Page No.
Date 23 01 23

Data Warehousing & Mining (BTCT602T)

* Course Outcomes :→

After completing of the course, students will be able to:

1. Explain the data warehousing component and design a data warehouse for any organization.
2. Learn data mining concept and working.
3. Explore functionality of the various data mining techniques
4. Discuss the data mining task like classification, clustering, association mining and extract knowledge using data mining techniques.
5. Apply data mining techniques in trending domain such as web mining and solve real-world problems in business and scientific information using data mining.

Date

Unit 1

* Syllabus :→

Introduction : characteristic, operational database systems and data warehouse (OLTP and OLAP), multi dimensional data models, Data warehouse architecture, OLAP operations, design and ~~decision~~ construction of data warehouses.

OLTP - Online Transactional process.

OLAP - online Analytical process.

* Textbooks :→

Jiawei Han and micheline kamber

"Data mining concept and techniques"

Second Edition Elsevier.

(store) → store historical data (old data)

* Data warehouse and characteristics :→

• A data warehouse is a subject oriented, Integrated, Non-volatile and time variant collection of data in support of management's decision making process.

• The four keywords :
  - Subject oriented
  - Integrated
  - Time Variant
  - Non-volatile

distinguish data warehouse from other data repository system such as Relational DB system, transaction processing system and file system.

① Subject Oriented :→

A data warehouse is organized around major subject such as customers, suppliers, product and sales rather than concentrating on the day to day operations and transaction's processing of an organization.

② Integrated :→

A data warehouse is usually constructed by integrating multiple heterogeneous sources such as relational database (e.g. RDBMS, file, excel & etc.) DB, flat file and online transactional record.

③ Time - Variant :→ (data can not be modified)

Data are stored to provide information from a historical perspective (the past 5-10 years).

Page No.
Date

use only SELECT query

(ii) Non-Volatile :→ (do not modified)

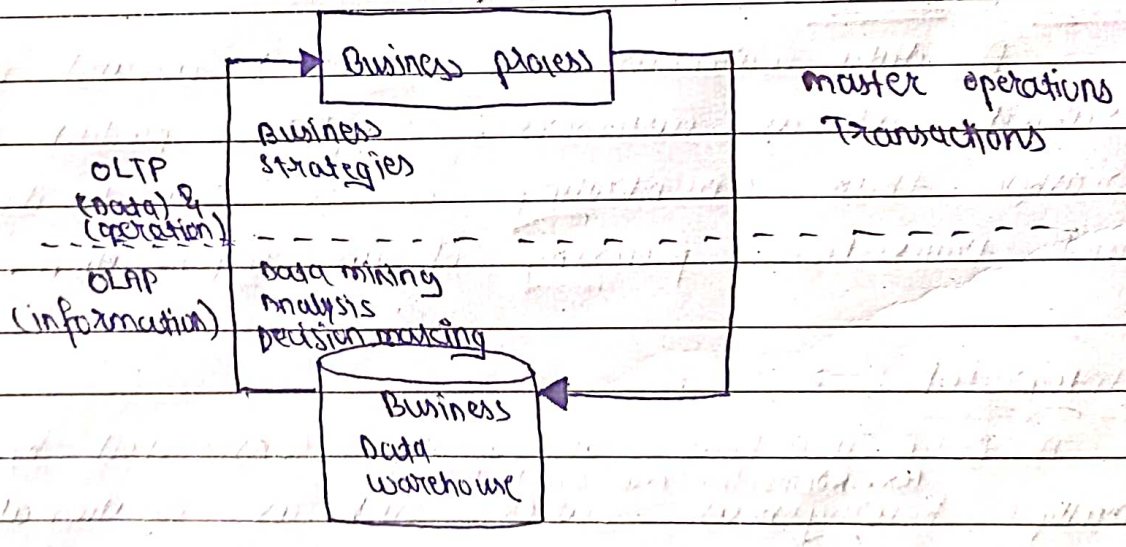
A Data warehouse is always a physically seperate store data, transform from the application data found in the operational environment. Due to this seperation the data warehouse does not require transaction processing, recovery and concurrency control mechanism.

Data warehouse

It usually requires only two operations in data Accessing ie, initial loading of data and access of data

25/10/2022

* OLAP V/S OLTP :→



7 to 8 marks

* Difference between OLAP & OLTP :→

Features	OLTP	OLAP
characteristic	operational processing	Informational processing
orientation	Transaction	Analysis
user	clerk, DBA, Database professionals	knowledge workers (eg. manager, executive, Analyst)
function	day to day operations	long term information requirements, decision support
DB design	ER diagrams, Application oriented	start / Snow flake, Subject oriented
Data	current i guranteed up to date	Historical data
Access	Read / write	mostly read
Focus	data in	information
No. of Record Acces	Tens	millions
No. of users	~~100 KB~~ Thousands	Hundreds
DB size	100 MB TO GB	100 GB to TB

features	OLTP	OLAP
priority	high performances high availibility	high flexibility, end user autonomy
metric	Transaction throughput	query throughput, response - time ...