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 a) Datastream **E1a** contains measured (Link to Datastream D) values of gas phase pollutants (e.g. Ozone, NO2, SO2, CO) measured in 2017 over Germany hourly ([https://datahub.uba.de/server/rest/directories/arcgisforinspire/INSPIRE/aqd\_MapServer/Daten/AQD\_DE\_E1a\_2017.zip](https://datahub.uba.de/server/rest/directories/arcgisforinspire/INSPIRE/aqd_MapServer/Daten/AQD_DE_E1a_2017.zip" \t "_blank)). b) Prevalence of Heart failures in the 2017: [https://www.versorgungsatlas.de/fileadmin/excel/data\_id\_97\_kreis11\_2\_j\_1483228800.xlsx](https://www.versorgungsatlas.de/fileadmin/excel/data_id_97_kreis11_2_j_1483228800.xlsx" \t "_blank)

### Technology Choice

Please describe what technology you have defined here. Please justify below, why. In case this component is not needed justify below.  
 The datasets are choosen in xml format for Datastream **E1a, and xlsx format for Heart Failure rates**

### Justification

Please justify your technology choices here.

These formats (xml and xlsx) are available and could be easily parsed by means of pandas.

## Enterprise Data

Pre-processed data set, including pollutant values and heart failure rates

### Technology Choice

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

Pandas data frame

### Justification

Please justify your technology choices here.

Easy to pre-process and perform exploratory analysis

## Streaming analytics

### Technology Choice

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

Not needed

### Justification

Please justify your technology choices here.

New data becomes available in one chunk pro year.

## Data Integration

### Technology Choice

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

Spark RDS object with SQL connector

### Justification

Please justify your technology choices here.

Usage of IBM Watson Studio and getting familiar with the toolbox.

## 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

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.

## 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.