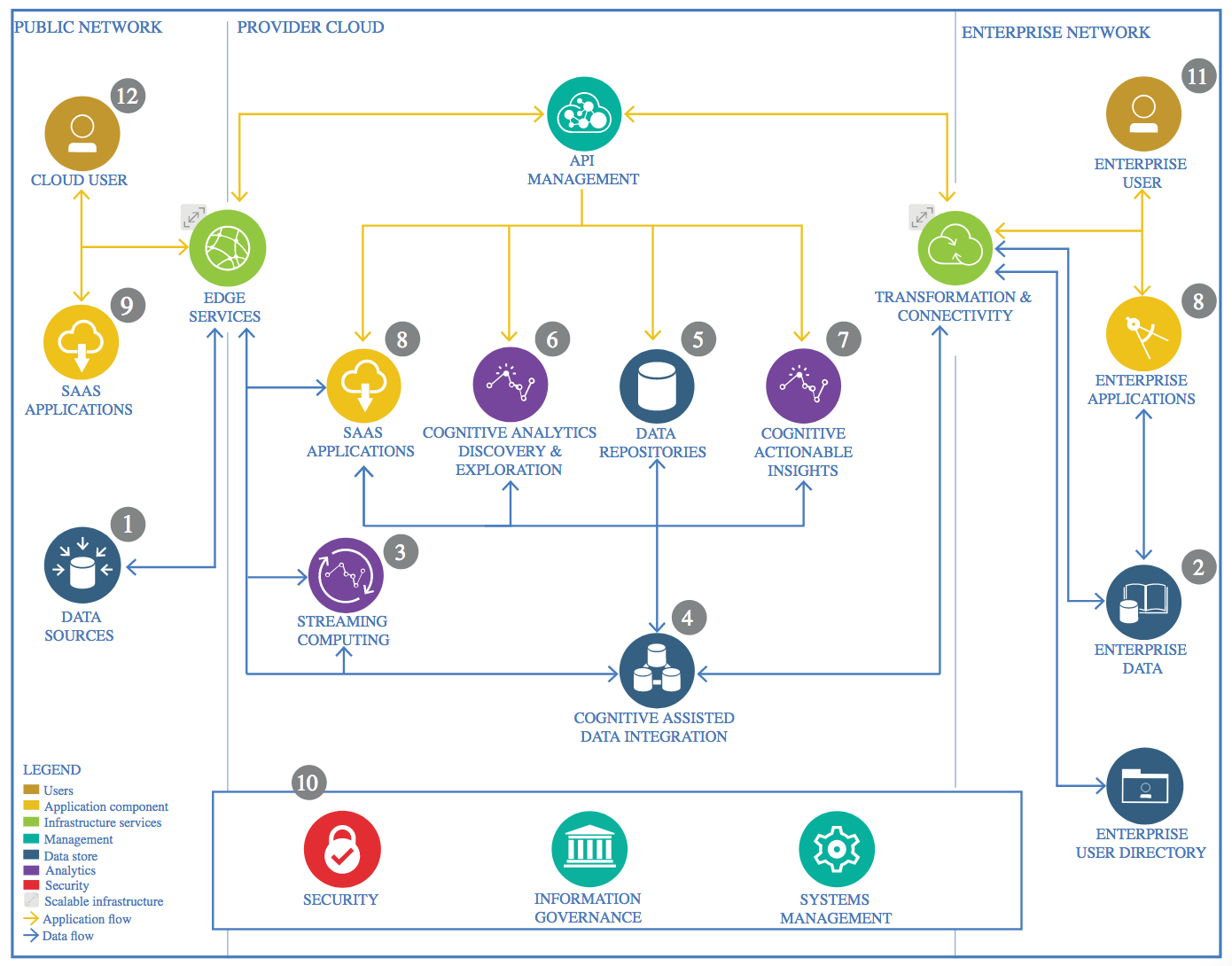
The Lightweight IBM Cloud Garage Method for Data Science

Architectural Decisions Document Template

# Architectural Components Overview



IBM Data and Analytics Reference Architecture. Source: IBM Corporation

## Data Source

### Technology Choice

Data source we get from the yahoo finance stock data. Use case we use is predicting the stock open position for the next day.

### Justification

I use yahoo finance stock data because many article on the internet say, it is the place to download csv for reliable stock data. The data it self is so clean, so i have less housekeeping to do, regarding to data cleansing. But i also do the data transforming and scaling.

## Enterprise Data

### Technology Choice

We use the open market stock price of the public traded enterprise. The data it self is from the finance.yahoo.com it in the section of historical data.

### Justification

The data I use is open source mean that it sharable for the public mass. I don’t use entreprise data because the stock data is open for public, because the company that traded in stock market is public company, so it is common sense to know the company stock price.

## Streaming analytics

### Technology Choice

I don’t use any streaming analytic technology. My prediction still not dynamic. It just still read from csv file from the finance.yahoo.com

The IBM Watson stream analytic is the best i know for this. It make a lot of easier for internet of thing device. I will make it soon.

### Justification

I use the IBM watson, so it is on the ibm environment. Meanwhile IBM watson give great stream analytic we have here. It is great tool for expert and it give a lot of important parameter to measure the process modelling, from raw data to final model.

## Data Integration

### Technology Choice

I take the data from one source. It form is csv file not sql. It is from finance.yahoo.com. It already represent enough for the prediction needed.

I don’t use any other data other than that and not integrate any either.

### Justification

The datasource from one source is trustable enough to predict the actual stock price. It enough to predict the open price of the stock. So it doesn’t need integration from other comparison stock history price list.

## Data Repository

### Technology Choice

Use the local repository, integrated to cloud storage and the ibm asset data storage

### Justification

When i run and remodel ML model or just try to do different cleansing, when i do offline, i need local storage to store the file, model or dataset. So for that reason i also repos on the local beside ibm watson cloud.

## Discovery and Exploration

### Technology Choice

I do data discovery by browsing through search engine and seek on the yahoo finance for particular stock data desired to predict (e.g googl, amzn etc). I explore 5 months data history from the day i generate the csv. The yahoo finance platform then generate downloadable csv and i just extract it to the machine learning dataset.

I use LSTM to train the model. The percentage of test is 20% and train 80%. i use 1 neural network only.

I use RMSE to evaluate the model.

### Justification

I use yahoo finance because it reliable data source for stock history data and it also have clean unambiguous csv file.

I use LSTM because the stock market data have the trait of sequence timed series of data. The RNN is not good fit because it may make the machine exhausted because lack of forget gate. In that case the LSTM that learn from the past and pass the data to the next model and with the forget gate is very suitable.

## Actionable Insights

### Technology Choice

Actionable insight we use here is as the plot jpeg file that show the comparison of test data and prediction data.

The other thing is the jupyter notebook it self and also pdf

We give the pdf resulting of plot of several publicly traded company stock and plot the comparison of test and prediction data.

### Justification

Business stakeholder need to easy open and read the plot of visualization of the prediction.

It must be understandable without telling. It needed for fast business decision.

The successor data scientist or developer need to improve or rengineering or visualize more the model need to have jupyter notebook file itself.

## Applications / Data Products

### Technology Choice

Application used in this case is to predict open price of the stock price using other parameter as label

Data product is open price of daily stock basis

### Justification

The prediction of open price give the stakeholder or investor the ability to decide wheter they buy or sell or hold the stock to internal benefit of the enterprise.

Data product of open price of the daily stock. The open market quite sensitive to price. So it is better predict the open price rather than close price.

## Security, Information Governance and Systems Management

### Technology Choice

Security

I believe that when we upload the project in the IBM watson studio, it has been keep as security matter.

Information Governance

Information governance used in the project follow the common security governance practice. If want to use, we must give proper governance policy like who can acess the data, what level of use they can use the model, legal aproach to use the data of consumer etc. I still don’t use any information governance, but will use ARMA approach

System Management

I use IBM watson platform to manage my model to delivery.

### Justification

Security

IBM watson, i believe have great security to protect from hacker or cracker that want the data or the model.

Information Governance

Information governance of ARMA approach proven internationally adaptive and efficient in the enterprise

System management

IBM watson system is quite great. It offer many automated and efficiency system.