The Lightweight IBM Cloud Garage Method for Data Science

Architectural Decisions Document Template

# Architectural Components Overview



IBM Data and Analytics Reference Architecture. Source: IBM Corporation

## Data Source

### Technology Choice

The data I am working on was downloaded from Kaggle :

<https://www.kaggle.com/mlg-ulb/creditcardfraud>

### Justification

Primary reason to download from Kaggle was availability and ease of use.

## Enterprise Data

### Technology Choice

GitHub repository.

### Justification

Up-to-date data would be available on the repository.

## Streaming analytics

### Technology Choice

No need.

### Justification

N/A.

## Data Integration

### Technology Choice

No need.

### Justification

N/A.

## Data Repository

### Technology Choice

Object storage.

### Justification

Cloud object storage makes it possible to store practically limitless amounts of data.

## Discovery and Exploration

### Technology Choice

Jupyter, Python 3.6, pyspark, scikit-learn, pandas, Matplotlib, seaborn

### Justification

Performant and open source tools.

## Actionable Insights

### Technology Choice

* Python, pandas and scikit-learn;
* Keras
* TensorFlow

### Justification

Python is the programing language and pandas and scikit-learn are powerful libraries to implement data and learning models and Neural network-based algorithms were used to train the model, thus we used Keras as a Deep Learning network since it is Easy to use and Fast to implement while TensorFlow is the backend.

## Applications / Data Products

### Technology Choice

A Jupyter notebook-based report was generated.

### Justification

As only the correlating factors needed to be identified Jupyter notebook-based report was consider sufficient.

## Security, Information Governance and Systems Management

### Technology Choice

IBM App ID.

### Justification

Identity Management allows for cloud-based user and identity management for web and mobile applications, APIs, and back-end systems.