

Introduction to BusinessIntelligence

* Operational & Decision Support System:

Data System
Available

Operational
System

Information system/
Decision Support System

Eg : Sales, order
processing,

Eg : Customer segmentation

human resource
planning

Reservation
System

OLAP

OLTP
RDBMS

DW.

Operational System : [This system contains operational data which is a detailed record of a company's daily business transactions.]

- [It provides information to run the day-to-day operations.]

- It is used to run a business in a real time, based on current data and process large volumes of relatively simple read/write transactions.

Eg: Sales order processing, reservation system

* Benefits:

- Quick retrieval,
- The information can be retrieved on continuous basis.
- The ability to share information across company.

Information System / Decision

Support System:

- It is designed to support decision making based on historical data and prediction data.)

- Design for complex queries and data mining applications]

- It contains summarization of transaction that users can make business decisions more easily.]

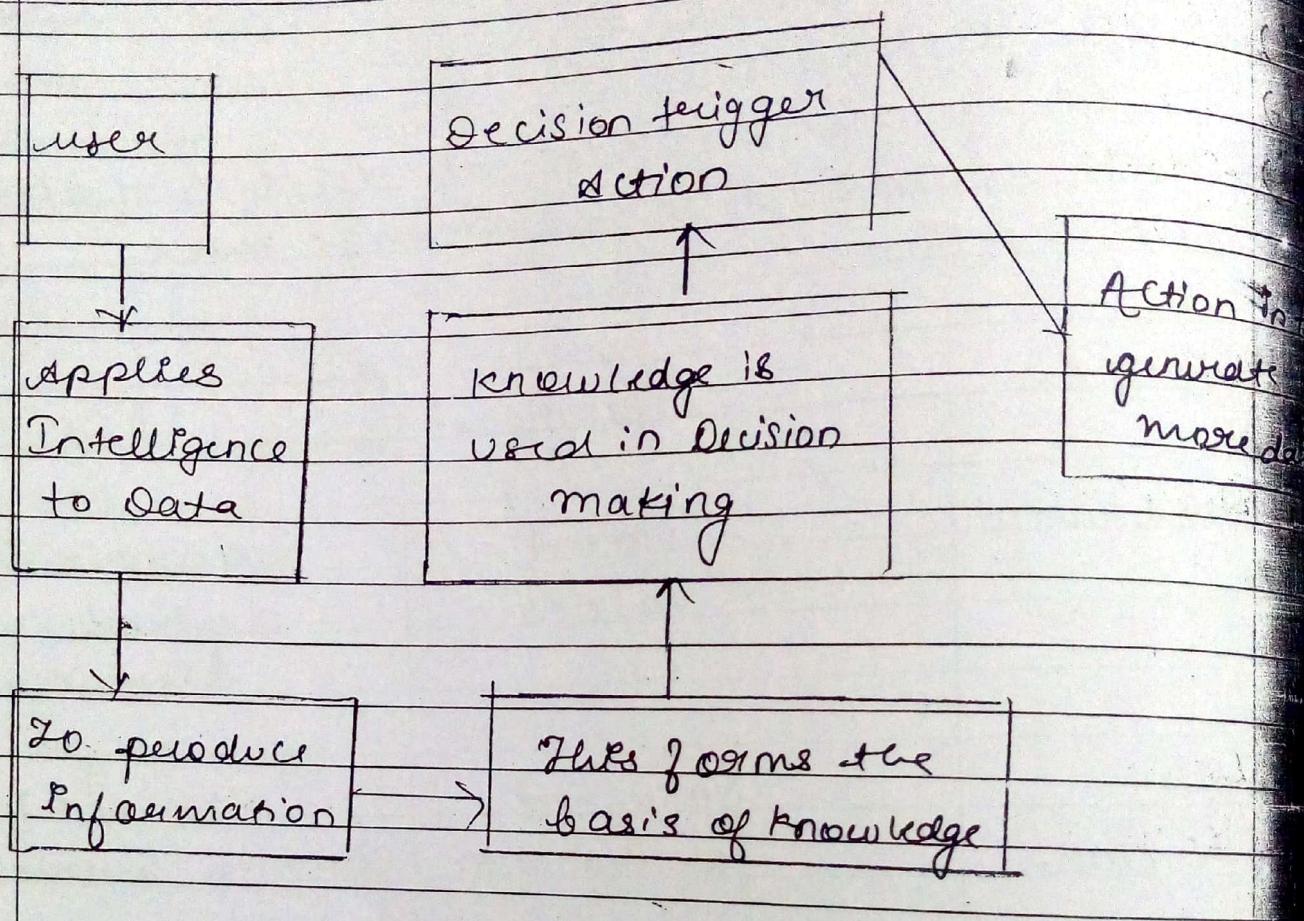
Eg: Customer segmentation, human resource planning.

* Benefits:

- i) Improve efficiency
- ii) Speed up the process of decision making.
- iii) Create competitive advantage over competition.

Difference	Operational	D.S.S.
1. Data Content	Current values	Summarized, derived, historical
2. Data Structure	Optimized for transaction Small data	Optimized for complex queries.
3. Access frequency	High	medium or low
4. Response Time	Sub - seconds	Several seconds to minutes.
5. Users	Large no's	Relatively small nos.
6. Primary Users	Clerks, Sales persons, DBA	Managers, Business analysts.

Data - Information - Knowledge decision making action cycle.



- 1) Data: Data is composed of the basic, unrefined and generally unfiltered ^{data} information
- 2) Information: Information is much more refined data that has evolved to the point of being useful for some form of analysis
- 3) Knowledge: Knowledge resides in the user, happens only when human experiences & insights are applied to the data and information.

4.) Decision-Making: Business Intelligence helps extract crucial facts from vast amount of unstructured data and transform them into information that enables companies to make informed strategic decisions, improving operational efficiency and business productivity.

Business Intelligence :-

* Definition:

- i.) B.I is an umbrella that combines architecture tools, databases, analytical tools, applications, and methodologies.]
- ii) It is a major objective is to enable interactive access to data, to enable manipulation of data and to give business managers and analysts, the ability to conduct appropriate analysis.]
- iii) By analysing historical and current data, decision makers make more informed and better decisions]
- iv) The process of B.I is based on the transformation of data to information then to decisions and finally to actions.

Performance & Strategy

Business Analysis

Environment

Data

Warehouse environment

Technical staff
Build the DW

Data Sources

Organising
Summarizing
StandardisingBusiness
AccessManipulation,
ResultsManagers,
executives

BPM Strategies

Further components
Intelligent system

User Interface
Browser
Portal
Dashboard

Architecture of

Business Intelligence

The B.I System has 4 major components:

1.) Data warehouse Environment:

- It is the corner stone of the BI system.
- DW include historical data that were organised and summarised to get easily view of information
- In addition to that some DW include current data
- So that they can provide real time decision support.

2.) Business Analytics:

- End user can work with data and information in a DW, by using variety of tools and techniques, which fits into 2 major categories : i.) Reports and Queries:

↳ B.I. include static and dynamic reporting, all types of queries, discovery of information, multidimensional view, drilled down to details and so on.

ii.) Data texts and web mining & other Sophisticated and mathematical, & statistical tools:

↳ Data mining is a process of searching for unknown relationships or information in large database or datawarehouse.

→ Estimating revenue from entertainment products such as music, movies, and so on, & for that DM tools are used.

3) BPM: Business (Corporate) Performance Management:

- i.) It is an application that contains evolving architecture and tools.
- ii.) It extends the monitoring, measuring and comparing of sales, profit, cost, profitability & other performance indicator by introducing the concept of management & feedback.

4) User Interface (Dashboard & other information broadcasting tools):

- i.) Dashboard provides visual view of corporate performance measures trends and exceptions. They integrate information from multiple business area. Dashboard presents graphs that show actual performance, compared to design metrics. i.e. view of the health of the organization.
- ii.) Broadcasting tools:
Other broadcast informations are corporate portals and other visualization tools such as multidimensional cube presentation to virtual reality.
- iii.) GIS : Geographical Information System.
It plays an increasing role in decision support.

* Styles of BI *

There are 5 Styles of BI:

- 1) Report, Delivery & Alerting
- 2) Enterprise Reporting using Dashboard
- 3) Cube Analysis (Slice & Dice Analysis)
- 4) Adhoc Queries
- 5) Statistics & Data Mining

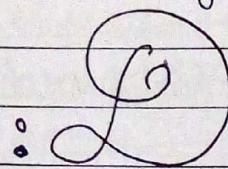
* Benefits of BI *

- i) BI provides accurate information to the company when needed, including real time view of corporate performances.
- ii) BI provides information for all types of decisions, for strategic planning, & for company survival.
- iii) BI used for general reporting, sales & marketing analysis, planning and forecasting, budgeting and profitability analysis.
- iv) Organization need accurate data to support decision making in order to improve business operations.

Need of BI:

- Managers need the right information at the right time and in right place.
- Organization have to work smart, because this, organizations need BI.

(4)



: Discretization:

- * Definition: [In machine learning, discretization refers to the process of converting or partitioning continuous attributes or variables, to discretize variables or attributes]
- * Discretization involves dividing the range of possible values into subranges called bins or bins]

* Discretization Methods:-

There are 2 discretization methods

i) Equal Width Binning:

It divides the range of possible values into N subranges of the same size

$$\text{Bin Width} = \frac{\text{max} - \text{min}}{N}$$

g: Data = 0, 4, 12, 16, 16, 18, 24, 26, 28

$$\text{Bin Width} = \frac{\text{max value} - \text{min value}}{N} = \frac{28 - 0}{9}$$

bin 1 = 0, 4 [- , 10)
bin 2 = 12, 16, 16, 18 [10 , 20)
bin 3 = 24, 26, 28 [20 , +)

ii.) Equal Depth / Equal height / Equal frequency binning :

It divides the range of possible values into N bins, each of which holds the same number of training instances
\\ same no of instances

g: Data: 0, 4, 12, 16, 16, 18, 24, 26, 28

bin 1 = 0, 4, 12 [- , 14)
bin 2 = 16, 16, 18 [14, 21)
bin 3 = 24, 26, 28 [21, +)

* Equal Width Binning : Example.

1.) If the observed values are all between 0 - 100 we could create 5 bins

$$\text{width} = \frac{100 - 0}{5} = \frac{100}{5} = 20$$

Bins are : [0, 20]
[20, 40]
[40, 60]
[60, 80]
[80, 100] 5 bins

* Example of Equal Depth Binning:

- 1) If we have values containing 10 training samples:

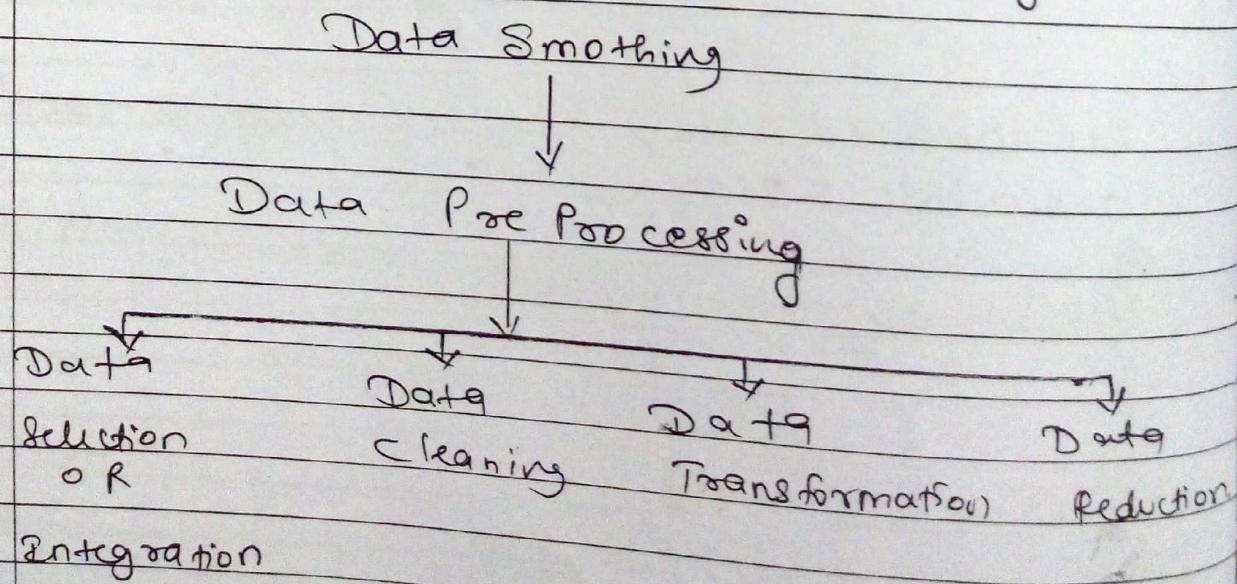
5, 7, 12, 35, 65, 82, 84, 88, 90, 95

→ To create 5 bins, we would divide up the range of values, so that each bin holds 2 of the training examples.

5, 7	12, 35	65, 82	84, 88	90, 95
------	--------	--------	--------	--------

Data Smoothing

- * It is same as the data processing.



* Data Normalization in Data Warehousing and Data Mining *

- Data Normalization in Data warehouse and mining is completely different from 1st, 2nd & 3rd form of normalizations used in RDBMS.
- The normalization in DW, is based on Euclidean distance.

→ Consider 2 points in a 2 dimensional space (x_1, y_1) & (x_2, y_2) . The distance between these 2 points is given by the formula:

$$\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

- The above formula is called the Euclidean distance between the points.
- The same concept can be extended to multi dimensional Space.
- If the points are $(a_1, b_1, c_1, d_1, \dots)$ and $(a_2, b_2, c_2, d_2, \dots)$, then the distance between these points is given by the formula:

$$\sqrt{(a_1 - a_2)^2 + (b_1 - b_2)^2 + (c_1 - c_2)^2 + \dots (d_n - d_n)^2}$$

* Sample Data :-

Emp_id	Salary	Age	Experie
1	25000	24	4
2	40000	27	5
3	55000	32	7
4	27000	25	5
5	53000	30	5

The Euclidean distance between points 1 & 2
 (Emp_ids, 1 & 2) can be calculated as below:-

$$\sqrt{(25000 - 40000)^2 + (24 - 27)^2 + (4 - 5)^2}$$

$$= \sqrt{50000.000333}$$

	1	2	3	4	5
1	0	15000.00033	30000.001217	2000	28000
2	15000.00033	0	10000.00145	13000.00154	13000.00
3	30000.001217	15000.000967	0	28000.00154	2000
4	2000	13000.000154	28000.0009	0	26000
5	28000.0006	13000.0005	2000.002	26000.00487	0

Min-Max Normalization:-

Date 1/1
Page 73
STUDY BUDDIES

The attribute data is scaled to fit in a specific range. There are many type of normalization available, out of which one technique is Min-Max Normalization.

- It transforms a value $A \rightarrow B$, which fits in the range $[C, D]$

$$B = \left[\frac{(A - \text{Minimum value of } A)}{(\text{maximum value of } A - \text{minimum value of } A)} \right] * (D - C) + C$$

Consider the below example, the salary value is 50,000. We want to transform this into the range $[0.0, 1.0]$. The maximum value of salary is 55,000 & minimum value of salary is 25,000. So the new scaled value for 50,000 will be:

$$B = \left[\frac{(A - \text{minimum value of } A)}{(\text{maximum value of } A - \text{minimum value of } A)} \right] * (D - C) + C$$

$$= \left[\frac{50000 - 25000}{55000 - 25000} \right] * (1 - 0) + 0$$

$$= 0.8333 \approx 1.0$$

Now let us apply normalization technique to all the attribute in our eg. Data set

* consider the following max & min. value:

$$\hookrightarrow \text{Max Salary} = 55001$$

$$\text{Min Salary} = 24999$$

$$\text{Max Age} = 33$$

$$\text{Min Age} = 23$$

$$\text{Max Experience} = 8$$

$$\text{Min Experience} = 3$$

The attributes needs to scaled to fit into the range $[0, 1]$, applying the min-max normalization to get the normalized eg. data set as given