

1. Conclusion

This poster summarises the research and development undertaken to produce a weather database platform that could replace legacy software currently used within Stantec. Firstly, a literature review was undertaken to understand the potential solutions for each component of the project that would need to be implemented. Using the Scrum development methodology the project was broken down into four development sprints; Data Acquisition, Cloud Deployment, ETL Development and Power BI Development.

Overall, the platform that was developed showcases a successful proof of concept that meets the objective of the project.

2. Introduction

The main aim of this project is to design and develop a cloud-based solution to replace a current deprecated piece of software called Infinity. It was used to collect and store the environment agency's open-source weather data across England from 2017 onwards. **This data is invaluable in the work that we do at Stantec, as most of the work we do in the UK is for the water industry.**

Objective

To research the feasibility of creating a replacement for Infinity, and with that, develop and implement the historical data load aspect of the platform.

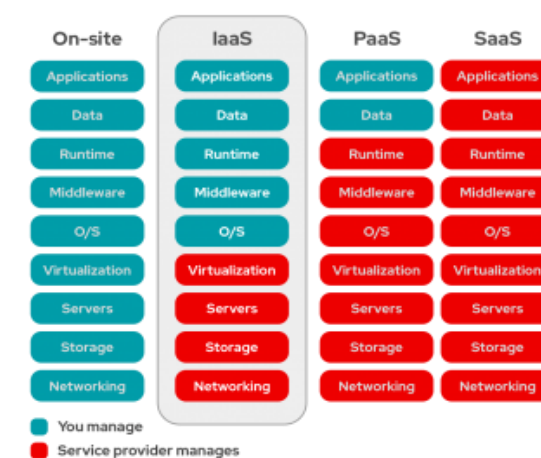
Aims

1. Acquire Data
2. Deploy Cloud Environment
3. Develop ETL Pipeline
4. Develop BI Front End

3. RESEARCH

a. Cloud vs On-Premise

Utilising the cloud for prototype development has several benefits and the most significant one is the reduction in upfront costs and development time. Microsoft Azure cloud platform was utilised as it the industry accepted cloud provider.



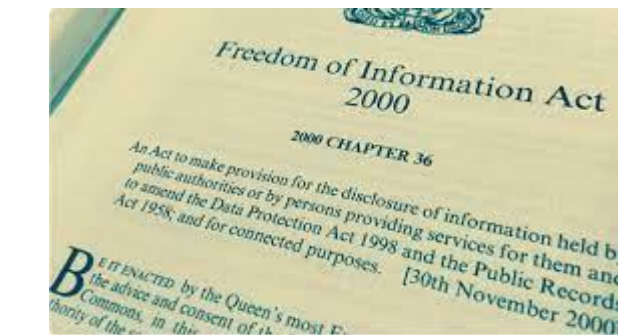
b. Power BI

The aim of data visualisation is to visually communicate information clearly and efficiently to users. Power BI offers a flexible data visualisation platform that has been developed to work directly with the Azure cloud environment and other Microsoft products such as the power platform.



c. Freedom of Information

Anyone has the right to ask for data that is held by public authorities. Public authorities include government departments, local councils, schools, the NHS, publicly owned companies, and the emergency services. The Freedom of Information Act (FOIA) gives you the right to see this information.



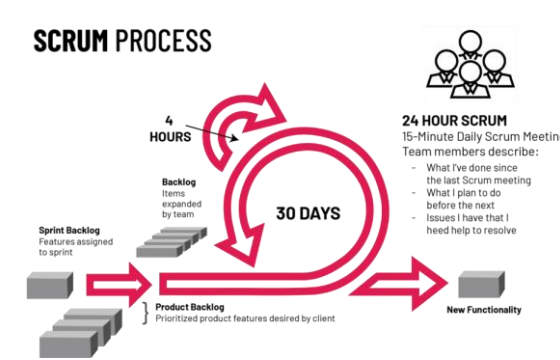
d. Python for ETL

ETL (Extract Transform and Load) is a process that data professionals use to extract information from different sources, transform the data into a new structure, and load that data into different systems. Python is an open source programming language that can be utilised for ETL.



4. METHODOLOGY

Although most of this project was undertaken as an individual, the Scrum methodology was adopted. This is because it lends itself perfectly to the type of project that the objective is set out to achieve in terms of the sprint development iterations.



5. DEVELOPMENT

Sprint 1 – Data Acquisition

A file containing 15-minute totals for all Environment Agency rain gauges for the period of 01/01/2010 to 31/10/2020, nearly 10 years of historical rainfall data was obtained.

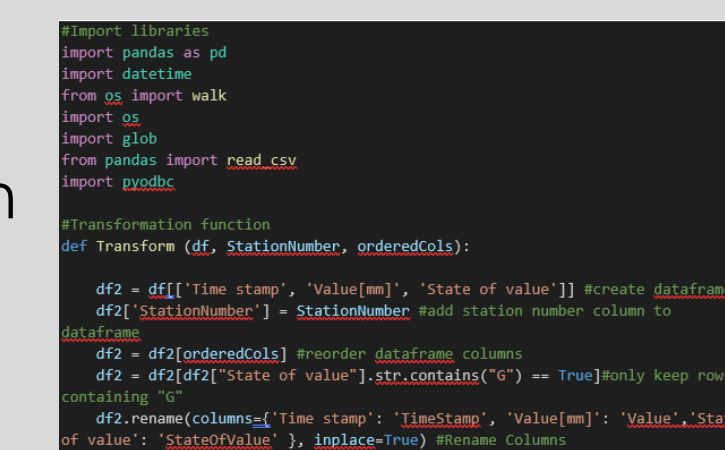
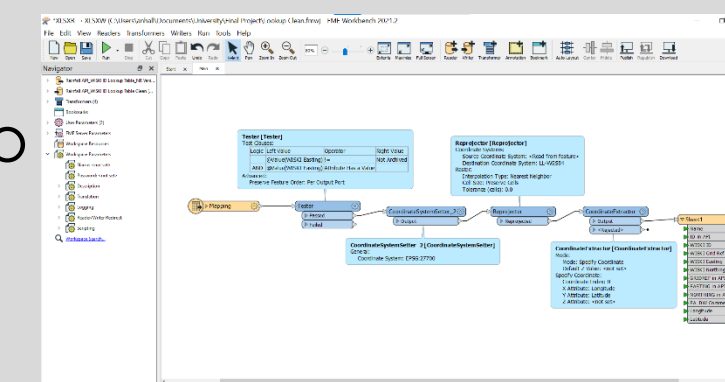
909 . Csv Files

10 year Period – 2010 to 2020

Potentially 345,216,384 rows

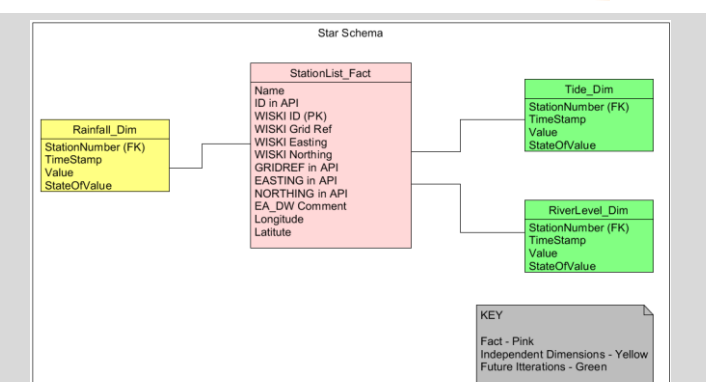
Sprint 3 – ETL

Sprint three consisted of developing a suitable process to extract the raw data from the 909 .csv files provided by the Environment Agency to the database deployed in the cloud. There are two separate processes developed, a more straightforward process to clean up the Station List table using GIS and a more complicated process to Transfer the raw rainfall data using Python.



Sprint 2 – Cloud Environment

A resource group was created where an SQL server and SQL Database were deployed within the Azure cloud environment. Firewall settings were created that allowed external connection to the Database. A Star Schema was Developed and tables were created using SQL.



```
CREATE TABLE [dbo].[Rainfall]([
    StationNumber [varchar](255) NOT NULL,
    [TimeStamp] [datetime] NOT NULL,
    [Value] [float] NOT NULL,
    [StationValue] [varchar](255) NULL
])
CREATE TABLE [dbo].[StationList]([
    Name [varchar](255) NOT NULL,
    ID [int] [varchar](255) NOT NULL,
    [Type] [varchar](255) NOT NULL,
    [Location] [varchar](255) NOT NULL
])
```

Sprint 4 – Front End Development

A dashboard was developed in Power BI so that users could interrogate any of the data for the 909 sites across England in graphical, tabular, and mapping visualisations.

