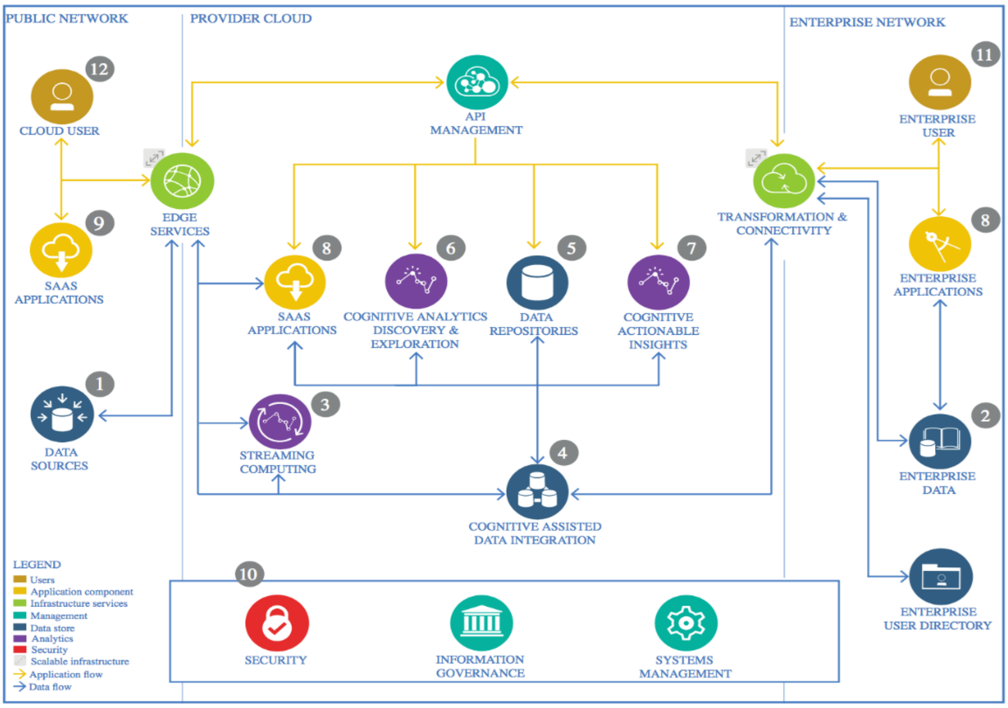
Architectural Decisions Document

Project Title	Analyze and Forecast Electricity Market in Germany
Author	Mehmet Kadri Is
Date	19/05/2024

Architectural Components Overview

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.



IBM Data and Analytics Reference Architecture. Source: IBM Corporation

1. Data Source

The Germany's electricity market data can be found in SMARD website. SMARD is a information platform that belongs to "Bundesnetzagentur". This platform makes it possible to look up data on electricity generation. Consumption, price, imports, exports and other stuff. It aims to increase transparency and in our case is the best data source we can find.

2. Enterprise Data

2.1. Technology Choice

Data can be download in pdf, xml, csv or excel formats with different time intervals from SMARD website. In this case, we use csv files.

2.2. Justification

These datasets are not huge. Therefore, they can be easily downloaded and stored in csv format, which is very easy to read and to work with in python.

3. Streaming analytics

3.1. Technology Choice

SMARD websites provides data on Electricity Market nearly in real-time. 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 simplicity. But it can be implemented at any time. For example, using IBM Streaming Analytics service.

3.2. Justification

IBM Streaming Analytics provides fast streaming application delivery using Python. Data Scientists and Developers can use existing Python code for building Streams applications without starting from scratch.

4. Data Integration

4.1. Technology Choice

All the datasets have been downloaded to a local machine. In the ETL jupyter notebook can be seen that data is cleaned, merged and get ready for building a model. (In case of a real project or huge data, they can be loaded into a data warehouse for example IBM Object Storage)

4.2. Justification

Jupyter notebooks and python are now mostly used by data scientists and it they are easy technologies to work with. That's why everything is done using python.

5. Data Repository

5.1. Technology Choice

Part of the job is done locally. So, there's a directory with all the data on local machine. Moreover, they are pushed regularly to a GitHub repository as backup. The other part of the job, which includes training RNNs, is done on cloud, specifically on IBM Watson studio. The models are then stored to IBM Object Storage, and finally downloaded to local machine.

5.2. Justification

IBM Watson Studio provides an environment with 16 vCPU and 64 GB RAM for free. Specially, for the RNNs, which take a lot of time for training, this works much better than a local machine. Also, storing the models in Object Storage can be used later for building an interactive product using a REST API for example.

6. Discovery and Exploration

There is a Jupyter notebook especially for EDA. In these notebooks, data is explored. The Electricity consumption and generation in Germany in the last 5 years is visualized using matplotlib.

7. Applications / Data Products

Electricity producers need to know in advance when electricity will be needed from their installations, so electricity suppliers report the demand they expect from their customers. These consumption forecasts allow electricity producers to plan the use of their generating capacity. The electricity producers also produce generation forecasts for their power plants. Although consumption and generation forecasts are always improving, they are still not perfect. The actual figures are not the same as the forecast ones. Forecasting errors may occur because consumers need more electricity than expected, for example, or because the wind is stronger than forecast.