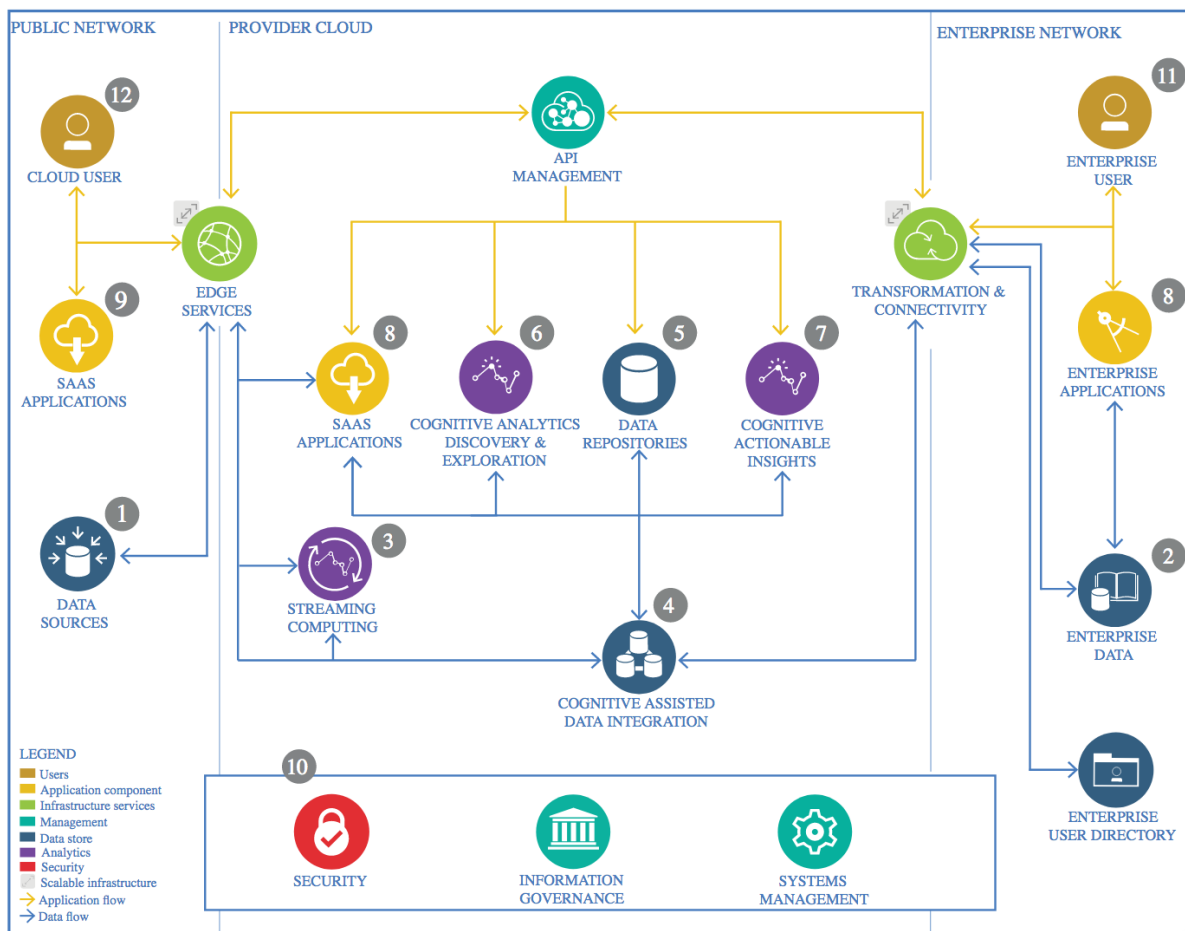


The Lightweight IBM Cloud Garage Method for Data Science

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

Open source Kaggle Data Set.

1.1.2 Justification

Kaggle was recommended by course instructor Romeo Kienzler as potential source of data for interesting Data Science problem.

1.2 Enterprise Data

1.2.1 Technology Choice

N/A.

1.2.2 Justification

The data did not originate from the Enterprise which may be a misnomer in this case. The data was obtained from the public Kaggle platform.

1.3 Streaming analytics

1.3.1 Technology Choice

N/A.

1.3.2 Justification

The data from the Kaggle website is fixed data and not changing with time. Hence a Streaming Platform is NOT required for this steady-state data.

1.4 Data Integration

1.4.1 Technology Choice

N/A

1.4.2 Justification

The data is all available on Kaggle and uploaded as a CSV file on IBM Watson Studio. The data set is not too large and hence this justifies our method. For larger data sets, one would use COS on IBM Watson Studio and other sources followed by integrating all the different sources into one large data set.

1.5 Data Repository

1.5.1 Technology Choice

CSV data file uploaded on IBM Watson Studio.

1.5.2 Justification

Data Set Not Too large. In case of large data set one would use COS. IBM Watson Studio is secure.

1.6 Discovery and Exploration

1.6.1 Technology Choice

Matplotlib and Scipy packages. Also used Pandas dataframe to clean data and remove duplicate data rows.

1.6.2 Justification

Efficiency of use.

1.7 Actionable Insights

1.7.1 Technology Choice

As per recommendation from Coursera's courses taught by IBM professionals.

1.7.2 Justification

Course taught by IBM on Coursera.

1.8 Applications / Data Products

1.8.1 Technology Choice

Apache Spark ML, Keras DL (on TensorFlow), Pandas, Matplotlib ... all implemented on IBM Watson Studio using Python programming language.

1.8.2 Justification

Easy of use, computational power and efficiency.

1.9 Security, Information Governance and Systems Management

1.9.1 Technology Choice

Coding on IPYTHON notebooks on IBM Watson Studio. Products posted on GitHub which is the gold standard for sharing. Identifying information removed from Notebooks.

1.9.2 Justification

Trust IBM and GitHub secure website.