The Lightweight IBM Cloud Garage Method for Data Science

Architectural Decisions Document Template

The project “Analyze and Forecast Electricity Market in Germany” uses the lightweight IBM Cloud Garage Method process model. The lightweight IBM Cloud Garage Method for data science includes a process model to map individual technology components to the reference architecture. This method does not include any requirement engineering or design thinking tasks. Because it can be hard to initially define the architecture of a project, this method supports architectural changes during the process model.

# Architectural Components Overview



IBM Data and Analytics Reference Architecture. Source: IBM Corporation

# Data Source

Germany’s and its neighbouring countries market data can be found on the SMARD website. SMARD is a information platform that belongs to “Bundesnetzagentur” which is a federal bureau of the German state. This platform enables anyone to look up data on electricity generation, consumption, price, imports, exports and more. It aims to increase transparency and is very comprehensive.

## Technology Choice

Data can be either downloaded as csv and xlsx files from the website directly. Moreover, a RESTful API is provided. The details for this API can be found at smard.api.bund.de. At the time of composing of this document the API version is 0.0.1. OA S3.

# Enterprise Data

## Technology Choice

Data can be downloaded as csv or xlsx (or pdf). Alternatively, it can be downloaded by a REST API, which is more practical for automated downloading. There latter solution is preferred.

## Justification

Python’s request API is easy to use and capable of accepting Json files. Moreover, Python scripts can be easily be adjusted for a later automated download of new data and updating the model..

# Streaming analytics

## Technology Choice

SMARD provides data on Electricity Market without about 30 to 60 minutes delay. Certain data about market prices and stock market data take a day or two days till they are listed.

Granularity of the data is 15 minutes. That means, the prediction model can be fed in almost real-time and we can have a streaming analytics and forecasting.

In this project, streaming analytics is not used, for there is no streaming API provided by SMARD.

## Justification

No data streaming available. Data is retrieved by Restful API.

# Data Integration

## Technology Choice

Use of Pandas or rather Spark and Pandas on Spark for cleansing of data. And later in the application for adding to the existing data.

## Justification

Python is already used, so no additional language is need for use here.

# Data Repository

## Technology Choice

Please describe what technology you have defined here. Please justify below, why. In case this component is not needed justify below.

## Justification

Please justify your technology choices here.

# Discovery and Exploration

## Technology Choice

Python with Pandas, Matplotlib and plotly

## Justification

Python is already used plus Plotly is for free.

# Actionable Insights

## Technology Choice

Please describe what technology you have defined here. Please justify below, why. In case this component is not needed justify below.

## Justification

Please justify your technology choices here.

# Applications / Data Products

## Technology Choice

Please describe what technology you have defined here. Please justify below, why. In case this component is not needed justify below.

## Justification

Please justify your technology choices here.

# Security, Information Governance and Systems Management

## Technology Choice

Please describe what technology you have defined here. Please justify below, why. In case this component is not needed justify below.

## Justification

Please justify your technology choices here.