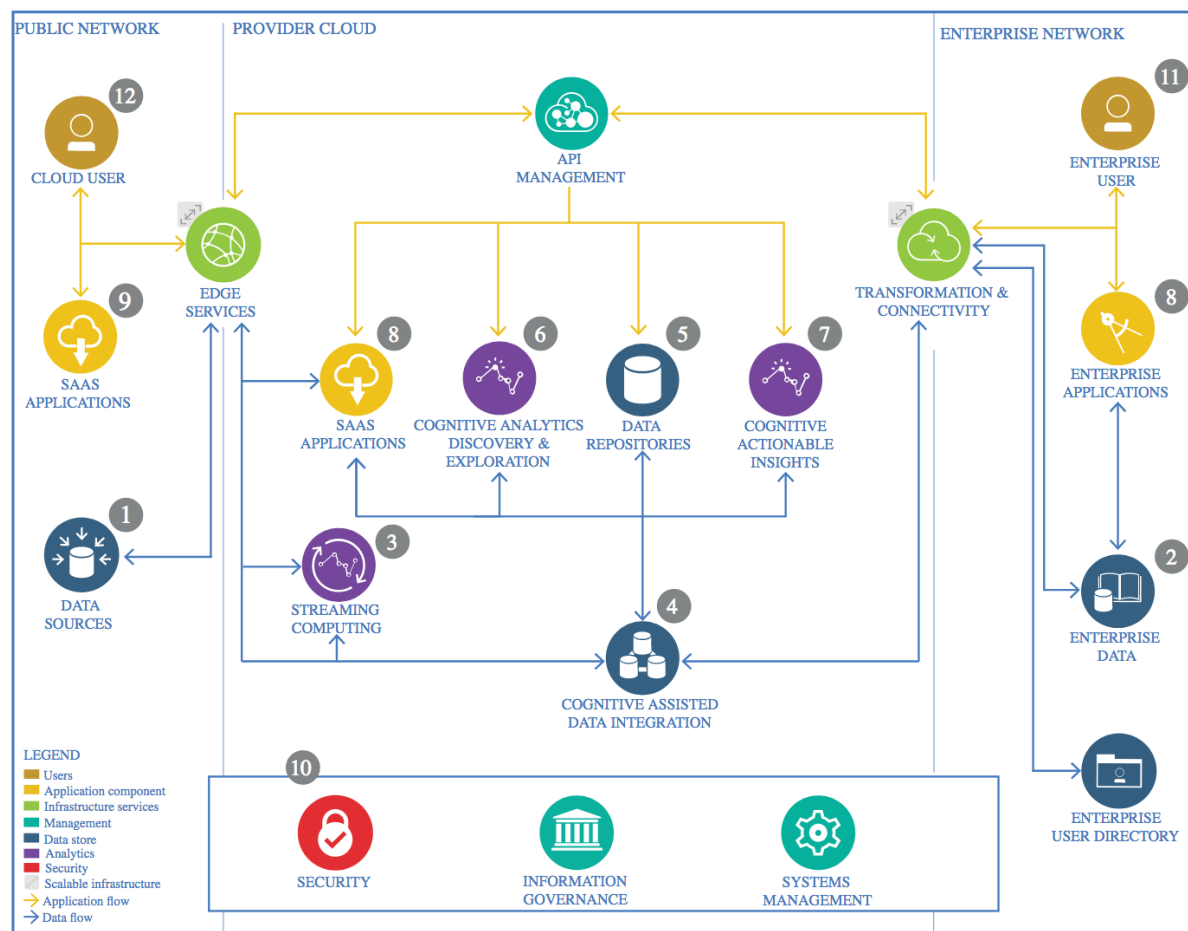


The Lightweight IBM Cloud Garage Method for Data Science

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

1.1.2 Justification

Forest Cover Type dataset CSV file from Kaggle ,Tree types found the Roosevelt National Forest in Colorado.

1.2 Data Integration

1.2.1 Technology Choice

IBM Watson Studio, IBM Cloud Object Storage.

1.2.2 Justification

IBM Watson Studio Jupyter Notebooks, scikit-learn, pandas, matplotlib, seaborn.

1.3 Data Repository

1.3.1 Technology Choice

IBM Cloud Object storage.

1.3.2 Justification

IBM Cloud Object Storage is used as the Data storage solution.

1.4 Discovery and Exploration

1.4.1 Technology Choice

Apache Spark , Jupyter Notebook with Python 3.5 and Spark 2.3 , IBM Watson Studio and IBM Object Storage as our data repository

1.4.2 Justification

all are an open-source web application.

1.5 Actionable Insights

None as of now.

1.5.1 Technology Choice

Loss VS Accuracy.

1.5.2 Justification

There are no actionable insights at present. However this can be relevant in the future as we enhance our model.

1.6 Applications / Data Products

1.6.1 Technology Choice

None as of now.

1.6.2 Justification

There are no applications in scope of our project at present. This is something that will be

addressed as part of the enhancement scope in the future.

1.7 Security, Information Governance and Systems Management

1.7.1 Technology Choice

Not applicable

1.7.2 Justification

At present our model is stand-alone model which does not interact or integrate with any external application or systems. Hence there is no real need or scope for implementing Security, Information Governance and Systems Management. This is also something which will need to be considered in the future as we enhance our model and start streaming data and also start interacting with external systems.