Assignment 3 (DMW)

Shashwat Shah
60004220126
C-22 Div B

1 Types of Meta Data

(1) Operational metadata - As you know, data for the data warehouse comes from several operational systems of the enterprise. These source systems contain different data structures. The data elements selected for the data warehouse have various fields lengths and records, combine parts of records from different source files and deal with multiple coding schemes and field lengths. When you deliver information to end users you might be able to tie that back to the original source data set.

Eg Operational metadata in retail data warehouse efficiently manages sales data, optimizes queries with indexes and monthly partitioning.

(2) Extraction and Transformation MetaData: This contains data about the extraction of data from the source systems, namely time extraction frequencies, methods of extraction and business rules for the data extraction transformations that take place in the data staging area.

Eg. In retail data warehouse it could document process of extracting daily sales record from various online and in store sources.

(3) End user metadata. - It is the navigational map of data warehouse. It enables the end users to find information from the data warehouse. The end user metadata allows the end users to use their own business terminology and look for information.

2)	OLTP	OLAP
1)	It is online transaction processing	It is online analytical processing.
2)	It is well known as online database modifying system	It is well known as online database query system
3)	Consists of current data	Consists of only operational or historic data
4)	It makes use of DBMS	It makes use Data warehouse
5)	It is application Oriented used for business tasks	It is subject oriented, used for data mining, analytics, decision making
6)	In OLTP, database tables are normalised.	The tables are not normalised
7)	Size of data is small.	A large amount of data is stored

3) Type 2 - Preservation of history.

In handling changes like kristin samuelsons marital status and address, which are type 2 changes in the data warehouse its crucial to preserve historical data. For marital status orders before October 1, 2000 should be categorized under single and those on after should be labelled married, similarly for a change in address from newyork to california on November 1, 2000, orders before that date are associated with 'CA'. Type 2 changes reflect true changes in source systems, and partitions the data warehouse history, ensuring every change for same attribute is retained.

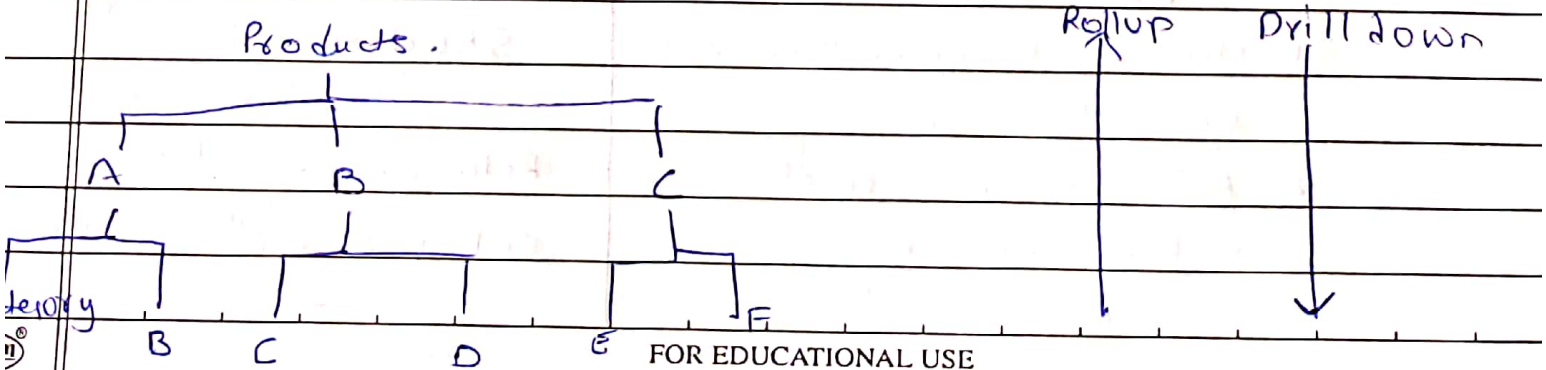
Type 3 - Tentative soft revisions

These changes common than type 1 or type 2 changes. serve a specific purpose. Unlike type 2 changes that cut off history at a certain date. Type 3 changes are more flexible. For instance, in the scenario of reassigning a a salesperson like robert smith from new England to Chicago territories, a type 3 change allows tracking orders under both alignment temporarily. This kind of change is temporily soft, often used for scenarios like company performance before and after a proposed transition. Key principles for type 3 changes include their association with soft or tentative alterations in source systems.

4 OLAP operations.

1) Drill down - It involves moving from higher level summary of data to more detailed level.
2) Rollup - It is opposite of drill down. It involves moving from a detailed level of data to higher more summerised level.

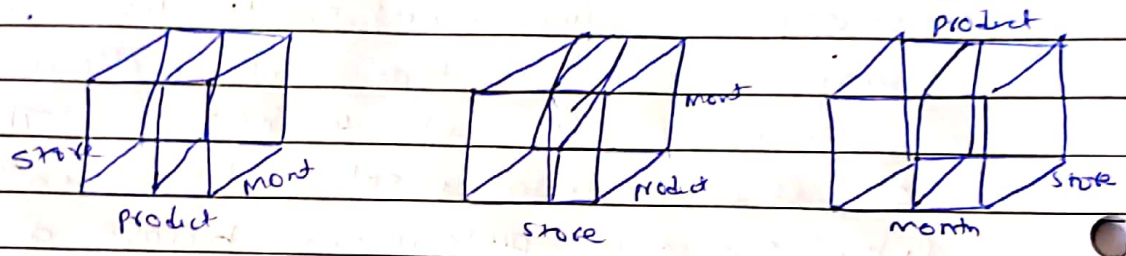
Eg : Dimension - Time (Year, quaterly, months)
- Product (category, subcategory, product)
- location (Region, State)



3) Slice : It involves selecting a single level from one dimension and viewing a subset of data at that level.

4) Dice : It involves selecting a subset values from two or more dimensions and viewing the data at the intersection of those values.

Eg.



5)

Data Warehouse	Data Marts
1) Corporate / Enterprise wide	Departmental.
2) Union of all data marts.	A single business process.
3) Data received from staging are	Star join (facts & dimensions)
4) Queries on presentation resource	Technology optimal for data access and analysis.
5) Structure for corporate view of data.	5) Structure to suit the dept view of data.
6) Organised on ER model	

Data Warehouse	Data Lake
1) structured, semi-structured, unstructured.	Structured.
2) Relational, non-relational	Relational.
3) schema on read	schema on write

4) Raw and Unfiltered format	Processed and vetted format
5) Easy to scale at a low cost	Difficult and expensive to scale
6) Data scientist & data engineers use this	Data warehouse professionals and business analysts use this.

Top down	Bottom up
1) Focuses on breaking down the problem into smaller parts	focus to solve smaller problems and complete solution.
2) In this communication is less among modules.	In this module must have communication
3) It is used in debugging module documentation	It is basically used in testing
4) Here decomposition takes place.	Here composition take place.
5) top function is hard to identify.	sometimes we cannot build a program we have started