

Unit-2

Designing and Implementation of Data Warehousing

❖ What is Data Warehouse Design?

- Data warehouse for business intelligence allows your organization to query data obtained from trusted & use the answer to gain a competitive in your business.
- Data warehouse design is the process of building a solution to integrate from data multiple sources & that support analytical reporting & data analysis
- Data warehouse is a relational database that is designed for query & analysis greater than transaction processing.
- Data model methodology & system architecture provides a design, basic form, data warehouse that emphasizes core data quality, quantity & ability to support enterprise while data provision needs.
- Data warehouse design for a custom software development consulting, satisfying, training companies.
- There are two types of designing in data warehouse.

1) Logical Design

2) Physical Design

- A data warehouse is a single data repository where a record from multiple data sources is integrated for online business analytical processing (OLAP).
- Data warehouse design takes a method different from view materialization in the industries.
- The target of the design becomes how the record from multiple data sources should be extracted, transformed, and loaded (ETL) to be organized in a database as the data warehouse.

- Data warehouse design is the process of building a solution to integrate data from multiple sources that support analytical reporting and data analysis. A poorly designed data warehouse can result in acquiring and using inaccurate source data that negatively affect the productivity and growth of your organization.
- Data warehouse design is a time consuming and challenging endeavor.
- Data warehouse is process of constructing & using a data warehousing.

❖ Logical Design for Data Warehouse

- The process of logical designing invoice arranging data into series of logical relationship called entity & attributes.
- Logical design for logical relationship between objects.
- Entity Relationship (ER) modeling technique can be used for logical design of data warehouse.
- Logical design for a data warehouse part of implementation a data warehouse with Microsoft SQL Server.
- Logical design often starts with a conceptual schema & then generates relation structure.
- Entity relationship involves identify the entity, attribute & relationship.
- An entity is chunk information which map to a table in database.
- An attribute is a part of an entity that map to a column in a database.
- A unique identifier can be used to make sure the data is consistent.
- A logical design is conceptual & abstract you don't deal with the physical implementation yet.
- You deal only with defining the types of information that you need.
- One technique you can use to model your organization logical information requirement & entity relationship modeling.
- During process business requirement analysis, business analyst talks to the user & examine the user model that exists in the user with the co-operation of the user this is formalized into an entity relationship model which essentially forms the logical model.

- It is important to realize that this logical model is based entirely upon the user requirement.
- Once the logical model is complete it is over to DBD (Database Design).
- At this point a decision is made about the database model & the database engine that will be used.
- A logical design is conceptual and abstract. You do not deal with the physical implementation details yet. You deal only with defining the types of information that you need.
- Schema design elements such as tables and views are considered a database's **logical database model**. These objects provide information about available data elements.

Logical Data Warehouse Use Cases

- Almost any business or industry could benefit from connecting all of its data and allowing access across the organization for better analysis and decision making. The following are just a few of the use cases where LDW can be applied:
 - Risk management
 - Monitoring KPIs
 - IoT edge analytics
 - Predictive Analysis
 - Self-service analytics

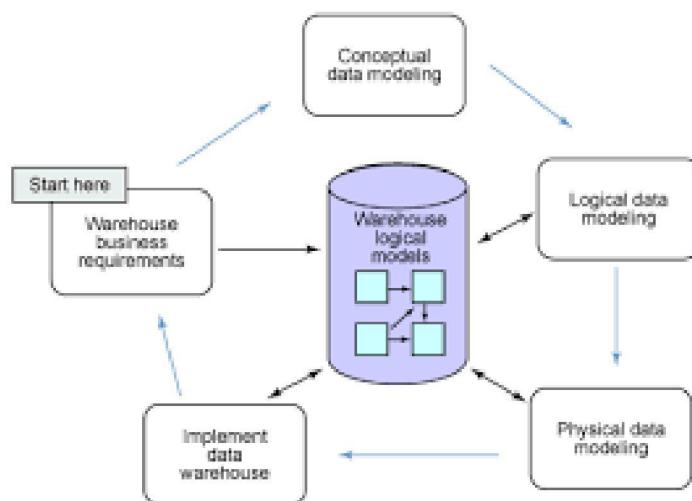
Features of a logical data model

- It involves all entities and relationships among them.
- All attributes for each entity are specified.

- The primary key for each entity is stated.
- Referential Integrity is specified (FK Relation).

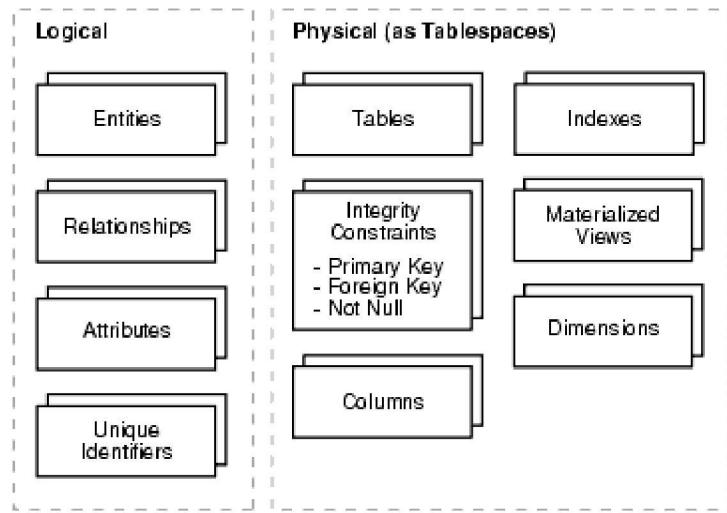
The phase for designing the logical data model which are as follows:

- Specify primary keys for all entities.
- List the relationships between different entities.
- List all attributes for each entity.
- Normalization.
- No data types are listed

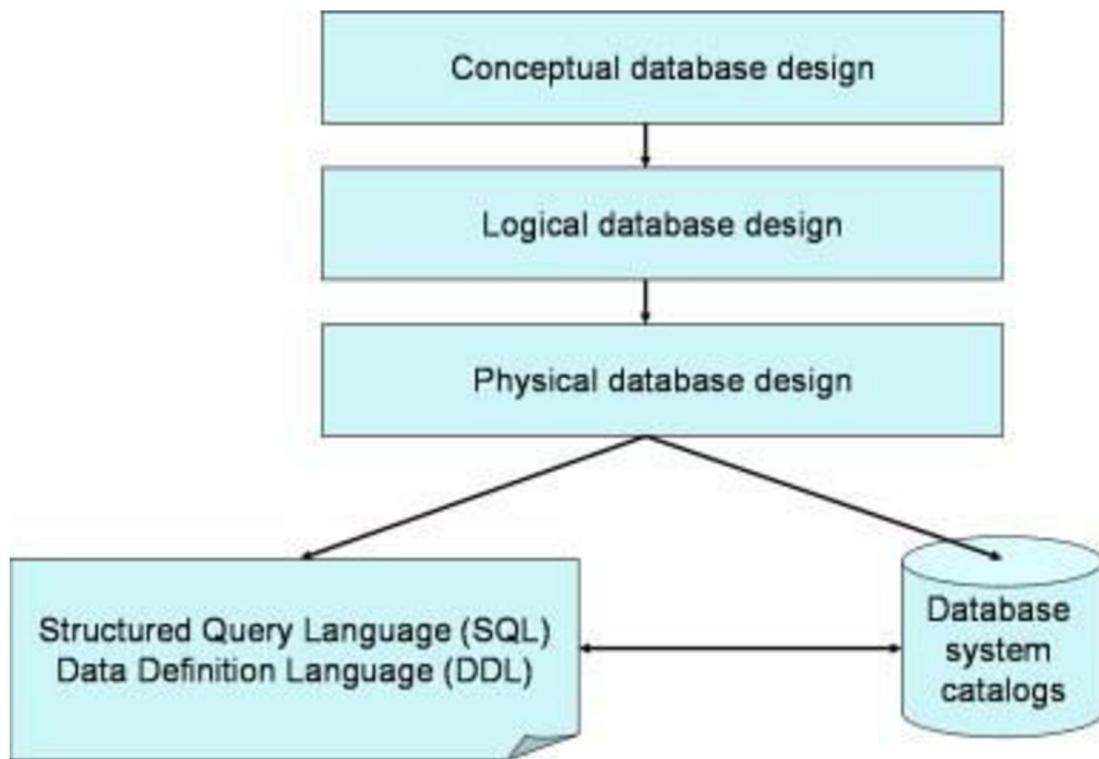


Physical Design for Data Warehouse

- During the logical design phase, you defined a model for your data warehouse consisting of entities, attributes, and relationships. The entities are linked together using relationships.
- Physical design deals with the effective way of storing and retrieving the data. In the physical design, the logical design needs to be converted into a description of the physical database structures.
- Physical design involves creation of the database objects like tables, columns, indexes, primary keys, foreign keys, views, sequences etc.



- During the physical design process the expected schema into database structure at this time you have to map like:
 - Entities to tables
 - Relationships to foreign key constraints
 - Attributes to columns
 - Primary unique identifiers to primary key constraints
 - Unique identifiers to unique key constraints
- Physical decision that may overall development time of a data warehouse such as tables, vertical & horizontal partition of a table, ETL process (Extract Transfer Load) & so on.
- Physical design is a important & highly influences the overall performance of the data warehouse & maintenance.
- The physical design procedure of an information warehouse has to be significant & comparative to its physical storage & public presentation
- Physical design of warehouse by using the components diagram & deployment diagrams of UML (Unified Modeling Language).
- The physical design deal with the affective way of storing & retrieving the data.
- Physical design decision is mainly driven by query performance & database maintenance aspects.

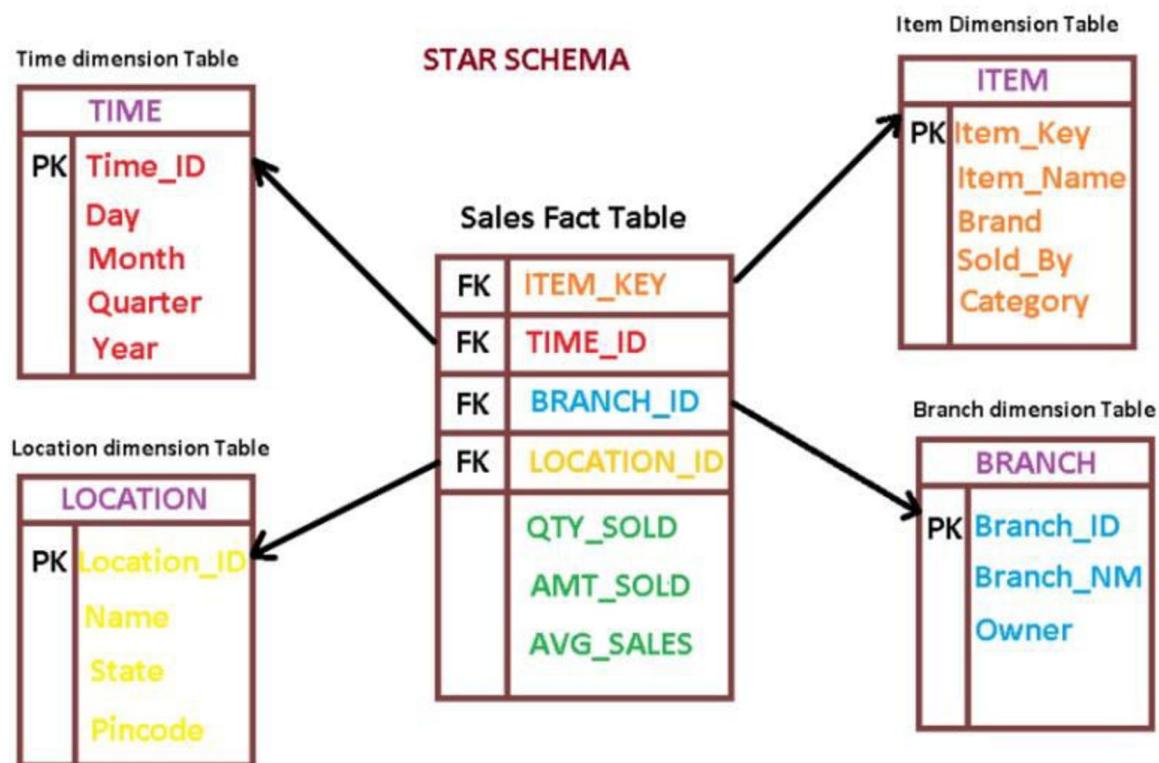


✚ Design Dimension Table/Fact Table For Data Warehouse

➤ What is Dimension Table?

- A dimension table is a table in a star schema of data warehouse.
- Data warehouse built using dimension table or data model which consists of fact & dimension tables, dimension tables are used to describe they contain dimension key, values & attributes.
- The time dimension would contain every hours, day, week, months, quarter, year that has occurred since you started your business operations.
- Dimensional table typical small from a few to more thousand rows (records), the occasionally dimension can grow.
- Dimension table can be utilized access for the reports & it's about reusability,
- **Example:** An E-Commerce Company can create a dimensional table with various columns depending on different subject that would like gain information from like name of person, address, date of order, shipping information, etc.
- A dimensional table has a primary key column that uniquely identifies each dimension record or row.

- The dimension table is associated with fact table using this key.
- The data warehouse organizes descriptive attributes in columns dimensional record.
- Dimension table are referred by fact table using key when extracting a dimension table in a data warehouse a system generated key is used to uniquely identify a row in the dimension.
- Typically any dimension table has a primary that link all the dimension records to the individual primary records.
- Each dimension is defined by a **single primary key**, which serves as the basis for referential integrity with any given fact table to which it is joined
- Dimension attributes serve as the primary source of the query constraints, groupings, and report labels.
- Attributes are identified as the *by* words. For example, when a user wants to see dollar sales by brand, brand must be available as a dimension attribute.



➤ What is Fact Table?

- In data warehouse a fact table consists of the measurement matrixes as fact table of a business process.
- A fact table is a primary table in a dimensional model.
- A fact table typically has two types of columns those that contain a fact & that are a foreign key to dimension table.
- Fact table provides the usual values that act as independent variable (by which dimensional attributes are analyzed).
- A fact table holds the data to be analyzed & a dimensional data stores the data about a way in which the data in the fact table can be analyzed.
- The foreign key column allow join with dimensional table & the major columns contain the data that has been analyzed.
- The fact table consists of fact of a particular business process like sales revenue by month of product.
- Facts are also known as measurement or matrix.

Measure types

Fact table can store different types of measures such as additive, non-additive, semi-additive.

- **Additive** – As its name implied, additive measures are measures which can be added to all dimensions.
- **Non-additive** – different from additive measures, non-additive measures are measures that cannot be added to all dimensions.
- **Semi-additive** – semi-additive measures are the measure that can be added to only some dimensions and not across other.

Designing fact table steps

Here is overview of four steps to designing a fact table described by Kimball:

- **Choosing business process to model** – The first step is to decide what business process to model by gathering and understanding business needs and available data
- **Declare the grain** – by declaring a grain means describing exactly what a fact table record represents
- **Choose the dimensions** – once grain of fact table is stated clearly, it is time to determine dimensions for the fact table.
- **Identify facts** – identify carefully which facts will appear in the fact table.

❖ Difference between Dimension table vs. Fact table

Parameters	Fact Table	Dimension Table
Definition	Measurements, metrics or facts about a business process.	Companion table to the fact table contains descriptive attributes to be used as query constraining.
Characteristic	Located at the center of a star or snowflake schema and surrounded by dimensions.	Connected to the fact table and located at the edges of the star or snowflake schema
Design	Defined by their grain or its most atomic level.	Should be wordy, descriptive, complete, and quality assured.
Task	Fact table is a measurable event for which dimension table data is collected and is used for analysis and reporting.	Collection of reference information about a business.

Parameters	Fact Table	Dimension Table
Type of Data	Facts tables could contain information like sales against a set of dimensions like Product and Date.	Every dimension table contains attributes which describe the details of the dimension. E.g., Product dimensions can contain Product ID, Product Category, etc.
Key	Primary Key in fact table is mapped as foreign keys to Dimensions.	Dimension table has a primary key column that uniquely identifies each dimension.
Storage	Helps to store report labels and filter domain values in dimension tables.	Load detailed atomic data into dimensional structures.
Hierarchy	Does not contain Hierarchy	Contains Hierarchies. For example Location could contain, country, pin code, state, city, etc.
Attribute	Fact table can have data in numeric as well as textual format.	Dimension table always contains attributes in textual format
Table Size	Fact table grows Vertically.	Dimension table grows horizontally.

❖ Design & Implementation Effective Physical Data Structure Or Data Warehouse

- The physical data modeling apply physical constraint such as space performance of data.
- The physical data model is related to database system & data warehouse tools that you will use.
- The purpose of physical data structure to design actual physical implementation.
- Once a data warehouse is implemented & your customer begins using it, they will generate new request & requirements.

Physical Data Design

- Physical design translates the logical data model into a set of SQL statement that defines the database.
- Physical data model database independence & other techniques may be applied relational database engine,
- Physical data models describe how the system will be implemented using a specific DBMS system.
- This model is typically created by DBA developers the purpose of actual designing & implementation of the database.
- Physical data & design structure also help to visualize like columns, key, constraints, indexes, function, procedure & other RDBMS objects.
- The physical data design contains relationship between tables that which address connectivity on established other objects.
- The physical set of table sitting between the operational system & data warehouse or a specially administrative. The data warehouse it.
- The data design should take advantage of the database administrator capability like database partition, multi dimension clustering, table partition & materialize query tables.

 **How To Implement Physical Data In Data Warehouse**

- Implement physical data model transform the physical data into a physical database by generating the SQL DDL (Data Definition Language) script to create all the objects in the data.
- You implement physical data model in a production environment & populate it with data the ability to change the implementation is limited because of the data volume in a data warehouse.
- Physical data warehouse design is good query performance & this is achieved by proper implementation across all the data in the databases.
- Physical data model describe data specific implementation of the data model.
- It extracts the database & helps to generate schema.
- Physical data model implementation depends on the hardware & software used by the companies.
- Physical data can implement objects such as tables & columns are created based on entities & attributes.
- Physical data model describes data need for a single project or application through it may be integrated with other physical data models based on project scope.
- Physical data develop for a specific version of a DBMS, location, data storage or technology to be used in project.
- Physical data model describe the database specific implementation of the data model.
- The physical data model includes all required tables, columns, relationships, database properties for the physical implementation of database.
- Database performance, index strategy physical storage & de-normalization are important parameter of a physical structure.