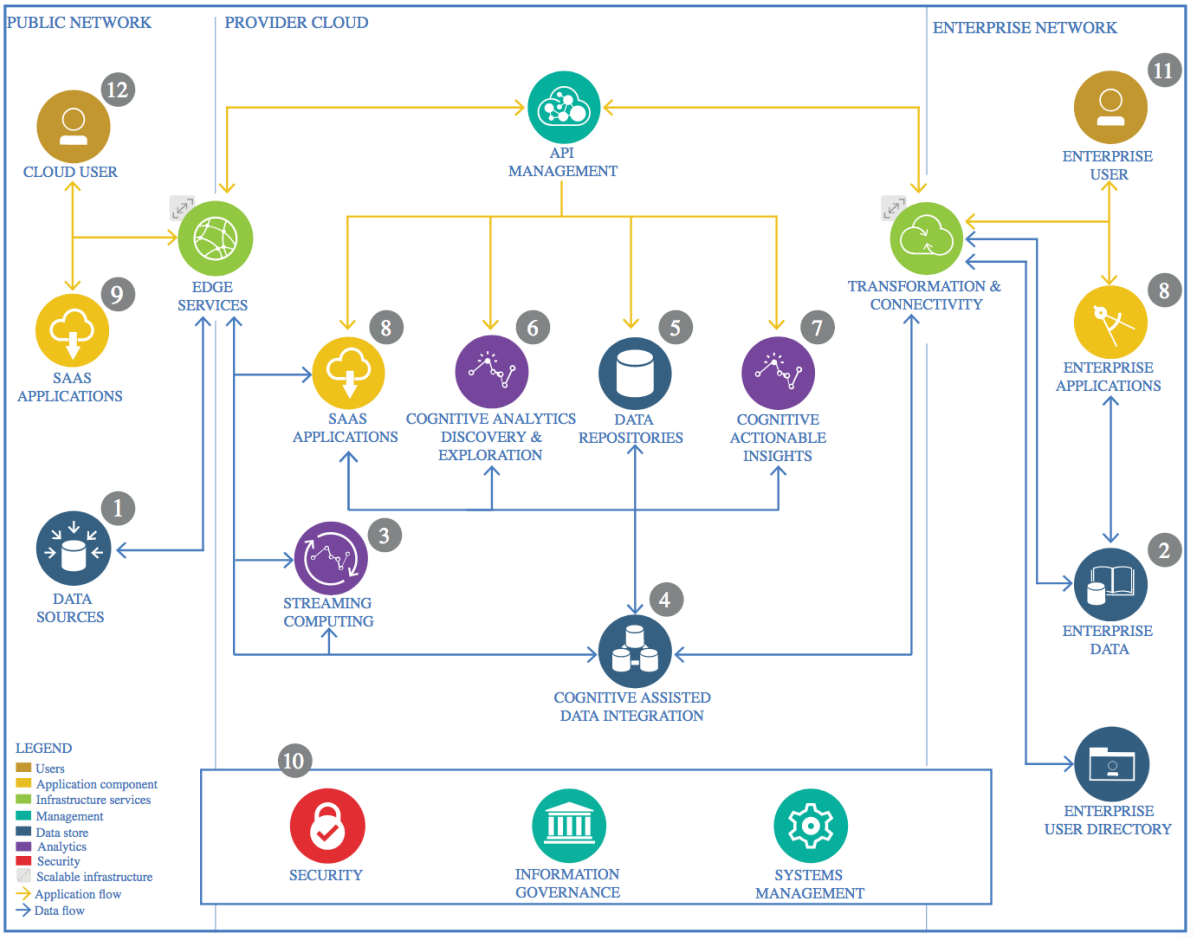
Dog breed identification using convolutional neural network

Architectural Decisions Document

Architectural Components Overview:



IBM Data and Analytics Reference Architecture. Source: IBM Corporation

1.1 Data Source

1.1.1 Technology Choice

This projects data source is Stanford Dogs Dataset, hosted on the Stanford Vision and Learning Lab website.

1.1.2 Justification

Data source is chosen based on its availability.

1.2 Enterprise Data

1.2.1 Technology Choice
GNU Wget is used for external communication to retrieve content from web server and integrate it into the project.

1.2.2 Justification
GNU Wget is chosen because of its stability and robustness.

1.3 Streaming analytics

1.3.1 Technology Choice
This component is not used.

1.3.2 Justification
This component is not used since our data source is static and not real-time.

1.4 Data Integration

1.4.1 Technology Choice
Retrieved data is presented as Pandas dataframe.

1.4.2 Justification
Pandas is chosen since it is the most popular data science library used in Python programming language for data wrangling and analysis.

1.5 Data Repository

1.5.1 Technology Choice
IBM Cloud Object Storage is used as a data repository solution.

1.5.2 Justification
IBM Cloud Object Storage is chosen because it is designed for durability, resiliency and security.

1.6 Discovery and Exploration

1.6.1 Technology Choice
Matplolib Python library is used for data exploration.

1.6.2 Justification
Matplotlib is chosen because of the project maturity and stability.

1.7 Actionable Insights

1.7.1 Technology Choice
Actionable insights would be extracted by applying machine learning algorithms with the help of Keras Python library.

1.7.2 Justification
Keras is chosen because it supports multiple backend engines and does not lock project into one ecosystem.

1.8 Applications / Data Products

1.8.1 Technology Choice
This project uses Jupyter Notebooks with Python programming language.

1.8.2 Justification
Jupyter Notebooks were chosen because software development style using markdown text punctuated with code blocks is an approach suitable for conducting Data Science experiments. Python programming language is chosen because it is easy to use when it comes to analytical and quantitative computing.

1.9 Security, Information Governance and Systems Management

1.9.1 Technology Choice
The project uses IBM Watson platform for its information governance.

1.9.2 Justification
IBM Watson is chosen because it allows to explicitly choose collaborators and change the security permissions of a collaborator or service ID.