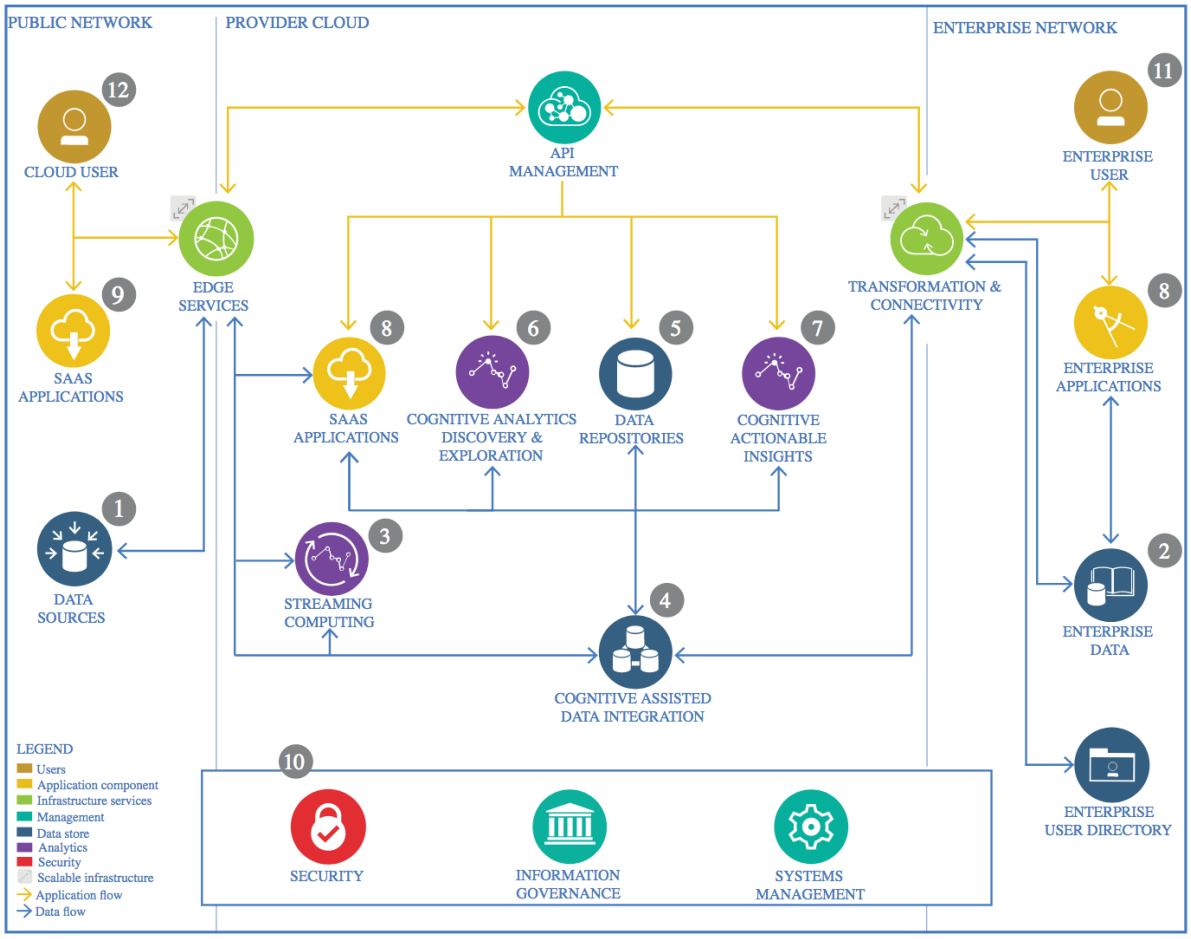
The Lightweight IBM Cloud Garage Method for Data Science

Architectural Decisions Document

1 Architectural Components Overview



IBM Data and Analytics Reference Architecture. Source: IBM Corporation

1.1 Data Source

1.1.1 Technology Choice

CSV File from Dubai Government Land Department website.
https://dubailand.gov.ae/en/

1.1.2 Justification

It's simple and easy to work.

1.2 Data Repository

1.2.1 Technology Choice

GitHub.

1.2.2 Justification

It's free and easy to use because our data is csv

1.3 Discovery and Exploration

1.3.1 Technology Choice

- Jupyter notebooks.
- Python.
- Pandas.
- Numpy.
- Scipy.
- Ploty.
- Matplotlib.
- Seaborn.
- Prince.
-

1.3.2 Justification

- Jupyter and python because they are very useful with data science work, it's fast easy to work and have many features.

- Pandas for working with data as a data frames, it's very powerful tool and has all features needed to manipulate data.

- Numpy, scipy for mathematics work, it's very powerful with math and statistics work.

- Plotly, matplotlib and seaborn for visualization.

- Prince for corresponding analysis, it does a very good job in feature reduction.

1.4 Applications / Data Products

1.4.1 Technology Choice

- Sikit-learn
- Keras
- flask
- html
- java script

1.4.2 Justification

- Sikit-learn because it has everything related to machine learning like machine learning models, model selection, feature selection and metrics.

- Keras for neural networks, it's high level tool built on tensor flow and easy to work with.

- Flask for build the back end server, because it's easy and fast.

- Html and java script for front end.