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 is externally sourced from Kaggle in CSV format which contains brain-related signals and artifacts from multi-channel EEG data .

### Justification

Analysing this data and to produce ML models is much simpler given the completed and more structured dataset collected.

## Enterprise Data

### Technology Choice

The data will be handled in the cloud.

### Justification

Fewer datapoints not requiring complex data handling systems.

## Streaming analytics

### Technology Choice

Real-time data processing.

### Justification

An an increased value of data is expected, hence, real-time analytics capabilities is needed.

## Data Integration

### Technology Choice

IBM Data Stage on Cloud.

### Justification

Low learning curve which does not require programming skills.

## Data Repository

### Technology Choice

Object storage

### Justification

To support unlimited amount of data.

## Discovery and Exploration

### Technology Choice

upyter, Python, pyspark, scikit-learn, pandas, Matplotlib, PixieDust

### Justification

These are open source and can be supported within the IBM Cloud.

## Actionable Insights

### Technology Choice

IBM Watson Analytics for automatic modelling capabilities. Python, pandas and scikit-learn can also be considered as these are widely used and hence support can be readily available.

### Justification

Relevance and support availability.

## Applications / Data Products

### Technology Choice

Node-RED allowing for fast creation of user interfaces.

### Justification

Modularity.

## Security, Information Governance and Systems Management

### Technology Choice

IBM App ID for cloud-based user and identity management.

### Justification

Integration with other commonly used platforms such as Google and Facebook.