the top navbar. From the Services drop-down, select S3 from the Storage section. This should display the S3 dashboard.
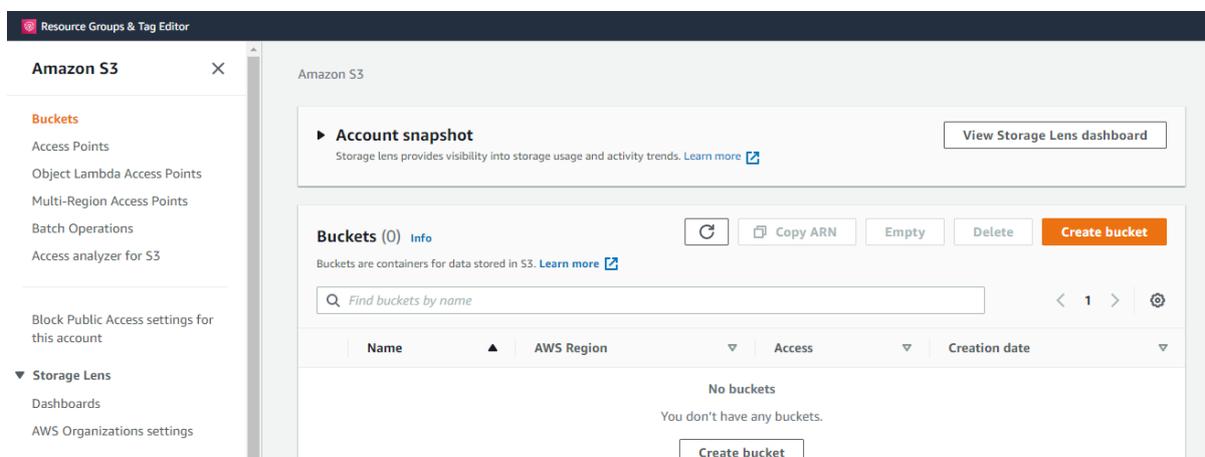
Figure 0.23. Create S3 Bucket 1

From the S3 dashboard, click on Create bucket. Give the bucket a unique name, the name you choose must be globally unique. Next, choose your preferred AWS Region from the drop-down.
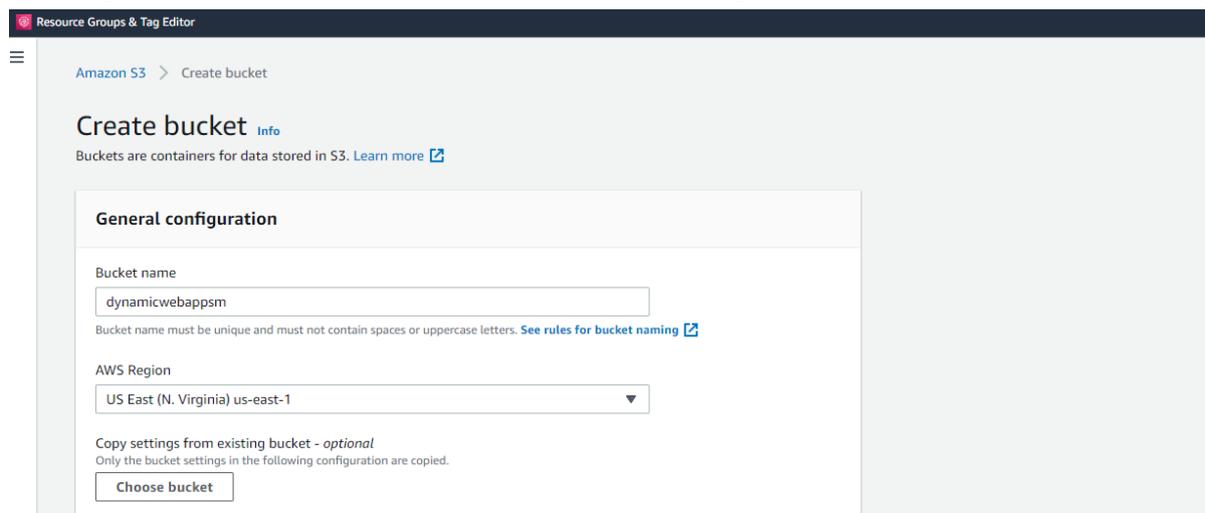
Figure 0.24. Create S3 Bucket 2

Under Block Public Access settings for this bucket section, check the Block all public access checkbox. This is done to make the bucket not accessible to the public.
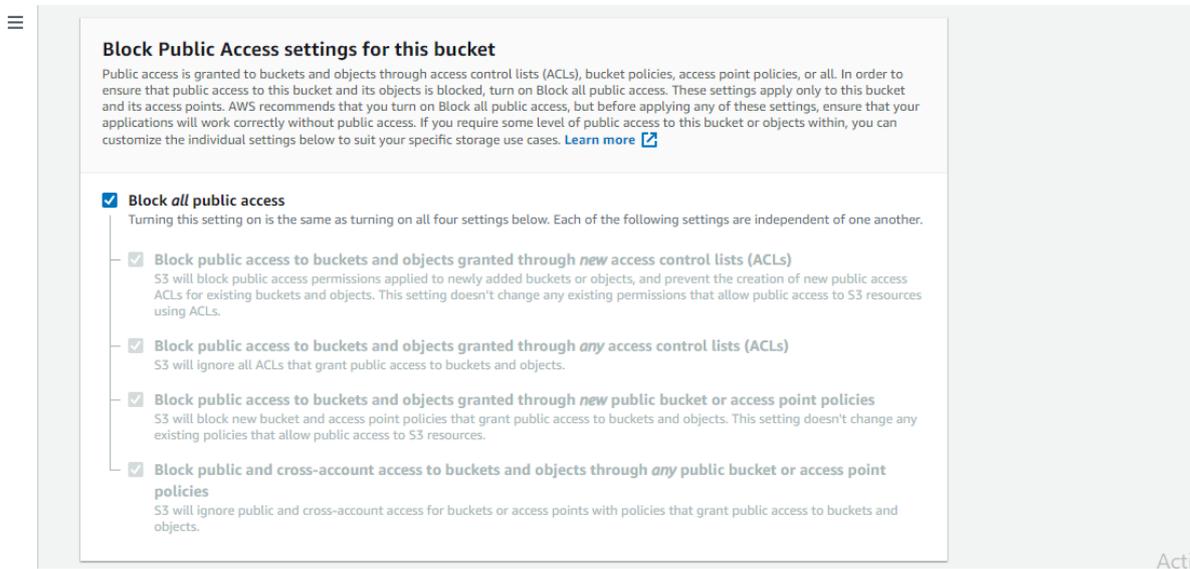
Figure 0.25. Create S3 Bucket 3

Click on Disable for Bucket Versioning. You can also Add tag to the bucket for easy identification.
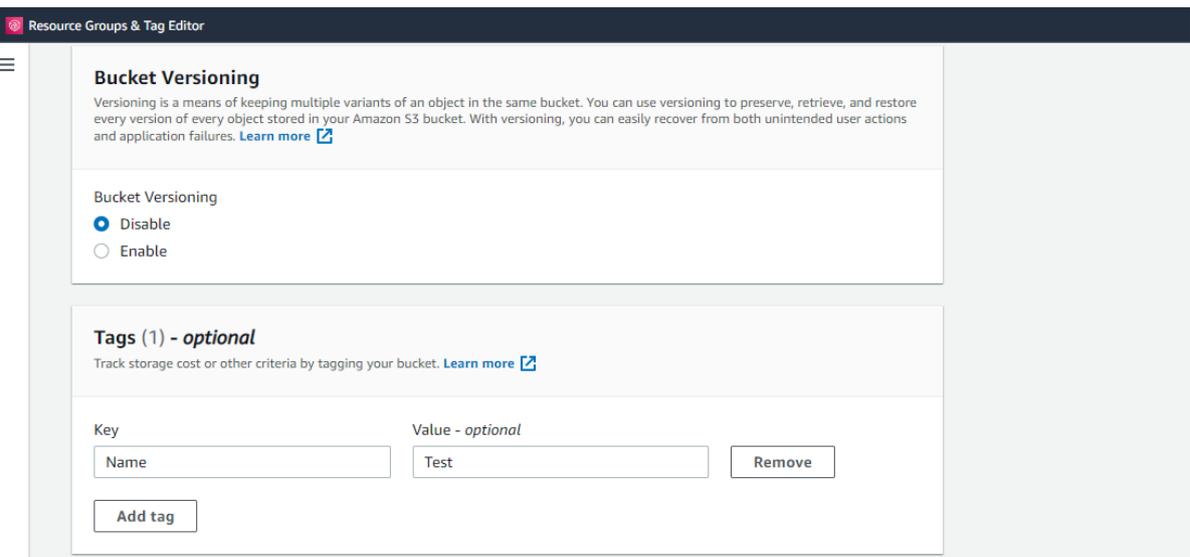
Figure 0.26. Create S3 Bucket 4

Under Default encryption section, click on Enable for Server-side encryption. Then check Amazon S3 Key (SSE-S3).

2021-1-SI01-KA220-VET-000034641
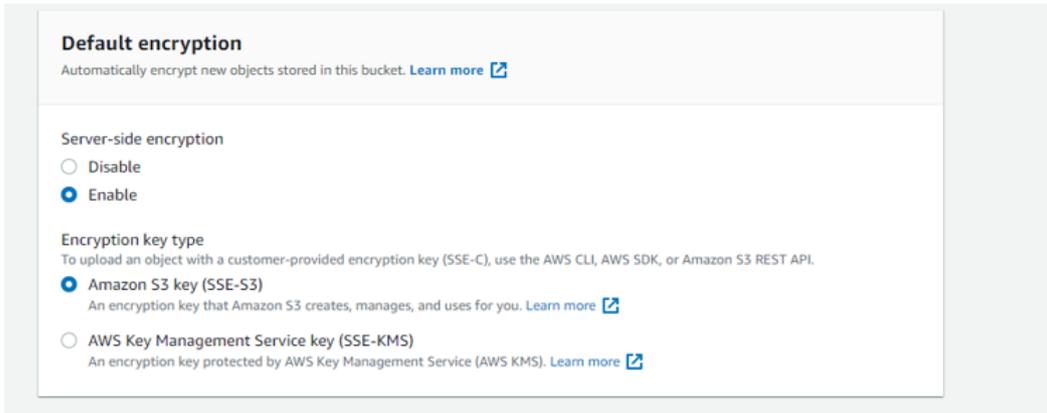
Figure 0.27. Create S3 Bucket 5

Then click on Create bucket.
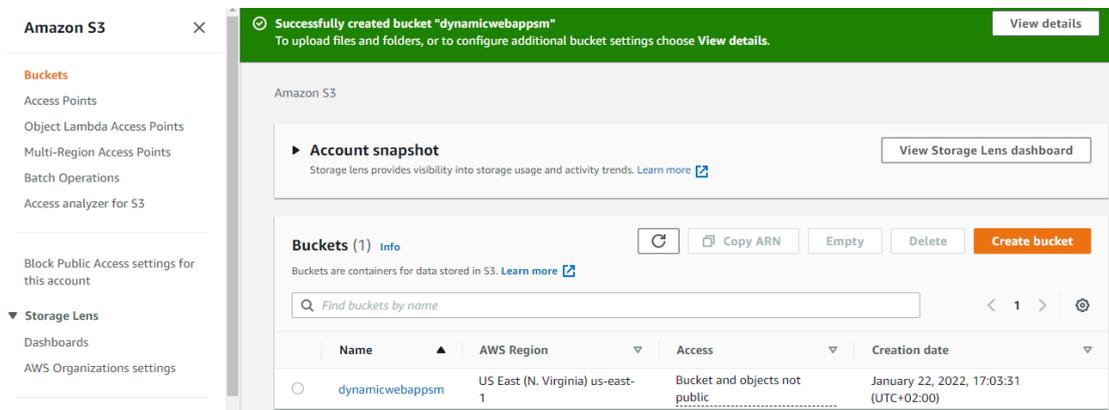
Figure 0.28. Create S3 Bucket 6

Step 2: Upload web files to S3 bucket

After creating the bucket, you need to upload your website's files and folders into it.
From the S3 dashboard, click on the name of the bucket you just created. On the Objects tab, you can see that the bucket is currently empty, click on the Upload button.
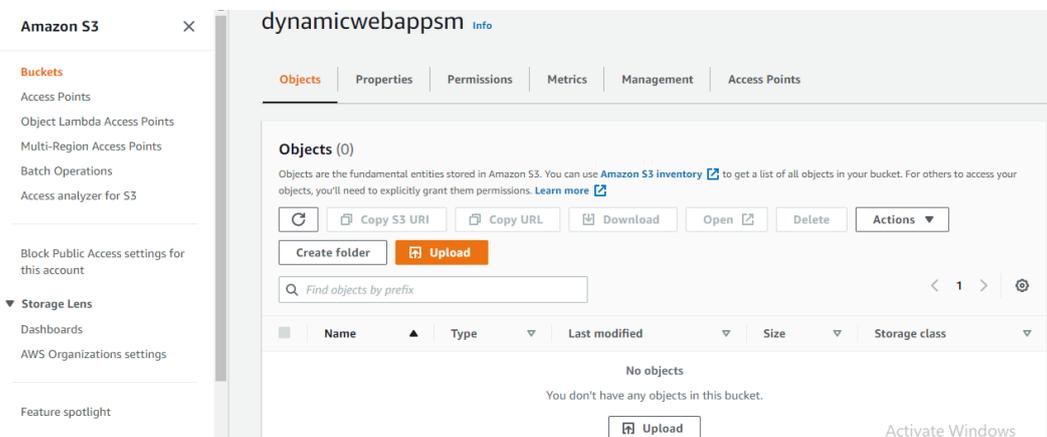
Figure 0.29. Create S3 Bucket 7

2021-1-SI01-KA220-VET-000034641

This should take you to the Upload page.

Figure 0.30. Create S3 Bucket 8

Figure 0.31. Create S3 Bucket 9

After the necessary files and folders have been added, scroll down, and click on Upload. The uploading should be done in a few minutes depending on your network and content size. Also, please do not close the tab while the upload process is going on.

Step 3: Create IAM Role

Now, EC2 want to pull code from S3. So, you want to create IAM Role to give EC2 permission to access S3. To do this, from the Services drop-down, select IAM from the Security, Identity& Compliance section. From the IAM dashboard, click on Roles. Then Click on Create role.
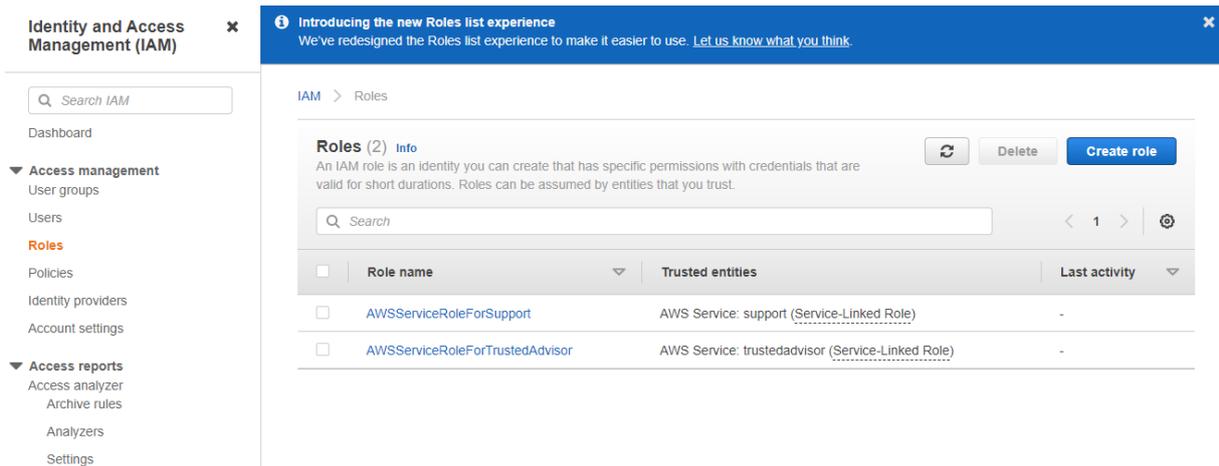
Figure 0.32. Create LAM role 1

Choose EC2 and click Next: Permissions.
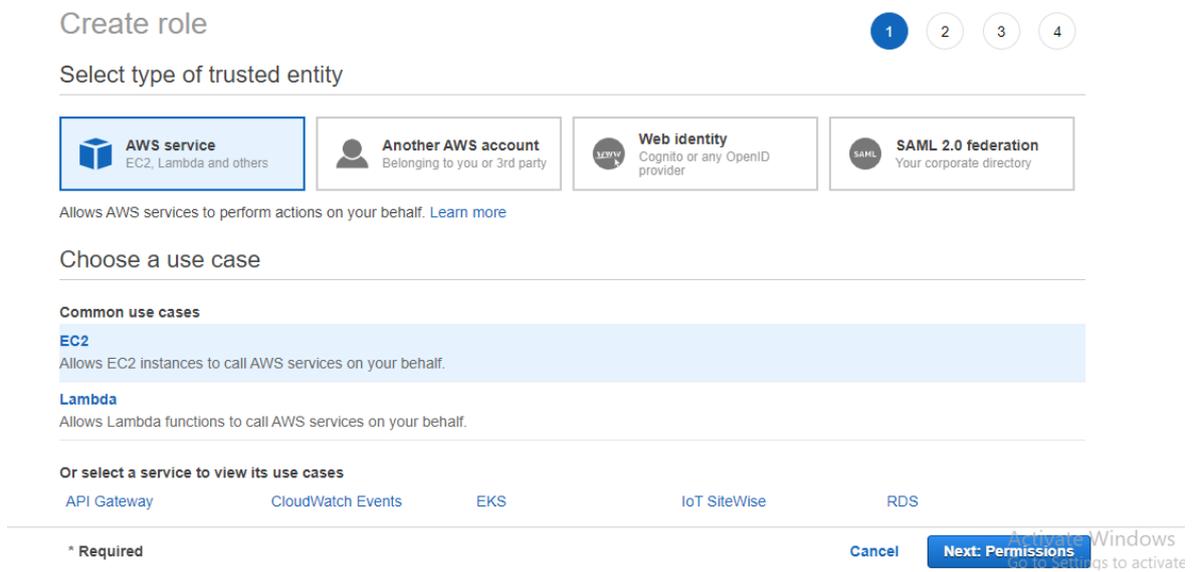
Figure 0.33. Create LAM role 2

Search for S3 and check AmazonS3FullAccess. Then click Next: Tags.

Figure 0.34. Create LAM role 3

Click on Next: Review.

Figure 0.35. Create LAM role 4

Give the role name and description. Then click on Create role.

Figure 0.36. Create LAM role 5

Now, the role has been created successfully.

Figure 0.37. Create LAM role 6

Step 4: Create an EC2 instance

You will need to create an EC2 instance to install Apache (/var/www/html) and copy the content of S3 to html directory. To do this, from the Services drop-down, select EC2 from the Compute section. This should display the EC2 dashboard. From the EC2 dashboard, click on Launch instance.
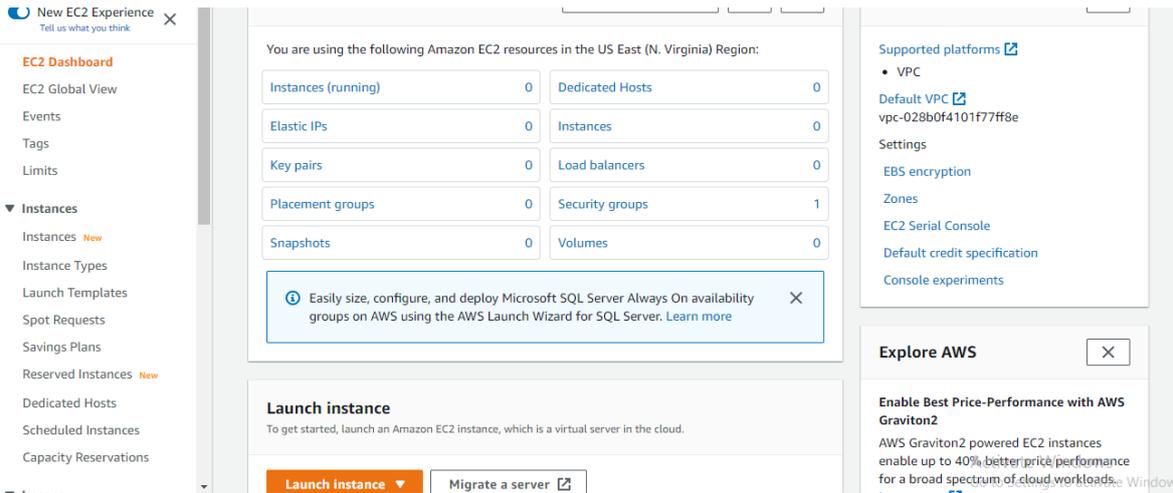
Figure 0.38. Create an EC2 instance 1

For AMI, choose Quick Start and click on Select for Amazon Linux (Free tier eligible).
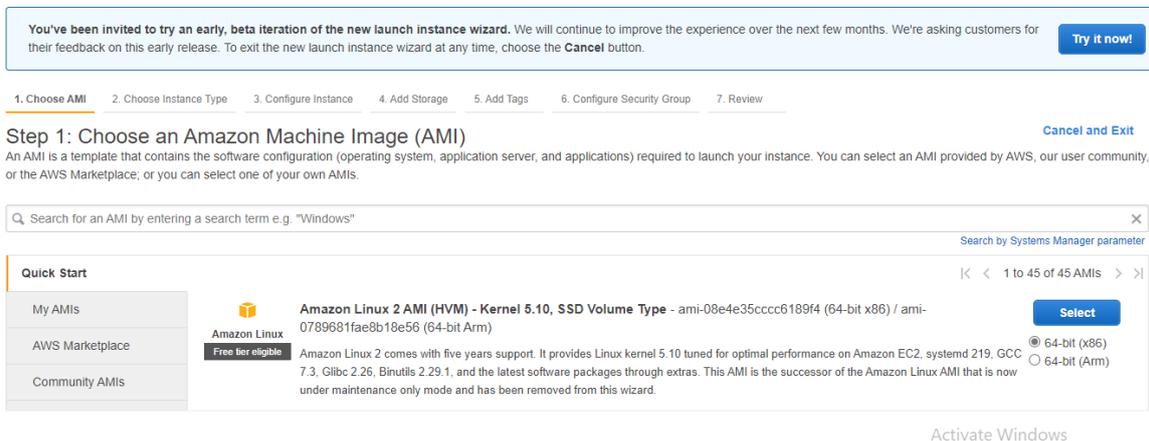
Figure 0.39. Create an EC2 instance 2

For an instance type, choose t2. micro (Free tier eligible). And click on Next: Configure Instance Details.
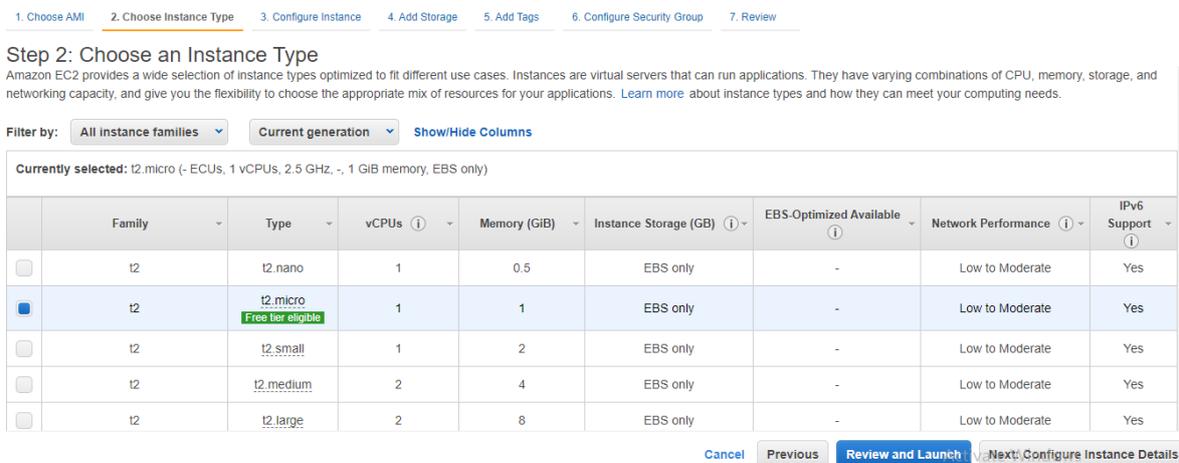
Figure 0.40. Create an EC2 instance 3

Determine 1 for Number of instances, default VPC for Network and Default in us-east-1a for Subnet.

Figure 0.41. Create an EC2 instance 4

Choose ec2s3role or whatever you named for IAM role. And Terminate for Shutdown behavior. Then click on Next: Add Storage.

Figure 0.42. Create an EC2 instance 5

Click on Next: Add Tags.

Figure 0.43. Create an EC2 instance 6

You can add tag Name: DynamicSite. Then click on Next: Configure Security Group.
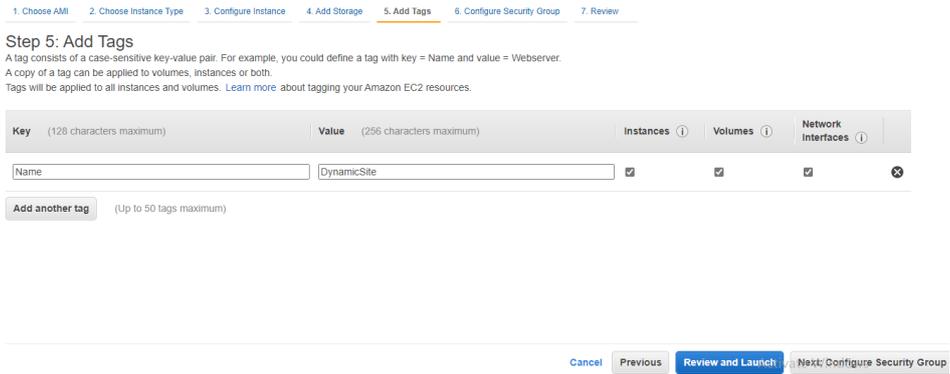
Figure 0.44. Create an EC2 instance 7

Select Create a new security group. Give it Name: DynamicWebsiteSG and description: SG for DynamicWebApp. For SSH rule select My IP for Source. Click on Add Rule and select HTTP for Type and Anywhere for Source. Last rule select HTTPS for Type and Anywhere for Source. Click on Review and Launch.
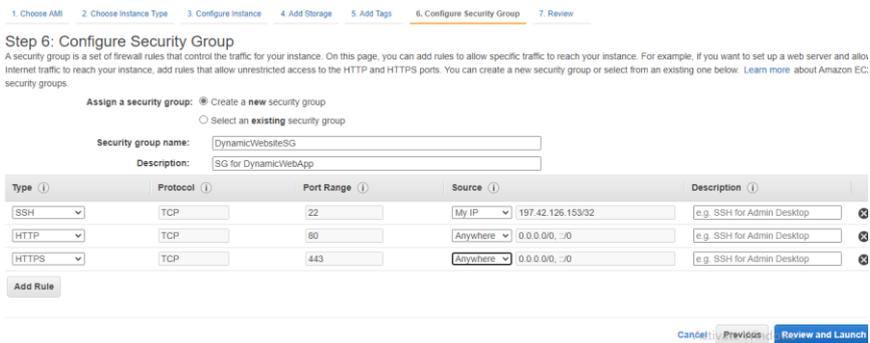
Figure 0.45. Create an EC2 instance 8

Click on Launch.
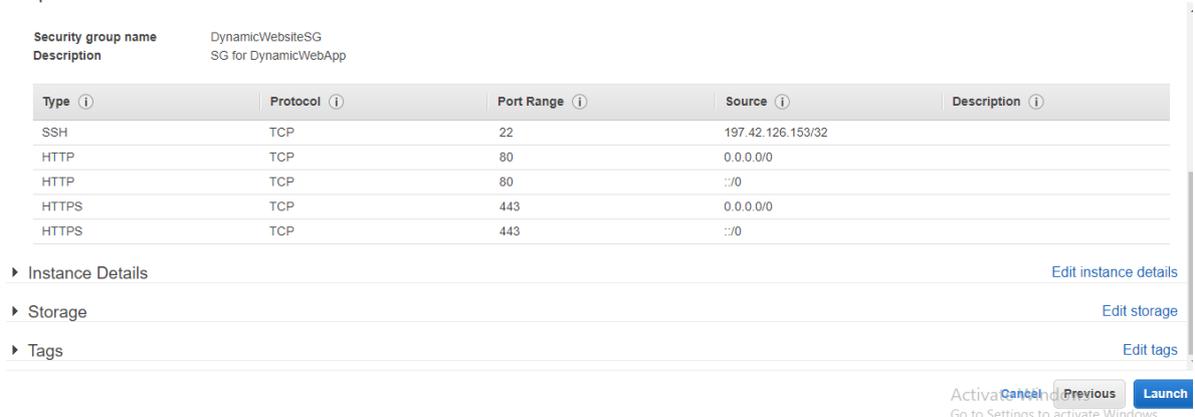
Figure 0.46. Create an EC2 instance 9

Select Create a new key pair and RSA for type. Give it name WebServerKey and click on Download Key Pair. Note: You should download the key to can ssh on EC2. Click on Launch Instances.
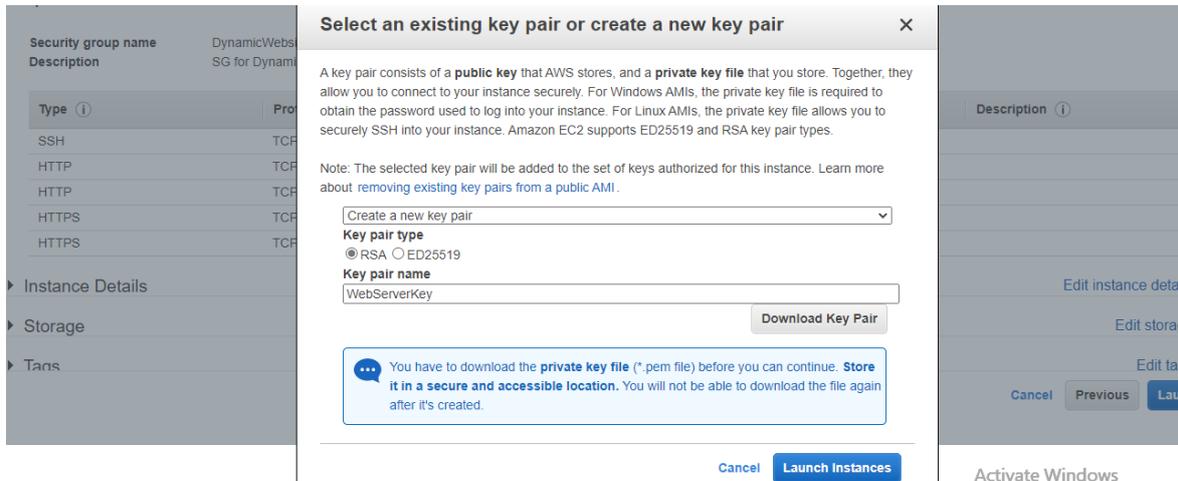
Figure 0.47. Create an EC2 instance 10

Now, instance is launching successfully.
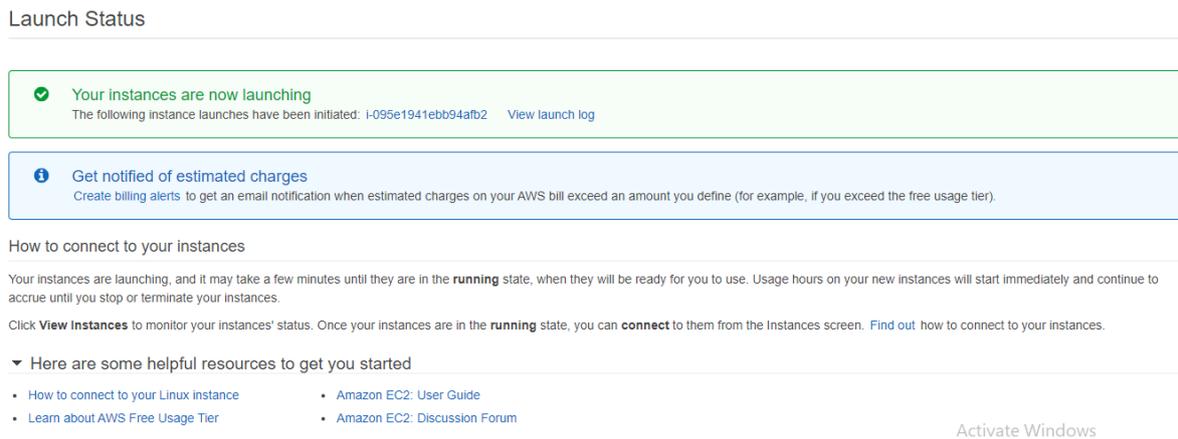
Figure 0.48. Create an EC2 instance 11

Click on Review Instance and wait Status check will be 2/2 checks passed.
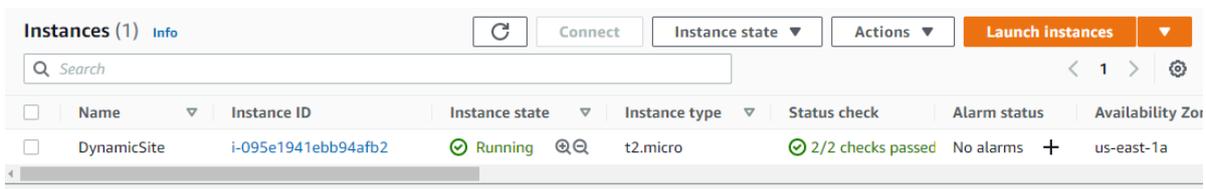
Figure 0.49. Create an EC2 instance 12

<u>*Step 5: SSH with MobaXterm*</u>

Now, you want to connect to EC2 by using MobaXterm. First you should copy public IPv4 address of EC2 instance.
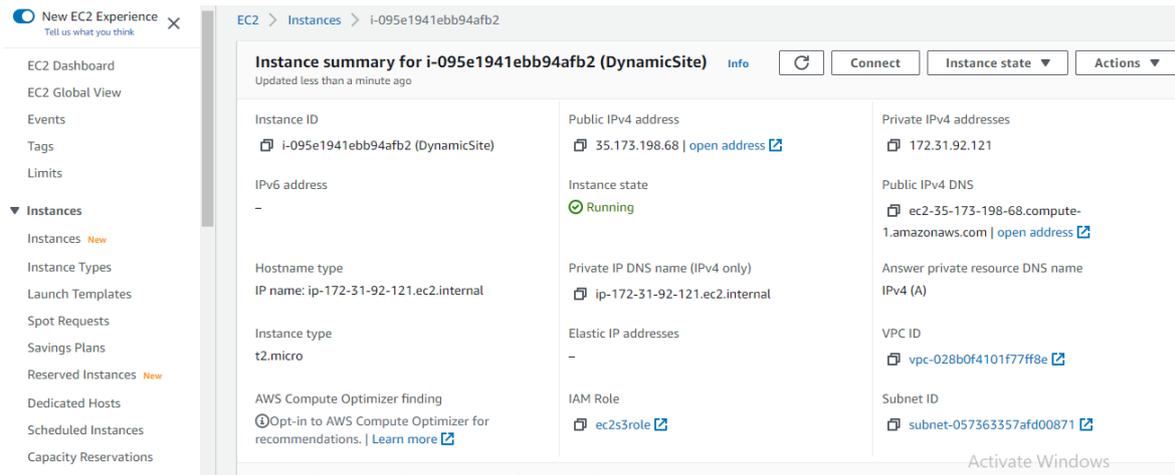
Figure 0.50. Create an EC2 instance 13

Open MobaXterm and start a new remote session by clicking on Session.
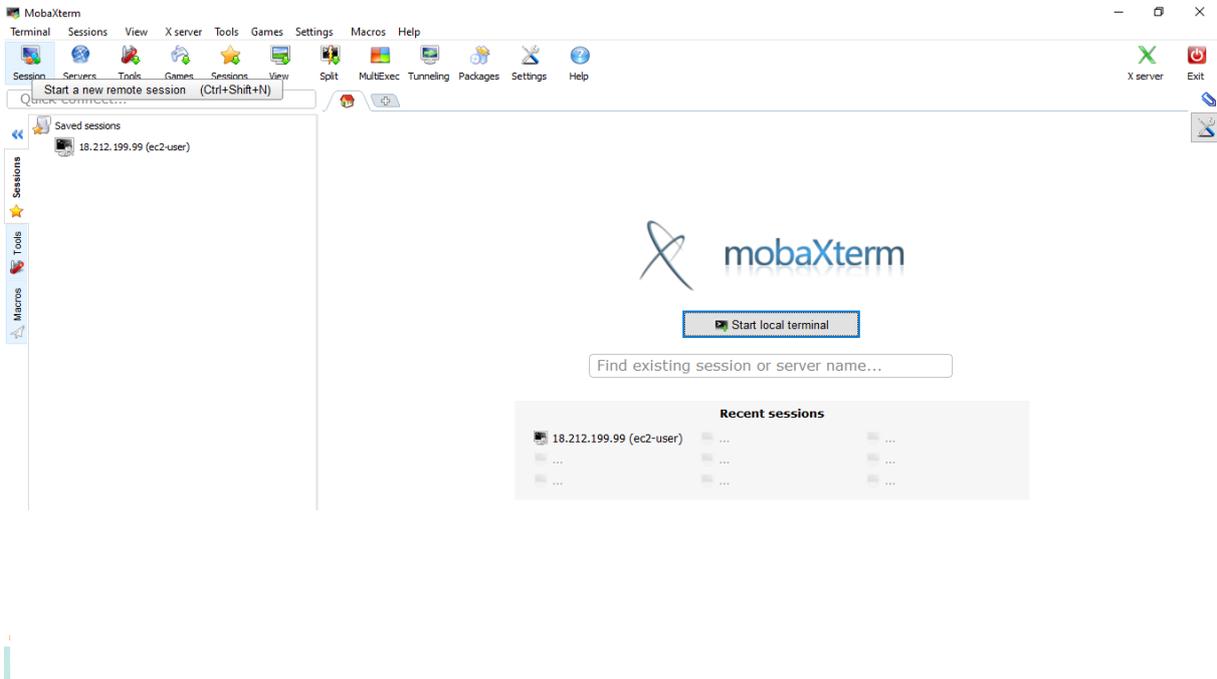
Figure 0.51. Create an EC2 instance 14

Click on SSH. Paste IP of your EC2 For example:(3.86.76.216). And ec2-user for Specify username. Click on Advanced SSH settings, check Use private key and browse location of key. Click OK.
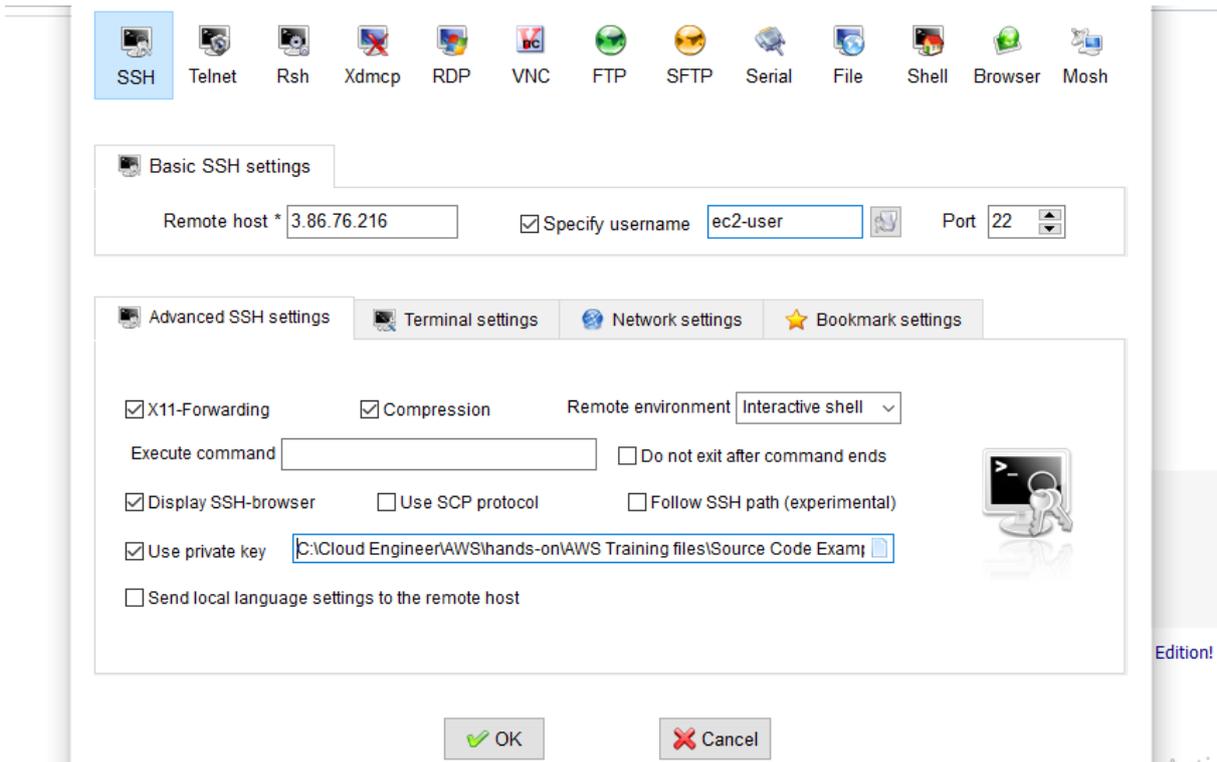
Figure 0.52. Create an EC2 instance 15

Now, you connected to EC2 successfully.
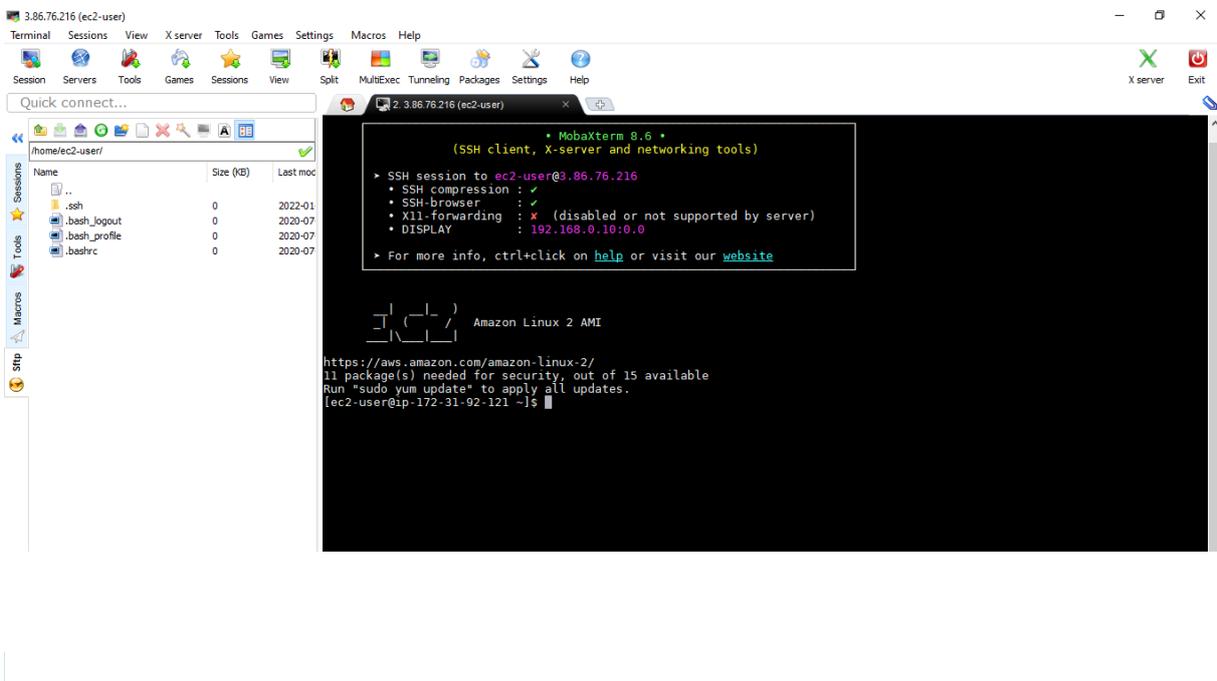
Figure 0.53. Create an EC2 instance 16

Step 6: *Install a LAMP web server on Amazon Linux 2*

The following procedures help you install an Apache web server with PHP and MariaDB. To ensure that all of your software packages are up to date, perform a quick software update on your instance.

Co-funded by the European Union

sudo yum update -y

Install the lamp-mariadb10.2-php7.2 and php7.2 Amazon Linux Extras repositories to get the latest versions of the LAMP MariaDB and PHP packages for Amazon Linux 2.

sudo amazon-linux-extras install -y lamp-mariadb10.2-php7.2 php7.2

Now, you can install the Apache web server, MariaDB, and PHP software packages.

sudo yum install -y httpd mariadb-server.

Start the Apache web server.

sudo systemctl start httpd.

Use the systemctl command to configure the Apache web server to start at each system boot.

sudo systemctl enable httpd.

You can verify that httpd is on by running.

sudo systemctl is-enabled httpd.

Now, you want to copy content of website from S3 to directory /var/www/html in EC2 . Make sure you copy your S3 bucket name.

sudo aws s3 cp s3://dynamicwebappsm --region us-east-1 /var/www/html/ --recursive

To verify that content is copied to /var/www/html .

cd /var/www/html

ls

Copy public IPv4 DNS and paste it in a new tab.

Figure 0.54. Install Lam -1

Congratulations, you have deployed a dynamic website on EC2 successfully.
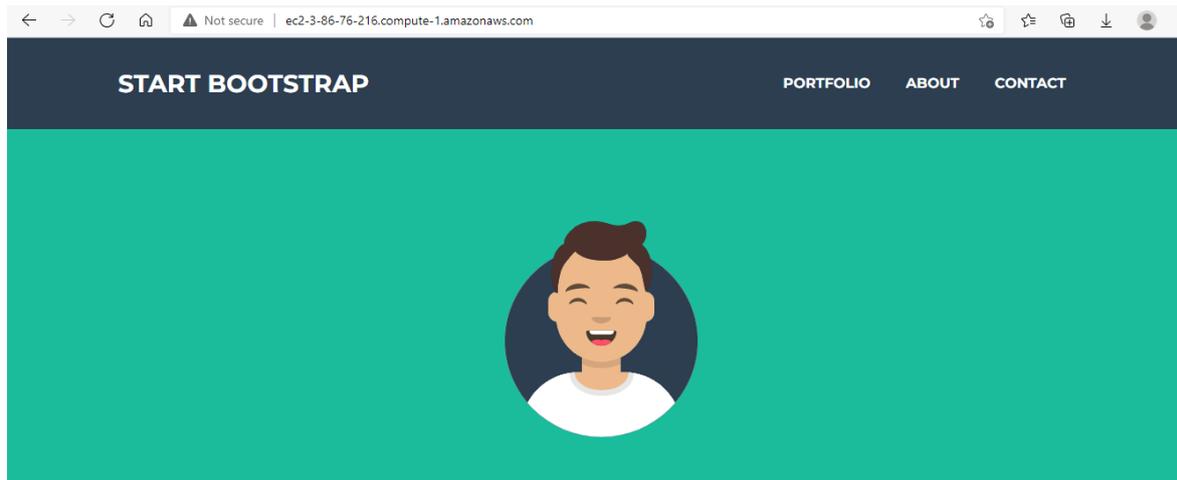
Figure 0.55. Install Lam -2

Inspired by: (Sara Mostafa, Host dynamic, LinkedIn, 2023)

Use case: Host a static website using AWS (or other clouds)

Step-by-Step Guide

<u>*Basic Configurations*</u>

- Go to the S3 console and create a new bucket with default settings.
- Go the properties of your bucket and choose the option "Static website hosting."
- Enable the option "Use this bucket to host a website."
- Provide the names of the HTML to be displayed as the homepage and the HTML file that will be displayed in case an error occurs on your site.

Optionally, provide redirection rules if you want to route requests conditionally according to specific object key names, prefixes in the request, or response codes to some other object in the same bucket or external URL.
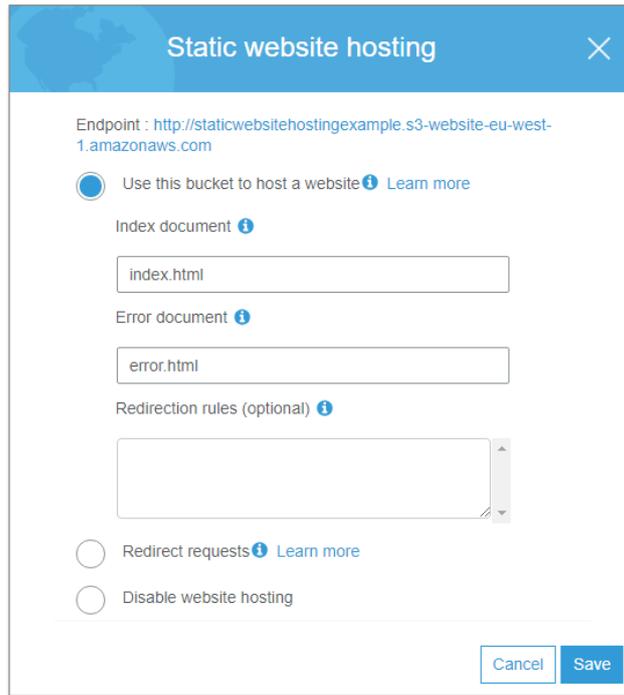
Figure 0.56. Basic Configurations 1

Now, go the Permissions section of your bucket and add the following into your Bucket Policy section.
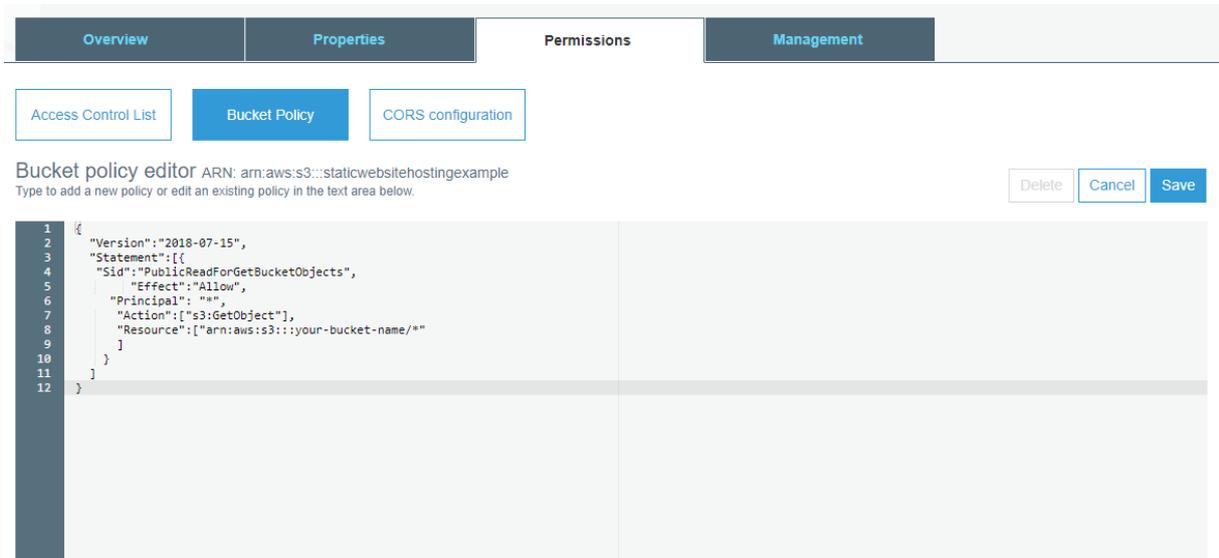
Figure 0.57. Basic Configurations 2

Replace *your-bucket-name* with the name of your bucket.

To enable your S3 static website to respond to requests like GET and POST coming from an external application hosted on a certain domain, you would need to configure CORS in your bucket settings. To do this, add the following into the CORS configuration section of Permissions.

```
1   <!-- Sample policy -->
2   <CORSConfiguration>
3       <CORSRule>
4           <AllowedOrigin>*</AllowedOrigin>
5           <AllowedMethod>GET</AllowedMethod>
6           <MaxAgeSeconds>3000</MaxAgeSeconds>
7           <AllowedHeader>Authorization</AllowedHeader>
8       </CORSRule>
9       <CORSRule>
10          <AllowedOrigin>*</AllowedOrigin>
11          <AllowedMethod>POST</AllowedMethod>
12          <MaxAgeSeconds>3000</MaxAgeSeconds>
13          <AllowedHeader>Authorization</AllowedHeader>
14      </CORSRule>
15  </CORSConfiguration>
16
```
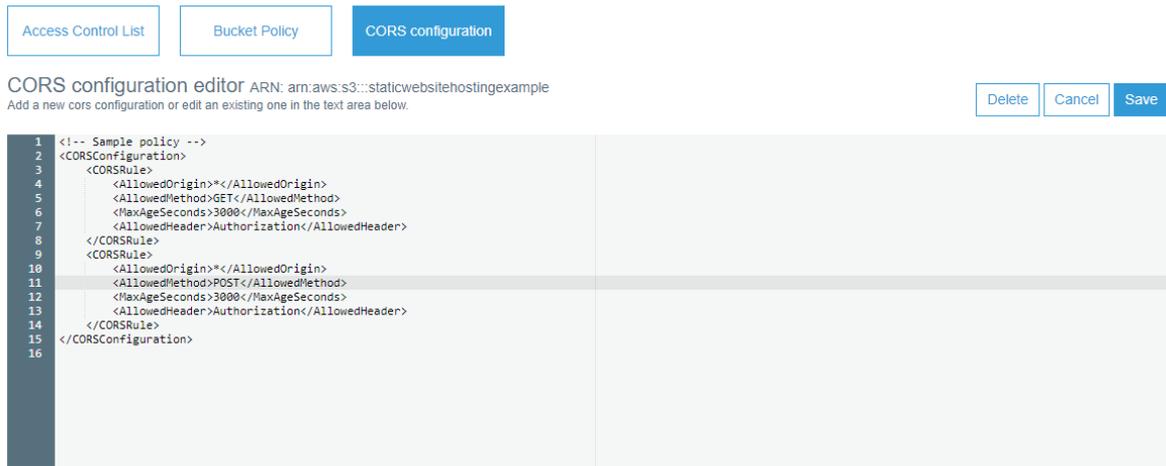
Figure 0.58. Basic Configurations 3

Upload your code. For this tutorial, create two simple HTML files by the name of index.html and error.html and upload them to the bucket.
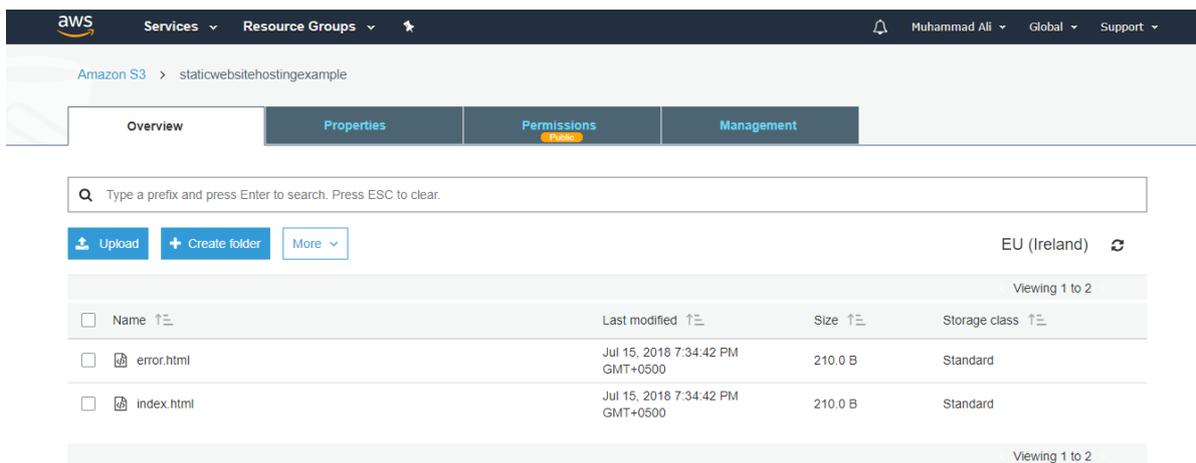
Figure 0.59. Basic Configurations 4

To launch and test the site, the endpoint can be retrieved from Properties > Static website hosting.

<ins>*Enrich Your Website by Adding Dynamic Behavior*</ins>

You can use a combination of HTML5 and CSS3 to graphically enrich your website. You can also use jQuery Ajax to call an API (microservice) and dynamically fetch data from a data source and display it on your website. Similarly, by invoking API endpoints using Ajax, you can store any kind of user's data back to your data source, like any other web application. If your requirement is to use AWS only for all your development needs, you can use a combination of API Gateway and Lambda to build APIs, a tutorial for which can be found here.

CORS Settings in API Gateway Endpoints

It is important to note that when developing APIs (microservices) using an API Gateway and Lambda, make sure to do the following.

Enable CORS in the API gateway at the time of creating a new resource.

Figure 0.60. Enable CORS in the API gateway at the time of creating a new resource.

When writing the lambda function (which you will integrate with the API Gateway endpoint to provide functionality to your microservice), make sure to add an additional parameter into the response header by the name of **Access-Control-Allow-Origin** with the value **"*".**

Appendix 2: The 2012 US Presidential Campaign and how AWS supported Obama

In this unit, we will look at how Amazon's cloud computing technology allowed President Obama's 2012 Presidential campaign to avoid an IT investment that would have run into the tens of millions of dollars.

A look at our case study:
The campaign's IT team used AWS to build, launch, run, and grow their apps. Following the election, they backed everything up to Amazon S3 and scaled way, far down. They created and operated over 200 AWS apps that could handle millions of people. On the final four days of the campaign, one of these applications, the campaign call tool, handled 7,000 concurrent users and placed over two million calls.

Here is a graph (figure) that shows the increase in call volume in the days leading up to the election:
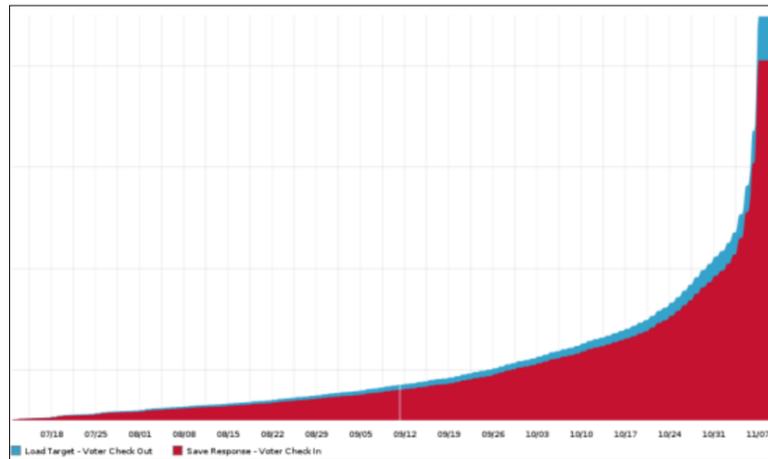
Figure 0.61. Call volume

Why use AWS?

Here are 3 key aspects that influenced why AWS would be used as the cloud computing provider in the Obama campaign:

1.   Security and Compliance

Elections draw some of the world's most aggressive information security threats. When it comes to election technology, information security is a major priority. AWS understands election administrators' responsibilities and meets or exceeds security and compliance standards at every level of our customers' cloud journey. AWS prioritizes data security, and our worldwide infrastructure is developed and managed in accordance with security best practices.

2.   Voter Engagement

In 2018, all millennials (those aged 18 to 29) were eligible to vote in the United States for the first time. Millennials prefer online transactions and have high expectations for tailored customer experiences. AWS offered building blocks that can be quickly assembled to support virtually any secure workload for targeted outreach.

3.   Elections Management:

Elections Management refers to back-office duties like voter registration that serve as operational efficiency drivers across multiple linked systems, applications, and local organizations spanning counties and districts. AWS offers a number of database services to help with voter registration. These fully managed systems may be launched in minutes with a few clicks. Furthermore, the AWS Database Migration Service facilitates a simple and cost-effective transition to the AWS Cloud.

How it was done:

- The primary registry of voter file information was a database hosted on Amazon RDS. This database combined data from various sources (including www.barackobama.com and donor information from the finance team) to provide campaign managers with a dynamic, fully integrated picture of what was going on.
- This collection of databases enabled campaign workers to target and segment prospective voters, shift marketing resources based on near real-time feedback on the effectiveness of specific ads, and

power a donation system that raised more than $1 billion (the 30th largest ecommerce site in the world).

The Obama campaign's apps are equivalent in extent and complexity to those seen in the largest companies and data-rich startups.

To give a point-by-point example of how the election campaign made use of apps that were available on the AWS cloud platform and performed tasks both complex and massive in scale:

- Vertica and Elastic MapReduce are used to model massive amounts of data.
- Multi-channel media management via TV, print, online, mobile, radio, and email with dynamic production, targeting, retargeting, and multi-variant testing, similar to what you'd find in a competent digital media agency.
- Coordination and collaboration of volunteers, contributors, and supporters on a social level.
- Large-scale transaction processing.
- Voter abuse prevention and protection, including incident collection and volunteer deployment.
- A comprehensive information distribution system for campaign news, polling, topic information, voter registration, and more.

Since the 2016 U.S. presidential election, Amazon Web Services has quietly increased its presence in state and local elections; more than 40 states now use one or more of Amazon's election offerings as do America's two major political parties, Democratic presidential candidate Joe Biden, and the federal agency in charge of enforcing federal campaign finance laws.

While it does not handle voting on election day, according to company documents and interviews, AWS now runs state and county election websites, stores voter registration rolls and ballot data, facilitates overseas voting by military personnel, and helps provide live election-night results.

Nonetheless, Amazon's growing presence in the elections industry may jeopardize what many officials regard as a strength of the US voting system: decentralization.

While most security experts agree that while Amazon's cloud is likely to be much more difficult to hack than the systems it is replacing, putting data from multiple jurisdictions on a single system raises the possibility that a single major breach could be disastrous. "It makes Amazon a more attractive target for hackers" and "increases the difficulty of dealing with an insider attack," said Chris Vickery, director of cyber risk research at cybersecurity startup Up guard.

The privatization of voting infrastructure is part of a larger trend that has swept across nearly every aspect of government in America, from parking tickets to prisons, and is continuing under the Trump administration.

According to companies that partner with both firms for government contracts, Azure, AWS's main competitor, has a sizable government business and offers some election services, but it has not focused on them and lags behind Amazon.

Questions to consider:

1. What are the advantages of putting elections on a cloud platform?

2.How is decentralization considered a threat?

3. Read and comment on how AWS used sentiment analysis to reflect on the inauguration speeches of Obama vs Trump and the conclusions made (Medium, 2023)