

IBM Coursera Advanced Data Science Capstone

Architectural Decisions Document

1. Architectural Components Overview

1.1. Data Source

1.1.1. Technology Choice

This is a R&D project for which a simple pandas read_csv method has been used to read a csv file.

1.1.2. Justification

Data has been stored as a CSV file in the local file system. This can be easily read using read_csv.

1.2. Enterprise Data

1.2.1. Technology Choice

Not required

1.2.2. Justification

Not required

1.3. Streaming analytics

1.3.1. Technology Choice

Not required

1.3.2. Justification

Not required

1.4. Data Integration

1.4.1. Technology Choice

Not required

1.4.2. Justification

Not required

1.5. Data Repository

1.5.1. Technology Choice

Local file system

1.5.2. Justification

Single CSV is used which contain pre-processed and engineered data.

1.6. Discovery and Exploration

1.6.1. Technology Choice

Jupyter notebook to run codes. Spark and Pandas to explore and manipulate data. Matplotlib and seaborn to visualise data.

1.6.2. Justification

This is an R&D project that requires exploring data and finding a good ML / DL model than can be trained on the data.

1.7. Actionable Insights

1.7.1. Technology Choice

Not Applicable

1.7.2. Justification

Not Applicable

1.8. Applications / Data Products

1.8.1. Technology Choice

Not Applicable

1.8.2. Justification

Not Applicable

1.9. Security, Information Governance and Systems Management

1.9.1. Technology Choice

Not Applicable

1.9.2. Justification

Not Applicable