

Chapter:-1

Introduction Of Data Warehousing

❖ What is data warehousing?

- Data warehouse is process of constructing & using a data warehousing.
- A data warehouse integrates by data from multiple heterogeneous sources that support analytical report, structure & adhoc queries & decision making.
- **“A data warehouse is a collection of corporate information & data derived from operational systems & external data source.”**
- Data is populated into data warehouse through the process of extraction, transformation & loading.

❖ History of data warehouse

- Data warehouse was first developed by **Bill Inmon** in **1990**.
- According to Inmon a data warehouse is a subject oriented, integrated, time variant & non-volatile collection of data.
- Data warehouse takes data from many different sources to a single location & translated into a format the data warehouse can process & store.
- Data warehouse also provides Online Analytical Process (OLAP).
- This tool helps us in interactive & effective analysis of data in a multidimensional space.
- A data warehouse help execute to organize, understand & use their data to take strategy decision.

❖ Features of data warehouse

- Subject Oriented.
- Integrated.
- Time Variant.
- Non- Volatile.

❖ Data warehouse application

- Financial Services.
- Banking Services.
- Consumer Goods.
- Retail Sector.
- Controlled Manufacturing.

❖ Types of data warehouse

- Information Processing.
- Analytical Processing.
- Data Mining.

❖ Data warehouse today

- Data warehouse technology is fast & new tool to take care of our future needs.
- Data warehousing combine information collected for multiple sources into one comprehensive database.
- Data warehouse subject oriented because it provides information around a subject rather than organization ongoing operations.
- Data warehouse is essential database & is kept separate from an organization operational database.

➤ Different form of a data warehouse

- 1) Information processing data warehouse.
- 2) Analytical processing data warehouse.
- 3) Data mining data warehouse.

1. Information processing data warehouse

- This specially allows processing of historical data which if store in it.
- There are many processing operations which can be performed like query, generating tables, chart or graph & basic operation under statistic analysis.

2. Analytical processing data warehouse

- This warehouse can be used for extensive analytical processing & this analytics performed on the data is stored in the warehouse.
- Which perform OLAP operational & few others like drill down & drill up with enhance the result of the analysis.

3. Data mining data warehouse

- This warehouse is dedicated to the data mining the discovery of information by uncovering, hidden pattern, prediction technique & analytical model construction.
- Data warehouse (DWH) today system follow updates approach rather than the traditional discuss.

❖ Data warehouse today many tools & utilities

- Data Extraction.
- Data Cleaning.
- Data Transformation.
- Data Loading.

❖ **Today data warehouse latest tool**

- Amazon red shift.
- Tera data.
- Panoply.
- Oracle 12C.
- Informatica.
- IBM Infopshere.
- Paracel.
- Cloudera.
- Marklogic.
- SAP.

❖ Future trends in data warehouse

- Data warehouse has Online Analytical Processing (OLAP) technology are now considered mature.
- The data warehouse has been the business insight work of enterprise computing.
- It means emerging technology in data warehouse & analytics in cloud.
- Data warehousing & big data analysis are emerging area in cloud today.
- Data warehouse one or more has a part of an enterprise data warehouse the hub is a subset of the data called the data core.
- Data warehouse to support two stage:
 - 1) **Refined data.**
 - 2) **Trusted data.**
- Data warehouse architecture cannot stay alone we must think purpose & position of the data warehouse in data management architecture.
- The major demand in data warehouse market for high speed data mining at lower hardware and implementation cost.
- In a future of data warehouse modernization we will need to consider cloud data warehousing with automation as well as architecture modernization.
- Data warehousing services lead the information management category in increase rate jumping for 24% to 34% in 2014.

- **Latest trends in data warehousing like:**

- 1) Rapid deployment of large scale cloud data warehouse.
- 2) Increase enable of self-services data access via cloud data integration services.
- 3) Continue growth of NoSQL adoption.
- 4) Big data analytics in the cloud.
- 5) Cloud base analytics & data dictionary.

- Data warehouse staying power because the concept of central data collects by dozens or hundreds of database, applications & system other source system.
- To the most efficient way of companies to get an enterprise wide view of their customers, supply chain, sales& operations.
- In today world of instant access by many different areas user & customers data is no longer nicely away in big warehouse.
- The trend is two words always on accessible & very open storage that is fast & friendly for customers yet complex & deep in of the most important data.

❖Data warehouse architecture

- A data warehouse is the source of business truth development by combined data from multiple separate sources.
- It supports analytical, reporting & both structure & adhoc queries.

- Data warehouse architecture includes the following layers:

- 1) **Data Source Layer.**
- 2) **Data Storage Layer.**
- 3) **Data Presentation Layer.**
- 4) **Data Staging Layer.**

➤ **Data Source Layer**

- Data source layer of data warehouse architecture is where original data is collected from a various internal & external sources, resides in the relational database.
- This store some type including like:
 - 1) **Operational Data.**
 - 2) **Social Media Data.**
 - 3) **Third Party Data.**

➤ **Data Storage Layer**

- Data storage layer is where data was cleaned in storage area as a single central process.
- Depending on your business & your data warehouse architecture requirements your data storage maybe a data warehouse, data mart & Operational Data Store (ODS).

➤ Data Presentation Layer

- The presentation layer is where users interact with cleaned and organized.
- This layer of data warehouse architecture provides users with the ability to query the data for product or services insight, analyzed & information to business scenarios & developed automated or adhoc reports.
- You may OLAP or reporting tool with a user friendly Graphical User Interface (GUI) to help users build their queries perform analysis or designed their reports.

➤ Data Staging Layer

- The data staging layer resides between data source and data warehouse.
- In this layer data is extracted from different internal and external data source.
- The data extraction layer will utilize multiple technologies & tools to extract the required data.
- The extract has been loaded it will be subjected to high level data quality data checks.
- The staging layer contain the following components:
 - 1) **Landing Database & Staging Area.**
 - 2) **Data Integration Tool.**
- A data warehouse relational database that is designed for query & analysis rather than for transaction processes.
- It usually contains historical data derived from transaction data but can include data from other sources.

- There are three types of architecture:

- 1) **Single Tier Architecture.**
- 2) **Two Tier Architecture.**
- 3) **Three Tier Architecture.**

➤ **Single Tier Architecture**

- The object of a single layer is to minimize the amount of data store.
- This goal is to remove data redundancy.
- This architecture is not frequently used.

➤ **Two Tier Architecture**

- Two tier architecture separate physically available sources & data warehouse.
- This architecture is not expandable & also not supporting a large number of end users.
- It also has connectivity problems because of network limitations.

➤ **Three Tier Architecture**

- This is the most widely used architecture.
- It consists of top, middle & bottom tier.

1) **Top Tier**

- The top tier is a front end client layer.
- Top tier is the tool & API that connect & get data out from the data warehouse.
- It could be query tools, reporting tools, manage query tool, analysis tool & data mining tools.

2) **Middle Tier**

- The middle tier in data warehouse in an OLAP server which is implemented using either ROLAP & MOLAP model.
- For a user this application tier presents and abstract view of database.
- This layer also acts as a mediator between the end user and database.

3) **Bottom Tier**

- The bottom tier of the architecture is the data warehouse database server.
- We use the back-end tools & utilities to feed data in the bottom tiers.
- This back-end tools & utilities perform the extract, clean, load & refresh function (DWH) tools.
- Data warehouse server fetch relevant information based on data mining & request.

❖ Data Flow Architecture

- A data warehouse system has two main architectures.
 - 1) **The Data Flow Architecture.**
 - 2) **The System Architecture.**
- The data flow architecture shows how the data stores are arranged within a data warehouse & how the data flows from the source system to the users through this data source.
- The system architecture is about the physical configuration of the server, network software & clients.
- The data flow architecture is a configuration of data stores within a data warehouse system.
- This includes how the data flow is controlled, logged & monitored as well as the mechanism to ensure the quality of the data in data stores.
- The data flow architecture is different from data architecture.
- Data architecture is about how the data is arranged in each data store & how a data is designed to reflect the business process.
- Data flow in the data warehouse describes which objects are needed at design time & which objects are needed at runtime to transfer data from a source to destination.
- The individual requirements of your company process are supported by number of ways to design the data flow.
- You can use any data source that transfers the data to destination or access the source data directly apply simple & complex methods & define data correspond to the requirements of your layer architecture.
- Data flow architecture transforms input data by a series of computational components into output data.

- Data flow architecture reduces development time & can move easily between design & implementation.
- Data flow architecture is a computer architecture that directly contrasts the traditional architecture or control flow architecture.
- Data flow architecture the data can be input graph topology with a cycle.
- Data flow architecture doesn't have a program counter or execution of instructions is determined based on availability.
- There are benefits and drawbacks to each type of data.
- **For Example:-**The data that you can access in the data warehouse is more complex & can represent a greater number & several of relationship but it can take longer time to collect & access that live data.
- Data flow architecture that are determined in nature enable program to manage complex task such as processor load balancing synchronization & access to common resources.
- A several of process enable reports to access application data in order to produce report output & live reporting data is accessed directly from the application using the API.