­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

We use the data from the NASA Ames Prognostics Data Repository.

More infos can be searched in this paper of A. Saxena and K. Goebel (2008). "Turbofan Engine Degradation Simulation Data Set", NASA Ames Prognostics Data Repository ([https://ti.arc.nasa.gov/tech/dash/groups/pcoe/prognostic-data-repository/#turbofan)](https://ti.arc.nasa.gov/tech/dash/groups/pcoe/prognostic-data-repository/" \l "turbofan)), NASA Ames Research Center, Moffett Field, CA .

### Justification

The simulated data is focuses exclusively on prognostic problem of aircraft engine, i.e., data sets that can be used for development of prognostic algorithms. Engine degradation simulation was carried out using C-MAPSS. Four different sets were simulated under different combinations of operational conditions and fault modes, however, we only utilized the first set of data. Records several sensor channels to characterize fault evolution.

## Enterprise Data

### Technology Choice

Python 3.6

Tensorflow-gpu 2.1.0

sklearn 0.22.1

matplotlib 3.0.2

pandas 0.23.4

### Justification

Stable versions of python and sklearn stacks, moreover, tensorflow-gpu is used for better run time of feed forward network and recurrent neural network.



## Streaming analytics

### Technology Choice

At the moment, no streaming were setting up. However, in the future we can built an online algorithm, where it will use live data and predict on potential issue with the aircraft.

### Justification

No need for now.

## Data Integration

### Technology Choice

For bringing this to production, we still need to have a bigger dataset to test on, and more hyper-parameter tuning will make the model more reliable.

### Justification

A realistic dataset will be critical for the application in practice.

## Data Repository

### Technology Choice

Github, local server and dvc can be used for the version control of the data and code.

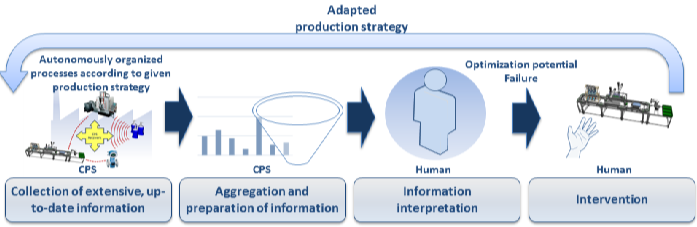
### Justification

These are the best open-source available tools in the industry.

## Discovery and Exploration

### Technology Choice

We might incorporate the human-in-the-loop to the process to validate the potential tricky validation.



### Justification

Human and machine interaction will benefit for both, hence, bring a better results if we do it properly. Moreover, we need the model users to understand the model and know when something wrong happened.

## Actionable Insights

### Technology Choice

The model is rather sensitive on the two thresholds of w0 and w1.

### Justification

Automate the process of parameter tuning and understand the sensitivity analysis results will be important for the next step.

## Applications / Data Products

### Technology Choice

Prognostic and predictive maintenance for aircraft engine.

### Justification

There are many current engineering approach, however, a data-oriented approach will be beneficial for the airline.

## Security, Information Governance and Systems Management

### Technology Choice

The application in practice will require a high standard of data security. Therefore, different types of users will have different level of access to the system.

### Justification

The project needs security domain experts to be able to classify what is good and what is bad and be able to provide insights into why that is the case that we must use some technology and not the others.