

UNIT 4 (PAST YEAR SOLVED QUESTIONS)

Q. Discuss security issues involved in database design.

OR

Q. List and discuss different issues related to database security.

OR

Q. Explain briefly the database security for integrity of database.

Database Security: Database security refers to the range of tools, controls, and measures designed to establish and preserve database confidentiality, integrity, and availability. It involves various types or categories of controls, such as technical, procedural/administrative and physical.

Database security can guard against a compromise of your database, which can lead to financial loss, reputation damage, consumer confidence disintegration, brand erosion, and non-compliance of government and industry regulation.

Database security safeguards defend against a myriad of security threats and can help protect your enterprise from:

1. Deployment failure
2. Excessive privileges
3. Privilege abuse
4. Platform vulnerabilities
5. Unmanaged sensitive data
6. Backup data exposure

7. Weak authentication

8. Database injection attacks

Security issues involved in database design are as follows:

- i. **Authentication:** Authentication is the process of confirmation that whether the user log in only according to the rights provided to him to perform the activities of data base. A particular user can login only up to his privilege but he can't access the other sensitive data. The privilege of accessing sensitive data is restricted by using Authentication. By using these authentication tools for biometrics such as retina and figure prints can prevent the data base from unauthorized/malicious users.
- ii. **Authorization:** Authorization is a privilege provided by the Database Administrator. Users of the database can only view the contents they are authorized to view. The rest of the database is out of bounds to them. The different permissions for authorizations available are:
 - Primary Permission - This is granted to users publicly and directly.
 - Secondary Permission - This is granted to groups and automatically awarded to a user if he is a member of the group.
 - Public Permission - This is publicly granted to all the users.
 - Context sensitive permission - This is related to sensitive content and only granted to a select users.

The categories of authorization that can be given to users are:

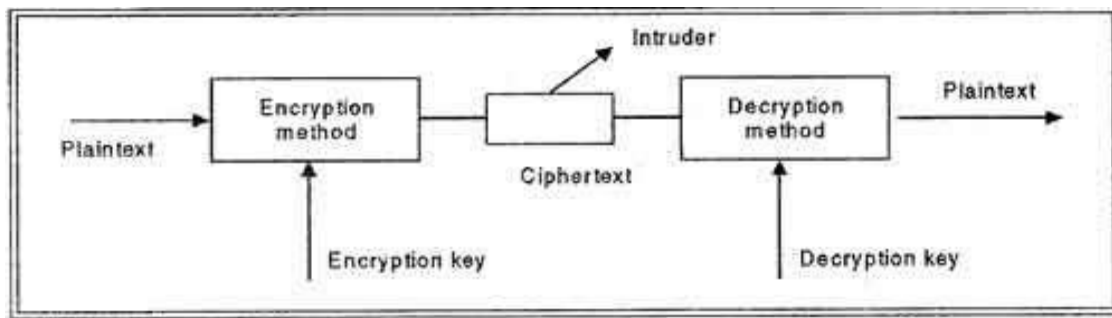
- System Administrator - This is the highest administrative authorization for a user. Users with this authorization can

also execute some database administrator commands such as restore or upgrade a database.

- **System Control** - This is the highest control authorization for a user. This allows maintenance operations on the database but not direct access to data.
- **System Maintenance** - This is the lower level of system control authority. It also allows users to maintain the database but within a database manager instance.
- **System Monitor** - Using this authority, the user can monitor the database and take snapshots of it.

iii. **Access Control:** The security mechanism of DBMS must include some provisions for restricting access to the data base by unauthorized users. Access control is done by creating user accounts and to control login process by the DBMS. So, that database access of sensitive data is possible only to those people (database users) who are allowed to access such data and to restrict access to unauthorized persons. The database system must also keep the track of all operations performed by certain user throughout the entire login time.

iv. **Encryption:** A DBMS can use encryption to protect information in certain situations where the normal security mechanisms of the DBMS are not adequate. For example, an intruder may steal tapes containing some data or tap a communication line. By storing and transmitting data in an encrypted form, the DBMS ensures that such stolen data is not intelligible to the intruder. Thus, encryption is a technique to provide privacy of data.

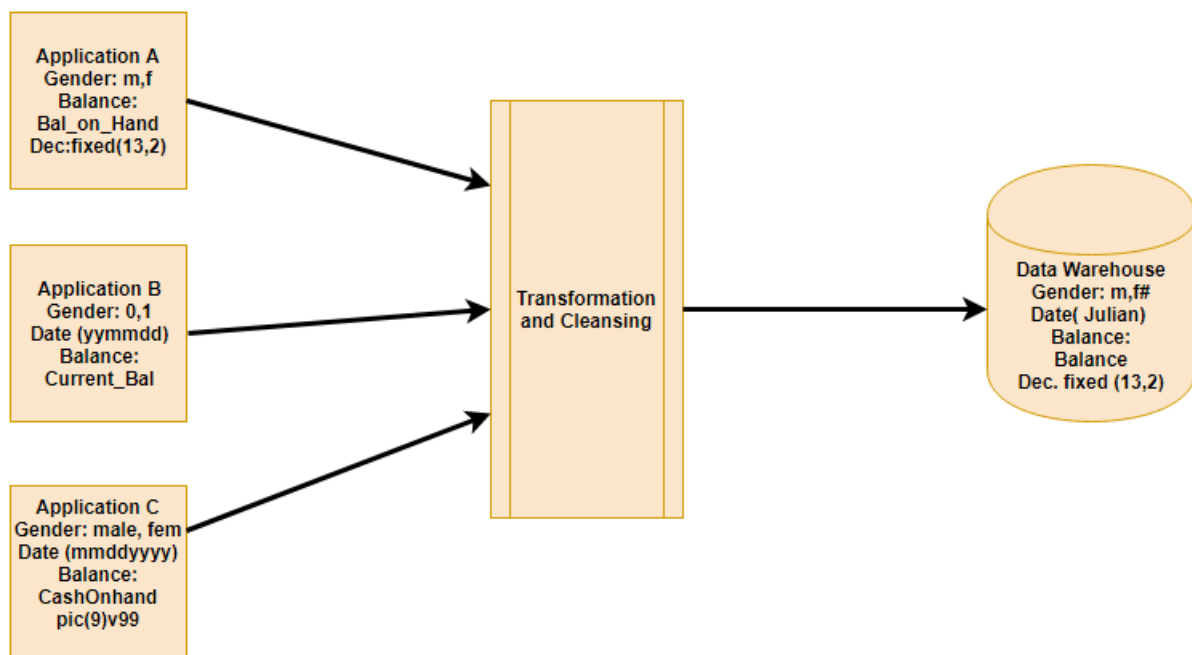


Q. Write about the four types of data warehouse schemas.

A datawarehousing is a process for collecting and managing data from varied sources to provide meaningful business insights.

A datawarehouse is typically used to connect and analyse business data from heterogeneous sources. It is electronic storage of a large amount of information by a business which is designed for query and analysis instead of transaction processing. It is a process of transforming data into information and making it available to users in a timely manner to make a difference. Datawarehouse benefits users to understand and enhance their organisation's performance.

Example:



Schema is a logical description of the entire database. It includes the name and description of records of all record types including all associated data-items. Datawarehouse also requires to maintain a schema.

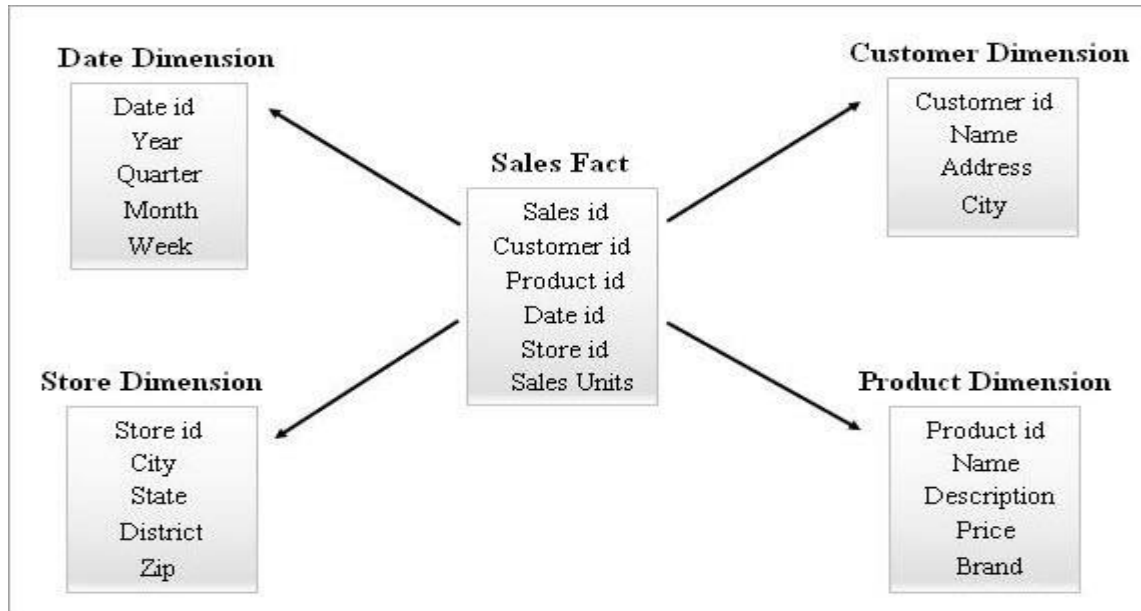
The schemas used in a datawarehouse:

1) Star Schema

This is the simplest and most effective schema in a data warehouse. A fact table in the center surrounded by multiple dimension tables resembles a star in the Star Schema model.

The fact table maintains one-to-many relations with all the dimension tables. Every row in a fact table is associated with its dimension table rows with a foreign key reference.

Example:



Benefits of Star Schema

- Queries use very simple joins while retrieving the data and thereby query performance is increased.
- It is simple to retrieve data for reporting, at any point of time for any period.

Disadvantages of Star Schema

- If there are many changes in the requirements, the existing star schema is not recommended to modify and reuse in the long run.
- Data redundancy is more as tables are not hierarchically divided.

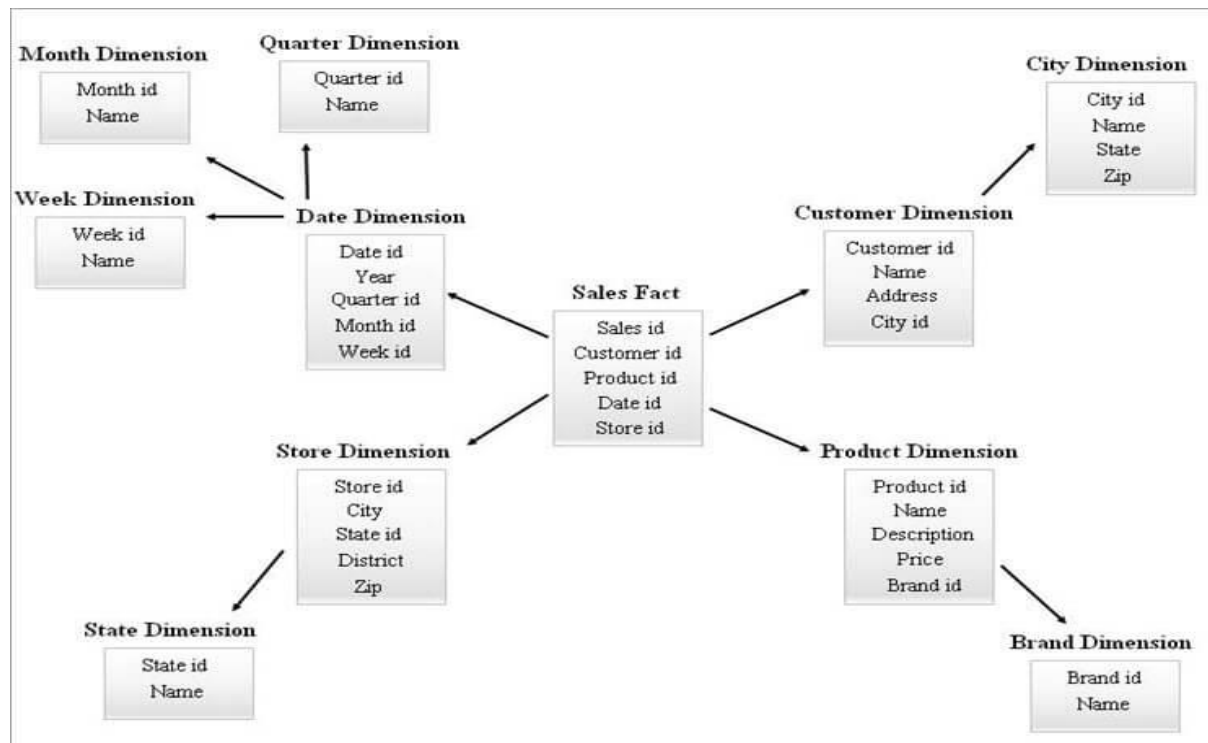
2) Snowflake Schema

Star schema acts as an input to design a Snowflake schema. Snow flaking is a process that completely normalizes all the dimension tables from a star schema.

The arrangement of a fact table in the center surrounded by multiple hierarchies of dimension tables looks like a Snowflake in the Snowflake schema model. Every fact table row is associated with its dimension table rows with a foreign key reference.

While designing Snowflake schemas the dimension tables are purposefully normalized. Foreign keys will be added to each level of the dimension tables to link to its parent attribute.

Example:



Benefits of Snowflake Schema:

- Data redundancy is completely removed by creating new dimension tables.
- It is easy to update (or) maintain the Snow Flaking tables.

Disadvantages of Snowflake Schema:

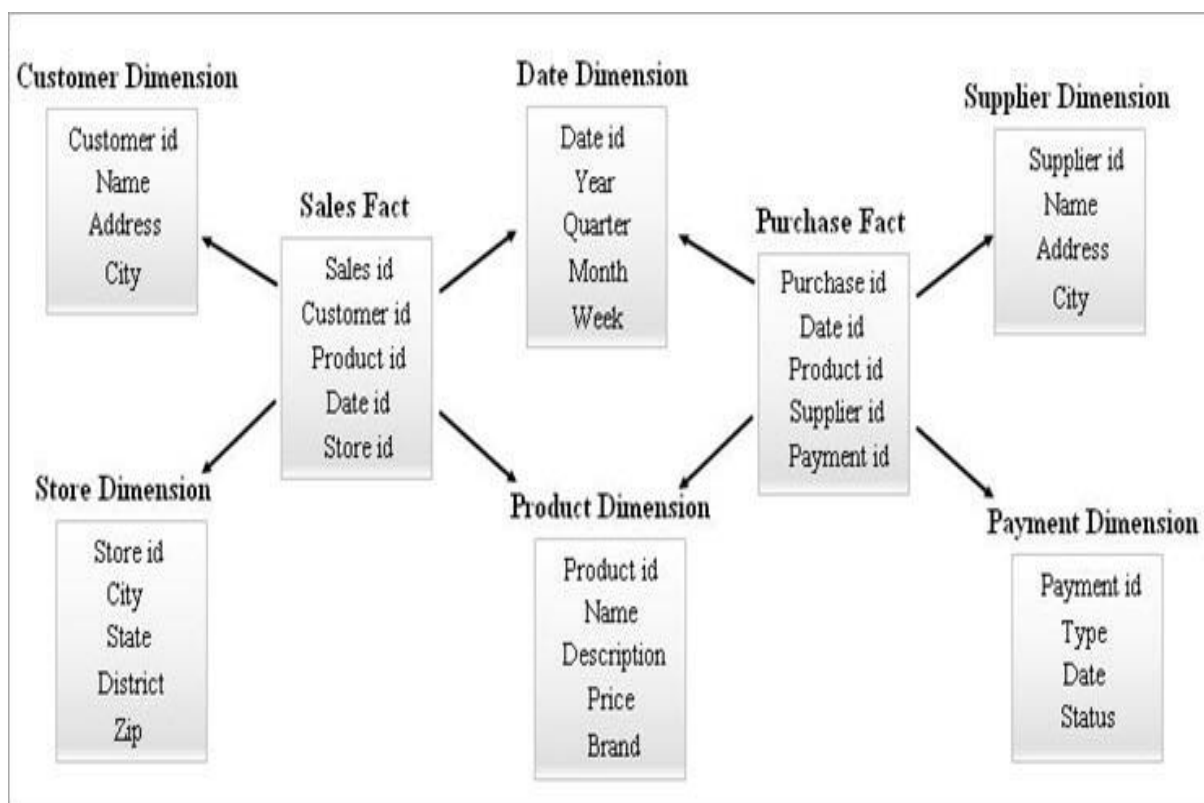
- Due to normalized dimension tables, the ETL system has to load the number of tables.
- You may need complex joins to perform a query due to the number of tables added. Hence query performance will be degraded.

3) Galaxy Schema

A galaxy schema is also known as Fact Constellation Schema. In this schema, multiple fact tables share the same dimension tables. The arrangement of fact tables and dimension tables looks like a collection of stars in the Galaxy schema model.

The shared dimensions in this model are known as Conformed dimensions.

Example:

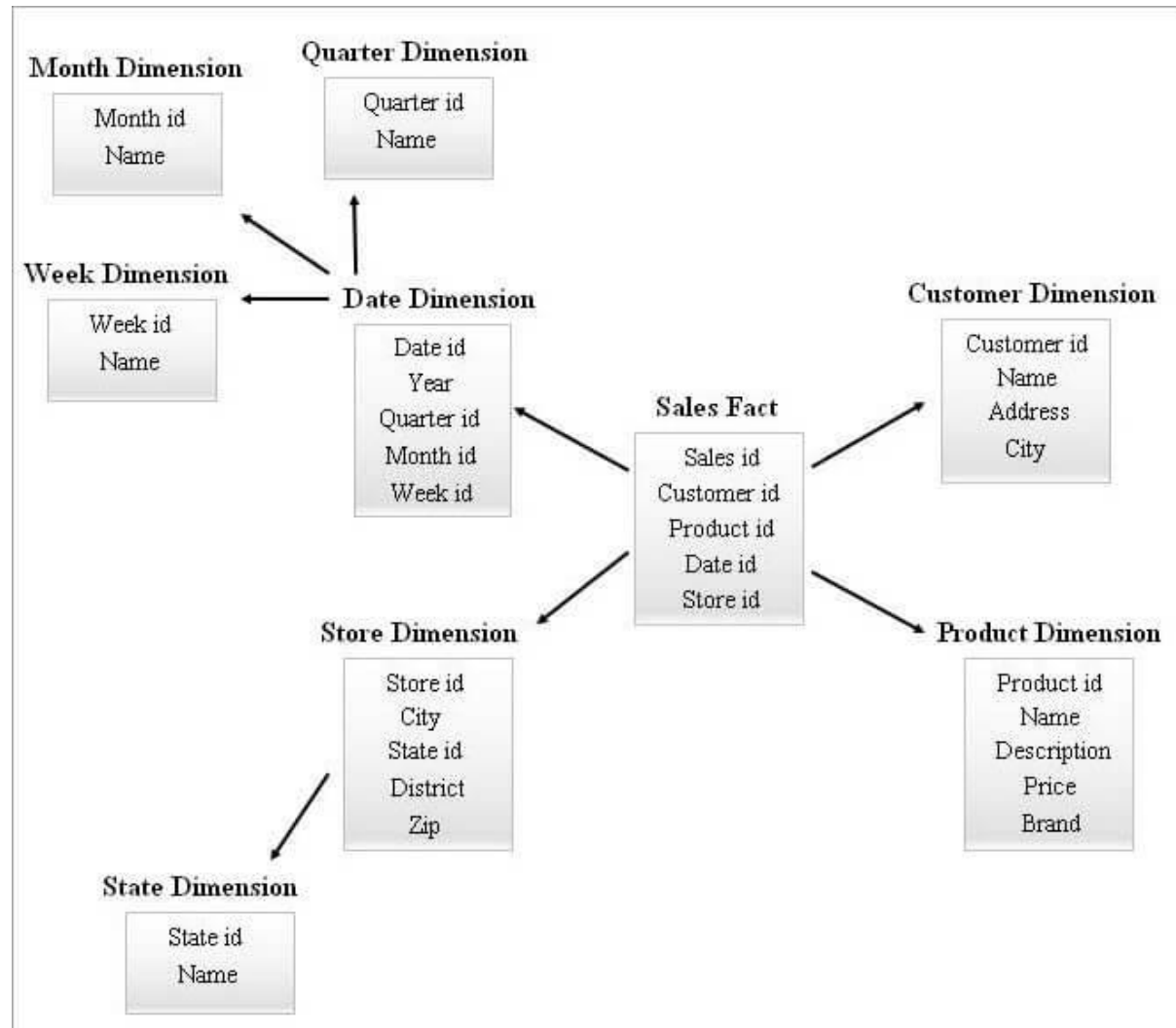


4) Star Cluster Schema

A Snowflake schema with many dimension tables may need more complex joins while querying. A star schema with fewer dimension tables may have more redundancy. Hence, a star cluster schema

came into the picture by combining the features of the above two schemas.

Example:



Q. What is data warehouse and its usage?

A Database Management System (DBMS) stores data in the form of tables, uses ER model. For example a DBMS of college has tables for students, faculty, etc.

A Data Warehouse is separate from DBMS, it stores huge amount of data, which is typically collected from multiple heterogeneous source like files, DBMS, etc. The goal is to produce statistical results that may help in decision makings.

Applications of Data Warehousing

Data Warehousing can be applicable anywhere where we have huge amount of data and we want to see statistical results that help

Social Media Websites: The social networking websites like Facebook, Twitter, Linkedin etc. are based on analyzing large data sets. These sites gather data related to members, groups, locations etc. and store it in a single central repository. Being large amount of data, Data Warehouse is needed for implementing the same.

Banking: Most of the banks these days use warehouses to see spending patterns of account/card holders. They use this to provide them special offers, deals, etc.

Government: Government uses data warehouse to store and analyze tax payment which is used to detect tax thefts. There can be many more applications in different sectors like E-Commerce, Telecommunication, Transportation Services, Marketing and Distribution, Healthcare and Retail.

Data Warehouse Architecture

Data Warehouse Architecture is complex as it's an information system that contains historical and commutative data from multiple sources. There are 3 approaches for constructing Data Warehouse

layers: Single Tier, Two tier and Three tier. This 3 tier architecture of Data Warehouse is explained as below.

Single-tier architecture

The objective of a single layer is to minimize the amount of data stored. This goal is to remove data redundancy. This architecture is not frequently used in practice.

Two-tier architecture

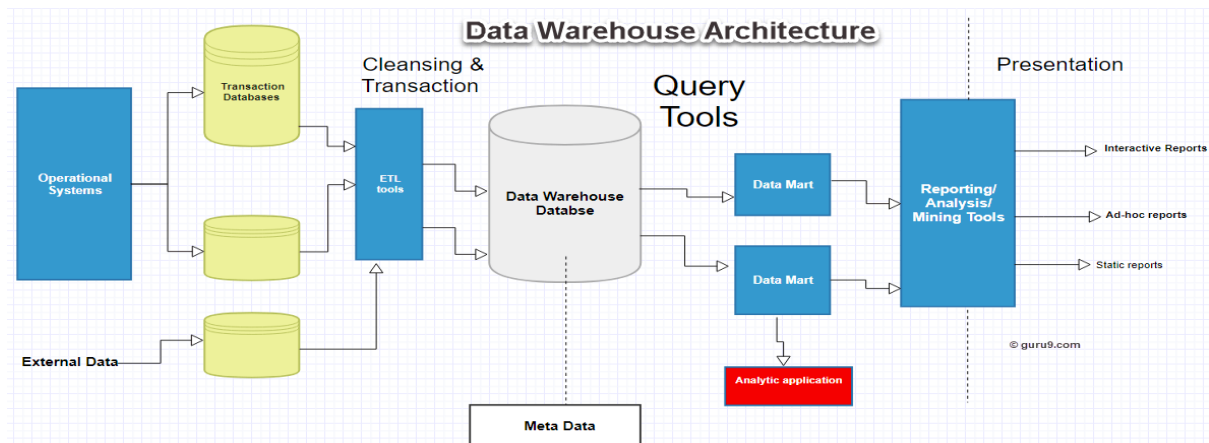
Two-layer architecture is one of the Data Warehouse layers which separates physically available sources and data warehouse. This architecture is not expandable and also not supporting a large number of end-users. It also has connectivity problems because of network limitations.

Three-Tier Data Warehouse Architecture

This is the most widely used Architecture of Data Warehouse.

It consists of the Top, Middle and Bottom Tier.

1. Bottom Tier: The database of the Datawarehouse servers as the bottom tier. It is usually a relational database system. Data is cleansed, transformed, and loaded into this layer using back-end tools.
2. Middle Tier: The middle tier in Data warehouse is an OLAP server which is implemented using either ROLAP or MOLAP model. For a user, this application tier presents an abstracted view of the database. This layer also acts as a mediator between the end-user and the database.
3. Top-Tier: The top tier is a front-end client layer. Top tier is the tools and API that you connect and get data out from the data warehouse. It could be Query tools, reporting tools, managed query tools, Analysis tools and Data mining tools.



To design Data Warehouse Architecture, you need to follow below given best practices:

- Use Data Warehouse Models which are optimized for information retrieval which can be the dimensional mode, denormalized or hybrid approach.
- Choose the appropriate designing approach as top down and bottom up approach in Data Warehouse.
- Need to assure that Data is processed quickly and accurately. At the same time, you should take an approach which consolidates data into a single version of the truth.
- Carefully design the data acquisition and cleansing process for Data warehouse.
- Design a MetaData architecture which allows sharing of metadata between components of Data Warehouse.
- Consider implementing an ODS model when information retrieval need is near the bottom of the data abstraction pyramid or when there are multiple operational sources required to be accessed.
- One should make sure that the data model is integrated and not just consolidated. In that case, you should consider 3NF data model. It is also ideal for acquiring ETL and Data cleansing tools.

Uses of Datawarehouse

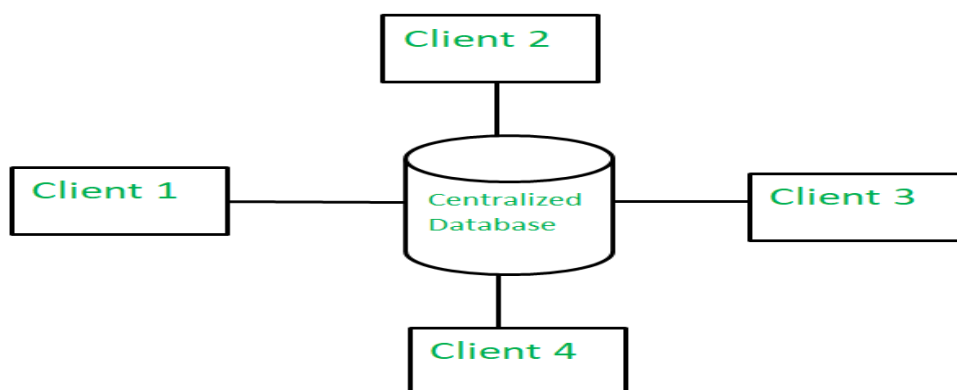
Here, are most common sectors where Data warehouse is used:

1. Airline: In the Airline system, it is used for operation purpose like crew assignment, analyses of route profitability, frequent flyer program promotions, etc.
2. Banking: It is widely used in the banking sector to manage the resources available on desk effectively. Few banks also used for the market research, performance analysis of the product and operations.
3. Healthcare: Healthcare sector also used Data warehouse to strategize and predict outcomes, generate patient's treatment reports, share data with tie-in insurance companies, medical aid services, etc.
4. Public sector: In the public sector, data warehouse is used for intelligence gathering. It helps government agencies to maintain and analyze tax records, health policy records, for every individual.
5. Investment and Insurance sector: In this sector, the warehouses are primarily used to analyze data patterns, customer trends, and to track market movements.
6. Retail chain: In retail chains, Data warehouse is widely used for distribution and marketing. It also helps to track items, customer buying pattern, promotions and also used for determining pricing policy.

7. Telecommunication: A data warehouse is used in this sector for product promotions, sales decisions and to make distribution decisions.
8. Hospitality Industry: This Industry utilizes warehouse services to design as well as estimate their advertising and promotion campaigns where they want to target clients based on their feedback and travel patterns.

Q. What do you understand by distributed data processing?

A distributed database is a collection of multiple interconnected databases, which are spread physically across various locations that communicate via a computer network. i.e, Telephone, network applications-peer to peer networks.



- Databases in the collection are logically interrelated with each other. Often they represent a single logical database.
- Data is physically stored across multiple sites. Data in each site can be managed by a DBMS independent of the other sites.
- The processors in the sites are connected via a network. They do not have any multiprocessor configuration.
- A distributed database is not a loosely connected file system.

Types of Distributed Databases

1. Homogeneous Distributed Databases

In a homogeneous distributed database, all the sites use identical DBMS and operating systems. Its properties are:

- The sites use very similar software. i.e. same database system software.
- The sites use identical DBMS or DBMS from the same vendor.
- Each site is aware of all other sites and cooperates with other sites to process user requests.

2. Heterogeneous Distributed Databases

In a heterogeneous distributed database, different sites have different operating systems, DBMS products and data models. Its properties are –

- Different sites use dissimilar schemas and software.
- The system may be composed of a variety of DBMSs like relational, network, hierarchical or object oriented.
- Query processing is complex due to dissimilar schemas.

Advantages of Distributed Databases

- **More Reliable**– In case of database failures, the total system of centralized databases comes to a halt. However, in distributed systems, when a component fails, the functioning of the system continues may be at a reduced performance. Hence it is more reliable.
- **Better Response** – If data is distributed in an efficient manner, then user requests can be met from local data itself, thus providing faster response.
- **Lower Communication Cost** – In distributed database systems, if data is located locally where it is mostly used, then the communication costs for data manipulation can be minimized. This is not feasible in centralized systems.

Disadvantages of Distributed Databases

- Need for complex and expensive software– it demands complex and often expensive software to provide data transparency and co-ordination across the several sites.
- Overheads for improper data distribution – Responsiveness of queries is largely dependent upon proper data distribution. Improper data distribution often leads to very slow response to user requests.
- Data integrity –The need for updating data in multiple sites pose problems of data integrity.

Q. Explain Data Mining and Data Warehousing.

Data Mining

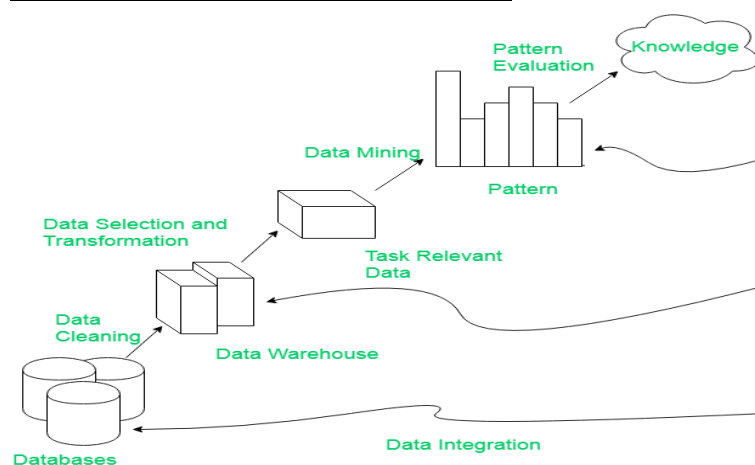
Data Mining refers to the extraction of useful information from a bulk of data or data warehouses.

Data Mining is the process of discovering or mining knowledge from a large amount of data.

The result of data mining is the hidden patterns and the knowledge that we gain at the end of the extraction process.

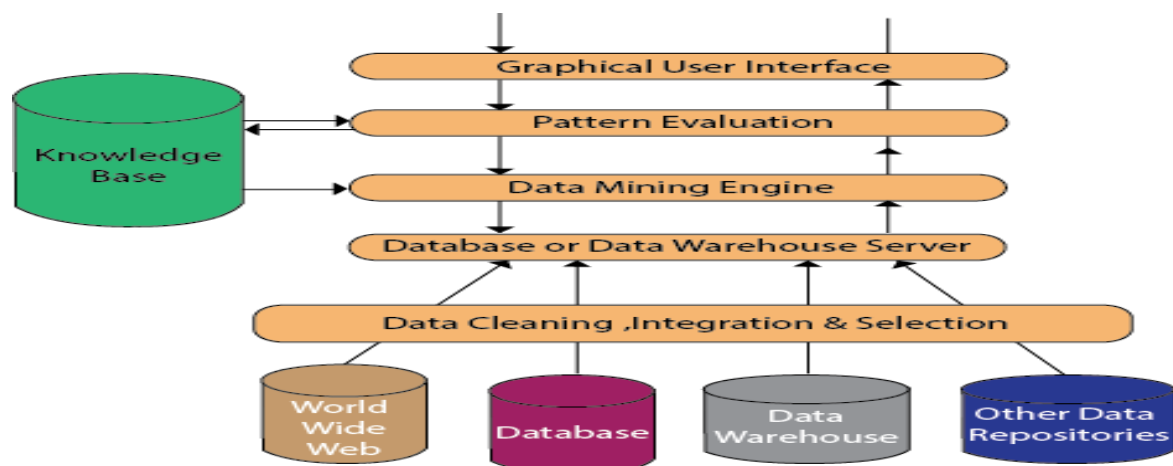
Data mining is also known as Knowledge Discovery from Data (KDD).

STEPS IN DATA MINING PROCESS



- 1) Data Cleaning: In this step, the noisy and irrelevant data is removed from the collection.
- 2) Data Integration: In this step, the heterogeneous data from multiple sources is combined in a common source.
- 3) Data Selection: In this step, the data relevant to the analysis is decided and retrieved from the data collection.
- 4) Data Transformation: In this step, the data is transformed into appropriate form required by the mining procedure.
- 5) Data Mining: In this step, intelligent techniques are applied to extract meaningful information that is useful for us.
- 6) Pattern Evaluation: In this step, strictly increasing patterns are identified that represent knowledge based on the given measures.
- 7) Knowledge representation: Knowledge representation is defined as technique which utilizes visualization tools like tables, reports etc to represent data mining results.

DATA MINING ARCHITECTURE



- Data Sources: Database, World Wide Web(WWW), Data warehouse and other data repositories are data sources. The data in these sources may be in the form of plain text, spread sheets or in other forms of media like photos or videos. WWW is one of the biggest sources of data.
- Database Server: The database server contains the actual data ready to be processed. It performs the task of handling data retrieval as per the request of the user.

- **Data Mining Engine:** It is one of the core components of the data mining architecture that performs all kinds of data mining techniques like association, classification, characterization, clustering, prediction, etc.
- **Pattern Evaluation Modules:** They are responsible for finding interesting patterns in the data and sometimes they also interact with the database servers for producing the result of the user requests.
- **Graphic User Interface:** Since the user cannot fully understand the complexity of the data mining process so graphical user interface helps the user to communicate effectively with the data mining system.
- **Knowledge Base:** Knowledge Base is an important part of the data mining engine that is quite beneficial in guiding the search for the result patterns. Data mining engine may also sometimes get inputs from the knowledge base. This knowledge base may contain data from user experiences. The objective of the knowledge base is to make the result more accurate and reliable.

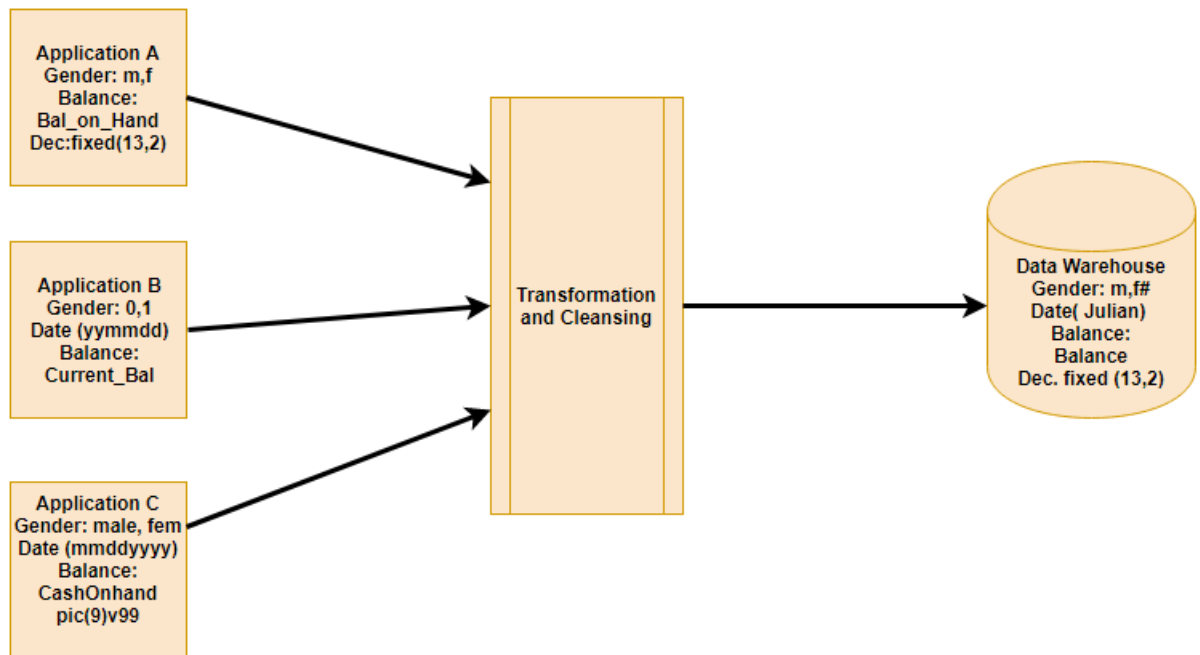
Data Warehouse

Characteristics of Data warehouse

Data Warehouse Concepts have following characteristics:

- **Subject-Oriented:** A data warehouse is subject oriented as it offers information regarding a theme instead of companies' ongoing operations. These subjects can be sales, marketing, distributions, etc. A data warehouse never focuses on the ongoing operations. Instead, it put emphasis on modeling and analysis of data for decision making. It also provides a simple and concise view around the specific subject by excluding data which is not helpful to support the decision process.
- **Integrated:** In Data Warehouse, integration means the establishment of a common unit of measure for all similar data from the dissimilar database. The data also needs to be stored in the Datawarehouse in common and universally acceptable manner. A data warehouse is developed by integrating data from varied sources like a mainframe, relational databases, flat files, etc. Moreover, it must keep consistent naming conventions, format, and coding. This integration helps in effective analysis of data. Consistency in naming conventions, attribute

measures, encoding structure etc. have to be ensured. Consider the following example:



In the above example, there are three different application labeled A, B and C. Information stored in these applications are Gender, Date, and Balance. However, each application's data is stored different way.

- In Application A gender field store logical values like M or F
- In Application B gender field is a numerical value.
- In Application C application, gender field stored in the form of a character value.
- Same is the case with Date and balance

However, after transformation and cleaning process all this data is stored in common format in the Data Warehouse.

- **Time-Variant:** The time horizon for data warehouse is quite extensive compared with operational systems. The data collected in a data warehouse is recognized with a particular period and offers information from the historical point of view. It contains an element of time, explicitly or implicitly. One such place where Datawarehouse data display time variance is in the structure of the record key. Every primary key contained with the Data Warehouse should have either implicitly or explicitly an element of time. Like the day, week month, etc. Another aspect of time variance is that once data is inserted in the warehouse, it can't be updated or changed.

- Non-volatile: Data warehouse is also non-volatile means the previous data is not erased when new data is entered in it. Data is read-only and periodically refreshed. This also helps to analyze historical data and understand what & when happened. It does not require transaction process, recovery and concurrency control mechanisms. Activities like delete, update, and insert which are performed in an operational application environment are omitted in Data warehouse environment.

Data Warehouse Architecture

Data Warehouse Architecture is complex as it's an information system that contains historical and commutative data from multiple sources. There are 3 approaches for constructing Data Warehouse layers: Single Tier, Two tier and Three tier. This 3 tier architecture of Data Warehouse is explained as below.

Single-tier architecture: The objective of a single layer is to minimize the amount of data stored. This goal is to remove data redundancy. This architecture is not frequently used in practice.

Two-tier architecture: Two-layer architecture is one of the Data Warehouse layers which separates physically available sources and data warehouse. This architecture is not expandable and also not supporting a large number of end-users. It also has connectivity problems because of network limitations.

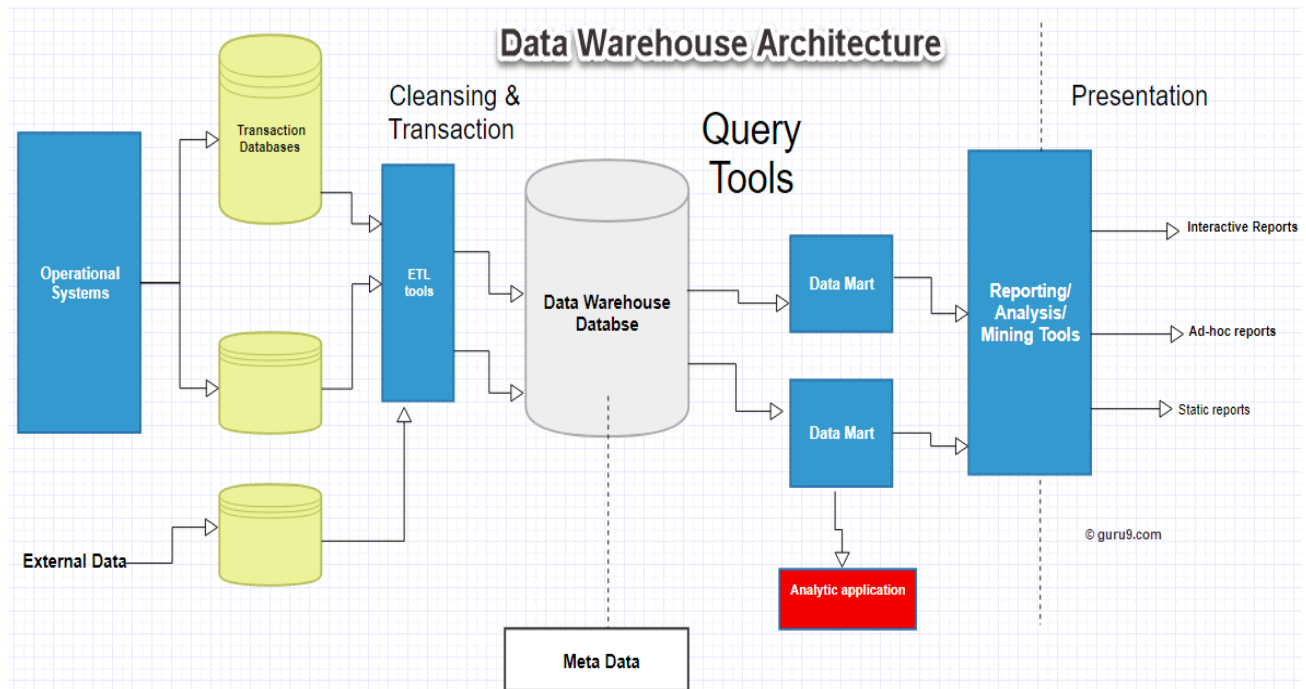
Three-tier architecture: This is the most widely used Architecture of Data Warehouse. It consists of the Top, Middle and Bottom Tier.

1. Bottom Tier: The database of the Datawarehouse servers as the bottom tier. It is usually a relational database system. Data is cleansed, transformed, and loaded into this layer using back-end tools.

2. Middle Tier: The middle tier in Data warehouse is an OLAP server which is implemented using either ROLAP or MOLAP model. For a user, this application tier presents an abstracted view of the database. This layer also acts as a mediator between the end-user and the database.

3. Top-Tier: The top tier is a front-end client layer. Top tier is the tools and API that you connect and get data out from the data warehouse. It could be Query

tools, reporting tools, managed query tools, Analysis tools and Data mining tools.



Q. Explain web database.

A web database is a wide term for managing data online. A web database gives you the ability to build your own databases/data storage without you being a database guru or even a technical person.

Examples: banks, airline and rental car reservations, university course registration and so on

- The Web is a distributed information system base on hypertext.
- Most Web documents are hypertext documents formatted via HTML
- HTML Documents contain

- Text along with font specifications, and other formatting instructions
- Hypertext links to other documents, which can be associated with region of the text.

Application of Web Database

The manner in which many organizations use a database in conjunction with a website is a good example.

It help us to serve up banner advertisements on web pages. We don't like them any better than you do, but the fact remains that they are a popular application for web databases, which can be used to store advertisements and retrieve them for display by a web server.

It helps you to make the website more useful in ways not even related to the membership list. You may publish an electronic newsletter that has a children's section in each issue containing a relevant quiz.

Advantages of Web Database

1. Save Money: One of the advantages of online database software is that it can save your business money. When you don't need to buy a software program for your business, this could result in a major savings overall. In most cases, businesses pay for a software program and then pay for a licensing fee for each computer that uses it. Using an online database may prove cheaper, depending on the number of computers you use. You also don't need to invest in servers to store the data at your business.

2. Flexible Use: Another benefit of using an online database program is that it allows your business to be flexible. You only pay for the amount of storage that you use. You need not worry about purchasing servers as you go or eliminating them when they are no longer needed. If your business grows or shrinks, you do not need to

be concerned about the costs of database management software or servers.

3. Technical Support: Another advantage of using a Web-based database program is that you can shift the technical support burden to someone else. Paying a company for access to an online database includes technical support. If the database has problems, you simply contact the company and the staff handles it. You don't need to pay for an information technology professional for this purpose. If you already have an IT department, your employees can focus on other things.

4. Access: Having access to the database at all times from multiple locations is another major advantage of this type of database. With an online database, you could theoretically access the information in the database from any computer. The information is also available 24 hours a day, seven days a week. This means that all employees have access to the same information and can collaborate with one another on projects — regardless of location. This advantage can increase productivity and improve efficiency.

Disadvantages of Web Database

1. No internet connection = no access: Online means... online. If you don't have a reliable internet connection to support your online database software, you don't have a reliable access to your data. It's sad but true.

If you are dealing with very sensitive data and you can't ever be out of touch with it, I suggest having a backup internet connection from a different provider. An installed solution sure can be considered, but you won't have access to it unless the software is installed on the computer you are working on.

2. Subscription and per user pricing: In terms of pricing, an installed solution represents one big lump sum when you acquire the software. You sure need to prepare a while in advance to fit this in your budget but once it's paid, you are good until you have to buy the next version. When it comes to online database software, there is no huge initial cost of acquisition but you still have to plan ahead since the fees are recurrent on a monthly or yearly basis. The bigger the team, the pricier it gets. Same for your apps, the more you need, the more you pay.

3. Cost of extra storage: There is one last point to consider in the negative column in terms of pricing if you are going for an online database platform. With virtually all vendors, you will face a cap when it comes to storage space and/or number of items you can create.