

# SQL Presto Tech

Computer Science  
Technologies for Big Data Management

## SUPERVISOR

Massimo Callisto De Donato

## TEAM

Peddiboina Mohan Vamsi  
Vaibhav Mathukiya



# CONTENTS

**Project Description &  
Objectives**

---

**Methodology and  
Technologies**

---

**Technical Implementation**

---

**Approach**

---

**Achieved Results &  
Conclusion**

# 1. Project Description

## **SQL Presto Tech**

A tool that enables Real-Time Data Analytics on data from devices IoT devices using technologies such as Kafka, MongoDB and Presto

# Objectives

## ❖ Messages

Handle and publish  
messages in real-time  
on Kafka Topics

## ❖ Storing

Store IoT data in a  
NoSQL Database

## ❖ Presto

Study and use Presto's  
SQL engine to perform  
queries on a NoSQL  
Database

## ❖ Analytics

Create a tool for  
analytics

## 2. Methodologies & Technologies

### General Architecture



# Technologies

## Apache Kafka

It is a distributed data streaming platform that can publish, subscribe to, store, and process streams of records in real time.

## MongoDB

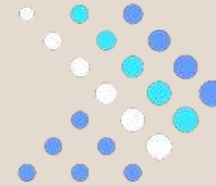
It is an open source NoSQL database that uses a non-relational, document-oriented data model to stores data objects

## Presto

It is a distributed SQL query engine that is open-source and optimized for high-speed analytic queries of data of any size.

## Power BI

It is a Microsoft business analytics service that enables users to visualize and share insights from their data through interactive dashboards and reports.





### **3. Technical Implementation**

# Apache Kafka

A distributed publish-subscribe messaging system used to stream messages that comes from the iot-simulator



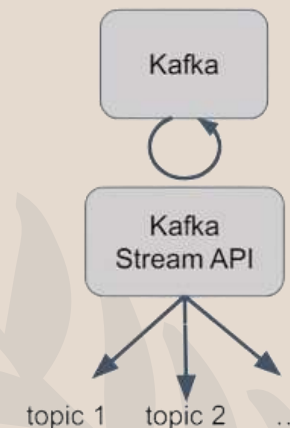
- In order to stream the messages is needed a MQTT-Source connector that connects to a MQTT broker and subscribes to the specified topics

```
{
  "name": "mqtt-source",
  "config": {
    "connector.class": "io.confluent.connect.mqtt.MqttSourceConnector",
    "mqtt.server.uri": "tcp://localhost:1883",
    "mqtt.topics": "#",
    "kafka.topic": "mqtt.echo",
    "value.converter": "org.apache.kafka.connect.converters.ByteArrayConverter",
    "key.converter": "org.apache.kafka.connect.storage.StringConverter",
    "confluent.topic.bootstrap.servers": "localhost:9092",
  }
}
```



## Apache Kafka

Kafka Stream API

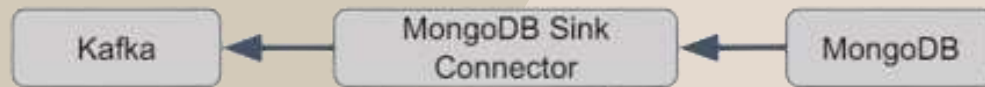


- It **simplifies** the development of real-time streaming applications.
- With Kafka Streams API, is possible to perform various operations on the data streams, such as **filtering**, **transforming**, **aggregating**, and **joining**, to enable **real-time analytics** and processing.



# Mongo DB

MongoDB is a document-oriented NoSQL database used to store messages streamed by Kafka.



- In order to address messages inside each mongodb collection is necessary to use the MongoDB Sink Connector, that reads data from Apache Kafka topic and writes data to MongoDB

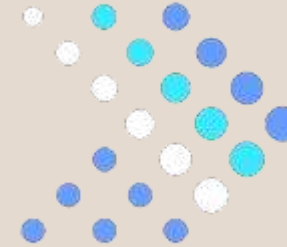
```
{
  "name": "mongodb-sink-2",
  "config": {
    "connector.class": "com.mongodb.kafka.connect.MongoSinkConnector",
    "topics": "mqtt.measures",
    "connection.uri": "mongodb://localhost:27017",
    "database": "tbdmproject",
    "collection": "measures",
    "key.converter": "org.apache.kafka.connect.storage.StringConverter",
    "value.converter": "org.apache.kafka.connect.json.JsonConverter"
  }
}
```

# Presto

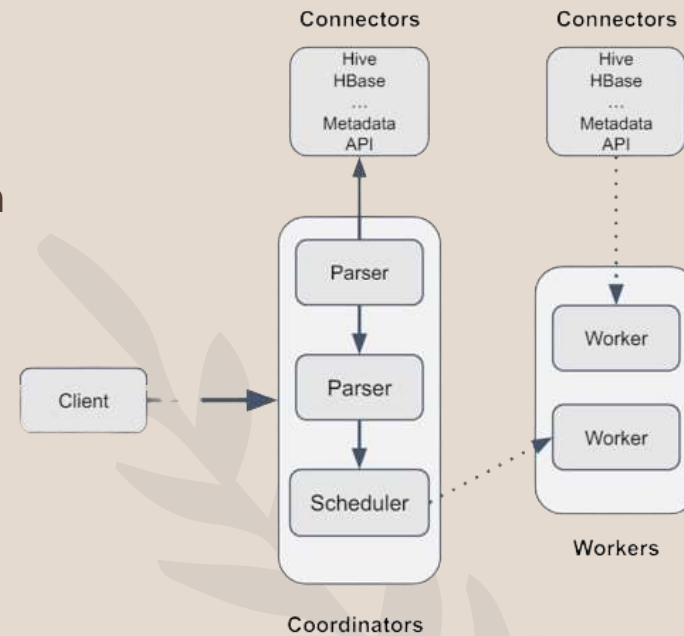
Presto is a distributed SQL query engine that supports non-relational sources.



- That is how Presto allows to perform SQL computation over MongoDB (NoSQL).

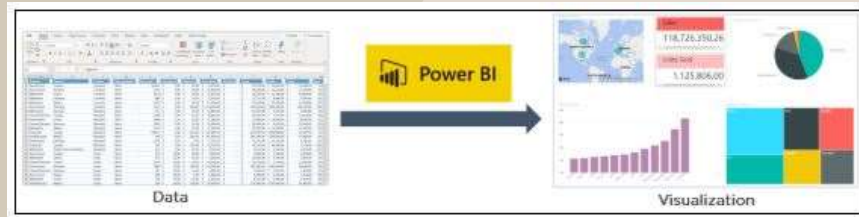


## Presto architecture



# Power BI

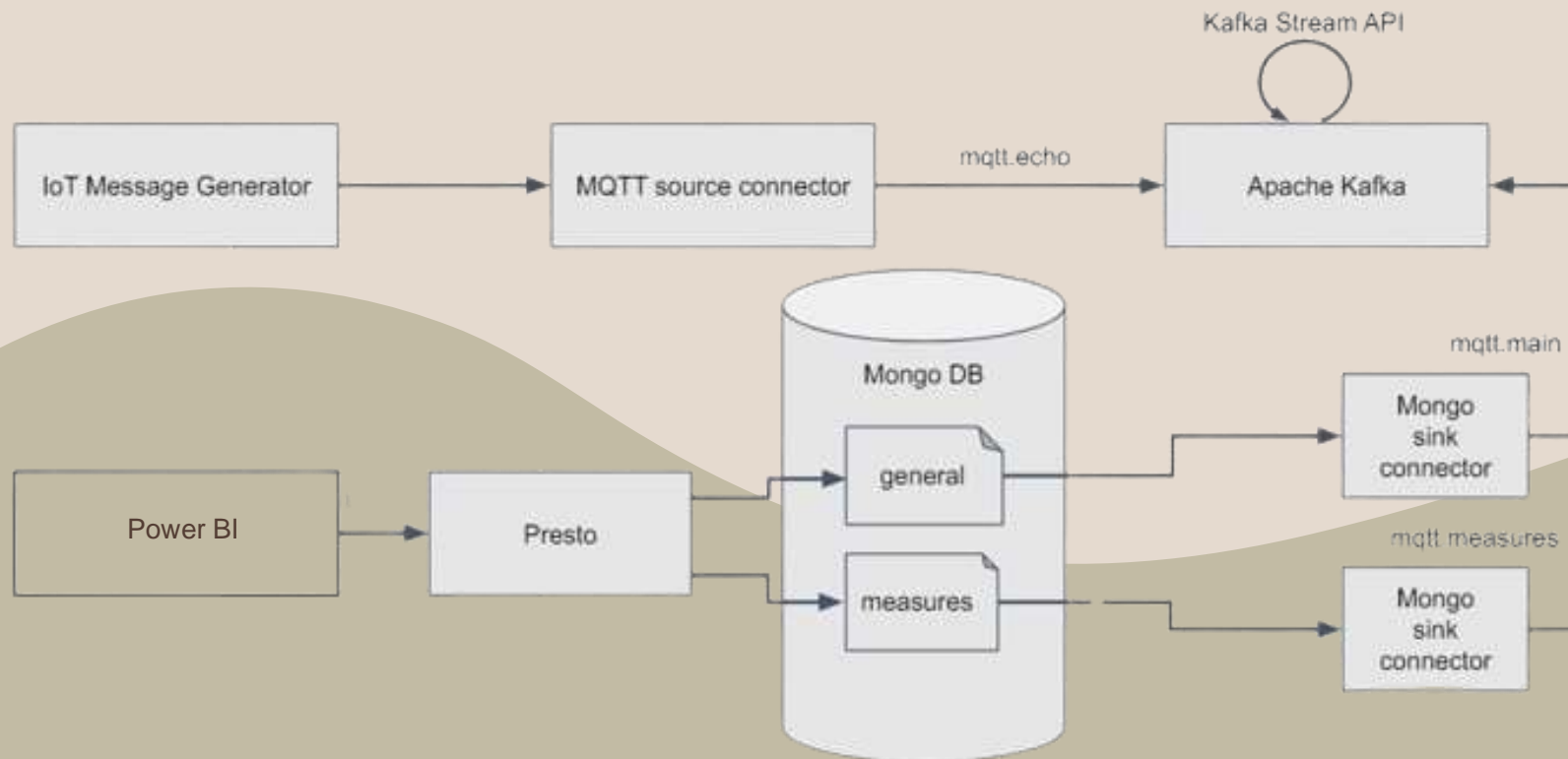
It is a Microsoft business analytics service that enables users to visualize and share insights from their data through interactive dashboards and reports.



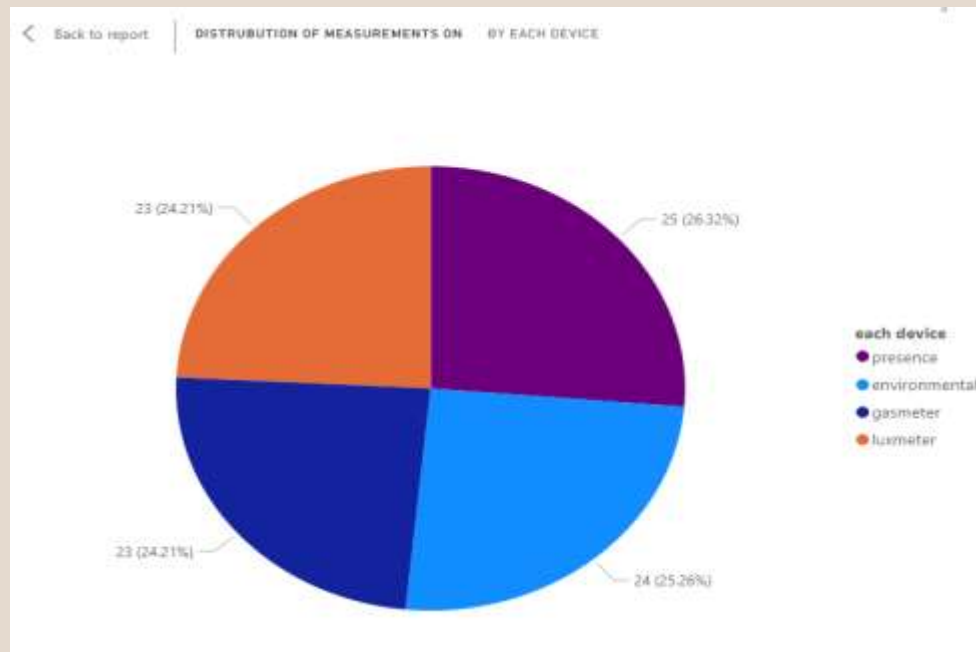
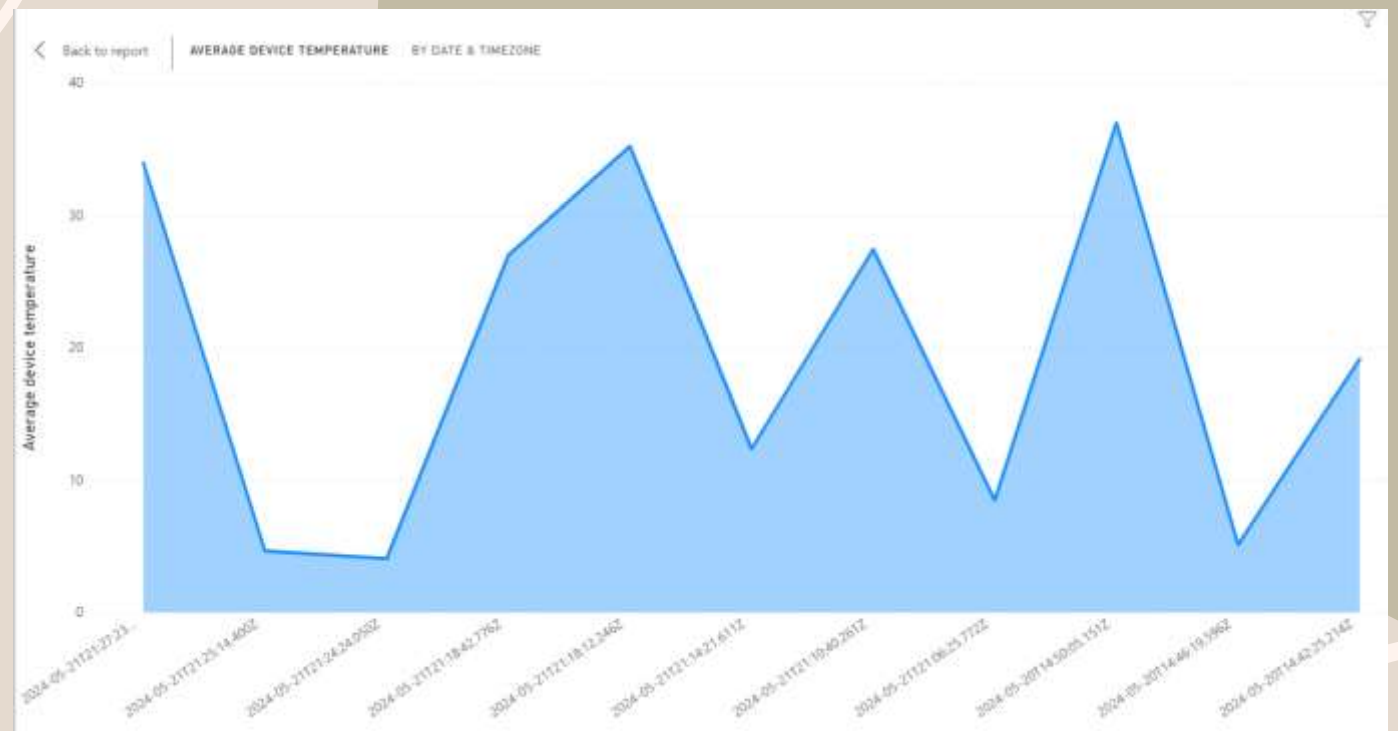
- While the platform is commonly used by business analysts, it is also designed to be easily accessible for those without any specialized data knowledge.



## 4. Approach



- +  Able to handle data in real-time
- +  Automatic Data flattening
-  Lack of community support



# Results

## Future implementation

- AI and Machine Learning Integration
- Advanced Analytics and Visualization
- Blockchain for Data Integrity



**thank you**