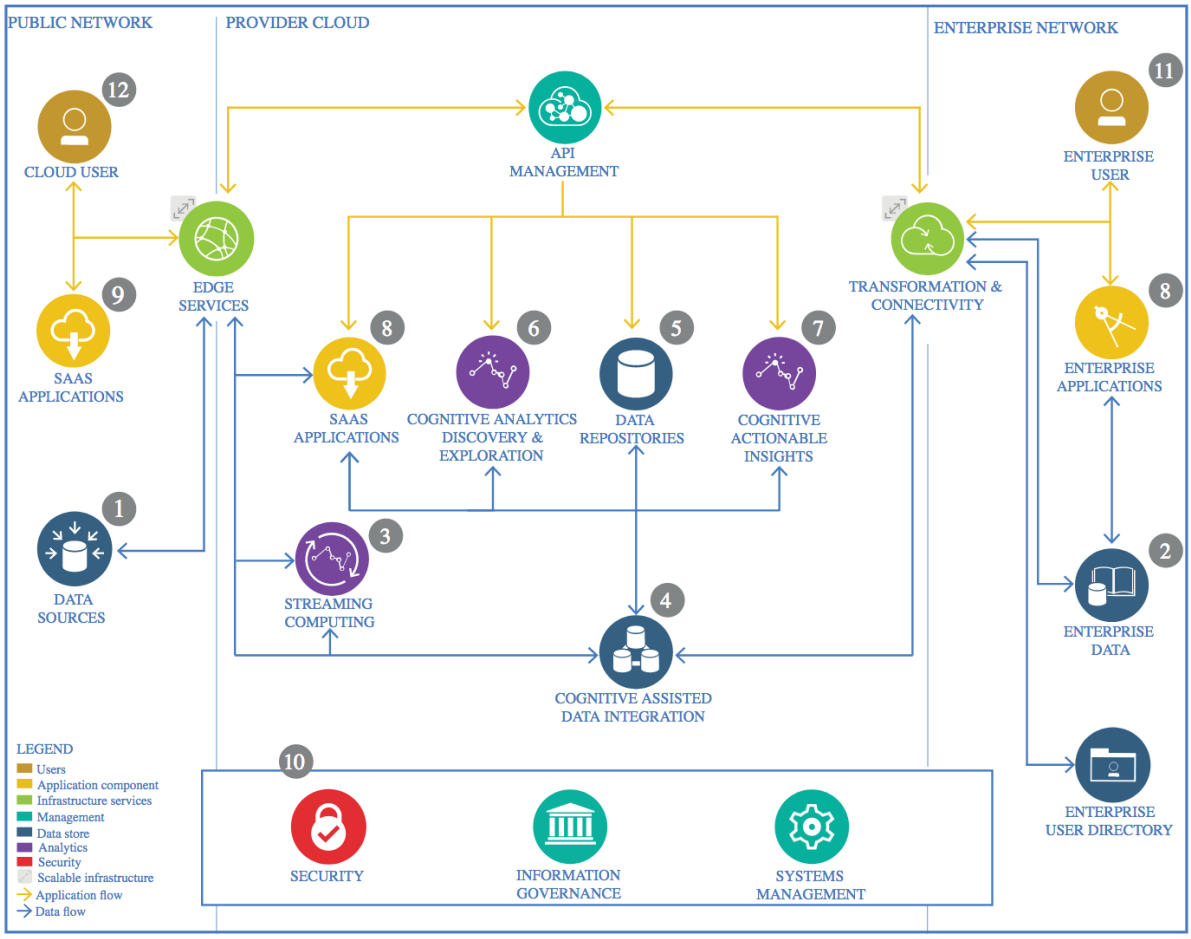
The Lightweight IBM Cloud Garage Method for Data Science

Architectural Decisions Document

Glass Type Prediction

1 Architectural Components Overview



IBM Data and Analytics Reference Architecture. Source: IBM Corporation

1.1 Data Source

1.1.1 Technology Choice

The data was downloaded from the following source:
https://archive.ics.uci.edu/ml/datasets/Glass+Identification

1.1.2 Justification

This was the official site for the availability of this dataset

1.2 Enterprise Data

1.2.1 Technology Choice
Not Used

1.2.2 Justification
Not Used

1.3 Streaming analytics

1.3.1 Technology Choice
Not Used

1.3.2 Justification
Not Used

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

1.5.2 Justification
Easy to perform version control and readily available

1.6 Discovery and Exploration

1.6.1 Technology Choice
Python 3.6 libraries used:

- Pandas
- Seaborn
- Matplotlib

1.6.2 Justification

The size of data was small enough to be visualized and explored by these tools and hence advanced tools like spark was not required

1.7 Actionable Insights

1.7.1 Technology Choice

Python 3.6 libraries used:

- Pandas
- Numpy
- Scikitlearn
- Keras
- Tensoflow

1.7.2 Justification

Scikit learn contains basic machine learning models and even some ensemble models. Tensorflow and Keras was used for Deep Learning implementation.

1.8 Applications / Data Products

1.8.1 Technology Choice

Report was generated using Jupiter Notebook

1.8.2 Justification

Jupyter Notebooks are portable and was inline with the requirements for the submission

1.9 Security, Information Governance and Systems Management

1.9.1 Technology Choice

Not Used

1.9.2 Justification

Not Used