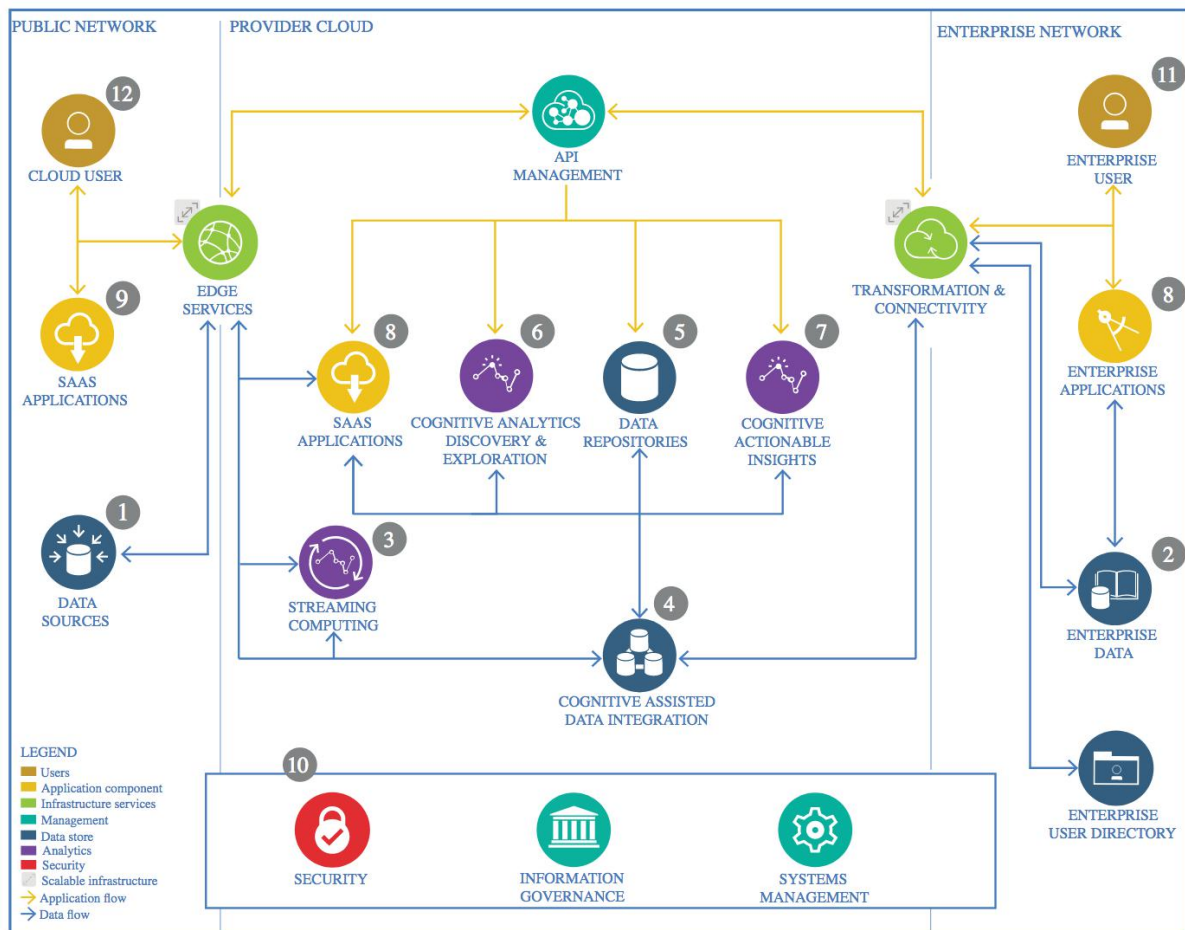# The Lightweight IBM Cloud Garage Method for Data Science

Architectural Decisions Document

Shipment Pricing Prediction
(Supply Chain Management System)

1   Architectural Components Overview

IBM Data and Analytics Reference Architecture. Source: IBM Corporation

1.1   Data Source

1.1.1 Technology Choice

Dataset Source:
https://data.usaid.gov/HIV-AIDS/Supply-Chain-Shipment-Pricing-Data/a3rc-nmf6

1.1.2 Justification
Primary reason to download from this website was availability, ease of use and relaibility of data as it is obtained from USA government website.

1.2 Enterprise Data

1.2.1 Technology Choice

NA

1.2.2 Justification

NA

1.3 Streaming analytics

1.3.1 Technology Choice

NA

1.3.2 Justification

NA

1.4 Data Integration

1.4.1 Technology Choice

Not Used

1.4.2 Justification

Not Used

1.5 Data Repository

1.5.1 Technology Choice

GitHub Repository.

1.5.2 Justification

Up-to-Date data would be available on the Repository.

1.6 Discovery and Exploration

1.6.1 Technology Choice

The following Python 3.7 and above libraries were used for Data Exploration and Visualization:

Pandas,
Numpy,
Matplotlib,
Seaborn

1.6.2 Justification

The size of the dataset was the key factor in deciding data exploration tools.
The current data small enough to be processed on a single computer ruling out the need for distributed processing (Spark, pyspark)

1.7 Actionable Insights

1.7.1 Technology Choice

The following Traditional Machine Learning Algorithms and Deep Learning Neural Network are used for model Training:
Linear Regression,
Decision Tree,
Random Forest,
Light-GBM,
Multi Layer Perceptron Neural Network (MLP).

1.7.2 Justification

To understand the Correlating features a white-box model was required. Tree based algorithms were identified as a good match. Thus Light-GBM was used. Neural network based algorithm was used as a reference for the Tree based model. Easiest and Fastest implementation is possible in keras. Tensorflow is the backend.

1.8 Applications / Data Products

1.8.1 Technology Choice

A Jupyter notebook based report was generated.

1.8.2 Justification

As only the correlating factors needed to be identified Jupyter notebook based report was consider sufficient.

1.9 Security, Information Governance and Systems Management

1.9.1 Technology Choice

None

1.9.2 Justification

None