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

CSV file from Kaggle Datasets

### Justification

* CSV files is common formats for storing structured data. They allow for easy import and manipulation using data science libraries in Python.

## Enterprise Data

### Technology Choice

N/A for this specific project, as the dataset used is not enterprise-level.

### Justification

The project uses a publicly available dataset and does not involve enterprise data integration.

## Streaming analytics

### Technology Choice

Not applicable.

### Justification

The project does not require real-time data processing or streaming analytics.

## Data Integration

### Technology Choice

N/A for this specific project.

### Justification

The project focuses on static data analysis and does not require integration of multiple data sources.

## Data Repository

### Technology Choice

Local storage for dataset files.

### Justification

For the scope of this project, local storage suffices to hold the dataset. This choice simplifies the data handling process during the development and testing phases.

## Discovery and Exploration

### Technology Choice

Pandas, Matplotlib, Seaborn.

### Justification

Pandas is used for data manipulation and basic exploration. Matplotlib and Seaborn are used for creating visualizations to better understand the data distribution, relationships, and potential anomalies.

## Actionable Insights

### Technology Choice

Scikit-learn for machine learning model development.

### Justification

Scikit-learn provides simple and efficient tools for data mining and data analysis. It is built on NumPy, SciPy, and Matplotlib, and it offers a wide range of algorithms for classification, regression, and clustering, which are essential for deriving actionable insights.

## Applications / Data Products

### Technology Choice

Not Applicable

### Justification

Not deployed the model on any application.

## Security, Information Governance and Systems Management

### Technology Choice

Basic data governance practices, secure handling of sensitive data, and version control with Git.

### Justification

Ensuring data security and proper version control is crucial for maintaining the integrity and confidentiality of the data. Git is used for version control, and best practices for data security are followed to protect sensitive information.