

# ISIT312

## Azure Database Solution

7068293	Mahek Sajid
6222560	Mohammed Maarij Uddin Khan
7057167	Rida Fatimah Asif
6866219	Rohan Sudhir
7068943	Samir Fazil
5840454	Kianoush Rahravan
7427359	Reyon Noronha
7352396	Ibrahim Siddiqui
6997053	Nageen Irfan Minhas



# Table of Contents



*Our company: FortyGuard*



Problem Statement



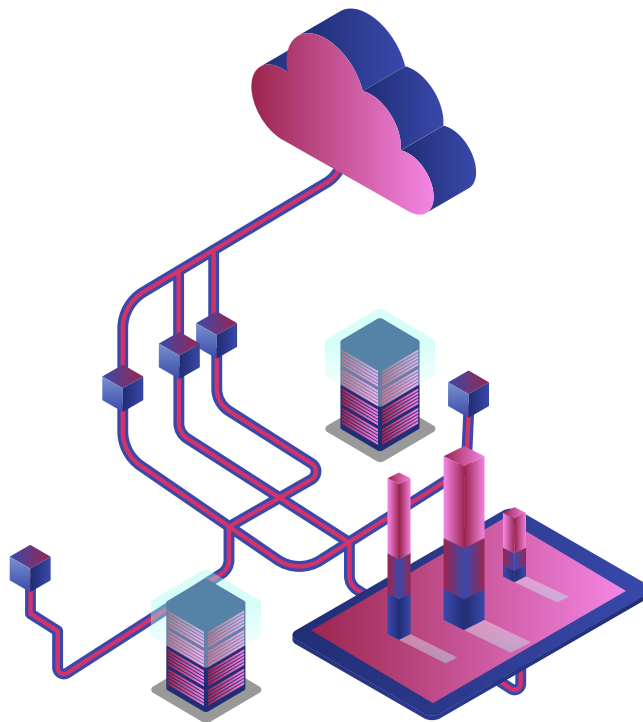
Implementation



Solution Demo



Evaluation



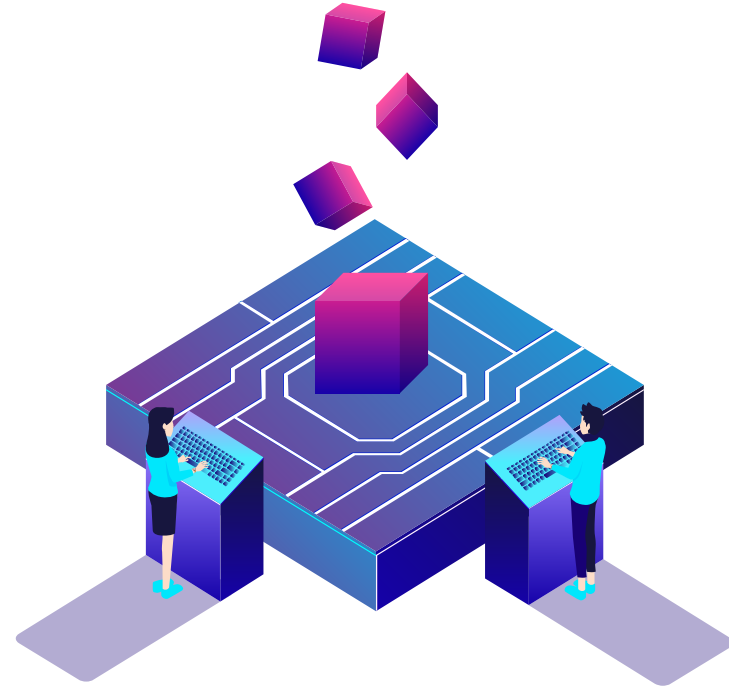
# *Our Company:* **FortyGuard**

FortyGuard, founded in 2020, provide expertise in **outdoor weather data analysis**.

They offer data science and predictive analysis, with personalized and dynamics visuals within their solution using their 'FortyEngine' which is their in-house AI cloud based platform.

The gathered analysis is made to use in fulfilling their mission “to cool cities”

*(FortyGuard, 2023)*

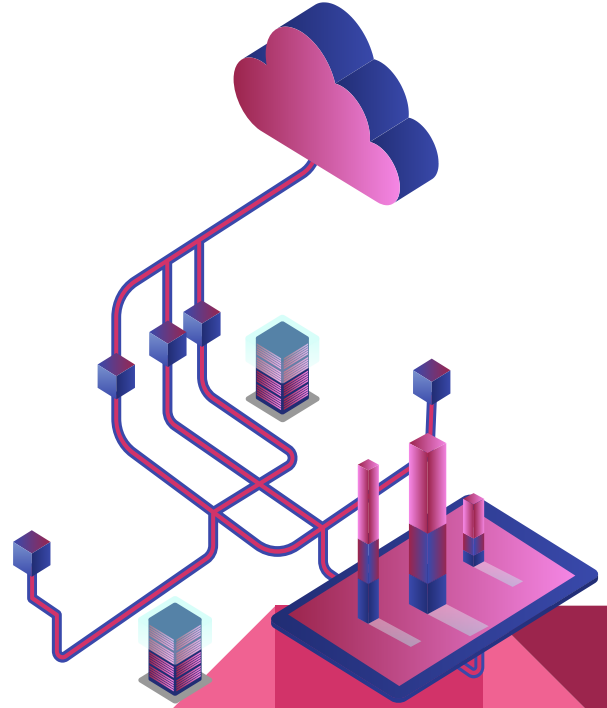


# Problem Statement

FortyGuard in building its solution requires an enormous amount of data collection and storage.

With multiple sources of incoming data and inconsistent data file formats in its collection, the problem domain relied on addressing the **batch processing issue with their data**.

They have data sources that provide monthly summaries of weather data and need a pipeline to handle this data stream.



# *Our Implementation:* Research Aspects



## Amount of data collected?

500 thousand data points over a period of 25-30 days.



## How is it collected?

From various sources through APIs from satellites, sensors and weather stations



## Where is it stored?

Hard drives, AWS Storage and Google Cloud



## How is it processed?

Stream processing at regular 5-minute intervals & monthly batch processing



## Tool for the integration?

Cloud based AI powered engine in combination with AWS services



## Potential issues

Multiple sources of incoming data and inconsistent data formats

## *Our Implementation:*

# Technical Requirements

We used **Microsoft Azure** for our proposed solution, as it hosts a variety of services, we needed to link our data lake architecture together.



Data Factory



Azure SQL Server



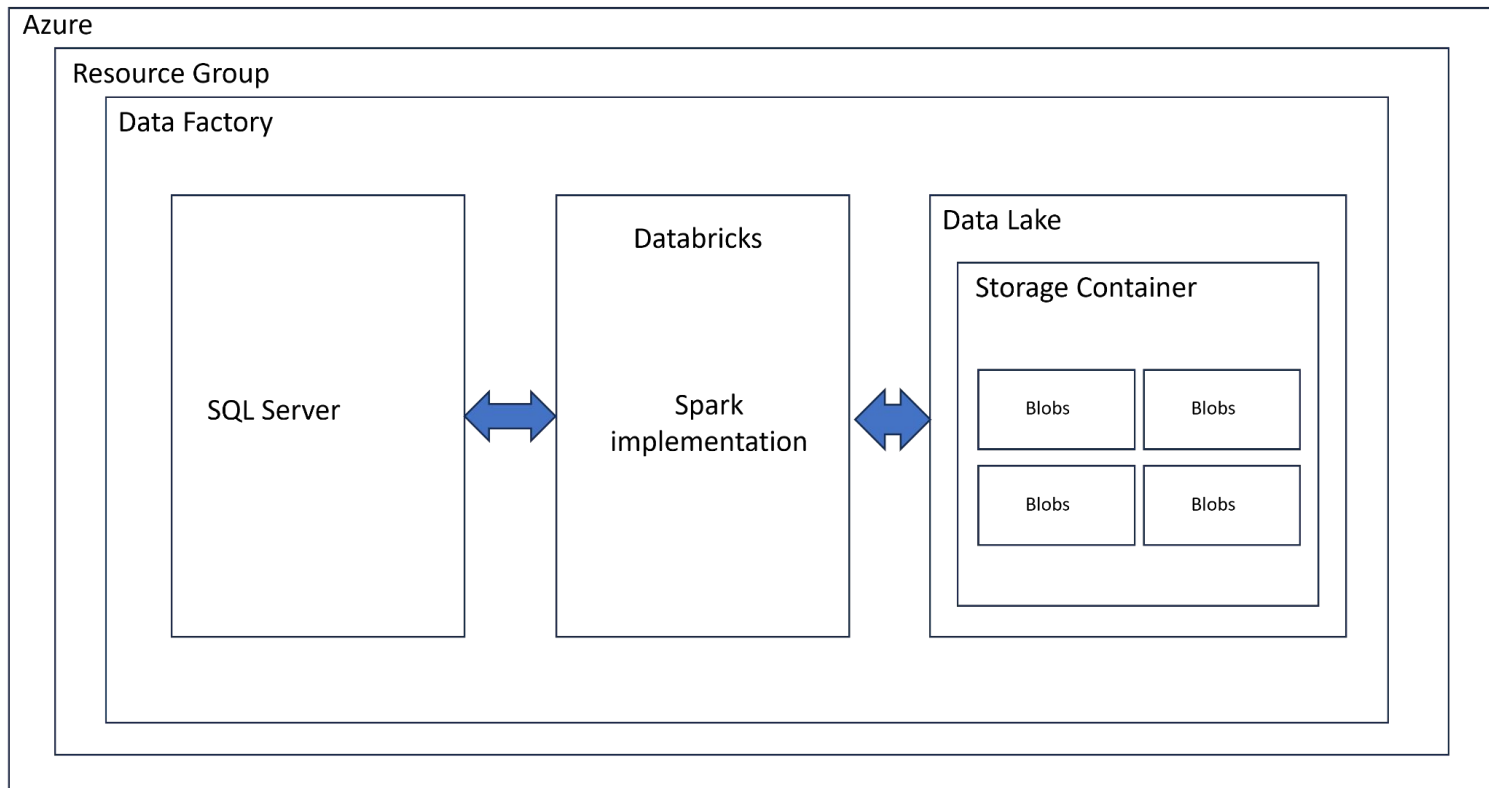
Azure Data Lake  
Storage Gen2



Azure Databricks



# *Our Implementation:* Solution Implementation



# *Our Implementation:* Solution Implementation

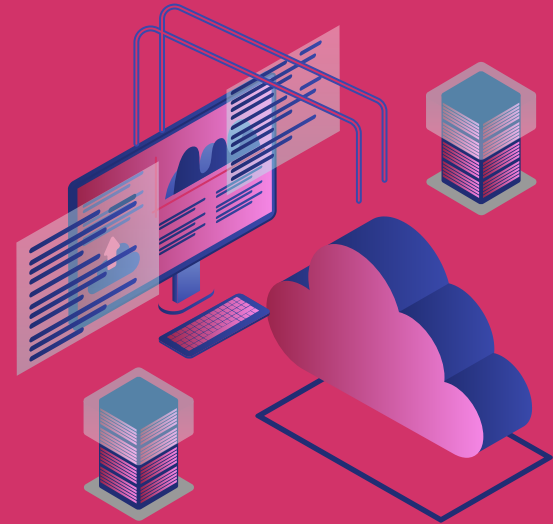
We implemented a multilayer solution using **Azure architecture services**.

- Using the **data factory** services, an **sql server** and a **data lake** was implemented to intake ***structured*** and ***unstructured data***. They are then linked to **Azure Databricks** which allows us to program specific operations on the data using python and spark.
- Using spark, we can code in ***Python*** to implement the necessary data operations such as predictive modelling, aggregations or transformations of the data for dashboarding. *For example purposes, we ran a simple operation of calculating aggregate data for multiple weather stations over a single month.*
- Azure Data Factory has a **trigger function** (*executes the pipeline according to a schedule*) that can ensure that this operation runs once a month, hence implementing batch processing. Furthermore, the trigger can be specified to run at smaller intervals in the case of stream processing data.



*Our Implementation:*

# Solution Implementation DEMO



# Evaluation:

1. Company found our implementation to be a comprehensive solution that was interesting and novel.
2. Achieved automation of batch processing
3. Scalability , security and efficiency
4. Valuable learning experience and potential for future exploration into Stream Data





# THANK YOU!

Any Questions?

# References

- FortyGuard, 2023, “On a mission to cool cities”, <<https://www.fortyguard.com/>>,  
Date Accessed: 21/06/2023
-