The Lightweight IBM Cloud Garage Method for Data Science

Architectural Decisions Document Template

# Architectural Components Overview



IBM Data and Analytics Reference Architecture. Source: IBM Corporation

## Data Source

### Technology Choice

Jupyter Notebook with Python 3.6

Datasets from Kaggle that will be merged for full dataset

### Justification

Easy to view workspace and analyses

Comprehensive census and presidential election dataset, combined to form dataset of county demographics and which party won that county

## Enterprise Data

### Technology Choice

N/A

### Justification

N/A

## Streaming analytics

### Technology Choice

N/A

### Justification

N/A

## Data Integration

### Technology Choice

Data integrated with Python pandas library

### Justification

Easy to use library that is able to handle the amount of data in this project

## Data Repository

### Technology Choice

Datasets are stored as csvs

### Justification

The datasets are not very large so can be easily read by Python pandas in csv form

## Discovery and Exploration

### Technology Choice

Python pandas, numpy and matplotlib to analyze and plot the data

IDE is Jupyter notebook

### Justification

Jupyter notebook is simple to use interface that works well with sci-kit learn’s statistics packages, listed above

## Actionable Insights

### Technology Choice

Found with Jupyter notebook and models are explored here as well

### Justification

IDE is simple to train models and to find insights in dataset

## Applications / Data Products

### Technology Choice

Models are trained with sklearn library in Python and stored as .sav files using Jupyter notebook IDE

Models are also evaluated with sklearn metrics module

### Justification

Sklearn is an efficient library to train ML models; datasets aren’t very big so do not need to use SystemML or Apache Spark machine learning library.

.sav file models are easy to read back into Python to make more predictions or do evaluation

## Security, Information Governance and Systems Management

### Technology Choice

N/A

### Justification

N/A