BI Unit 1- Part 3
Topic: Business Intelligence Architecture

Presented by: Drashti Shrimal

Topics :

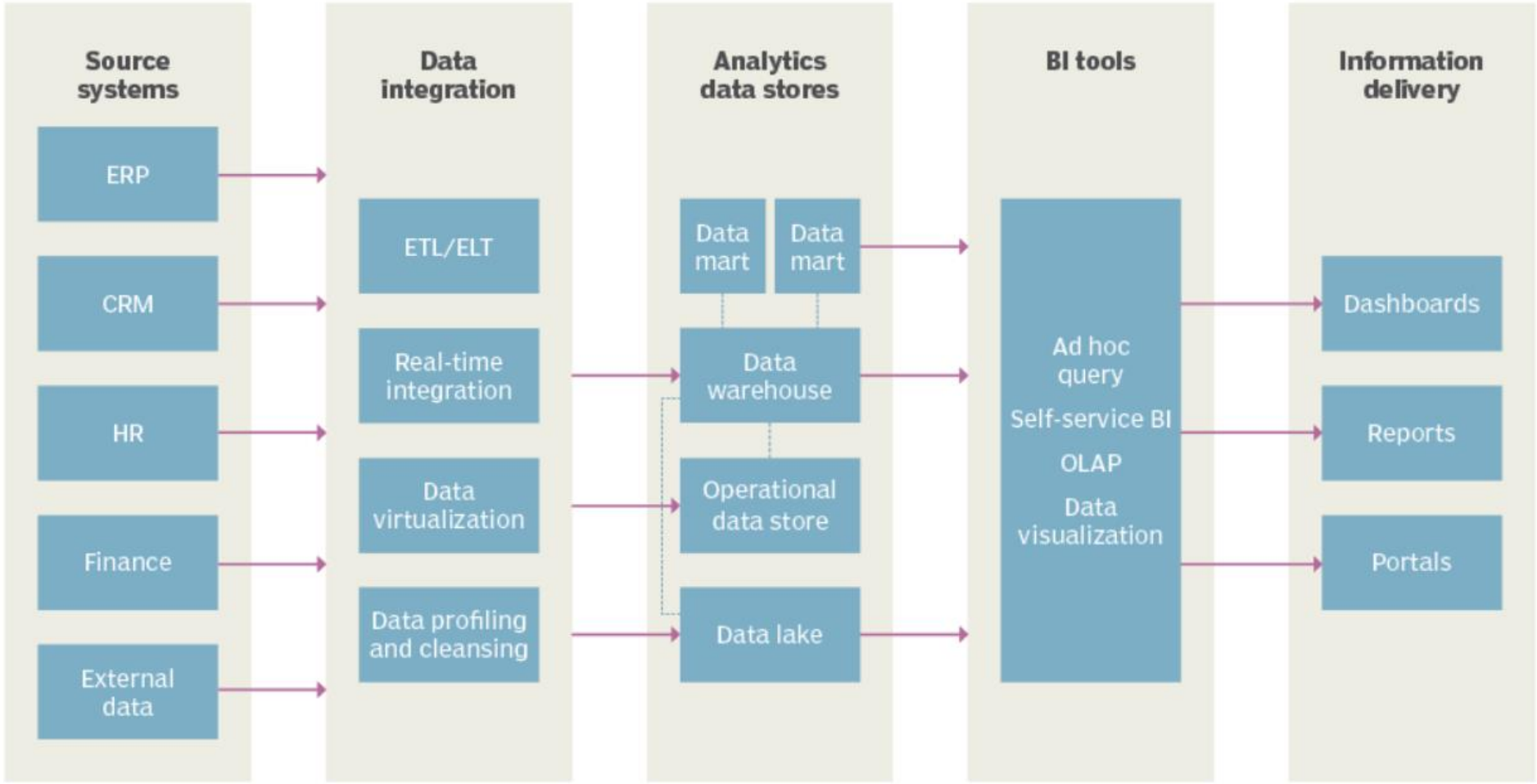
BI architecture

Data preprocessing: Cleaning - Missing values, inconsistent values, noisy data.

Data preprocessing: Transformation

Data preprocessing: Reduction

BI Architecture



BI Architecture

- Source Systems :

Systems from where the data is collected for analysis.

- Data Integration :

Merging the data collected.

- Analytic stores :

Creating various sub parts of whole data like Data Marts, Data Views.

- BI Tools :

Mining tools which perform analysis on the analytics stores.

- Information Delivery :

Various tools to display reports of analysis.

Data Pre-processing - Cleaning

- Missing values :

This situation arises when some data is missing in the data. It can be handled in various ways.

Some of them are:

a. Ignore the tuples:

This approach is suitable only when the dataset we have is quite large and multiple values are missing within a tuple.

b. Fill the Missing values:

There are various ways to do this task. You can choose to fill the missing values manually, by attribute mean or the most probable value.

Data Pre-processing - Cleaning

- Noisy values :

Noisy data is a meaningless data that can’t be interpreted by machines.It can be generated due to faulty data collection, data entry errors etc. It can be handled in following ways :

Binning Method:

This method works on sorted data in order to smooth it. The whole data is divided into segments of equal size and then various methods are performed to complete the task.

Data Pre-processing - Cleaning

- Inconsistent values :

Examples - Male/Female in one set and M/F in other.

- Redundant values :

Copies of data all over the set.

- Outliers :

An outlier is an observation that lies an abnormal distance from other values in a random sample from a population.

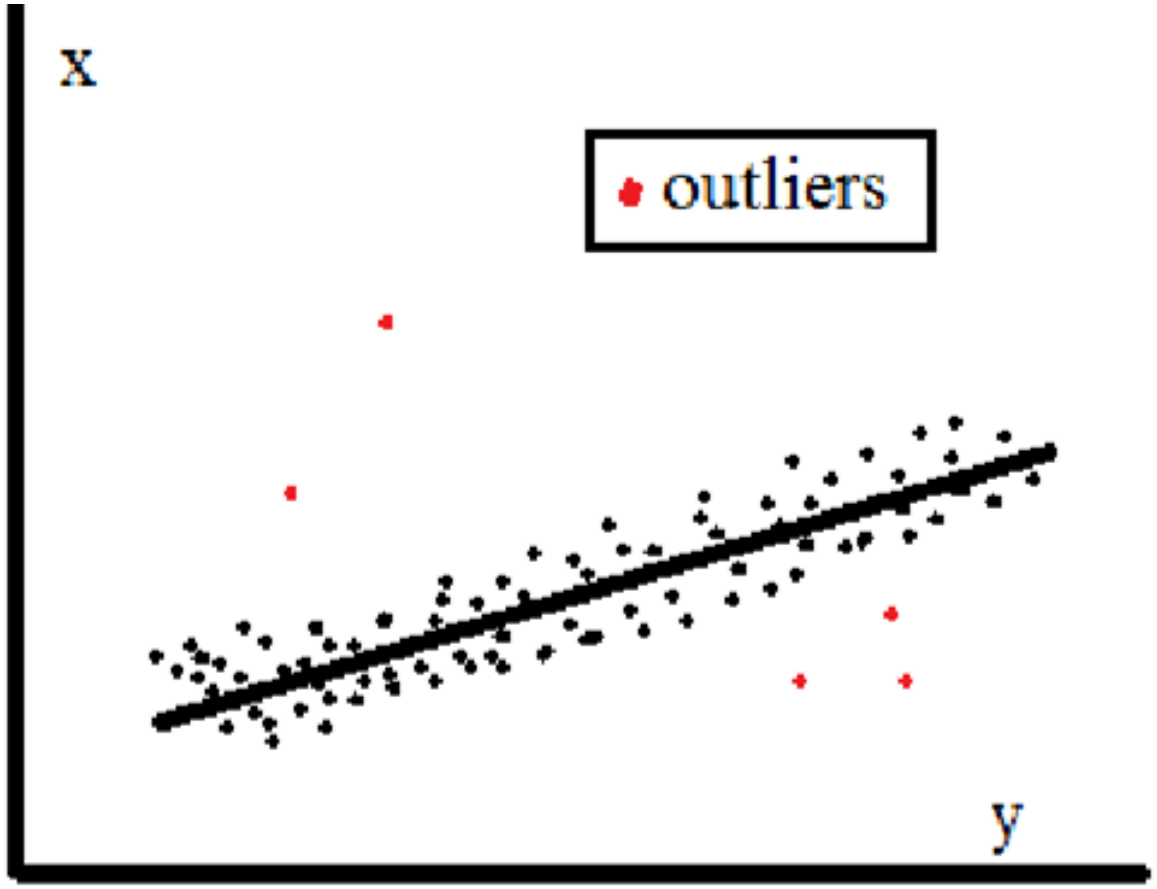
Data Pre-processing - Transformation

- This step is taken in order to transform the data in appropriate forms suitable for mining process.
- This involves following ways:

Normalization

Attribute Selection

Outliers



Data Pre-processing -Reduction

- Data Reduction:

Since data mining is a technique that is used to handle huge amount of data. While working with huge volume of data, analysis became harder in such cases. In order to get rid of this, we uses data reduction technique. It aims to increase the storage efficiency and reduce data storage and analysis costs.

Data Binning

- Data binning, bucketing is a data pre-processing method used to minimize the effects of small observation errors.
- The original data values are divided into small intervals known as bins and then they are replaced by a general value calculated for that bin.
- This has a smoothing effect on the input data and may also reduce the chances of overfitting in case of small datasets