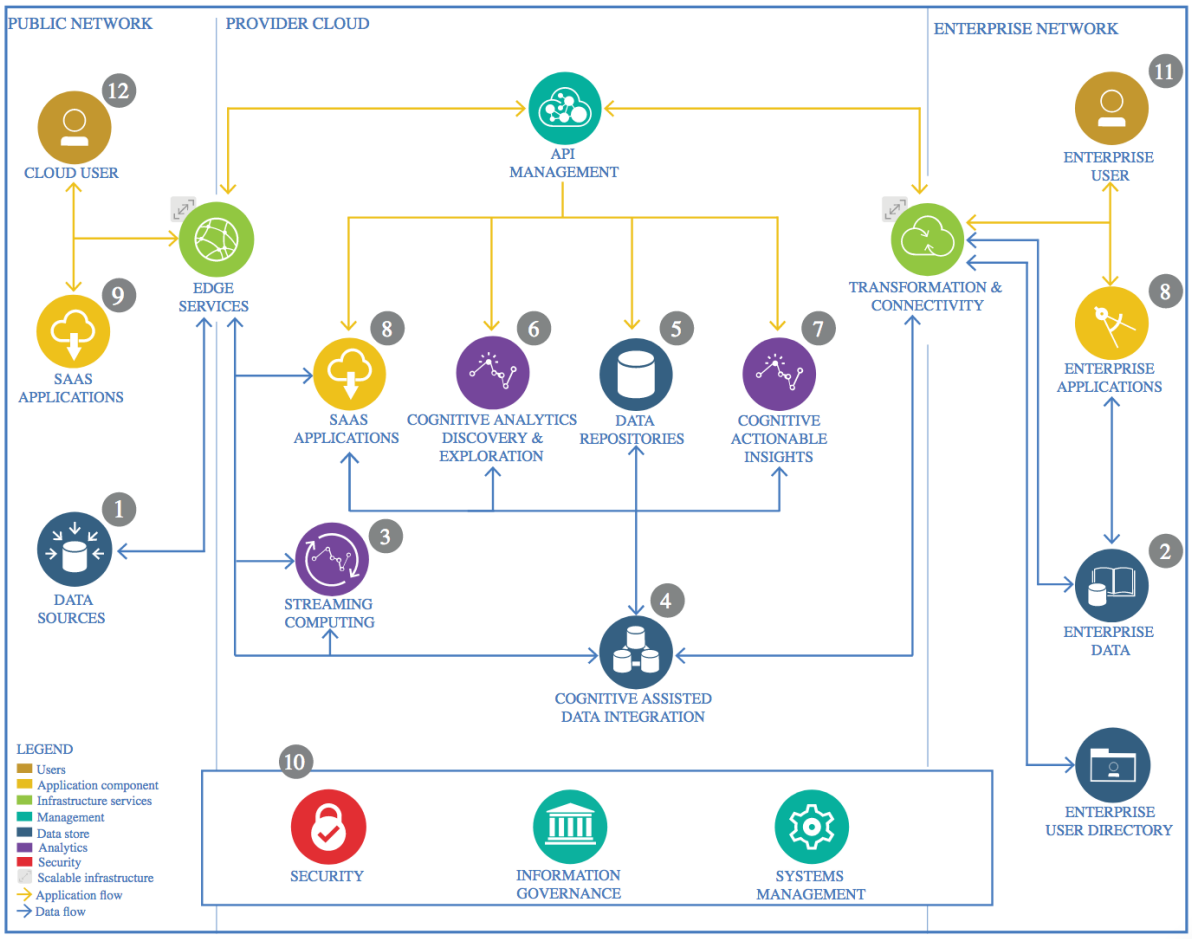
Architectural Decisions Document

Customer patterns 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 is obtained from Kaggle

1.1.2 Justification

Due to their open source of datasets and ease of using them

1.2 Enterprise Data

1.2.1 Technology Choice
Github repository.

1.2.2 Justification
First choice for any technical member because it is up-to-date features will be available.

1.3 Streaming analytics

1.3.1 Technology Choice
None

1.3.2 Justification

1.4 Data Integration

1.4.1 Technology Choice
None

1.4.2 Justification
None

1.5 Data Repository

1.5.1 Technology Choice
None

1.5.2 Justification
None

1.6 Discovery and Exploration

1.6.1 Technology Choice
Language – Python3
Matplotlib and Seaborn used for data visualization
Pandas for data analysis
Scikit and keras to build models

1.6.2 Justification
They are the best choice.

1.7 Actionable Insights

1.7.1 Technology Choice

Language – Python3
Matplotlib and Seaborn used for data visualization
Pandas for data analysis
Scikit and keras to build models

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

Jupyer – python notebook

1.8.2 Justification

Please justify your technology choices here.

1.9 Security, Information Governance and Systems Management

1.9.1 Technology Choice

None

1.9.2 Justification

None