

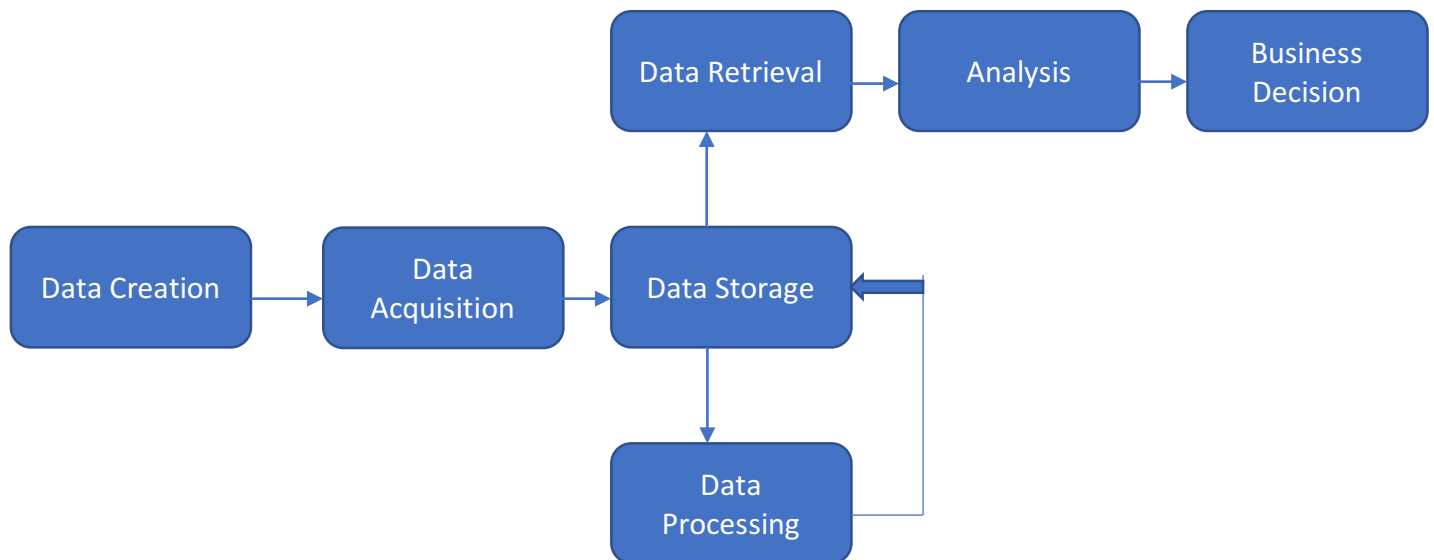
DATA ANALYTICS FUNDAMENTALS

DAY- 1 NOTES

DATE: 13-Jan-2018

COLLECTED BY- SMITHA PAMBALATH

Different Stages of Data Processing:



Data Creation/Origin: Data can be originated from many resources.

- It can be natural (i.e. Temperature of the city)
- It can be human created (i.e. No of product assembled in a factory)
- It can be non-human originated (i.e. quantity of honey made by bee)
- It can person specific (i.e. Sugar level, blood pressure of single person)

Data Acquisition: acquiring data from the origin

(I.e. taking temperature measurement from Thermometer)

Storage: Storing the data on different mediums (Like: excel, Database)

Extraction: Extracting desired data from storage

(I.e. Reports)

Analysis: Analysing the data and taking summary from that and to see what our data represent.

Decision: Taking decision based on data analysis and summary of data.

Conditions to be met for a good data set:

- 1) The data must be meaningful. It should not contain spurious co-relations.
- 2) The sampling of the data-set must be sizeable. Larger data-sets covering various patterns and collected over a large period makes for a lot more accurate data set.
- 3) All the exhaustive parameters that influence the data should be considered.
- 4) Confidence of data/Margin of error/Risk accounting must be done
- 5) Design effective experiments to collect the most relevant data

BIAS:

The error induced by wrong choice of data collection is termed as BIAS. The types of Bias are:

- 1)Temporal Bias: Bias introduced due to time factor
- 2) Spatial Bias: Bias introduced due to the location where the data is collected
- 3)Societal Bias: Bias introduced which is specific to societies.

Suggested Reference Books:

- 1) Kishore S Trivedi : for Probability
- 2) Forecasting by Makridakis : On Analytics
- 3) Ledolter and Abraham

