Employee Attrition Predictive Modelling

Architectural Decisions Document

# Architectural Components Overview



IBM Data and Analytics Reference Architecture. Source: IBM Corporation

## Data Source

### Technology Choice

The data source is provided by page Kaggle.com competition. Since the data is prepared in csv format then I choose Pandas library to load data in dataframe.

Technology choice:

* Jupyter Notebooks
* Pandas library

### Justification

Since data is not big then Pandas library is adequate.

## Discovery and Exploration

### Technology Choice

The following Python 3.6 libraries were used for Data Exploration and Visualization: -

Pandas,

Matplotlib,

Seaborn

### Justification

The size of the dataset was the key factor in deciding data exploration tools.

The current data small enough to be processed on a single computer ruling out the need for distributed processing (Spark, pyspark)

## Actionable Insights

### Technology Choice

The following Python 3.6 libraries were used for Actionable Insights: -

Pandas,

scikit-learn,

Keras,

Tensoflow.

### Justification

To understand the Correlating features a white-box model was required. Using scikit-learn for classifying models.

Neural network based algorithm was used as a reference for the Tree based model. Easiest and Fastest implementation is possible in Keras. Tensorflow is the backend.

## Applications / Data Products

### Technology Choice

A Jupyter notebook based report was generated

### Justification

As only the correlating factors needed to be identified Jupyter notebook based report was consider sufficient.

## Security, Information Governance and Systems Management

### Technology Choice

NA

### Justification

NA