**1. REQUIREMENTS ANALYSIS**

Identifying information needs will allow the user to make decisions in accordance with the scope of their objectives.

To do this, questions must be drafted, the answers to which can satisfy these requirements.

* 1. **Identifying Business Questions**

**General Perspective (Company/Branch):**

What is the overall attendance percentage across the organization?

Which branches have the highest percentage of absences or absences?

How does attendance compare by branch over a specific period?

Which day of the week has the highest number of absences or tardies?

How many hours of overtime work have been recorded by branch or group in a month?

**Individual Perspective (Person):**

Which individuals have the highest levels of accumulated tardiness?

How many absences did each person accumulate during a quarter?

How has a specific employee's attendance varied over time?

What percentage of an employee's attendance is on schedule?

Which employees exhibit recurring patterns of lateness?

**Organizational Perspective (Group/Section):**

Which groups perform the best in terms of attendance?

How are absences and tardies distributed across different groups?

Which departments have the most overtime hours?

How do absences and tardiness affect the overall performance of each area of ​​the organization?

**Time Perspective:**

What is the overall attendance trend throughout the year?

Which months have the most absences?

How are lateness distributed during peak hours (e.g., morning vs. evening entry)?

Which holidays have the greatest impact on attendance?

**Compliance Perspective:**

What is the schedule compliance rate by branch or group?

What percentage of scheduled hours are approved as actually worked?

How does the average lateness affect overall productivity?

**Cost Perspective:**

What is the cost associated with absences and lateness in terms of hours not worked?

What percentage of time worked is approved overtime?

How much do unapproved hours represent in the overall productivity of the company?

* 1. **Identification of Indicators and Perspectives**

**Key Indicators (KPIs):**

• Minutes worked.

• Minutes late.

• Minutes missing.

• Attendance percentage.

• Average hours worked per person.

• Schedule compliance rate.

**Perspectives:**

• Person: Individual performance.

• Group/Section: Internal comparisons.

• Branch: Regional analysis.

• Time period: Comparisons between periods (daily, weekly, monthly, yearly).

**1.3 Conceptual Model**

The conceptual model is based on a star diagram, with a central fact of attendance and several related dimensions that allow the indicators to be analysed from different perspectives (person, branch, group, section, etc.).

**2. DATA SOURCE ANALYSIS**

**2.1 Formation of Facts and Indicators**

**Main Fact:**

The central fact is Attendance, which records the data derived from marking and evaluating the established schedules.

**Fact Indicators:**

• Minutes worked.

• Minutes delayed.

• Approved and unapproved minutes.

• Missing minutes.

• Overtime.

• Compliance percentage.

**Relationship with Dimensions:**

The fact is related to dimensions such as Person, Branch, Group, Section, and Time.

**2.2 Mapping**

Mapping involves identifying the relationships between data in the tables in the source system and the entities in the data warehouse model.



**Signs:** Primary source for calculating actual hours.

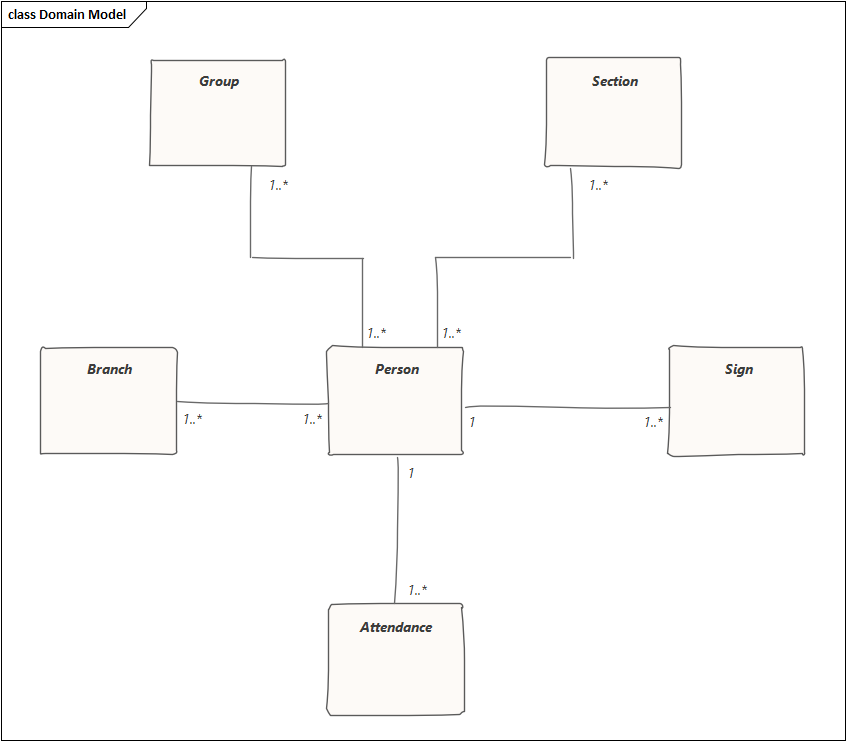
**Attendances:** Contains data based on punches.

**Person:** General data for each employee.

**Branch:** Identification of the location where the punching is done.

**Group and Section:** Organizational subdivisions of employees.

**Time:** Derived from the dates of punches and attendances.



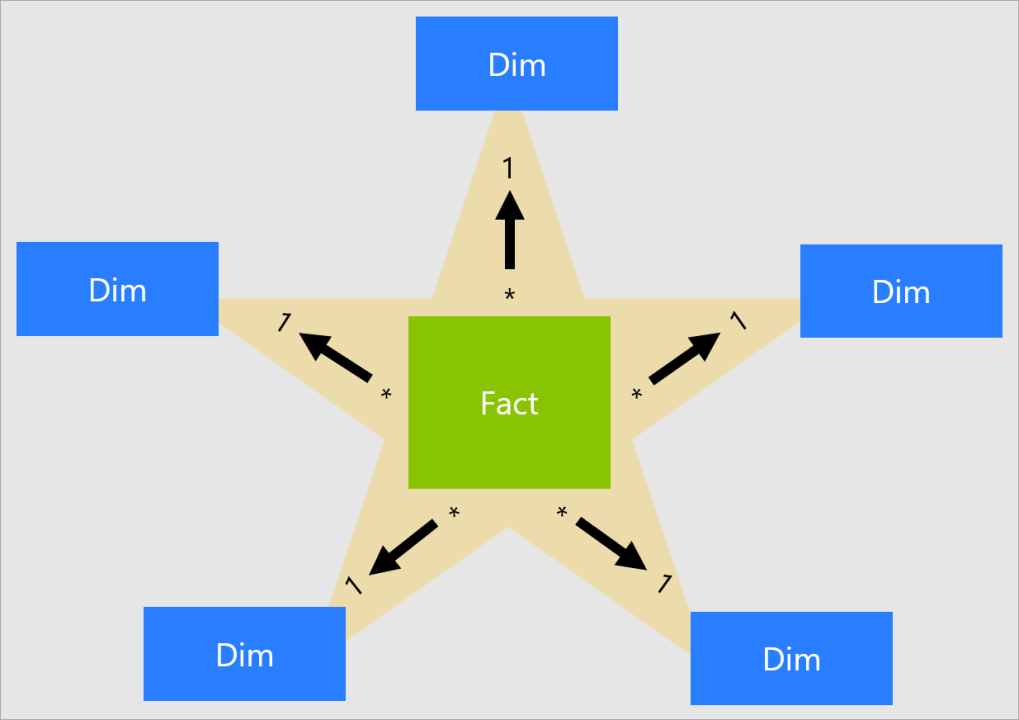
**3. DETERMINATION OF ARCHITECTURE**

**3.1 Logical Model Typology**

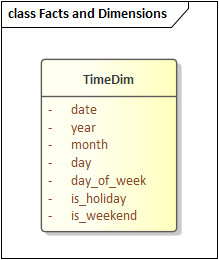
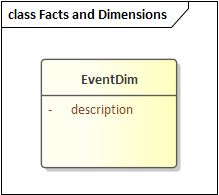
The topology is star-based, however when using Elasticsearch the architecture will be based on a denormalized model to optimize querying and visualization in Kibana. This implies:

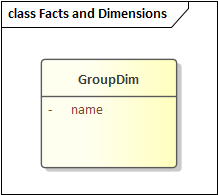
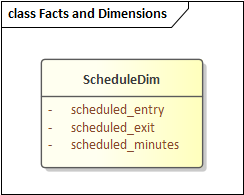
**Core fact:** Includes the key metrics and attributes of the denormalized dimensions.

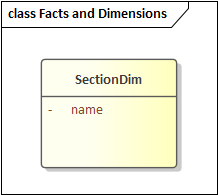
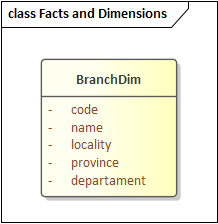
**Key dimensions:** Aggregated within the main documents of the fact index.

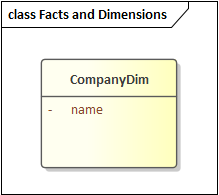
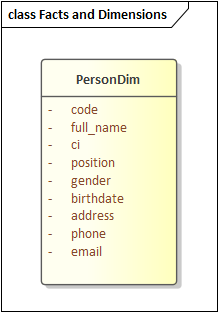


**3.2 Dimension Tables**

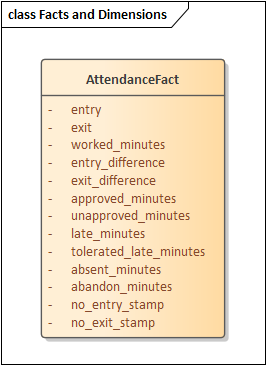
 

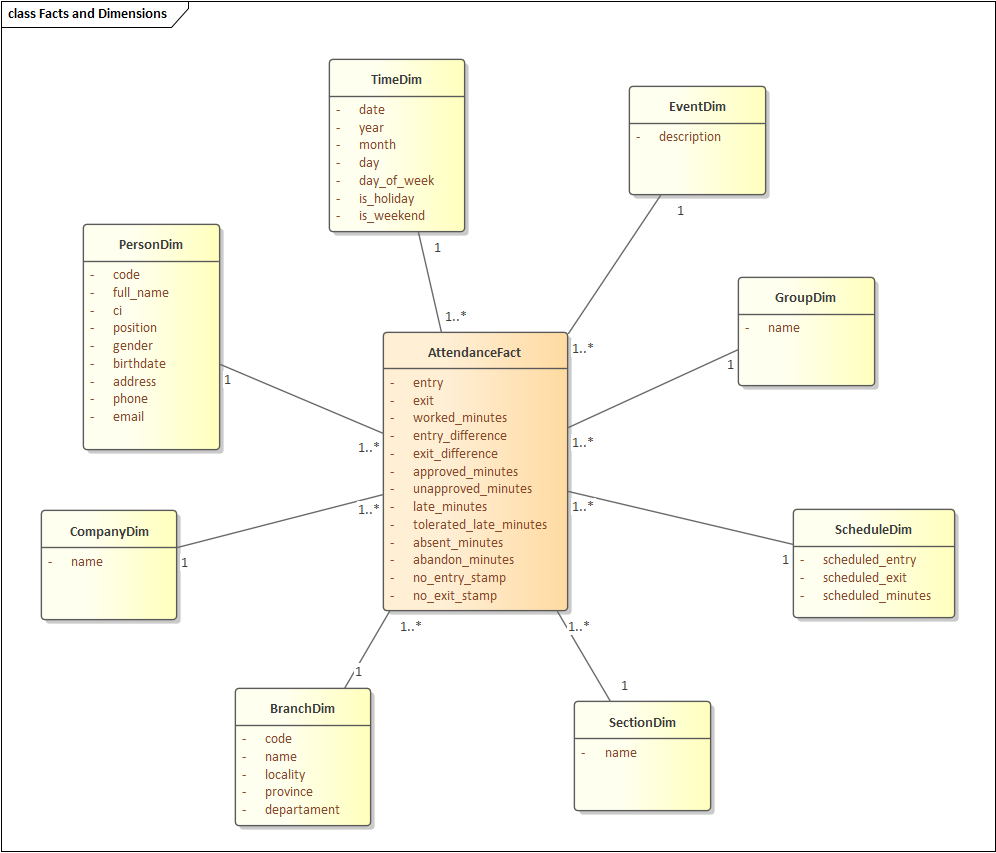
 

**3.3 Fact tables**

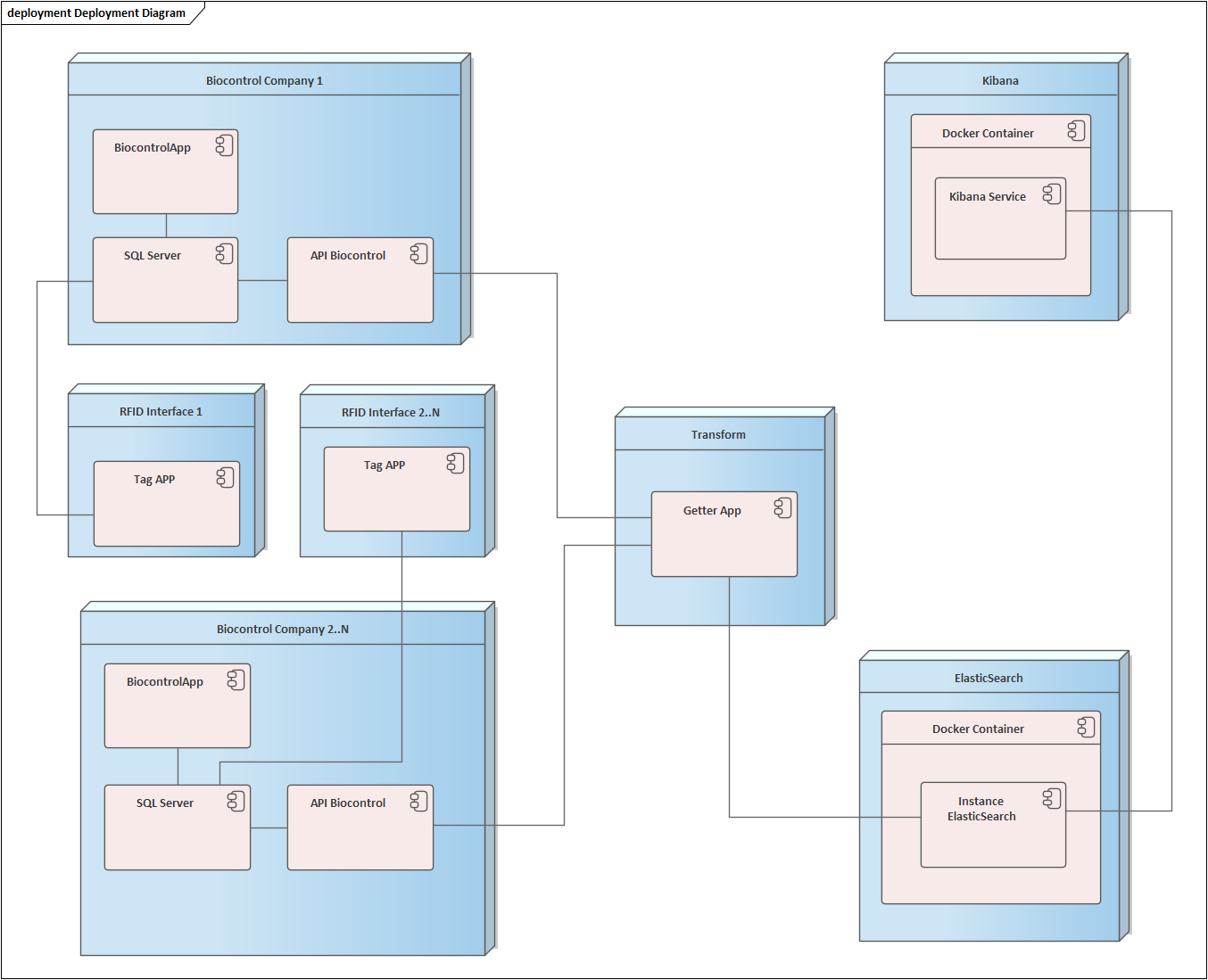


**3.4 Unions**



This design allows for fast queries in Kibana and is tailored to the needs of organizational analysis.

**4 ARCHITECTURE**



**Interface RFID (N)**

These are physical interfaces where company employees make their markings using RFID tags. Using integrated software that feeds the database in SQL Server.



**Biocontrol Company (N)**

It consists of:

**\* SQL Server:** SQL Server instance (Company database)

**\* BiocontrolApp:** Application where assistance is configured.

**\* Biocontrol API:** Interface to export data.

In the mentioned architecture there can be multiple companies, this is why our data warehouse design has support for multiple companies.

**Transform**

In this node, the transformation of the data coming from “Biocontrol API” to “Elasticsearch” is performed, using a program made in Python.

**Elasticsearch**

Within a Docker container, there is an instance of an Elasticsearch server, which is a highly effective non-relational database with an indexing algorithm.

The transformed data is saved so that Kibana can interpret it.

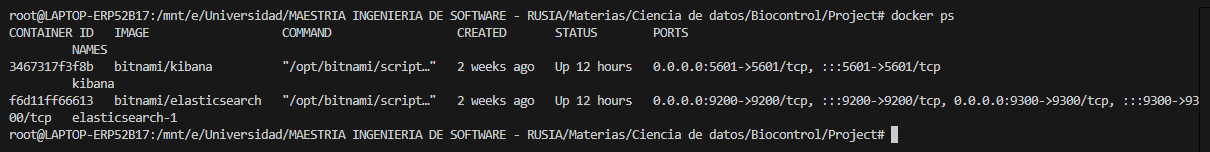
**Kibana**

Within a Docker container, there is the Kibana service which connects to the Elasticsearch database to perform the corresponding data analysis.

Kibana is designed to work with data stored in Elasticsearch and provides a graphical interface that allows you to explore, analyze and visualize large volumes of data in real time.

**5 IMPLEMENTATION**

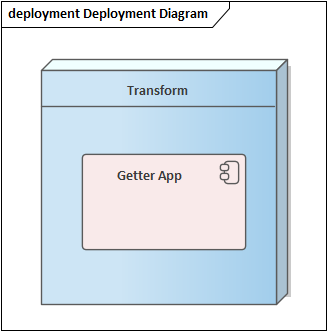
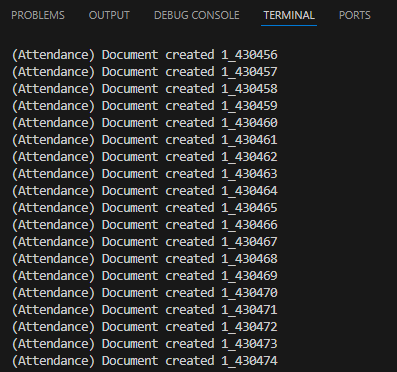
**5.1 Containerization of services**

****

Bitnami Docker images of Kibana and Elasticsearch were used

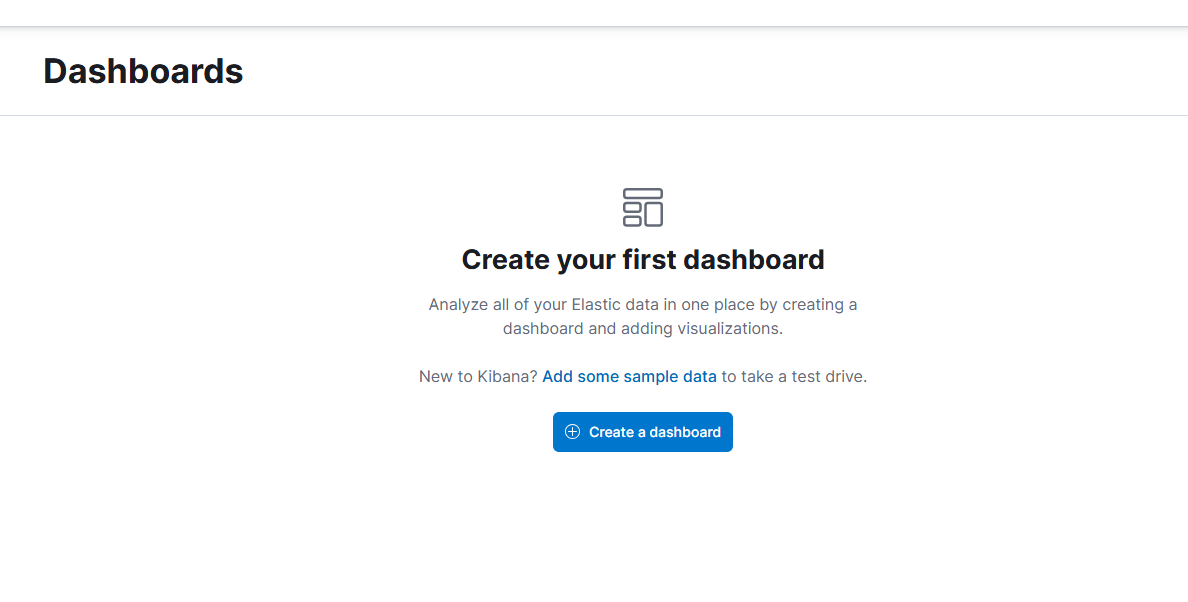
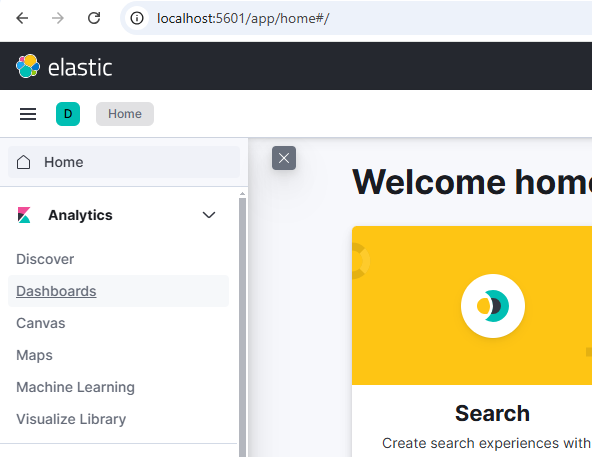
**5.2 Data transformation**

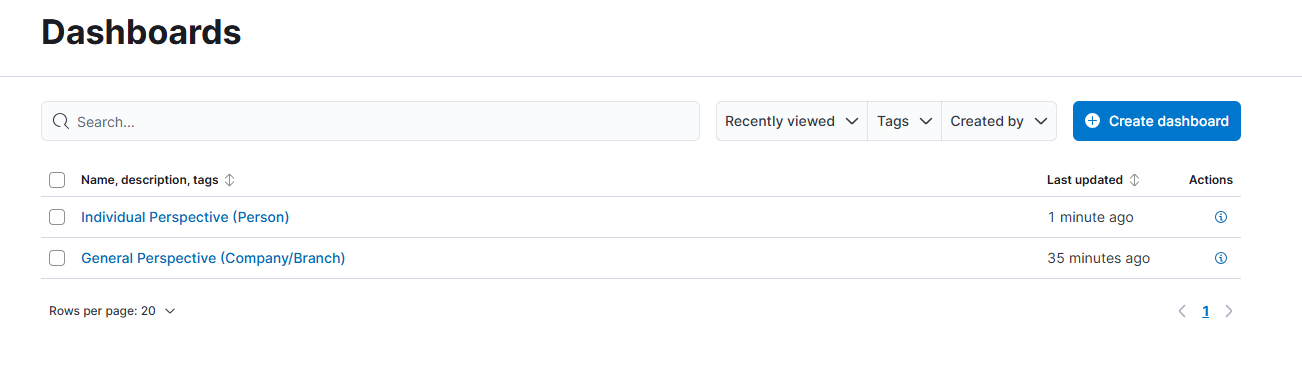
The data transformation was done through a Python program, which consumes the Biocontrol APIs, obtains the data and saves them in temporary dimension tables (Star Topology), and passes them to a denormalized model in Elasticsearch.

****

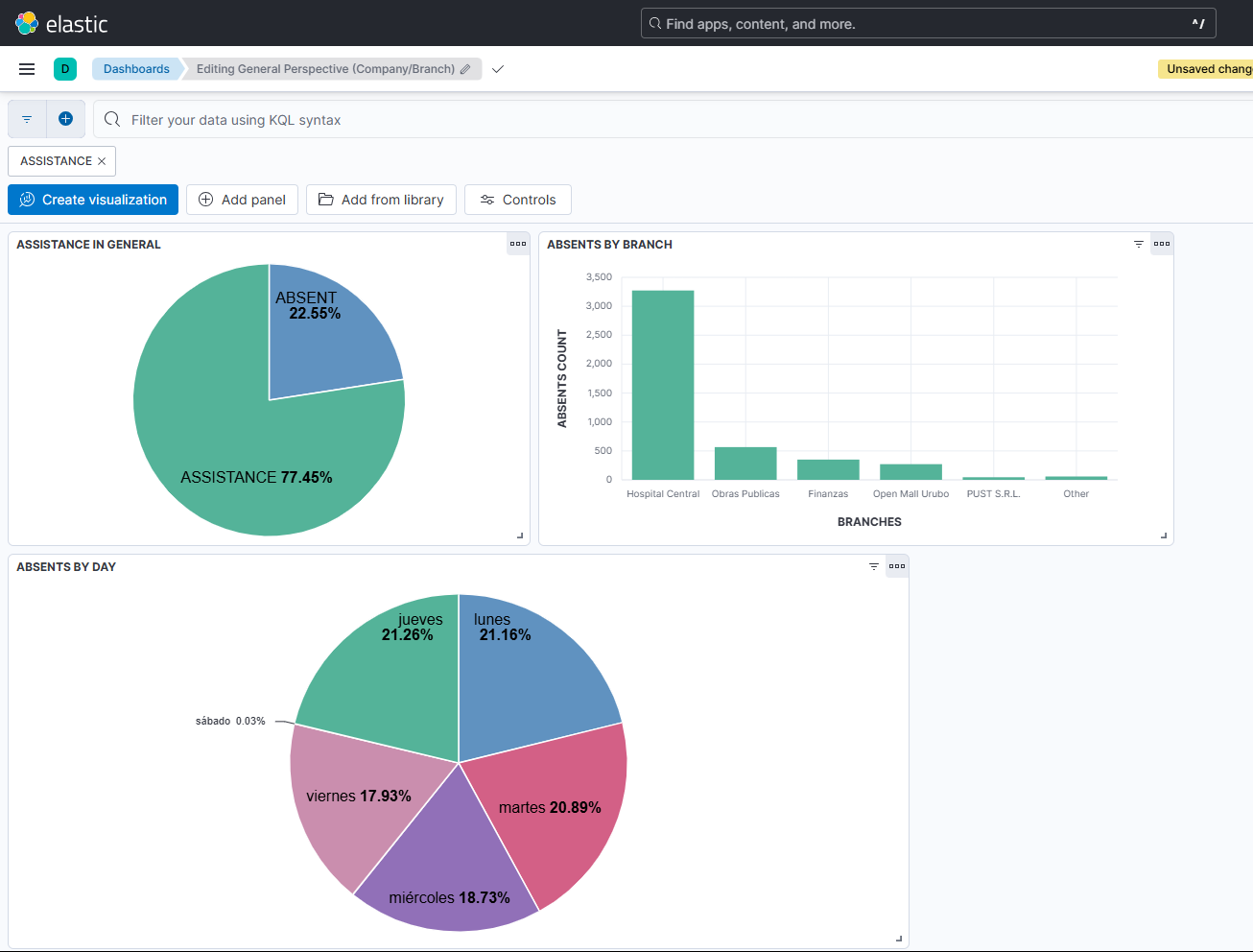
**5.3 Creation of Dashboards in Kibana**

Dashboards

****

Creating Dashboards according to business questions****

**General Perspective (Company/Branch)**

****

**Individual Perspective (Person):**

****