The Lightweight IBM Cloud Garage Method for Data Science

Architectural Decisions Document

# Architectural Components Overview



IBM Data and Analytics Reference Architecture. Source: IBM Corporation

## Data Source

### Technology Choice

CSV File from Dubai Government Land Department website.

<https://dubailand.gov.ae/en/>

### Justification

It’s simple and easy to work.

## Data Repository

### Technology Choice

GitHub.

### Justification

It’s free and easy to use because our data is csv

## Discovery and Exploration

### Technology Choice

* Jupyter notebooks.
* Python.
* Pandas.
* Numpy.
* Scipy.
* Ploty.
* Matplotlib.
* Seaborn.
* Prince.

### Justification

* Jupyter and python because they are very useful with data science work, it’s fast easy to work and have many features.
* Pandas for working with data as a data frames, it’s very powerful tool and has all features needed to manipulate data.
* Numpy, scipy for mathematics work, it’s very powerful with math and statistics work.
* Plotly, matplotlib and seaborn for visualization.
* Prince for corresponding analysis, it does a very good job in feature reduction.

## Applications / Data Products

### Technology Choice

* Sikit-learn
* Keras
* flask
* html
* java script

### Justification

* Sikit-learn because it has everything related to machine learning like machine learning models, model selection, feature selection and metrics.
* Keras for neural networks, it’s high level tool built on tensor flow and easy to work with.
* Flask for build the back end server, because it’s easy and fast.
* Html and java script for front end.