Top 50 Data Warehouse Interview Questions & Answers

1. What is Datawarehousing?

A Datawarehouse is the repository of a data and it is used for Management decision support system. Datawarehouse consists of wide variety of data that has high level of business conditions at a single point in time.

In single sentence, it is repository of integrated information which can be available for queries and analysis.

2. What is Business Intelligence?

Business Intelligence is also known as DSS – Decision support system which refers to the technologies, application and practices for the collection, integration and analysis of the business related information or data. Even, it helps to see the data on the information itself.

3. What is Dimension Table?

Dimension table is a table which contain attributes of measurements stored in fact tables. This table consists of hierarchies, categories and logic that can be used to traverse in nodes.

4. What is Fact Table?

Fact table contains the measurement of business processes, and it contains foreign keys for the dimension tables.

Example – If the business process is manufacturing of bricks

Average number of bricks produced by one person/machine – measure of the business process

5. What are the stages of Datawarehousing?

There are four stages of Datawarehousing:

- Offline Operational Database
- Offline Data Warehouse
- Real Time Datawarehouse
- Integrated Datawarehouse

6. What is Data Mining?

Data Mining is set to be a process of analyzing the data in different dimensions or perspectives and summarizing into a useful information. Can be queried and retrieved the data from database in their own format.

7. What is OLTP?

OLTP is abbreviated as On-Line Transaction Processing, and it is an application that modifies the data whenever it received and has large number of simultaneous users.

8. What is OLAP?

OLAP is abbreviated as Online Analytical Processing, and it is set to be a system which collects, manages, processes multi-dimensional data for analysis and management purposes.

9. What is the difference between OLTP and OLAP?

Following are the differences between OLTP and OLAP:

OLTP	OLAP
Data is from original data source	Data is from various data sources
Simple queries by users	Complex queries by system
Normalized small database	De-normalized Large Database
Fundamental business tasks	Multi-dimensional business tasks

10. What is ODS?

ODS is abbreviated as Operational Data Store and it is a repository of real time operational data rather than long term trend data.

11. What is the difference between View and Materialized View?

A view is nothing but a virtual table which takes the output of the query and it can be used in place of tables.

A materialized view is nothing but an indirect access to the table data by storing the results of a query in a separate schema.

12. What is ETL?

ETL is abbreviated as Extract, Transform and Load. ETL is a software which is used to reads the data from the specified data source and extracts a desired subset of data. Next, it transform the data using rules and lookup tables and convert it to a desired state.

Then, load function is used to load the resulting data to the target database.

13. What is VLDB?

VLDB is abbreviated as Very Large Database and its size is set to be more than one terabyte database. These are decision support systems which is used to server large number of users.

14. What is real-time datawarehousing?

Real-time datawarehousing captures the business data whenever it occurs. When there is business activity gets completed, that data will be available in the flow and become available for use instantly.

15. What are Aggregate tables?

Aggregate tables are the tables which contain the existing warehouse data which has been grouped to certain level of dimensions. It is easy to retrieve data from the aggregated tables than the original table which has more number of records.

This table reduces the load in the database server and increases the performance of the query.

16. What is factless fact tables?

A factless fact tables are the fact table which doesn't contain numeric fact column in the fact table.

17. How can we load the time dimension?

Time dimensions are usually loaded through all possible dates in a year and it can be done through a program. Here, 100 years can be represented with one row per day.

18. What are Non-additive facts?

Non-Addictive facts are said to be facts that cannot be summed up for any of the dimensions present in the fact table. If there are changes in the dimensions, same facts can be useful.

19. What is conformed fact?

Conformed fact is a table which can be used across multiple data marts in combined with the multiple fact tables.

20. What is Datamart?

A Datamart is a specialized version of Datawarehousing and it contains a snapshot of operational data that helps the business people to decide with the analysis of past trends and experiences. A data mart helps to emphasizes on easy access to relevant information.

21. What is Active Datawarehousing?

An active datawarehouse is a datawarehouse that enables decision makers within a company or organization to manage customer relationships effectively and efficiently.

22. What is the difference between Datawarehouse and OLAP?

Datawarehouse is a place where the whole data is stored for analyzing, but OLAP is used for analyzing the data, managing aggregations, information partitioning into minor level information.

23. What is ER Diagram?

ER diagram is abbreviated as Entity-Relationship diagram which illustrates the interrelationships between the entities in the database. This diagram shows the structure of each tables and the links between the tables.

24. What are the key columns in Fact and dimension tables?

Foreign keys of dimension tables are primary keys of entity tables. Foreign keys of fact tables are the primary keys of the dimension tables.

25. What is SCD?

SCD is defined as slowly changing dimensions, and it applies to the cases where record changes over time.

26. What are the types of SCD?

There are three types of SCD and they are as follows:

SCD 1 – The new record replaces the original record

SCD 2 – A new record is added to the existing customer dimension table

SCD 3 – A original data is modified to include new data

27. What is BUS Schema?

BUS schema consists of suite of confirmed dimension and standardized definition if there is a fact tables.

28. What is Star Schema?

Star schema is nothing but a type of organizing the tables in such a way that result can be retrieved from the database quickly in the data warehouse environment.

29. What is Snowflake Schema?

Snowflake schema which has primary dimension table to which one or more dimensions can be joined. The primary dimension table is the only table that can be joined with the fact table.

30. What is a core dimension?

Core dimension is nothing but a Dimension table which is used as dedicated for single fact table or datamart.

31. What is called data cleaning?

Name itself implies that it is a self explanatory term. Cleaning of Orphan records, Data breaching business rules, Inconsistent data and missing information in a database.

32. What is Metadata?

Metadata is defined as data about the data. The metadata contains information like number of columns used, fix width and limited width, ordering of fields and data types of the fields.

33. What are loops in Datawarehousing?

In datawarehousing, loops are existing between the tables. If there is a loop between the tables, then the query generation will take more time and it creates ambiguity. It is advised to avoid loop between the tables.

34. Whether Dimension table can have numeric value?

Yes, dimension table can have numeric value as they are the descriptive elements of our business.

35. What is the definition of Cube in Datawarehousing?

Cubes are logical representation of multidimensional data. The edge of the cube has the dimension members,and the body of the cube contains the data values.

36. What is called Dimensional Modelling?

Dimensional Modeling is a concept which can be used by dataware house designers to build their own datawarehouse. This model can be stored in two types of tables – Facts and Dimension table.

Fact table has facts and measurements of the business and dimension table contains the context of measurements.

37. What are the types of Dimensional Modeling?

There are three types of Dimensional Modeling and they are as follows:

- Conceptual Modeling
- Logical Modeling
- Physical Modeling

38. What is surrogate key?

Surrogate key is nothing but a substitute for the natural primary key. It is set to be a unique identifier for each row that can be used for the primary key to a table.

39. What is the difference between ER Modeling and Dimensional Modeling?

ER modeling will have logical and physical model but Dimensional modeling will have only Physical model.

ER Modeling is used for normalizing the OLTP database design whereas Dimensional Modeling is used for de-normalizing the ROLAP and MOLAP design.

40. What are the steps to build the datawarehouse?

Following are the steps to be followed to build the datawaerhouse:

- Gathering business requirements
- Identifying the necessary sources
- Identifying the facts
- Defining the dimensions
- Defining the attributes
- Redefine the dimensions and attributes if required
- Organize the Attribute hierarchy
- Define Relationships
- Assign unique Identifiers

41. What are the different types of datawarehosuing?

Following are the different types of Datawarehousing:

- Enterprise Datawarehousing
- Operational Data Store
- Data Mart

42. What needs to be done while starting the database?

Following need to be done to start the database:

1. Start an Instance
2. Mount the database
3. Open the database

43. What needs to be done when the database is shutdown?

Following needs to be done when the database is shutdown:

1. Close the database
2. Dismount the database
3. Shutdown the Instance

44. Can we take backup when the database is opened?

No, We cannot take full backup when the database is opened.

45. What is defined as Partial Backup?

A Partial backup in an operating system is a backup short of full backup and it can be done while the database is opened or shutdown.

46. What is the goal of Optimizer?

The goal to Optimizer is to find the most efficient way to execute the SQL statements.

47. What is Execution Plan?

Execution Plan is a plan which is used to the optimizer to select the combination of the steps.

48. What are the approaches used by Optimizer during execution plan?

There are two approaches:

1. Rule Based
2. Cost Based

49. What are the tools available for ETL?

Following are the ETL tools available:

Informatica
Data Stage
Oracle
Warehouse Builder
Ab Initio
Data Junction

50.What is the difference between metadata and data dictionary?
Metadata is defined as data about the data. But, Data dictionary contain the information about the project information, graphs, abinito commands and server information.