DATA WAREHOUSING AND MINING

T.E. CSE-Data Science, Sem V
Academic Year: 2022-23

Data Warehousing Fundamentals: Introduction Lecture 1

Poonam Pangarkar

Data vs Information

- Data is a collection of facts, while information puts those facts into context.
- Data is raw and unorganized, information is organized.
- Data points are individual and sometimes unrelated. Information maps out that data to provide a big-picture view of how it all fits together.
- Data, on its own, is meaningless. When it’s analyzed and interpreted, it becomes meaningful information.
- Data does not depend on information; however, information depends on data.
- Data, on its own, is meaningless. When it’s analyzed and interpreted, it becomes meaningful information.
- Data isn’t sufficient for decision-making, but you can make decisions based on information.

Data vs Information

Examples:

At a restaurant, a single customer's bill amount is data. However, when the restaurant owners collect and interpret multiple bills over a range of time, they can produce valuable information, such as what menu items are most popular

The number of likes on a social media post is a single element of data. When that's combined with other social media engagement statistics, like followers, comments, and shares, a company can intuit which social media platforms perform the best and which platforms they should focus on to more effectively engage their audience.

Why do we need Information?

Who needs Information

The executives and managers

Why?

- keeping the enterprise competitive
- to make proper decisions.
- formulate the business strategies
- establish goals
- set objectives and monitor results

Example:

1.Wednesday bazaar (Supermarket offer)

2.KFC Wednesday offer

3.BSNL night calling offer(9pm to 7am free calling)

4. Free movie ticket on Tuesday (Vodafone)

Strategic information is important for the continued health and survival of the corporation

Real time example of data warehousing and mining

Hong Kong world's most visited city for 9th consecutive year

Hong Kong retained its spot as the world's most visited city for the 9th consecutive year, with 26.5 million international travellers in 2016, according to market research firm Euromonitor International. It was followed by Bangkok with 21.2 million and London with 19.2 million international travellers. A total of 1.2 billion international trips were taken worldwide during the year.

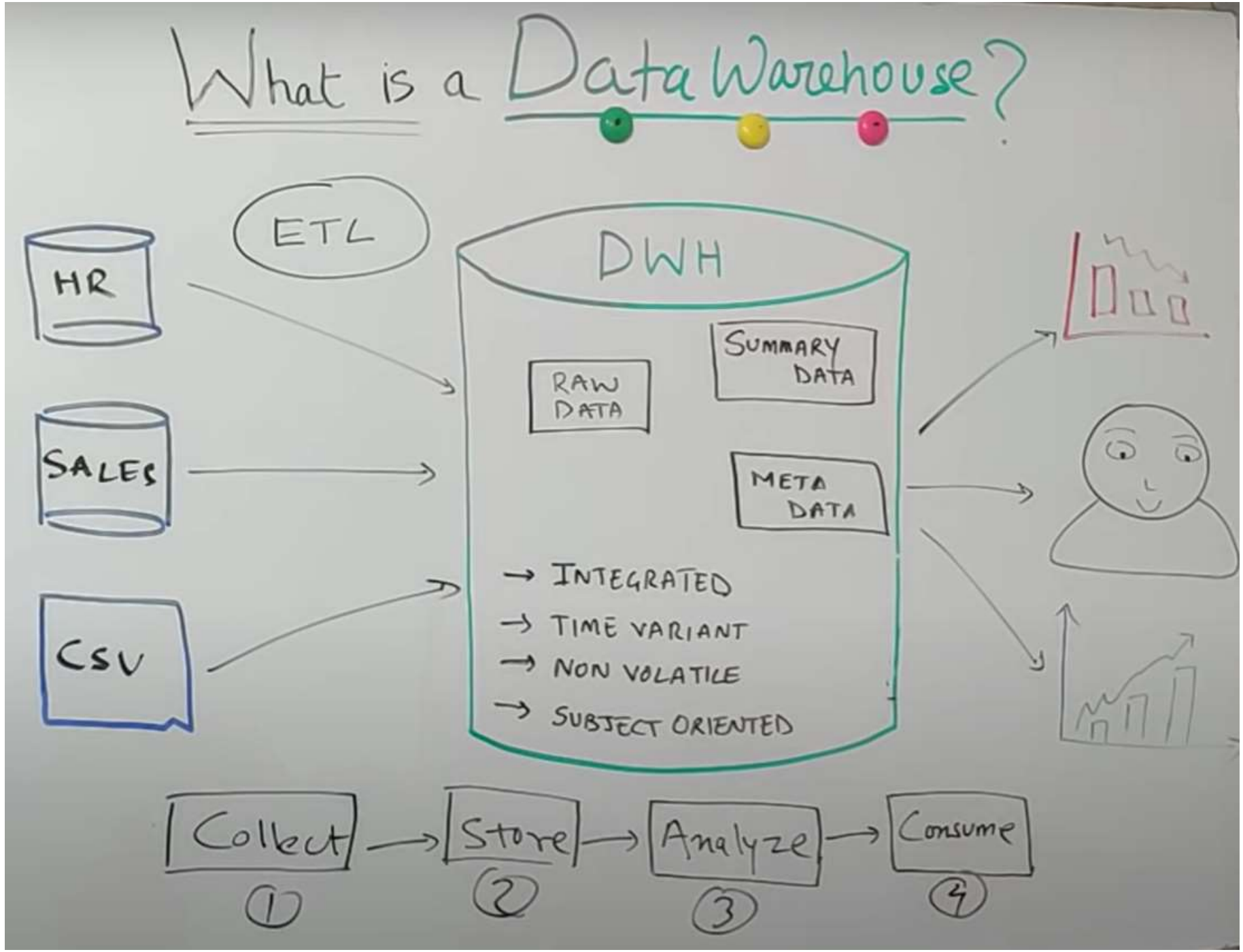
India to be high-middle income country by 2047: World Bank

Praising India's increasing per capita income, World Bank CEO Kristalina Georgieva on Saturday said she has no doubt India will be a high-middle income country by 2047, when it completes hundred years of Independence. She further lauded India for the 30-rank jump in Ease of Doing Business ranking, terming it the biggest leap ever in the history of the survey.

How do can they make such informed predictions??

Information Crisis

- Lots of information exists with the corporation
- The available data is not readily usable for strategic decision making.
- The data of an enterprise is spread across many types of incompatible structures and systems.
- The data in a corporation resides in various disparate systems, multiple platforms, and diverse structures.
- Data needed for strategic decision making must be in a format suitable for analyzing trends


What is a Data Warehouse?
ETL
HR
SALES
CSV
DWH
RAW DATA
SUMMARY DATA
META DATA
→ INTEGRATED
→ TIME VARIANT
→ NON VOLATILE
→ SUBJECT ORIENTED
Collect
1
Store
2
Analyze
3
Consume
4

The need for data warehousing

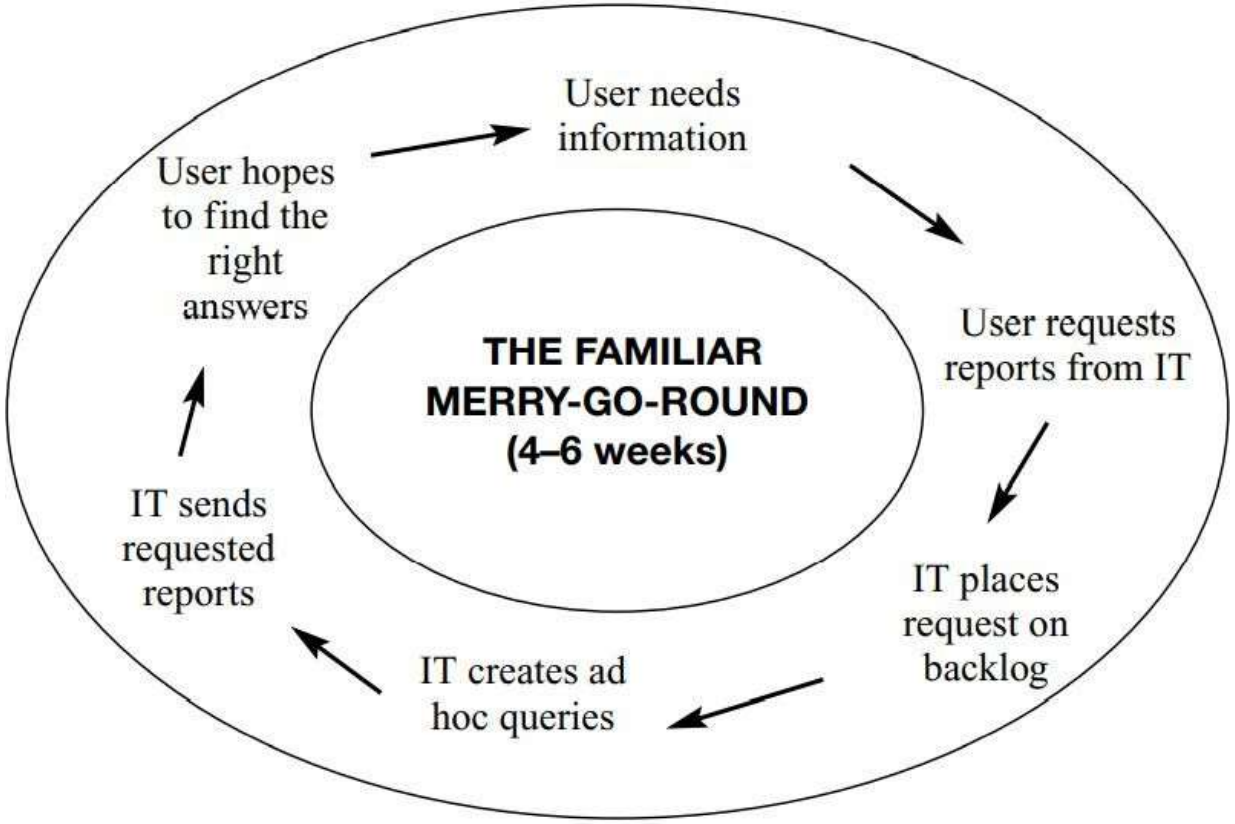
Suppose a CEO/Manager of a large garments retail chain having approximately 100 stores spread throughout the country.

- If he ask one of his employee to provide a status report on the business making an overall profit or loss.
- The question is not difficult but the problem is to collecting the relevant data from 100 stores.
- With great difficulty the employee contacts each and every store and asks the store manager to give a summarized figure of profit or loss
- After obtaining 100 such figures, he calculates the cumulative result and gives it to the CEO/Manager

The need for data warehousing

- The problem does not end here for the employee, the CEO/Manager now wants a detailed product report of the previous year as he wants to know which product sold well and their marginal sale
- Again the employee contacts each and every store and the entire process is repeated.
- Such situations are there in a non-datawarehouse environment. This is where the need of the datawarehouse comes into the picture.
- Increasing demand for strategic information(tpo, bp device) and inability of past (DSS)decision support systems leads to create need of datawarehouse.

What happens when your manager needs some adhoc information



Transaction Processing Systems are insufficient/incapable to fulfil the constant need for strategic information

Inadequate attempts to provide strategic information

Data Warehouse – A Need

- Need for different types of decision support systems to provide strategic information.
- The type of information needed for strategic decision making is different from that available from operational systems.
- We need a new type of system environment for the purpose of providing strategic information for analysis, trends, and monitoring performance.

Data Warehouse – A Need

The desired features of the new type of system environment are:

- ✓Database designed for analytical tasks
- ✓Data from multiple applications
- ✓Easy to use and conducive to long interactive sessions by users
- ✓Read-intensive data usage
- ✓Direct interaction with the system by the users without IT assistance
- ✓Content updated periodically and stable
- ✓Content to include current and historical data
- ✓Ability for users to run queries and get results
- ✓Ability for users to initiate reports

Operational vs Informational System

How are they different?

	OPERATIONAL	INFORMATIONAL
Data Content	Current values	Archived, derived, summarized
Data Structure	Optimized for transactions	Optimized for complex queries
Access Frequency	High	Medium to low
Access Type	Read, update, delete	Read
Usage	Predictable, repetitive	Ad hoc, random, heuristic
Response Time	Sub-seconds	Several seconds to minutes
Users	Large number	Relatively small number

Figure 1-7 Operational and informational systems.

The data warehouse essentially holds the business intelligence for the enterprise to enable strategic decision making.

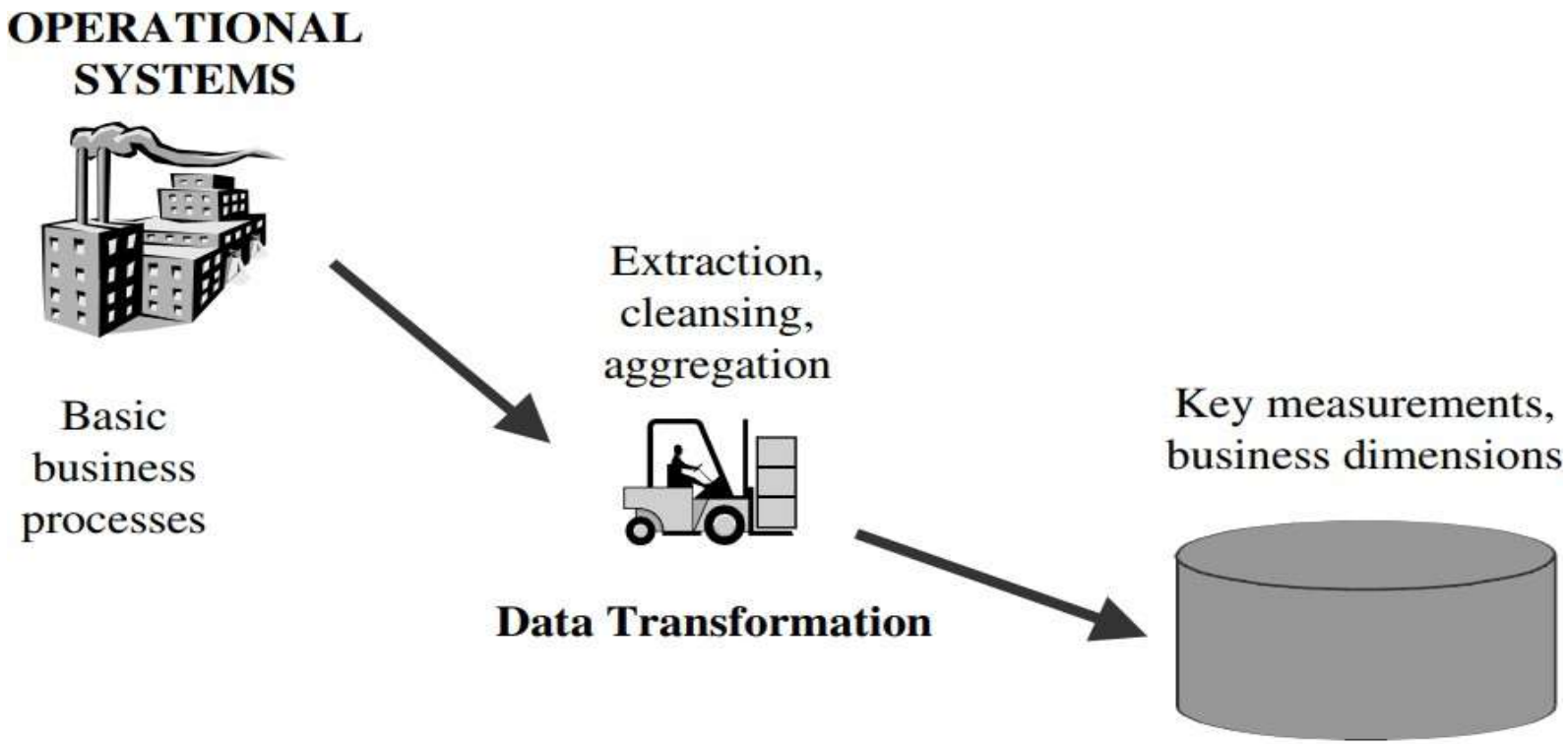
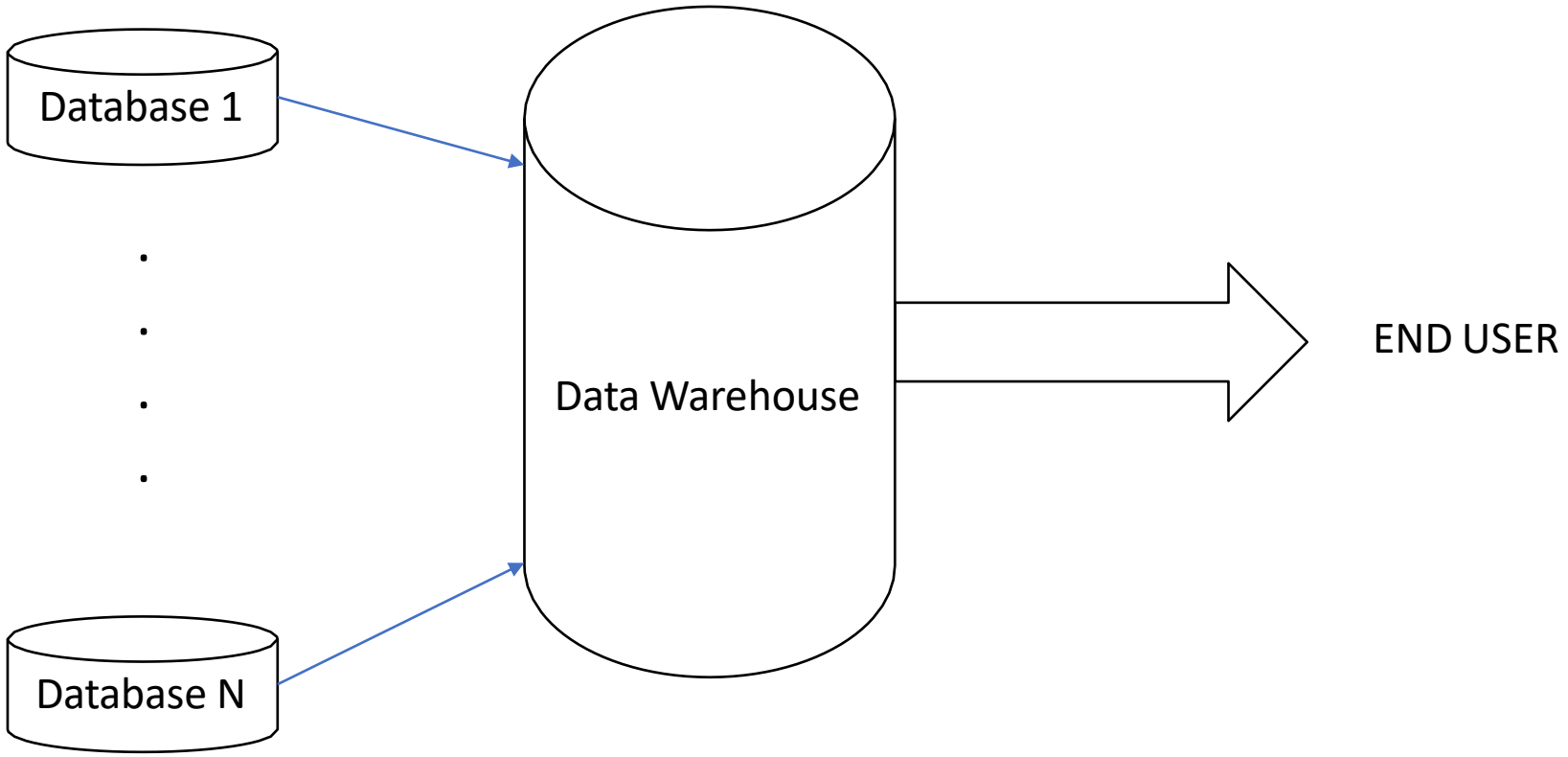


Figure 1-8 Business intelligence at the data warehouse.

Data warehouse



Exercise – Self study

1. For a commercial bank, name five types of strategic objectives.
2. Why are operational systems not suitable for providing strategic information? Give three specific reasons and explain.
3. Examine the opportunities that can be provided by strategic information for a medical center. Can you list five such opportunities?