Co-funded by the
Erasmus+ Programme
of the European Union

Pathway in Enterprise Systems Engineering (PENS)

Project Ref. No.: 586301-EPP-1-2017-1-PS-EPPKA2-CBHE-JP

http://www.pens.ps

Industrial Training Use-Cases

Executive Summary

This documents presents the case studies proposed to students, as part of the PENS project.

The case studies covers the following themes:

- Data integration
- Business transformation
- Big data
- Security
- Business process engineering
- Integration and Business Intelligence

The case studies are about an e-health application and waste management systems where IoT and Big data technologies are deployed and used.

Table of Contents

Table of Contents

List of Abbreviations

IoT......................Internet of Things

WSN...................Wireless Sensor Networks

1 Use Case 1: IoT and Big Data based use-case.

1.1 Introduction

This section introduces the projects' scope and describes the e-health use case context.

1.2 Purpose and Scope

The goal of this application is to allow students discover an industrial use-case where new IoT and Big Data technologies are used. They learn, through thisuse-case, how to design a software architecture of an IoT system ranging from data collection layer up to the processing layer. The application presents, additionally, some security challenges. Indeed, the students are asked to apply variant access control rules to the data base users.

The project deals with an IoT e-health system for a smart hospital. The system's specifications are described and the students are asked to design the system following the specifications. For implementation, they have to use the Big data and IoT technologies presented during the projects' training days. For confidentiality reasons, ProxymIT does not communicate client real data and proposes instead a public data from Kaggle web site.

1.3 Project's context

In e-health systems, multiple IoT (Internet of Things) devices called sensors are deployed to collect data. Data is processed to monitor and remotely control patients' state and environment. Indeed, data is stored and analyzed to allow for automated decision-making and to protect the well-being of patients. These decisions can take place at the edge of the network for minimal latency (at the gateways) or can be sent to server-side Big Data systems for storage and for remote diagnostics. In addition, in the case of alerts, notifications are sent in real time to the caregivers concerned.

Figure 1 shows an example of an e-health system deployed in a smart hospital. Body sensors are installed at the patient level to monitor their health status (e.g. heart rate, blood oxygen saturation and sleep patterns). Other temperature and pressure sensors are installed in his room.

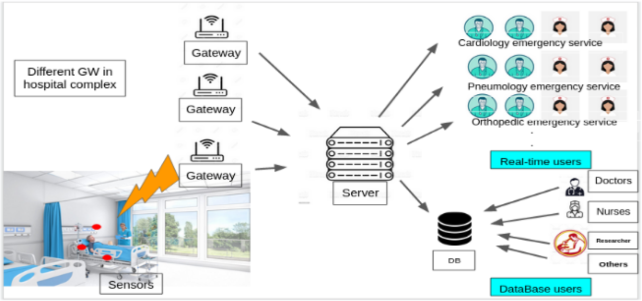


Figure 1 E-health system project components

1.4 Project's specification

This section describes the system's components and the dataset to use.

1.5 System's components

Gateways serve as intermediaries between the sensors and the remote server. They make it possible to collect several data from the sensors, filter, aggregate them and perform initial processing before sending them to the server.

The server deploys Big Data software components for collecting data from different gateways before disseminating this data to different types of users.

We distinguish two types of users:

• 'real-time' type users who are emergency department nurses and doctors who have to manage urgent interventions.

• Database users who are doctors and researchers who have access to paients' information in batch mode (after it has been recorded in the database).

1.6 Dataset description

We use the following dataset extracted from a Kaggle competition:

https://www.kaggle.com/datasets/johnsmith88/heart-disease-dataset

The dataset « heart.csv » dates from 1988 and groups data from four databases: Cleveland, Hungary, Switzerland, and Long Beach V. It contains 76 attributes but all published experiments refer to using a subset of 14 of them. The "target" field refers to the presence of heart disease in the patient. It is integer valued 0 = no heart disease and 1 = heart disease.

Dataset features' description:

- age - age in years (varies between 29 and 77)
- sex - (1 = male; 0 = female)
- cp - chest pain type :

  0: Typical angina: chest pain related decrease blood supply to the heart

  1: Atypical angina: chest pain not related to heart

  2: Non-anginal pain: typically esophageal spasms (non heart related)

  3: Asymptomatic: chest pain not showing signs of disease
- trestbps - resting blood pressure (in mm Hg on admission to the hospital) .The range of values is between 94 and 200. Anything above 130-140 is typically cause for concern
- chol - serum cholesterol in mg/dl : above 200 is cause for concern. The values varies between 126 mg/dl and 564 mg/dl .
- fbs - (fasting blood sugar > 120 mg/dl) (1 = true; 0 = false) ,above126' mg/dL signals diabetes
- restecg - resting electrocardiographic results :

  0: Nothing to note

  1: ST-T Wave abnormality: can range from mild symptoms to severe problems/ signals non-normal heart beat

  2: Possible or definite left ventricular hypertrophy > Enlarged heart's main pumping chamber
- slope - the slope of the peak exercise ST segment :

  0: Upsloping: better heart rate with exercise (uncommon)

  1: Flat Sloping: minimal change (typical healthy heart)

  2: Downsloping: signs of unhealthy heart
- thal -a blood disorder called thalassemia : 1 = normal; 2= fixed defect; 3= reversible defect.
- ca - number of major vessels (0-3) colored by fluoroscopy . It is an ordinal variable whose values are between 0 and 4.
- oldpeak - ST depression induced by exercise relative to rest looks at stress of heart during exercise. An unhealthy heart will stress more. The values varies between 0 and 6.2
- thalach - maximum heart rate achieved.The values vary between 71 and 202.

1.7 System's architecture

We simulate the gateways with a program that reads data from a file or generating data towards a message broker, Kafka.

Kafka contains at least two topics: 'normal' and 'alert' messages. Both messages go to the database throw a sink connector. We adopt MangoDB as a database.

Figure 2 shows the system architecture we target in this project.

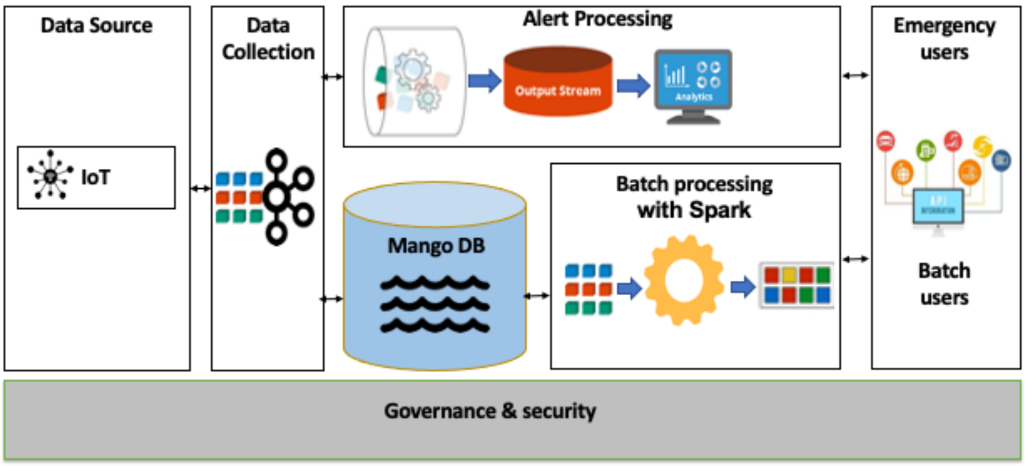


Figure 2 E-health system architecture

1.8 System's deployment

The system's deployment is based on Docker technologies. Docker allows to create lightweight virtual machines called containers. Each container can run a set of software and be independent of the containers. This technology is very adapted to test a Big Data environment in a local machine before production deployment.

For more information about Docker, please visit : https://www.docker.com/

First you need to install Docker on your machine.

Web sites to install Docker:

https://docs.docker.com/install/

https://docs.docker.com/compose/install/

To test that your installation is OK.

```
docker --version
Docker version 20.10.12, build e91ed57
```

```
docker-compose --version
docker-compose version 1.29.2, build 5becea4c
```

In Docker folder, you have the .yml file to use. Open this file to check all the software to install.

The following command allows Docker container installation.

```
docker-compose up -d --build
```

To check that containers are created

```
docker container ps
```

To get information about installed images

```
docker container ls -a
```

To connect to one container

```
docker exec -it <containerName> bash
```

Example

```
docker exec -it mango bash
```

To exit from the container, type the following command:

```
exit
```

To stop the docker containers

```
docker-compose down
```

NB: Sometimes, it is necessary to clean you machine completely from downloaded images and containers

```
docker container stop $(docker container ls -aq) && docker system prune -
af --volumes
```

```
docker container rm $(docker container ls -aq)
```

1.9 Work to do

1.10 Big Data storage

First log on the MangoDB DB and manually add a collection from the dataset.

Create a python program that copies from the dataset csv file to the database.

Check the content of the Database

1. Security roles and DB access control
2. Spark program to analyze data
3. Kafka topic creation and connectors' configuration

Each part will be explained into details at the next steps.

1.11 Documentation and presentation

Once the work is done, it is asked to prepare a presentation of 3 or 4 slides where the following questions are answered:

1. Explain why Big Data technologies are used in the realization of this type of system?

2. Which part is in 'batch' mode and which part is in 'streaming' mode?

3. Explain in the context of this e-health system the role of each layer and propose, for each layer, the Big Data technology to be used, justifying your choice.

4. The Message Broker is a very useful component in this type of system. Justify its usefulness.

5. What are the 'topics' that seem relevant to you for this system at the 'Message Broker' level?

After the system presentation, a demo is required where the work tasks are illustrated through examples.

1.12 References

[1] Kadhim, Kadhim Takleef, et al. "An Overview of Patient's Health Status Monitoring System Based on Internet of Things (IoT)." Wireless Personal Communications 114.3 (2020).

[2] Tawalbeh, Lo'ai, et al. "IoT Privacy and security: Challenges and solutions." Applied Sciences 10.12 (2020): 4102.

[3]. Chenthara, Shekha, et al. "Security and privacy-preserving challenges of e-health solutions in cloud computing." IEEE access 7 (2019): 74361-74382.

2 Use Case 2: Waste Management System (WMS) for Municipalities

2.1 Use Case Pre-requisites

- Students have successfully conducted course on Enterprise Systems and know what are the following tools: HRMS (Human Management System), Asset Management System (AMS) for inventory
- Existing functional VM or Container ODOO ERP offering the following capabilities:
    - HRMS with list of Municipalities employees sorted by category including Truck drivers.
    - Asset Management system (or AMS, allowing tracking of the available Waste Trucks with indication of their status (available, etc.)
    - The Municipality Employees Directory System (EDS) typically Microsoft Active Directory or Open-LDAP. This employee's directory holds details of each employee together with each role. We will assume the following roles are configured in the municipality directory: "Truck Driver" and "Supervisor".

    This could be the outcome of previous lab developed with the context of the "Enterprise Systems" course.
- Students have successfully conducted course of "Systems and Data Integration" and know what is a secured API and the process of securing APIs, etc.
- Secured API published to access the Municipality HRMS and Inventory management system described above.
    - This could be an outcome of a lab conducted within the framework of "System and Data Integration".

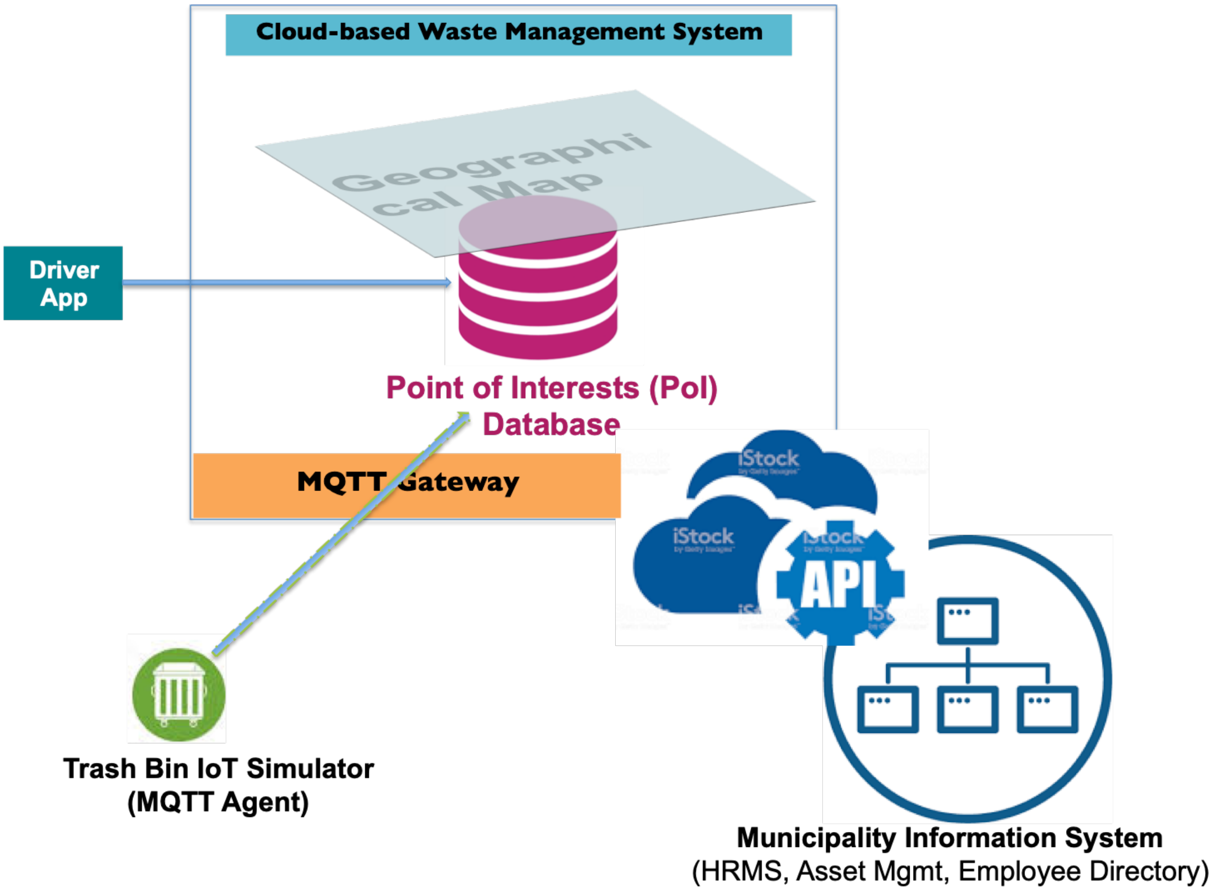
2.2 Use Case Objectives

- Build a complete use case for Waste Management system (WMS) for municipality leveraging existing enterprise assets (HRMS, EDS and AMS) and cloud-based new tools (to be developed as Lab in this training).
- Create awareness about the Bi-modal architecture and what needs to remain under-control of the Municipality (on-premise) and what makes sense to be deployed on the cloud.
- Define data exchange (API) contract between enterprise solutions of the Municipality Information System and the WMS solution
- Define and implement authentication methods on the mobile application and the WMS leveraging the Municipality Employees Directory System and various authentication options (REALM, OAuth, form-based authentication).
- Build mobile application for Truck drivers and WMS supervisors.
- Create simulators for smart Trash containers/bin with filling algorithm to be discussed based on knowledge from the "System Modeling and Simulation" course.

- Build control panel to be used within the WMS to plan “tourneys” of Truck drivers for the Day D+1.
- Implement Dashboard based on “Business Intelligence and Data Analytics” course to measure the following SLA and KPI Indicators and SLA (Service Level Agreement)
    - SLA of maximum duration a filled Trash ban has been emptied
    - Indicator of respect of Tourney by each Truck Driver
    - Cost of collection (per Tone) per zone
    - Visualize red zones where new trash bans need to be placed.
- Create awareness about artificial intelligence and data analytics potential to:
    - Automatically calculate Tourneys based on level of filling of Trash bans and the planned available resources (drivers & tracks)
    - Optimization (time and cost) of tourneys based on traffic hours, the past historical pattern of speed of filling of the bans.
- Learn role-based development: how to develop solution (the Driver mobile app) whose features presented to the connected user depends on the actual role of the connected user (Truck Driver or Supervisor).

2.3 Target Architecture

The following diagram presents the target architecture of the Waste Management Case Study.



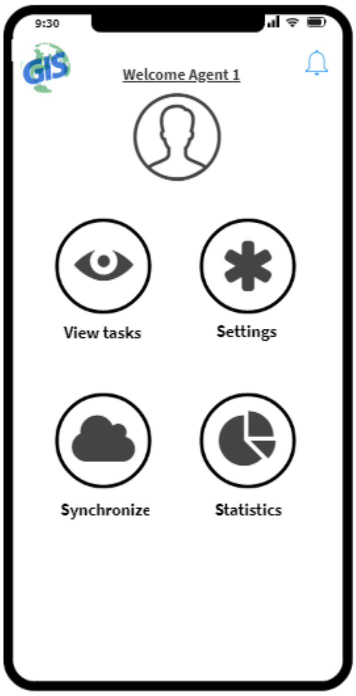
2.4 Scope and Solution Components

As outlined above the building blocks are:

- Geographical base for Point of Interests (named 'PoI-database'): This is a database powered by Postgres-GIS open-source software that will save all the places and information related to point of interests (PoI) of Waste Management such as: trash bins, waste deposit zones, etc.

- Driver Agent App: this is a mobile application that will be used by Municipality truck drivers on the field and their supervisors. This app is used in the case study to substitute GPS trackers usually on the waste collection trucks. Each Driver app is linked to particular truck. Features to be implemented in the Driver app:

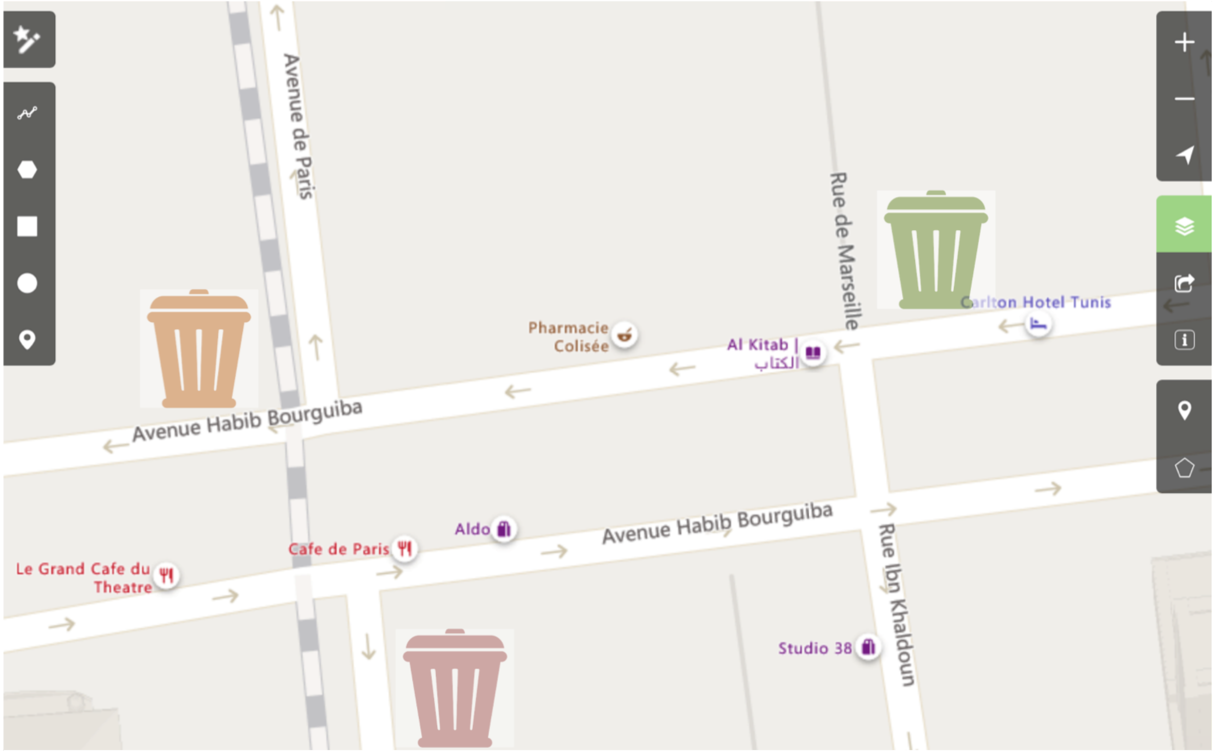
- Tracking GPS location and sharing it with the share with the waste management system.
- Connected driver makes ckeck-in with the app to locate nearby Trash bin and indicate its content has been emptied.
- When used by user having the 'Supervisor' role: the Driver app allows to add, remove or change characteristics (capacity, level of fullness, etc.) of trash bins on the municipality territory.



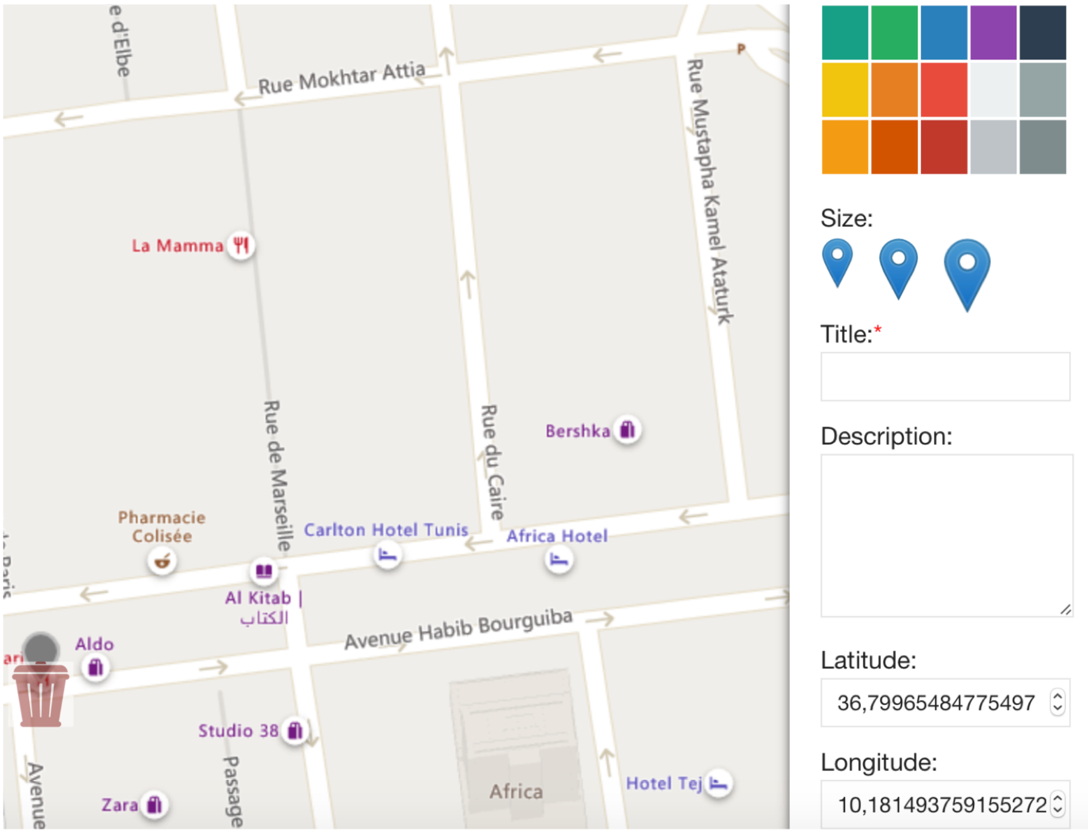
- Trash bin IoT simulator: these are MQTT simulator configured to send specific parameter to the WMS system such as location of the trash bin and the percentage of fullness (100 being 100% full). For each trash bin added on the WMS by supervisors, an associated instance of the IoT simulator will be created with an algorithm to simulate filling of a trash over-time. Various filling algorithms (such as Loi de Poisson, Fish Law) can be implemented with various speeds (such as trash bin in crowed areas are filled more often than less populated habitation areas). The IoT simulators will be provided for this lab.

- Geographical Map: this is a component that allows viewing by the Municipality supervisors the various Trash bins on the map with their status (filled, almost filled or empty). It basically allows viewing the PoI-database content on geographical map.

  The presentation of the trash bin should be presented as shown here-after with clear indication of the position of the trash container on the map and its fullness level.



- The Waste Management Web Console: it is a web portal that integrates the Geographical (above) together with menu to allocate the processing of a particular Trash bin to a particular Truck (driver) for the Day after D+1. The WMS Web console allows equally the management of the Trash bins as shown below.



.

2.5 Use Case Extension

The Use Case can be extended to cover additional courses and concepts. The following are possible suggestion of the work:

- Allocating work orders (assignments) for the Trash bins to be allocated to Truck Drivers can be modeled and made more complex using concepts of "Engineering of Business Processes" course. The engineering can be focused towards some real-life objectives such as:
  - Have maximum SLA for emptying an almost full Trash bin
  - Reduce cost of collection (cost of Tone of collected Waste)
  - Monitor the Quality of Service of outsourced Waste management of private companies (hence integrating new roles in the process).
- Dashboard and Business Intelligence to implemented using either Data visualization tools such as Tableau or Power BI or custom-made reports coded by students. This is an extension to the "Business Intelligence and Data Analytics" course.
- Data science analysis can be applied based on historical data of WMS to implement automatic (or semi-automatic) Work order allocation intelligence.

2.6 Resources

- Agreement on one or two target Cloud platforms should be agreed on (AWS, Azure, Google Cloud, IBM Cloud). Amazon AWS has already academic offers and agreements with Tunisian universities.
- Proxym-IT can undertake the development f necessary tools (Dockers images or VM) to make this Use Case executable by students. Offer to develop this can be financed by Staff costs pre-palnned on the project.
- IoT simulators (to simulate Smart Trash bin) exist in almost every cloud platform:
  - Amazon AWS IoT simulator: https://aws.amazon.com/fr/solutions/iot-device-simulator/
  - IBM Coud IoT device simulator: https://developer.ibm.com/iotplatform/2017/12/07/use-device-simulator-watson-iot-platform/