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

The data comes from NYC (Open Data for All New Yorkers).

### Justification

Open Data is free public data published by New York City agencies and other partners.

## Enterprise Data

### Technology Choice

### Justification

## Streaming analytics

### Technology Choice

The data contain documents of every trip along with the features of the people.

### Justification

We use the features of the people and travel time as input features, and predict the transportation method.

## Data Integration

### Technology Choice

### Justification

## Data Repository

### Technology Choice

The data is stored on GitHub.

### Justification

GitHub is an open-source community for managing Git repositories.

## Discovery and Exploration

### Technology Choice

Load and Handel data: Spark and Pandas.

Plot figures: Seaborn and matplotlib.

Classification: Scipy and Pytorch.

### Justification

They are all open-source libraries.

## Actionable Insights

### Technology Choice

First, we load the data with Spark and look into the structure. Then, the correlation and distribution of the features are performed by the figures. Later, after preprocessing the data with pandas, we use machine learning and deep learning method to build classifiers for this problem and evaluate them with accuracy and negative log loss.

### Justification

## Applications / Data Products

### Technology Choice

The classifiers for predicting the trip choice could be applicated by the government to analyze the demand for different transportation and traffic condition. For companies, the demand for taxis, buses, subways, and even shared bikes could be analyzed with further development.

### Justification

## Security, Information Governance and Systems Management

### Technology Choice

### Justification