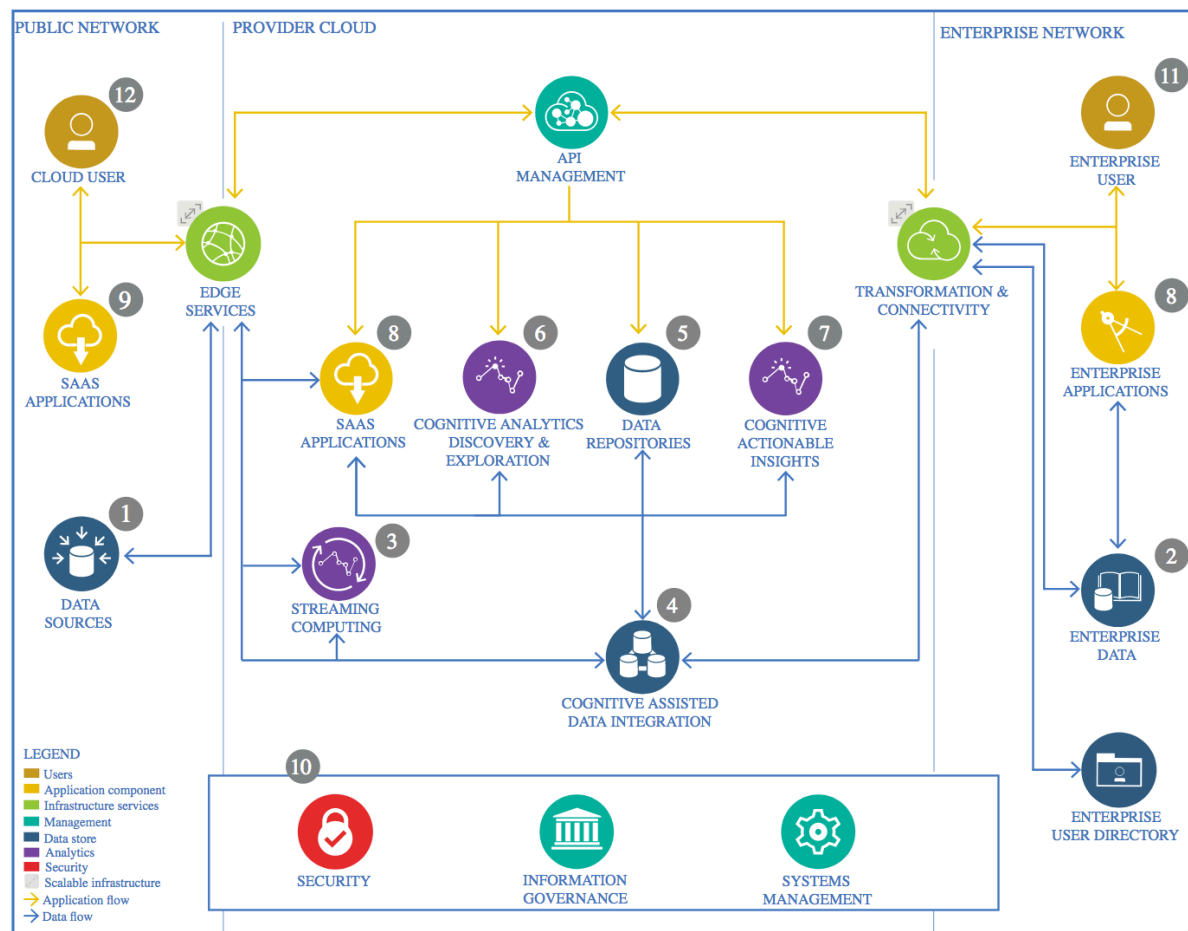


The Lightweight IBM Cloud Garage Method for Data Science

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

The data is tabular data which is why *.csv was chosen as a format and IBM COS as the storage technology.

1.1.2 Justification

*.csv is easily transferable across DB and storage technologies. Moreover, object store provides scalability and accessibility.

1.2 Enterprise Data

1.2.1 Technology Choice

This is not relevant for the current project but the data should reside in a cloud based instance of a relational DB such as DB2 or also a NoSQL DB such as Cloudant would be possible.

1.2.2 Justification

In an enterprise context data has to be stored in a way that integrity, security and privacy are maintained. That's why a proven DB technology would have to be selected to fulfill these requirements. A cloud instance also provides flexibility in terms of how the application scales.

1.3 Streaming analytics

1.3.1 Technology Choice

-

1.3.2 Justification

No streaming in this use case.

1.4 Data Integration

1.4.1 Technology Choice

The data is pulled from a public network and is ideally pulled by REST-API.

1.4.2 Justification

This is done for ease of use and for continuous updates of the data.

1.5 Data Repository

1.5.1 Technology Choice

As described above the data resides in IBM Object Store which can easily be integrated with IBM Watson Studio.

1.5.2 Justification

Using a cloud based repo and also an Enterprise development environment fosters collaboration while not impacting security.

1.6 Discovery and Exploration

1.6.1 Technology Choice

For EDA IBM Watson Studio with Jupyter Notebooks are used. The used programming language is Python.

1.6.2 Justification

As described above Watson Studio is great for collaborating in teams. Moreover, it gives you flexibility to easily scale to multiple compute nodes if required e.g. for training a complex neural network or performing large grid operations that scale exponentially which is also done here. Python has been used because it is an easy to read and widely used language. Additionally, it provides a great deal of data exploration and visualization tools.

1.7 Actionable Insights

1.7.1 Technology Choice

For modelling IBM Watson Studio with Jupyter Notebooks are used. The used programming language is Python.

1.7.2 Justification

Python has been used because it is an easy to read and widely used language. Moreover, it provides a host of machine and deep learning libraries and can also work with Apache Spark to scale the application to enterprise level.

1.8 Applications / Data Products

1.8.1 Technology Choice

The application has been deployed to the cloud-based IBM Watson Machine Learning.

1.8.2 Justification

To provide an easy to use interface and make it possible to scale.

1.9 Security, Information Governance and Systems Management

1.9.1 Technology Choice

This project is completely done in the IBM Public Cloud.

1.9.2 Justification

IBM Cloud Security's scalable suite of technologies and solutions are made more robust and complete through pervasive encryption, AI + automation and integration.

