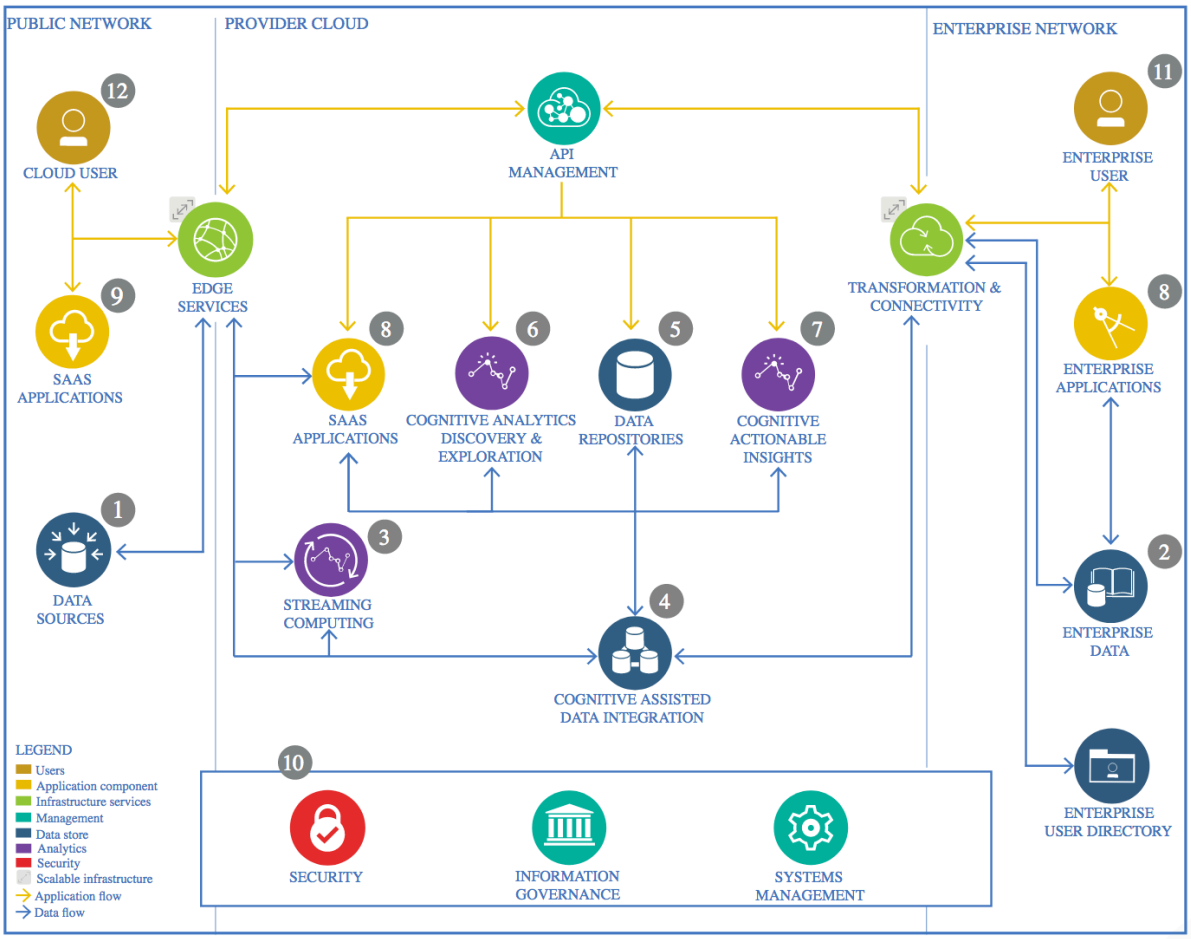
The Lightweight IBM Cloud Garage Method for Data Science

Architectural Decisions Document Template

1 Architectural Components Overview



IBM Data and Analytics Reference Architecture. Source: IBM Corporation

1.1 Data Source

https://www.kaggle.com/sulianova/cardiovascular-disease-dataset
You can find this dataset in Kaggle. Pretty clean dataset only need to eliminate some error values.

1.1.1 Technology Choice

I will use machine learning models to predict heart diseases. Will compare standart ML models and neuron ML models.

1.1.2 Justification

Will compare standart ML models and neuron ML models.

1.2 Enterprise Data

1.2.1 Technology Choice

I will use Kaggle dataset

1.2.2 Justification

I will use Kaggle dataset

1.3 Data Repository

1.3.1 Technology Choice

Kaggle as data repository.

1.3.2 Justification

Just use Kaggle as a perfect data storage.

1.4 Discovery and Exploration

1.4.1 Technology Choice

Use some ML models.

1.4.2 Justification

This is binary problem classification. Was used following models RandomForestClassifier,LogRegressionClassifiaer and SVM machine. As neuron model use MLPClassifier. All models from ScikitLearn package.
Best accuracy was in MLPClassifier (neuron model) – 0.7345. Demonstrate high results from the box on the default settings. F1 score on that model was – 0.74.

1.5 Actionable Insights

1.5.1 Technology Choice

The following features have high impact on the heart disease.

1.5.2 Justification

Please justify your technology choices here.

1.6 Applications / Data Products

1.6.1 Technology Choice

Please describe what technology you have defined here. Please justify below, why. In case this component is not needed justify below. Blood Pressure, Age and Cholesterol in blood.

1.6.2 Justification

It was surprising me but smoke, alcohol have not so much impact on the heart diseases.