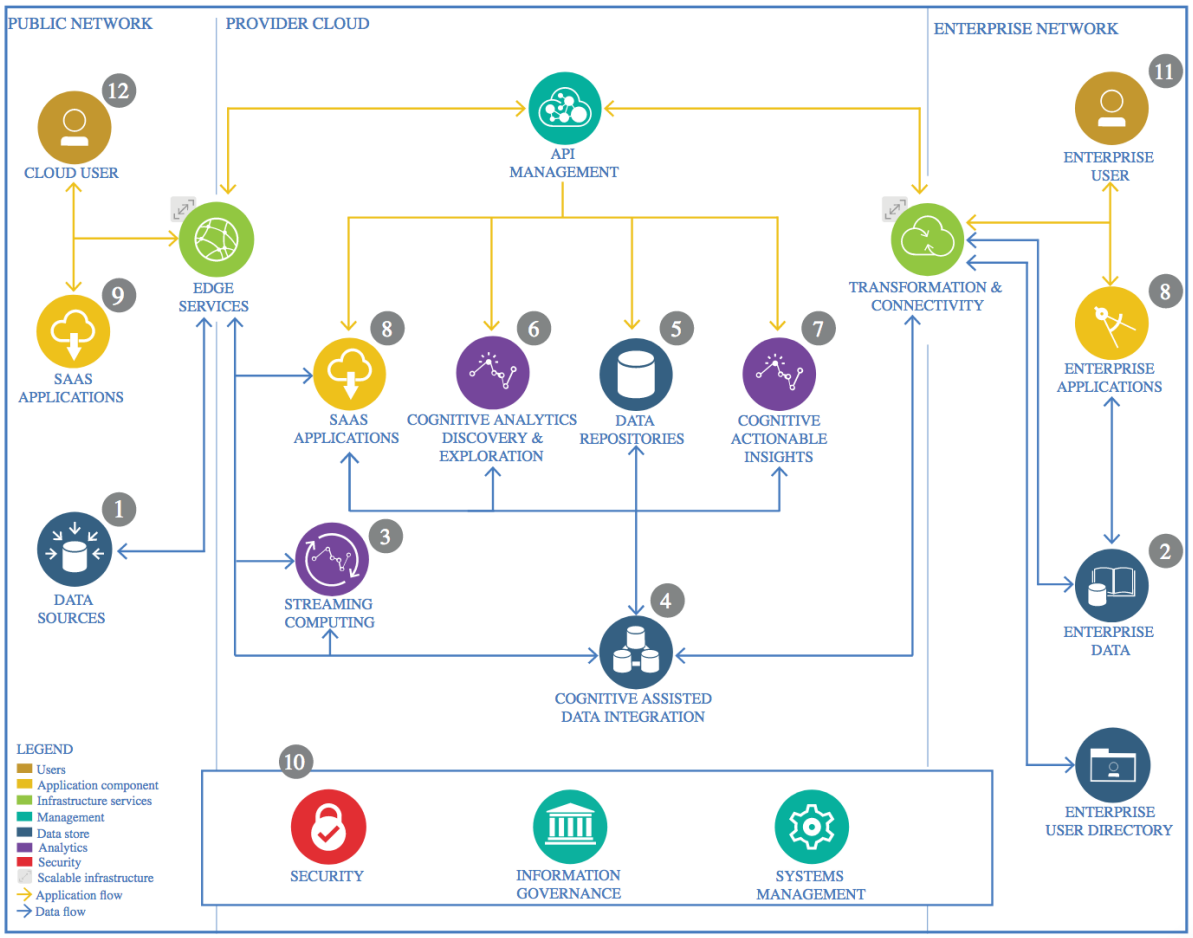
The Lightweight IBM Cloud Garage Method for Data Science

Architectural Decisions Document Template

1 Architectural Components Overview



IBM Data and Analytics Reference Architecture. Source: IBM Corporation

1.1 Data Source

1.1.1 Technology Choice

The data is avaible as public dataset on https://www.kaggle.com/mlg-ulb/creditcardfraud/download as CSV.

1.1.2 Justification

I was looking for unbalanced dataset that was public and was related to credit card.

1.2 Enterprise Data

1.2.1 Technology Choice

N/A

1.2.2 Justification

It is only for learning purposes.

1.3 Streaming analytics

1.3.1 Technology Choice

NA

1.3.2 Justification

NA

1.4 Data Integration

1.4.1 Technology Choice

NA

1.4.2 Justification

All dataset is available in one CSV file.

1.5 Data Repository

1.5.1 Technology Choice

NA

1.5.2 Justification

No need to store in a separate repository since the dataset was well describe in kaggle repository.

1.6 Discovery and Exploration

1.6.1 Technology Choice

Jupyter Notebook (Kernel Kaggle Python), pandas, seaborn, matplotlib and scikit-learn.

1.6.2 Justification

The methods provided by the imports above give us suficient tools to do exploration of the dataset.

1.7 Actionable Insights

1.7.1 Technology Choice

NA

1.7.2 Justification

We only need to be able to classifying the transactions given the features.

1.8 Applications / Data Products

1.8.1 Technology Choice

NA

1.8.2 Justification

The predictions are standalone data.

1.9 Security, Information Governance and Systems Management

1.9.1 Technology Choice

NA

1.9.2 Justification

The dataset is public.