Data Warehouse

A data warehouse is a repository of information collected from multiple sources, stored under a unified schema, and that usually resides at a single site. Data warehouses are constructed via a process of data cleaning, data integration data transformation, data loading, and periodic data refreshing. The data are stored to provided information from a historical perspective (such as from the past 5-10 years) and are typically summarized. For example:→ Storing the details of each sales transaction, the data warehouse may store a summary of the transactions per item type for each store or summarized. to a higher level, for each sales regions.

A data warehouse is usually modeled by a multidimensional database structure, where each dimension corresponds to an attribute or a set of attributes in the schema, and each cell stores the values of some aggregate measure such as count or sales-amount.

* Usage

Health care, Banking, Retail, Data Mining.

* Need of Warehouse

— Large data store
— Smart store
— efficent retrive
→ Hetrogenetic

* Willion .H. Inmon

→ Subject oriented
→ Integrated
→ Non-volatile
→ time-variant

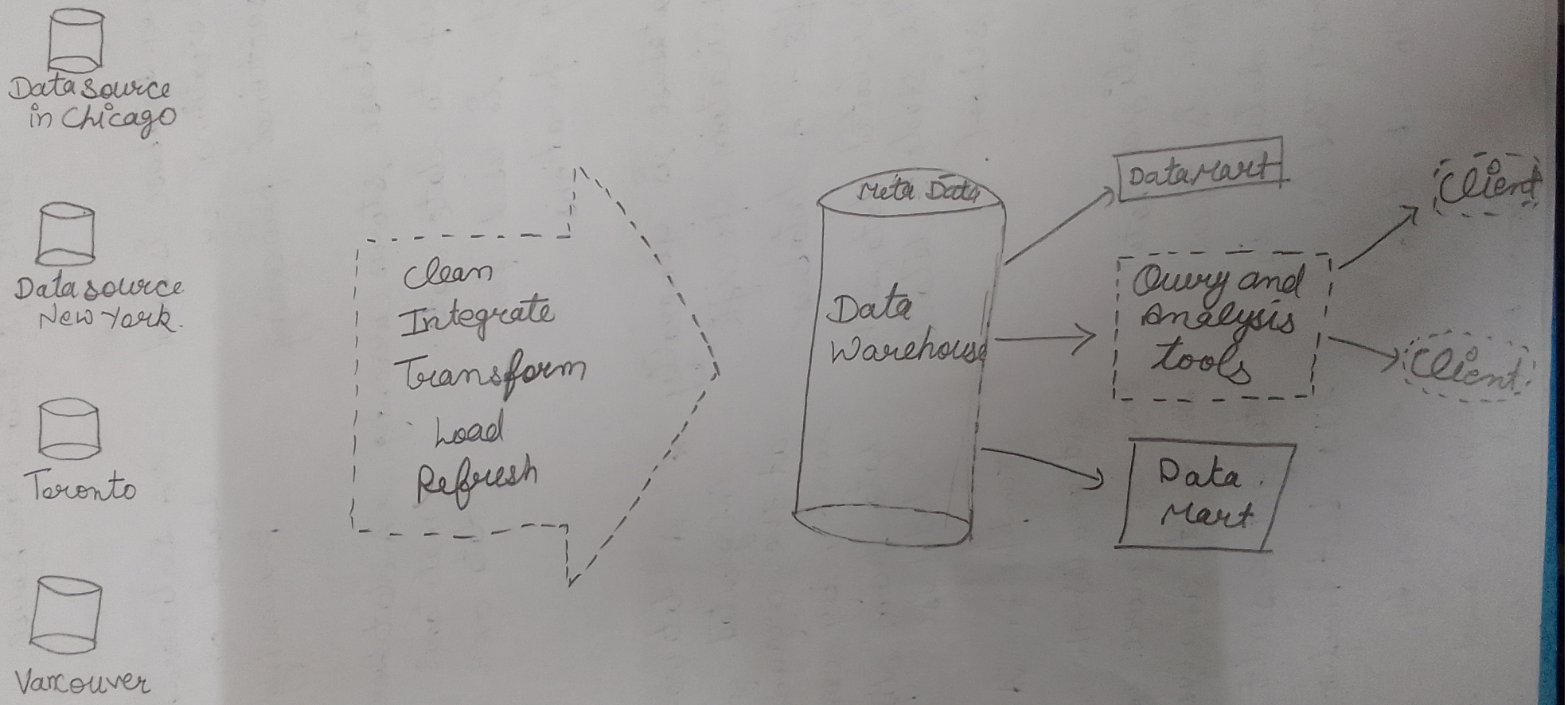
* Challenges face by design Warehouse

→ Data quality
→ Data Analytics
→ User exception
→ Cost
→ Performance
→ quality Accurance.

* Framework of a data Warehouse
* Advantage of data Warehouse

→ To clean data
→ Query processing multiple options
→ High query performance
→ Local processing at source warehouse.
→ Easy way of reporting access multiple system.

Fream work


Data source in Chicago
Data source New York
Toronto
Vancouver
Clean
Integrate
Transform
Load
Refresh
Meta Data
Data Warehouse
Data mart
Query and analysis tools
Data mart
Client
Client

DBMS V/S Data Warehouse

DBMS	Data Warehouse
① A common Database is based on operational or transactional processing.	* A Data warehouse is based on analytical processing.
* A database stores currently and up-to-date data. which is used for daily operations.	* A data warehouse maintains historical data over time.
* A database is generally application specific - Example - A database stores related data, such as the student details in a school.	* A data warehouse is integrated generally at the organization level, by combining data from different database. Example :- A data warehouse integrates the data from one or more databases, so that analysis be done to get result, such as the best performing school in a city
* Normalized data structure is there in a database in separate tables	* Dynamic and quick analysis of data is done
* 100 MB to GB data store.	* 100 GB to KB

Data Marts:- A data Mart contains a subset of corporate-wide data that is of value to a specific group of users. The scope is confined to specific selected subjects. For example:- A marketing data mart may confine its subjects to customer, item and sales. The data contained in data marts tend to be summarized.

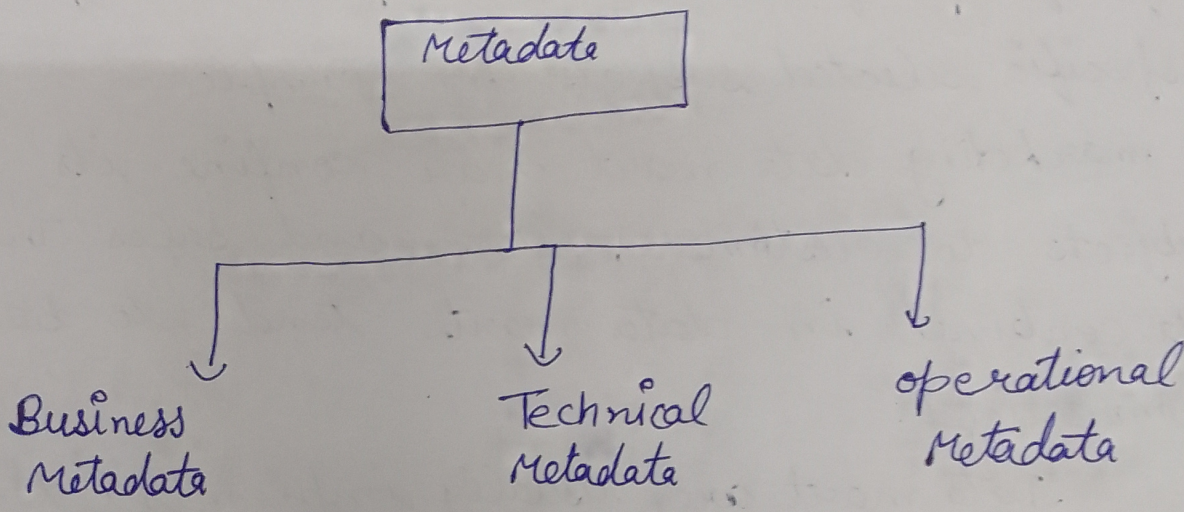
Data mart are usually implemented on low-cost departmental server that are Unix/Linux or Window based.

Depending on the source of data, data marts can be categorized as independent or dependent. Independent data marts are sourced from data captured from one or more operational system or external information providers, or from data generated locally within a particular department or geographic area.

Metadata:- Metadata is simply defined as data about data. The data that is used to represent other data is known as metadata. For example:- The index of a book serves as

a metadata for the contents in the book.

Categories of Meladata :→



① Business :→ It has the data ownership information, business definition, and changing policies.

② Technical Metadata :- It includes database system names, table and column names and size, data types and allowed values. Technical Metadata also includes structural information such as primary and foreign key attributes and indices.

③ operational Metadata :— It includes currency of data and data lineage. Currency of data means whether the data is active, archived, or purged.

Data warehouse and OLAP tool are based on multidimensional data model. In this model data is presented in the form of the data cube where as the data cube is define as represent of the data in multiple dimension It is define in terms of fact & Dimensions

Multidimensional mode → Data cube → facts & Dimension

↑

DWH

Data cube :-

When data is grouped or combined in multidimensional matricess called Data cubes. The data cube method has a few alternatives names or a few variants, such as multidimensional databases, materialized views, and OLAP

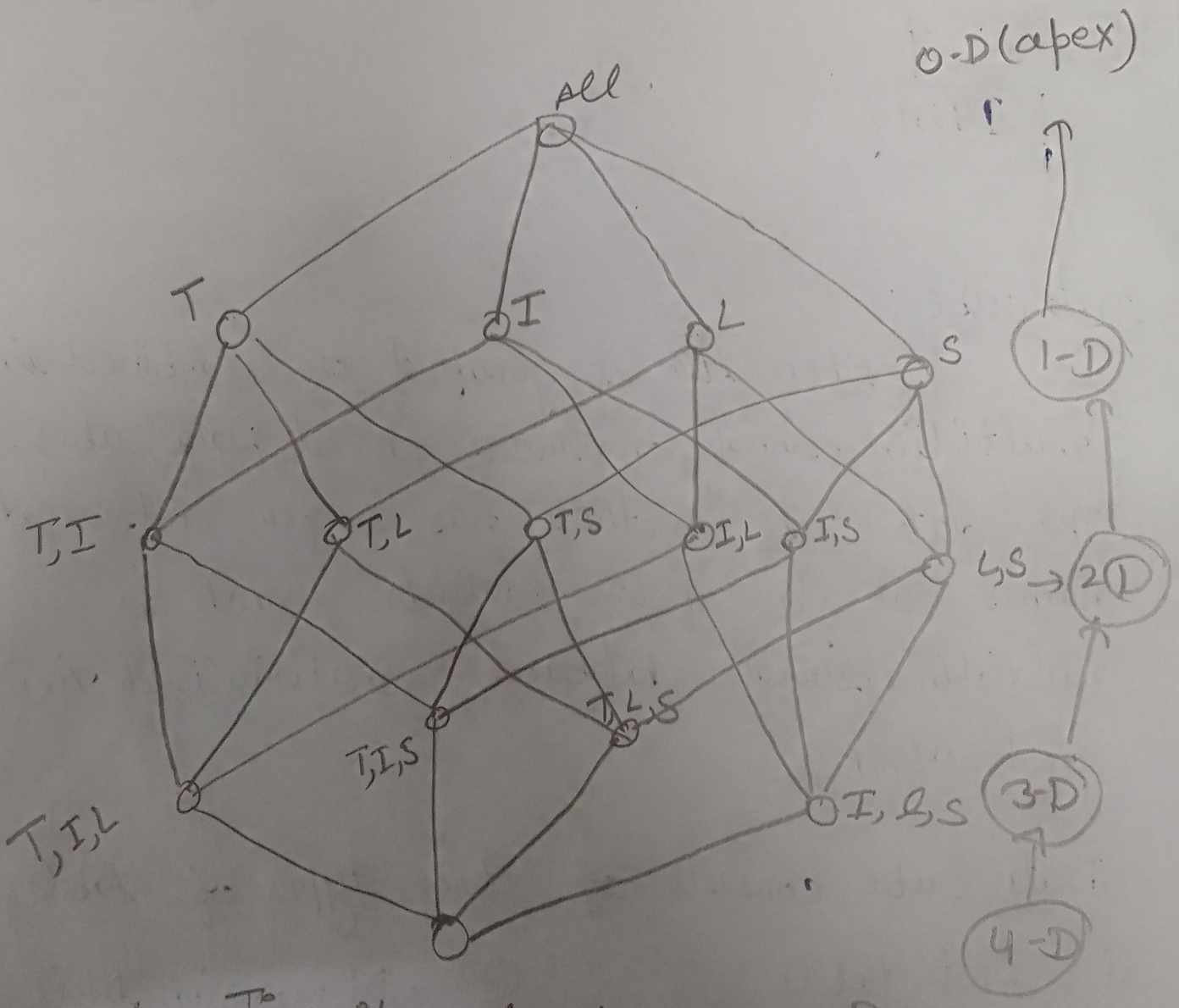
Data cube consists of two types of table

① fact table ② Dimension table

~~Data cube operations~~ :-

Based on different number of Dimensional or different subsets of data can be created from data cube. Such subset are known as data cubides. The arrangment of all possible cubides in the worker is known as data lattice.

Lattice consists of different level of summarization starting from based cuboids to apex cubid



Time, item, location, supplier