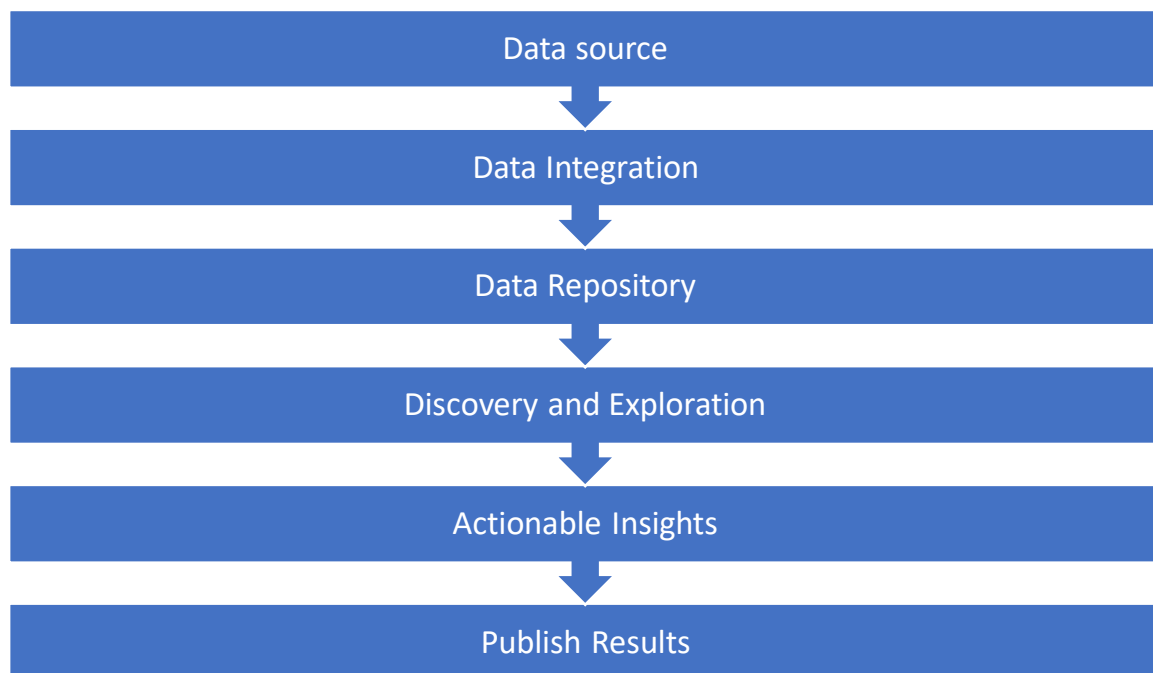# Architecture Design Document

Project title: Forecasting Global land Temperature and Uncertainty

Architectural Components Overview

```
Data source
    ↓
Data Integration
    ↓
Data Repository
    ↓
Discovery and Exploration
    ↓
Actionable Insights
    ↓
Publish Results
```

1.1   Data Source

1.1.1   Technology Choice

Text file containing of the land-surface average temperatures and uncertainty (95% confidence interval) produced by the Berkeley Averaging method over period of 1750 to 2019.

Source url: http://berkeleyearth.lbl.gov/auto/Global/Complete_TAVG_complete.txt

1.1.2   Justification

The text file is easy to input and can be cleaned up to create DataFrame representation of data.

1.2 Data Integration

1.2.1 Technology Choice

Python, Jupyter, libraries: pandas, NumPy, datetime

1.2.2 Justification

Python Jupyter notebook and python libraries are useful in reading a text file and manipulating it to create DataFrame using libraries pandas, NumPy, datetime and save the data-frame as .csv in cloud object storage for downstream analysis.

1.3 Data Repository

1.3.1 Technology Choice

Cloud Object Storage.

1.3.2 Justification

Cloud object storage provides a secure and easily accessible repository to integrate data into jupyter notebooks on IBM Watson.

1.4 Discovery and Exploration

1.4.1 Technology Choice

Jupyter, python, NumPy, pandas, matplotlib, Cloud object storage.

1.4.2 Justification

Working in Jupyter notebook in IBM Watson, input .csv file into data-frame from cloud object storage, perform necessary manipulations using libraries NumPy, pandas and create visualizations using matplotlib library.

1.5 Actionable Insights

1.5.1 Technology Choice

Jupyter, python, NumPy, pandas, matplotlib, scikit-learn, keras, TensorFlow, Cloud object storage.

1.5.2 Justification

Input the .csv file from cloud object storage in jupyter notebook, use NumPy and pandas libraries for data manipulations, scikit-learn library to create supervised machine learning model Gradient Boosted Regressor and performance metrics, keras and TensorFlow for creating LSTM (Long short term memory) Neural Network Model, matplotlib for performance metrics and results visualizations.

1.6 Publishing Results

1.6.1 Technology Choice

GitHub repository with access made public for everyone.

url: https://github.com/ViditAg/Global_temperature_Prediction

1.6.2 Justification

GitHub provides a platform to display and collaborate data-science projects.