CH-2 Designing And Implementing of data ware housing

Logical Design For DW

- When the client will define business requirements and functionality of business or need to translate it's requirement's into data ware house project then you have to create logical & physical design for DWH.
- First we have to talk about the logical design that will be create by using a pen& paper or by using the designing tools.
- The logical design is more conceptual and abstract and then the physical design.
- In the process of logical design include arranging into a series of logical relationships called entity's and attributes.
- In the logical design you look at the logical relationships among the objects.

The logical design include the following steps:

1) Define a model for your data warehouse.
2) Select entities, attributes & relationships in a model.
3) Select different types of information that you be needed in future.
4) Check the logical relationships of different objects.
5) Arrange the data into a series.
6) Set the attributes which are used to describe the entities.
7) Set the unique identifier (UID) to differentiate between an entity with another entity.

Logical design include different entity, attributes, relationship and unique identifier which are as under :

- An entity represents chunks of data warehousing structure information. it maps to the table.
- An attribute is a component of an entity.
- The relationship maintains the integrity and relationship among the objects. It represent the relational database and constraints .
- A unique identifier is used to define each records uniquely.
- It creates the difference between objects.

Explain Physical Design

Logical Design	Physical Design
Entity	Tables
Relationship	Primary Key , Foreign Key & Unique Key
Attributes	Columns & Data

- During the physical design process we convert the logical design into physical design by using database structures.

Sandip Upadhyay

- At the time of the conversion , you have to map:
  - Entity to tables.
  - Relationship with foreign key constraints.
  - Unique data with primary key constraints.
  - Attributes to columns.
  - Unique identifiers to unique key constraints.
- in the physical design you look at the most effective way of storing and retrieving the objects.

Physical Design Structure :

- Once you have converted your logical design to the physical design then you must create some or all of the following structures:

1) Table Spaces :

- It is the collection of data files & each tables belongs to a table space.

2) Tables and Partitioned Tables :

- the partitioned tables allows you to support very large volumes by partitioning them into smaller and manageable pieces.
- Single table or tables have the key problem for supporting large data.

3) Integrity Constraints :

- It is used to implement the constraints like NOT NULL & Relationships.

4) Views :

- It is the collection of one or more tables or views.

5) Dimensions :

- A dimension is a structure of objects that defines hierarchy key relationships between columns & set of data.
- By building the data warehouse in this way , we can make a design that allows for growth and changes as per the needs of users.

Give the difference Between Physical & Logical Design

Logical	Physical
It include entities , relationships and attributes.	It include tables , columns , keys , data types , validation rules , triggers , procedures and constraints.
It used the business names.	The names may be limited by the DBMS.
It include unique identifiers.	It include primary key & foreign key.
It is normalized to at least third normal form.	It may be de normalized for performed requirements.
It does not include any	It may be include any

Sandip Upadhyay

repeated data.	repeated data elements.
It does not include any derived data.	It may include the results of complex OS difficult codes to reused.
The business experts drive the modal.	The designs drive the model.
It capture's and record's information necessary for the business.	It include the technology and data elements for the DBMS.

What is Dimension Table?

- Dimensional table is a table which is in a stores schema (diagram) of data ware house.
- A dimension table stores attribute or dimension's that describe the objects in fact table.
- In data warehousing dimension is a collection of reference information about measurable events. This event's are known as fact & the facts are stored in fact table.
- Dimensions describe the DWH fact's as per categorized data. That support meaningful answer to business questions.
- A data warehouse organized a descriptive attributes as columns in dimension table for ex. Customer attribute can be include the dimension like customerid , fname , lname , gender , address etc.…
- A dimension table have primary key column that uniquely identifies each dimension record (row).
- The dimension table Is associated with a fact table using this key.
- The data in the fact table can be filtered and grouped by different combination at attributes.
- The dimension table are referenced by fact table's using the key's. when create the dimension table in a data warehouse. A system generated key is used to uniquely identify a row in the dimension

  For ex : login is a fact at customer in some website then the query about the visiting date's at that person in the website is a dimension.
- Like fact table the dimension table are also de normalized because this structure are not built to manage the transaction but it also enable the users to analyze data as easily as possible.

Key Points :-

- Dimension tables constraints dimensions of a fact.

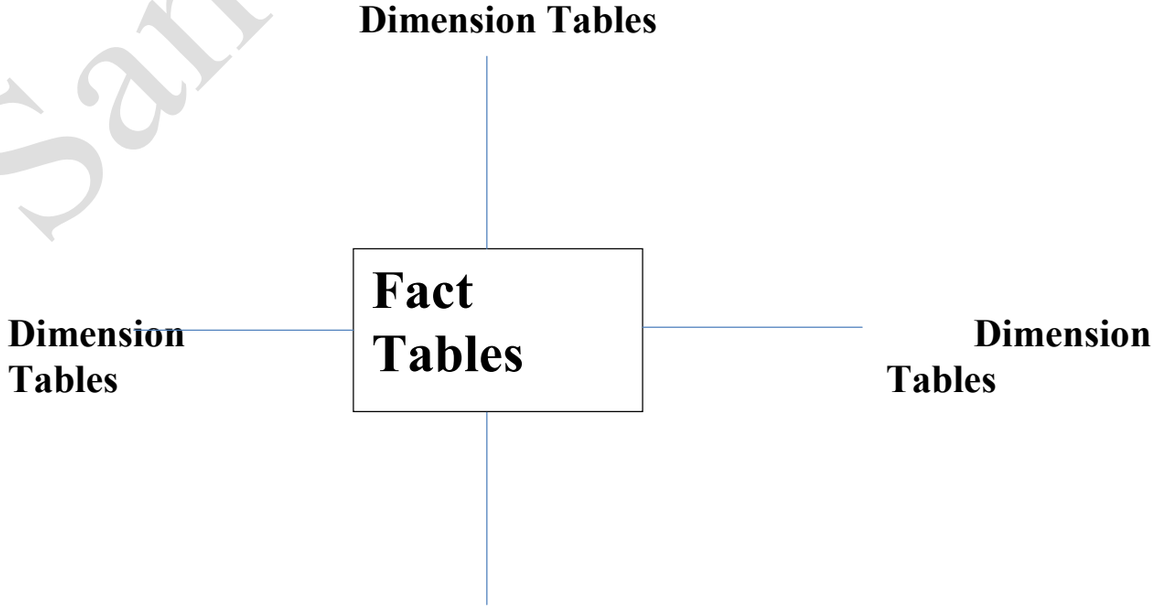
Sandip Upadhyay

- They are joined to fact table via foreign key.
- Dimension tables are de normalized tables.
- The dimension attributes are the different columns in a dimension table.
- Dimensions offers descriptive characteristics of the fact.
- The dimension can also contains one or more hierarchy.

Types :-

- **ConfirmedDimensions :**
- It is a special fact to which it is related.
- **OuTrigger Dimensions :-**
- When the dimensions have reference to another dimension table then the secondary dimension called out trigger dimension.
- **Dimension to Dimension Table :-**
- The dimensions may have reference to other dimension like relationship.
- **Role Playing Dimension :-**
- A single dimension helps to reference multiple times in a fact table.
- **Junk Dimensions :-**
- It is the collection of random transactional codes or text attributes.
- **Degenerate Dimension :-**
- Thisdimension does not have its dimension because it is derived from the fact table.
- **Step Dimensions :-**
- It is a sequential process like web page event.
- It has a separate row in a fact table for every row and process.

Diagram :-



Sandip Upadhyay

Dimension Tables

Star schema Example with Fact & Dimension Tables

Give the difference between Dimension Table and Fact Table.

Parameters	Fact Table	Dimension Table
Definition	It is measurement , matrix as fact about the business process.	It is related tables to the fact table it contains descriptive attributes.
Design	It design at atomic level.	It should be descriptive complete and quality assured.
Characteristics	It located at the center of a star schema and surrounded by dimension.	It Is connected to the fact table and located at the Edges at the star schema.
Task	Fact table is used to provide a measurable event where the dimension table data is collected and used for analysis and reporting.	It is a collection of reference , information about business.
Types of data	Fact tables can contain information like , salse , against a set at dimension like product and date.	Every dimension tables contains the attribute which describe the details at the dimension **For ex:** product dimension can contain productid , productcategory , price etc.
Key	The primary key in fact table is mapped as foreign key's to dimensions.	Foreign key's into the fact tables.
Storage	It help to store reports labels and filtered domains value in dimension table.	It load the detail data into dimension structure.
Hierarchy	Fact table does not contain hierarchy.	It contains hierarchies **For ex :** location can contain country that contain state , city ,

Sandip Upadhyay

		pincode etc.

Explain Fact Table.

- A fact table is a primary table in dimensional model.
- It is the central table in star schema at data warehouse.
- A fact table stores quantitive information for analysis.
- It include or hold the data which should be analyzed and the dimension table stored data about the way's in which the data in the fact table can be analyzed.
- The fact table is generally in the de normalized format.
- A fact table contains two types of columns.
  1) Measurements or facts
  2) Foreign key to dimension table
- In the first measurement column include the data that is fact and analyzed.
- In the foreign column it allows to join with dimension tables.
- **For ex.** One company sale the product to that customers then every sale is fact that happens and the fact table is used to record this facts.

Product id	customerId	Unit sold	Price	Date	Etc...
007	02	5	100	5/1/2018	
002	03	6	50	5/12/2018	
003	05	10	100	6/12/208	

Sales fact table

- Now we can add a dimension table about the customers.

Customer Id	Name	Gender	City	Mobile No.
01	A	Male	Rajkot	1236547890
02	B	Female	Surat	3214569870
03	C	Male	Junagadh	6321458972
04	D	Female	Jamnagar	9632587410
05	E	Male	Baroda	7896541230

- in this example the customer id column in the fact table is the foreign key that joins with the dimension table's primary key column.
- When building the fact table there are some physical and data limits. Here the ultimate size at the object as well as the access path should be considered.
- The table's should be built based on current and features requirements.

Diagram's :-

- All are in dimension table draw here……

Sandip Upadhyay

Types Of Fact :-

1) Additive :-

The measure's should be added to all dimension.

2) Semi Additive :-

Measure's may be added to some dimension.

3) Non additive :-

Measure's are not added in the dimension.

It stores only some basic unit of measurements of business process.

Design The Physical Data Structure For Data Warehouse :-

- ♣ To design any data or database data modeling is require.
- ♣ Data modeling is the process of creating a data model for the data to be stored in a database.
- ♣ The data warehouse stores many database so it must be stored in some structure.
- ♣ There are three different types of data model's used to store the data which are as under:

  1) Conceptual 2) Logical 3) Physical

1) Conceptual :-

- ♣ This data model define what the system contain. The purpose is to organize , scope and define the business concepts and rules.

2) Logical :-

- ♣ It defines how the system should be implemented the purpose to develop technical map of rules and data structures.

3) Physical :-

- ♣ It defines how the system will be implemented using specific DBMS. This model is created by DBA(Database Administrator) and developer.
- ♣ The purpose is actual implementation of the database.

Diagram Of Data Modelling :-

Feature	Conceptual	Logical	Physical
Entity names	✓	✓	-
Entity relationships	✓	✓	-
Attributes	-	✓	-
Primary key	-	✓	✓
Foreign key	-	✓	✓
Table names	-	-	✓
Column name	-	-	✓
Column data types	-	-	✓

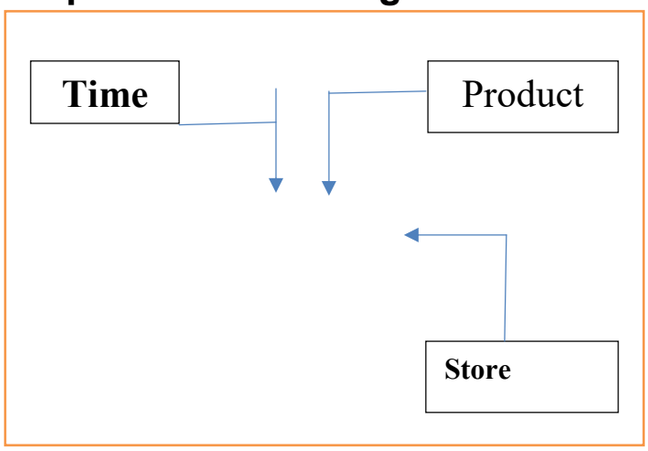
Example :-

Somdip Upadhyay

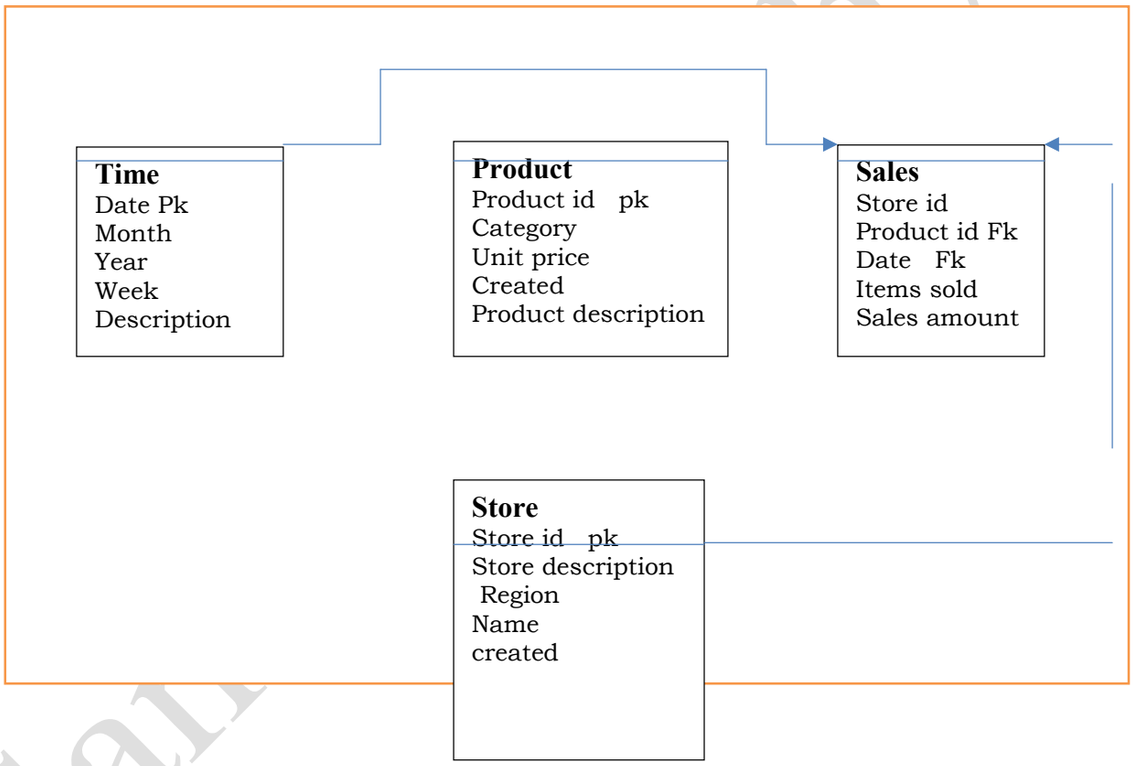
1)

2)

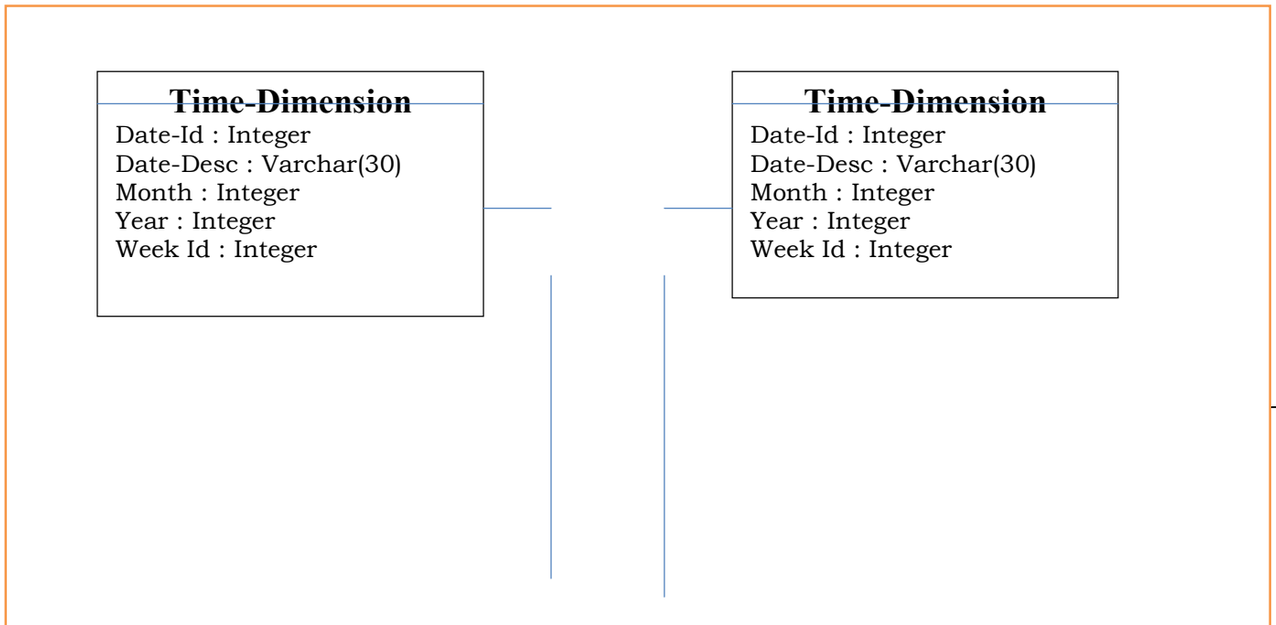
3) **Conceptual model Design**

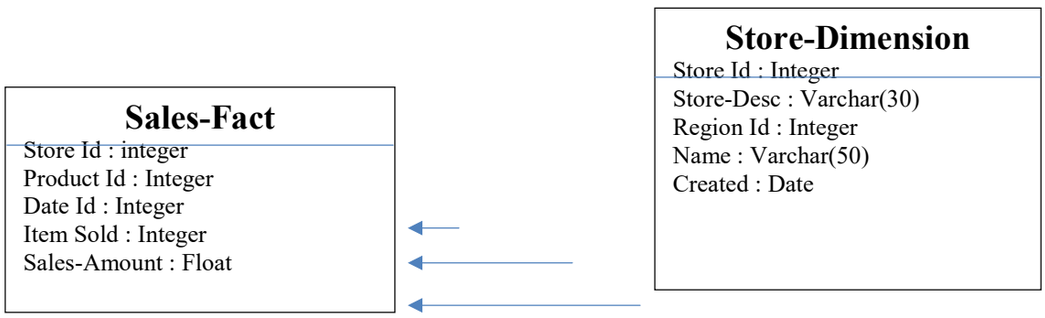


Logical Model Design :-



Physical model Design :-





- From the above diagram we can see that complexity increased from conceptual to logical and from logical to physical.
- When we start with conceptual data model then we understand above the entities in over data and how they are related to each other.
- During the logical data model we understand the details of over data. Without worry about actual implementation.
- In the physical data model we know the expert implementation of over data model in the database.
- Data warehousing project generally include the physical data model.
- The data warehouse physical data model describe how the model will be built in the database. It shows all table structure including column name , column data type , primary key , foreign key and relationships between tables.

Feature Of DWH Physical Data Model :-

1) It specify all tables and column that are available in the model.
2) You can use the foreign key's to identify the relationships between tables.
3) You can de normalized the tables based on user requirements.
4) You can speed up the data retrieval from the table.
5) The DWH physical model describes how the tables are actually created in the data warehouse environment.

Sandip Upadhyay