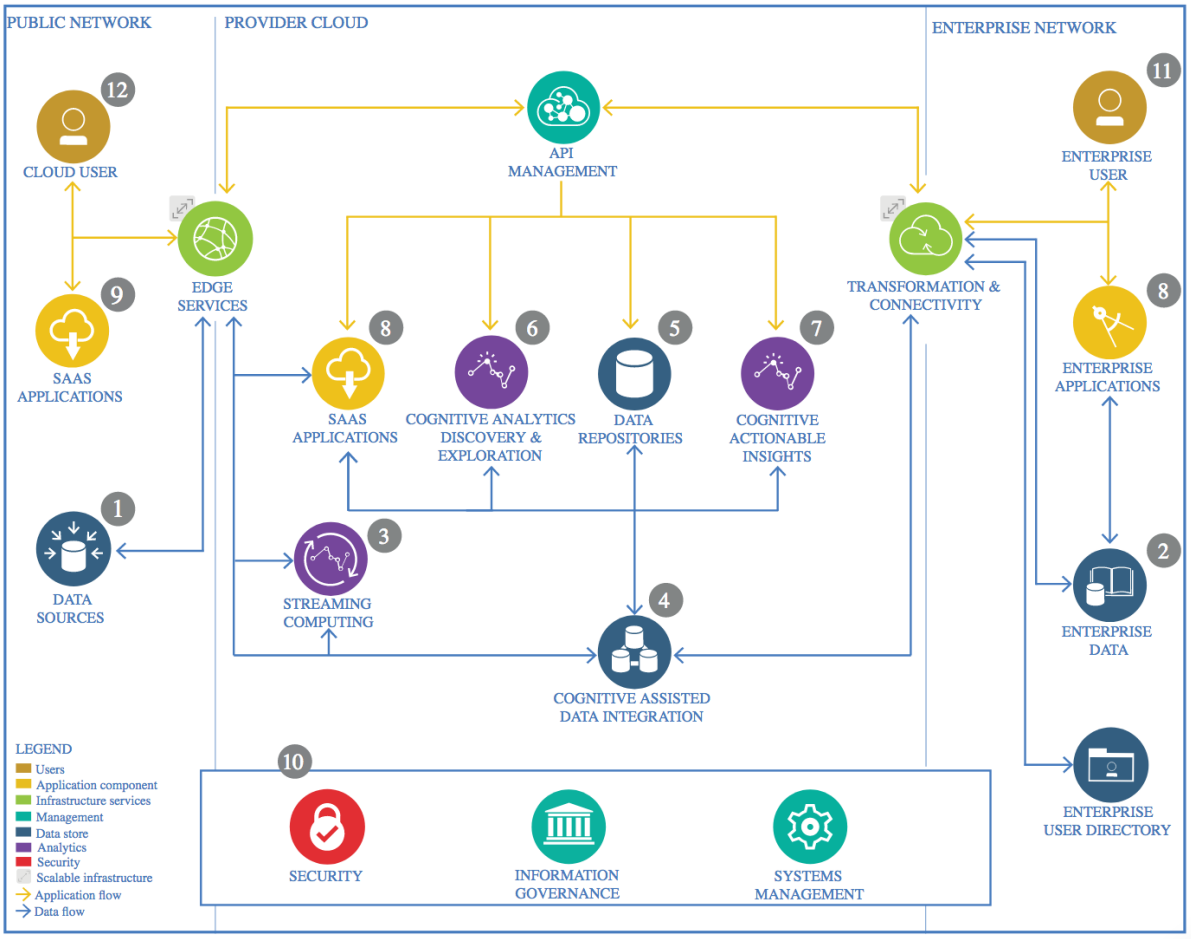
The Lightweight IBM Cloud Garage Method for Data Science

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

Pandas as it is easy to use.

1.1.2 Justification

CSV files are easy to process in pandas.

1.2 Enterprise Data

1.2.1 Technology Choice

Please describe what technology you have defined here. Please justify below, why. In case this component is not needed justify below.

1.2.2 Justification

Not needed

1.3 Streaming analytics

1.3.1 Technology Choice

Please describe what technology you have defined here. Please justify below, why. In case this component is not needed justify below.

1.3.2 Justification

Not needed.

1.4 Data Integration

1.4.1 Technology Choice

Pandas, ski-kit learn.

1.4.2 Justification

The dataset is medium size so parallelization will have little to no effect. Thus pandas and sklearn will suffice.

1.5 Data Repository

1.5.1 Technology Choice

Please describe what technology you have defined here. Please justify below, why. In case this component is not needed justify below.

1.5.2 Justification

Not needed.

1.6 Discovery and Exploration

1.6.1 Technology Choice

Pandas, numpy, matplotlib

1.6.2 Justification

Built in functions made everything much easier.

1.7 Actionable Insights

1.7.1 Technology Choice

PySpark.

1.7.2 Justification

To increase speed.

1.8 Applications / Data Products

1.8.1 Technology Choice

Please describe what technology you have defined here. Please justify below, why. In case this component is not needed justify below.

1.8.2 Justification

Not needed.

1.9 Security, Information Governance and Systems Management

1.9.1 Technology Choice

Please describe what technology you have defined here. Please justify below, why. In case this component is not needed justify below.

1.9.2 Justification

Not needed.