Cloud documentation:

Sprint 1:

Cloud and deployment team:

Identify the resources and limitations of each platform - Cloud and Deployment team

Compare AZURE and AWS to determine the better of the two cloud architectures

Sprint 2:

Cloud and deployment team:

Test deploy the required resources (Data lake, PostrgreSQL DB, Data Factory) and learn its functionalities

Create dummy scripts and trigger them using data factory peipelines and assess their effectiveness in data retrieval and storage

Sprint 3-4

Tasks:

Research and cloud support:

* Create dummy scripts and trigger them using data factory pipeline and assess their effectiveness in data retrieval and storage
* Research about ways to connect the instance of PostgreSQL DB to the front-end web application and test it

Cloud:

Discuss final reasoning behind switching the platform

* **Finalise the platform to base the system on.**
* Integrate and test automated scripts to make sure data is being updated of its own accord
* Connect Bot to DB instance to ensure SQL query requests being supported and retrieving the correct responses.

Final sprints:

-

Finally different datasets and pipelines were created on Azure Data Factory for the data in transit between different resources. These pipelines were triggered to completely automate this process.

* Ensure data pipeline automation and monitoring the platform to work on any data discrepancies.

**Challenges faced and how we tackled them**

1. We started with AWS as our cloud platform but since the student account of AWS does not allow us to create users and roles which we need to create pipelines and triggers, we moved to Azure
2. To transfer data from local file system to Azure data lake, we needed to have a secure way to keep the data safe during transit, so we created a self hosted Microsoft Integration Runtime
3. Once the data was stored on the lake as a csv file, we had a few discrepancies in the data since the column names contained commas. To deal with this issue, we converted all csv files as pipe (|) delimited files