

Introduction To Business Data Warehouse

Definition:

Many organisation produce data increasingly day by day which is complex to maintain and use.

- The no. of users needing to access information continues to increase.
- Challenge is how to manage ever increasing amount of data.
- Solution is data warehouse.
- Many organisation needs to create data warehouse, which is a massive data store of time series data for decision support.

Definition of Data Warehouse:

The Data warehouse is a decision support system to provide strategic information.

- A D.W is a subject oriented, integrated time variant and non-volatile collection of data in support of management's decision making process.
- It is a collection of current and historical data of potential interest of manager's throughout the organisation

* Characteristics of Data Warehouse

- There are 4 characteristics of DW:-

1.) Subject - Oriented:

- A DW is subject-oriented because it provides information around a subject rather than organizations or going operations
- These subjects can be product, customer, sales, loan

2.) Integrated:

- A DW is constructed by integrating data from heterogeneous sources such as relational databases, flat files etc.

3.) Time - Variant:

- The data collected in DW is identified with a particular time period.
- The data in data warehouse provides information from historical point of view
- & we can source out data according to the time.
- Non-v

4.) Non-Volatile:-

- It means the previous data is not erased when new data is added to it
- The data in D.W is not as volatile as in the operational databases

Some Additional Characteristics Of DW

1. Web-based:

- DW are typically designed to provide an efficient computing environment for web-based application.

2. Relational or Multidimensional:

- A DW uses either a relational structure or a multidimensional structure (OLAP) Online Analytical Processing

3. Client-Server:

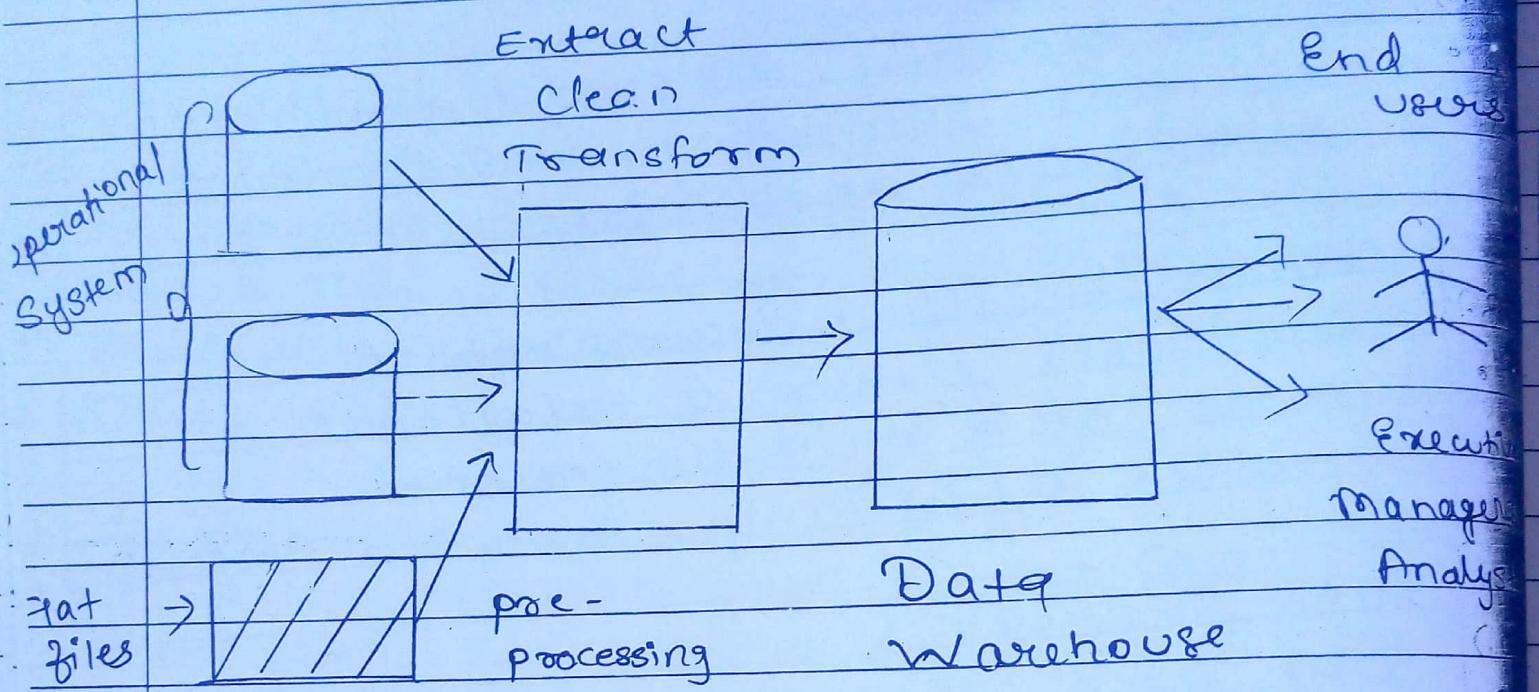
- A DW uses the client-server architecture to provide easy access for end users.

4. Real Time:

- Newer DW provide real time or active data access and analysis capabilities

5. Meta Data:

- A DW contains metadata (data about data) about how the data are organised and how to effectively use them



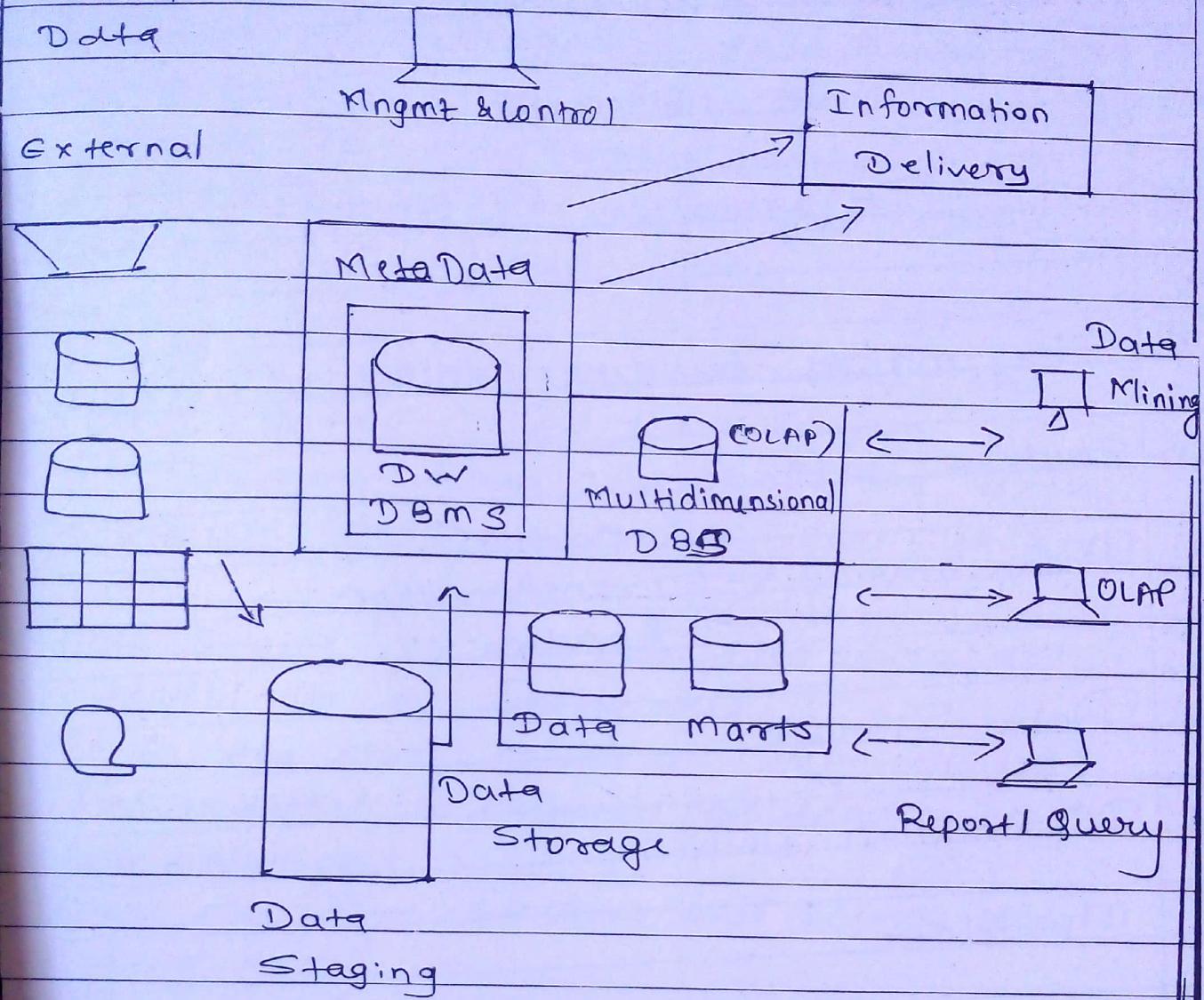
Logical Architecture of Data Warehouse

Date 11

Page 35

STUDY BUDDIES

Sourced
Data



Logical Architecture of Data Warehouse.

1) Data Warehouse Architecture:-

The structure that brings all the components of data warehouse together is known as architecture.

Architecture of DW contains three major areas.

(P.T.O.)

(i) Data Acquisition

(ii) Data Storage

(iii) Information Delivery

The major building blocks of DWH are -

(1) Sourced Data

(2) Data Staging

Extraction

Transformation

Loading

(3) Data Storage.

(4) Information Delivery

(5) Metadata

(6) Management & Control

Data Acquisition:-

[This area covers an entire process of extracting data from the data sources moving all the extracted data to the staging area & preparing the data for loading into the DW] repository.

* [2 major components of this area are

i) Source Data :  Data Staging

The primary data store consists of the enterprise operational system ; flat files etc.

ii) Data Staging : [The data from data sources are loaded into the staging area]

[It is the place where all extracted data put together, & prepared for loading into the data warehouse.]

[There are three functions and services of data staging area:]

i) Data Extraction:- Select the data sources and determine the type of ^{Powers} filters applied to be applied to individual sources.

ii) Data Transformation:-

1.) It maps i/p data to data for data warehouse repository.

2.) It cleans data , ^{remove} duplicate entries

3.) It resolves the missing value

4.) It persists the data

iii) Data Loading:

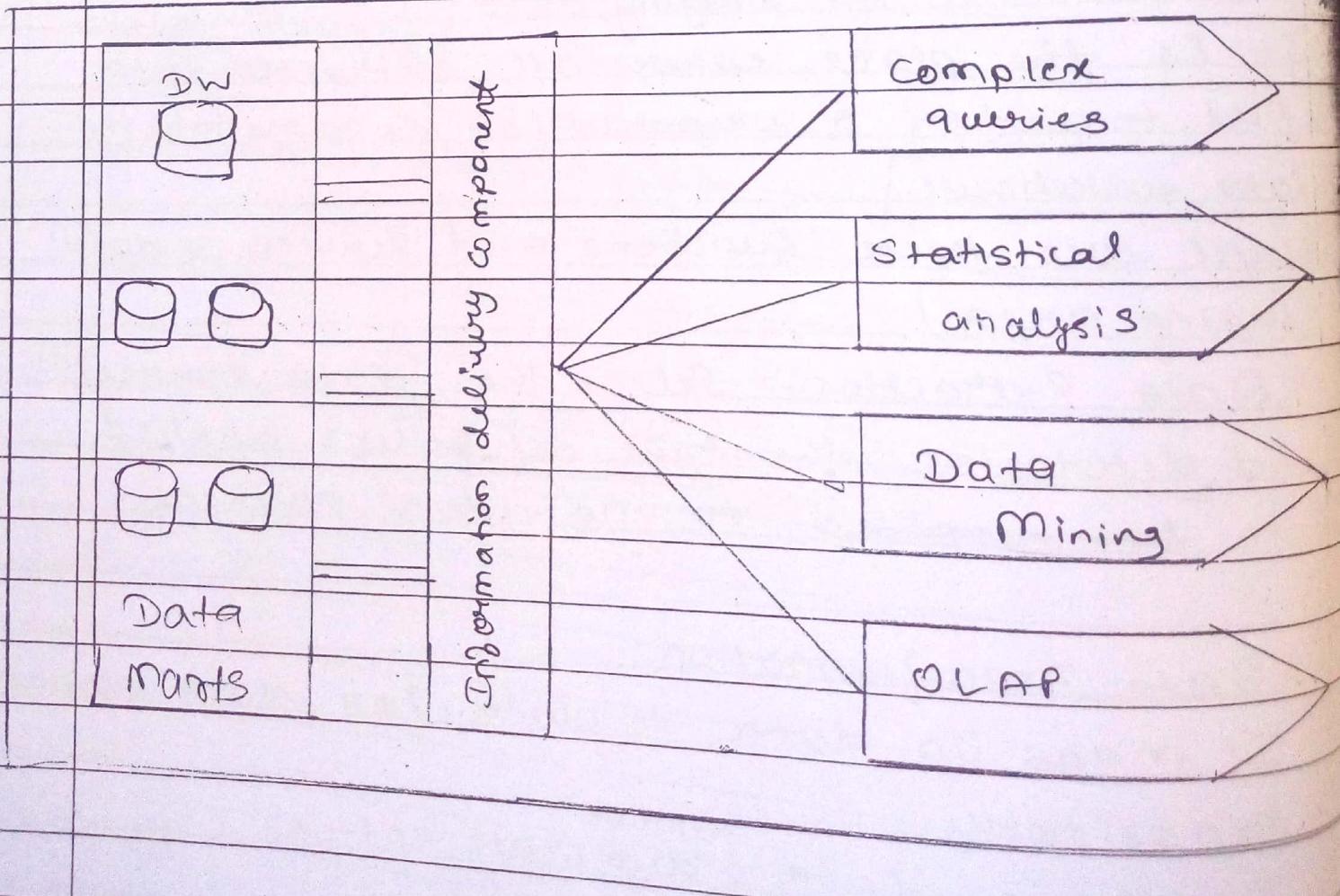
- Its function is to load the prepared data into the d.w. repository.

b) Data Storage:

- This area covers the storage process of loading the data from staging area into the DW repository.
- This area contains metadata IDWS, DataMart & multidimensional DB's.]

c) Information Delivery:

- This area of diff. methods of making information available to the end users]



Management & Control Module:

- It is an overall module, managing and controlling the entire DW environment.
- It is an umbrella component working at various levels, covering all the operations.
- That means it controls operations like data staging, data transformation, data loading and information delivery.

Metadata in the Data Warehouse:

- The metadata components serve as a dictionary of the contents of your data warehouse.
- It is the key architectural component of the data warehouse.
- It contains information like where DW resides., (location of DW), who are the end user, what are the transformation functions applied, and how many operational systems have been involved.

* Types of Metadata in DW:

i) Operational Metadata:-

- It contains all of the information about the operational data sources
- It stores information like data structures, data type and field length.

- ii) Extraction and transformation metadata:
- This contains data about the extraction of data from source system namely the extraction frequencies, extraction methods and business rules for the data extraction.
 - It also contains information about all the data transformation that takes place in the staging area. (Transformation information like cleaning data, remove duplicate entries, remove redundancies, remove missing values).

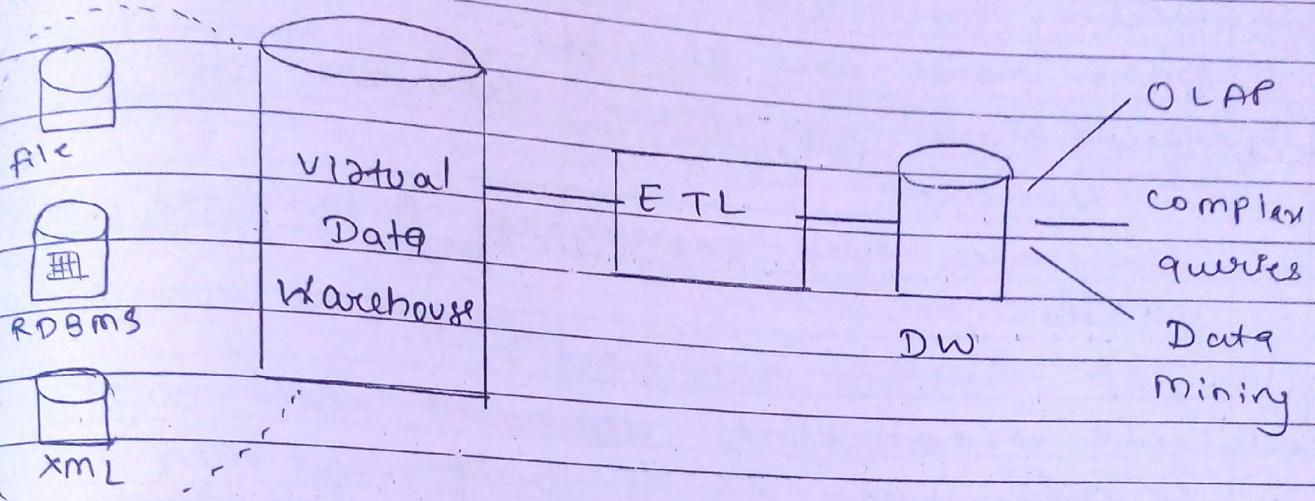
iii) End-user MetaData:

- It stores information like who are the end user of your DW. (Data Mining, OLAP, complex queries).
- It enables end user to find information easily from the data warehouse.

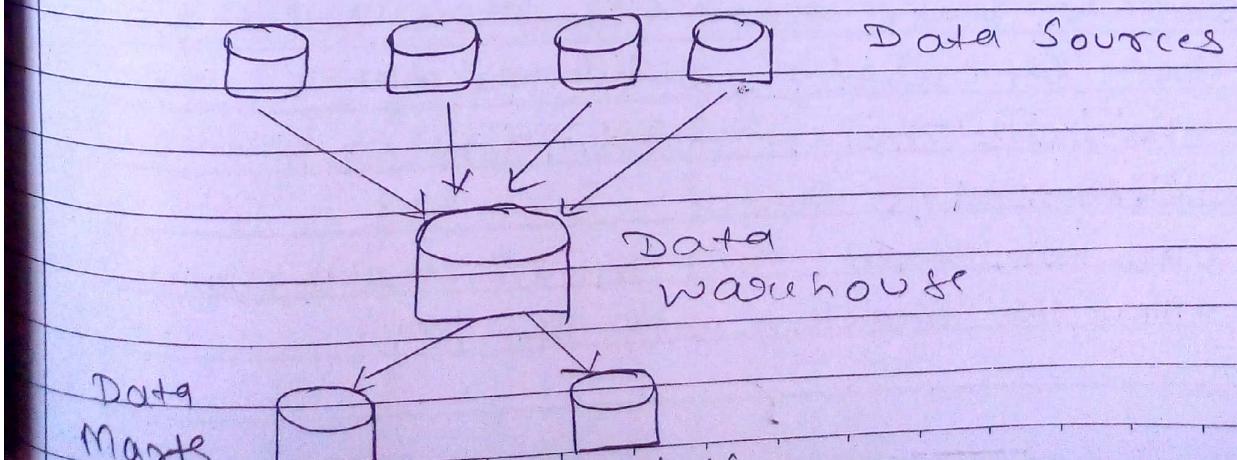
↑ Logical Architecture of DW. complete. ↑

* (1) Data Warehouse Model :-

Virtual DW (VDW)



- The view over an operational data warehouse is known as virtual data warehouse.
- It is easy to build virtual warehouse.
- A virtual warehouse is a container for components used to integrate data from multiple data sources on which E, T, L has been performed. In order to get persistent data to store in main DW.
- Building virtual W-H requires excess capacity on operational Data Mart (DM) database server

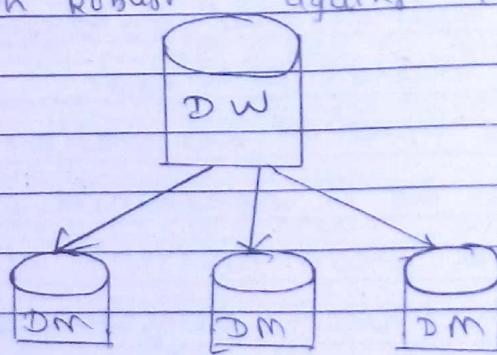


- Data marts customized by department
- The source of DM is data warehouse

- Data marts contains a subset of organisation wide data i.e. valuable to specific group of people in an organisation.
- For eg. Sales, products, customers.
- Data marts are small in size (subset of DW)
- Data marts are flexible
- Data marts can either dependent or independent
- There are two approaches to build the data mart.

(i) Top - Bottom approach:

- It represents dependent data mart
- In this, first dw is built and then various data marts are built from the DW
- This approach robust against changes.



(ii) Bottom - up Approach:

- It represents independent data mart
- In this, first data marts are built and then by merging, build the dw.
- This model contains consistent data marts & delivery delivered quickly
- DW has extended easily by just creating new DM & then integrating with other DM

Difference between DW & DM:-

DW	DM
i) DW is enterprise wide	i.) DM is departmental
ii) DW is union of all d.m.	ii.) DM is subset of DW (a single business process)
iii) D.W. receives data from data staging area	iii.) It receives data from fact & dimension table.

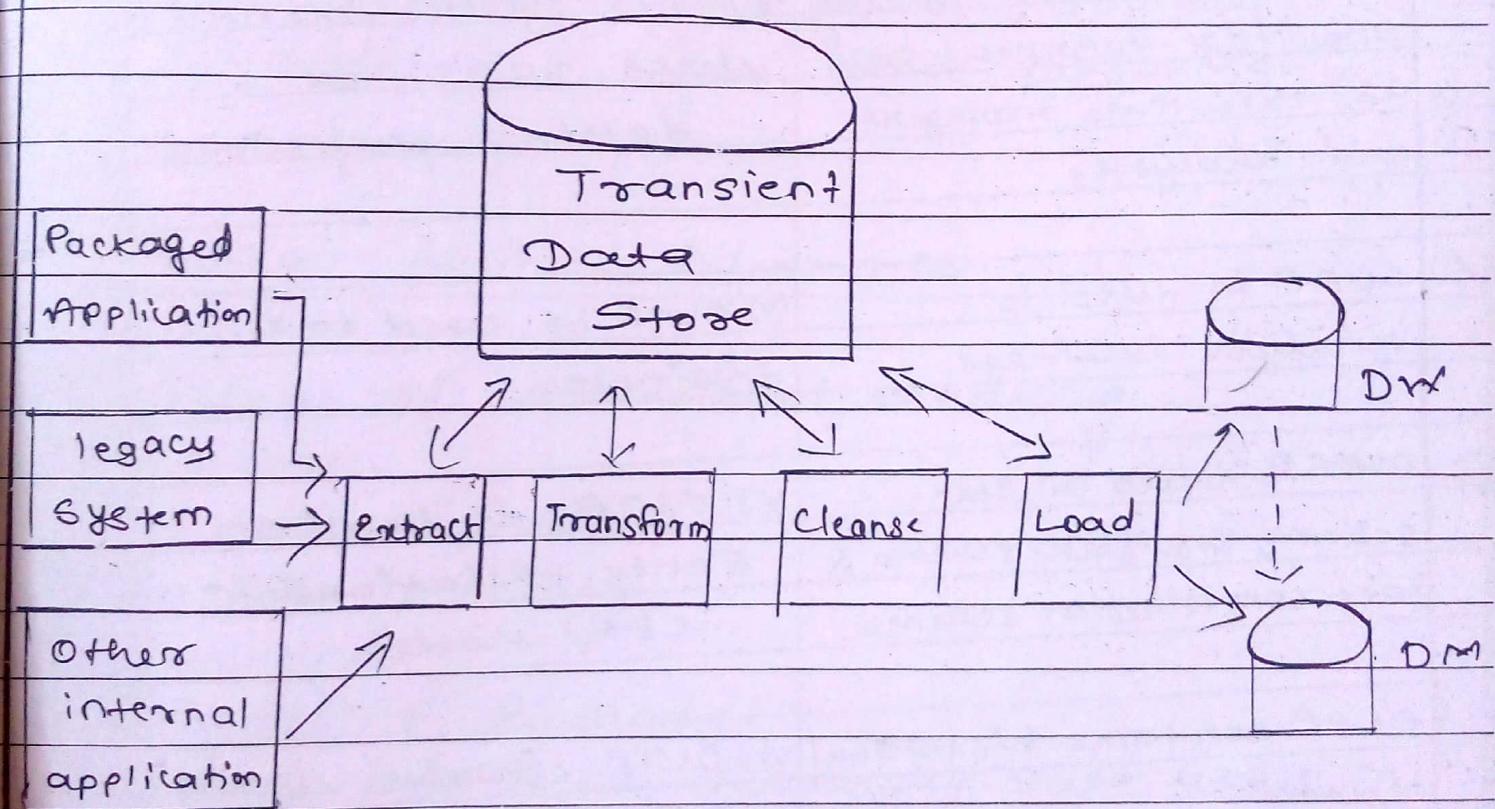
3) Operational Data Store (ODS):-

- It provides recent form of customer information file.
- It is database used as a source for staging data of DW.
- Unlike static content, the contents of ODS are updated throughout the course of business operation.
- It is used for short term decisions.
- It stores only very recent information.
- ODS consolidates data from multiple source data system and provides a real time integrated view of multiple current data.
- The ETL for ODS is similar for as DW.

Data Integration & ETL processes

- * **Data Integration:** It comprises 3 major processes to permit data to be accessed; and made available to BI & DW environment.
- 1: **Data access:** ability to access structured data from data source.
- 2: **Data Federation:** Integration of business views across multiple data store.
- 3: **Change capture:** Based on the identification, capture and delivery of the changes made to enterprise data store.
- * There are 4 technologies enable data metadata integration
 - i.) **Enterprise Application Integration (EAI):** It is used for transferring data from source system into dw.
→ It can be used to facilitate data acquisition directly into real time data warehouse.
 - ii.) **Service Oriented Architecture (SOA):** It is accomplished by using SOA, coarse-grained service (a collection of business processes/functions) that are well-defined & documented.
→ To implement SOA, we use web service.
 - iii.) **Enterprise Information Integration (EII):** -It is a tool that allow data integration from a variety of sources such as relational db's & multidimensional db, & web services. It is a mechanism for pulling data from source system to satisfy a request for information.
 - v.) **Extraction Transformation & Load (ETL):** at the heart of the technical side of dw process is ETL.
 - [The ETL process consists of a) Extraction
b.) Transformation ; c.) Load.]

Extraction: Reading data from 1 or more database
 Transformation: Converting the extracted data from its previous form into the form in which it needs to be, so that it can be placed into DW.
 Load:- Putting data into the DW.
 The purpose of ETL process is to load the warehouse, with integrated & cleansed data.



The ETL Process.