**BI & DATA WAREHOUSING INTERVIEW QUESTIONS**

**Question1:- What is Data Warehouse?**

**Data Warehouse is like a relational database designed for analytic needs.**

**It is a central location where consolidated data from multiple locations (databases) are stored.**

**Question2:- What is the very basic difference between a Data Warehouse and an operational Database?**

**Data Warehouse:-**

**Contains historical information which helps in analyzing business metrics.**

**Data warehouse is mainly used to read data.**

**End users are business analysts/ data analysts.**

**Operation Database:-**

**Contains current information that is required to run the business.**

**Database is mainly used to write the data.**

**End users are ops team members.**

**Question3:- What is Data Warehousing?**

**Data Warehousing is the act of organizing & storing data in a way so as to make its retrieval efficient and insightful.**

**It's also called as the process of transforming data into information.**

**Question4:- What is OLAP?**

**OLAP (Online Analytical Processing) is a flexible way to make complicated analysis of multidimensional data.**

**Data present in a Data Warehouse is accessed by running OLAP queries. DBs however, are queried by running OLTP (Online Transaction Processing) operations.**

**OLAP activities are performed by converting the multi-dimensional data in a Warehouse into an OLAP cube.**

**Question5:- What is OLTP? How different is OLAP from OLTP?**

**OLTP stands for Online Transaction Processing.**

**OLTP queries are used to perform DDL operations on a Database.**

**OLTP systems use data stored in the form of two-dimensional tables, with rows and columns.**

**OLAP is used to perform real-time analysis on multidimensional data in a Warehouse.**

**OLAP deals with De-normalized data.**

**Question6:- What is Dimension Table?**

**The tables that describe the dimensions involved are called Dimension tables.**

**Dividing a Data Warehouse project into dimensions provides structured information for analysis & reporting.**

**End uses fire queries on these dimension tables which contain descriptive information.**

**Question7:- What is FACT Table?**

**A fact table is a table containing the measure of the dimensions in a dimension table.**

**Fact is measured by summing, averaging or manipulation the data in a dimension table.**

**A fact table contains 2 kinds of data - a dimension key (foreign key) and a measure.**

**Note: Every Dimension table is linked to a fact table.**

**Question8:- What is the level of Granularity of a fact table?**

**The depth of data level is known as granularity.**

**A fact table is usually designed at a low level of Granularity.**

**Examples:**

**1. Lower level granularity of data dimension can be year, month, week, etc.**

**2. Lower level granularity of employee performance weekly.**

**Question1:- What is the difference between Additive, Semi-additive and Non-additive facts?**

**Additive fact:-**

**It is a measure in a fact table that can be fully summed across any of the dimensions associated with it.**

**Semi-additive fact:-**

**It is a measure in a fact table that can be summed across some dimensions associated with it, but not all.**

**Non-additive fact:-**

**It is a measure in a fact table that cannot be summed across any of the dimensions associated with it.**

**Question10:- What is factless fact table?**

**Factless fact tables are the fact tables which do not contain numeric fact column in the fact table.**

**Question11:- What is Conformed dimensions and Conformed facts?**

**A dimension table which is used by more than one fact table is known as a conformed dimension. Conformed dimensions can be used across multiple Data Marts in combination with multiple facts tables accordingly.**

**Similarly, Conformed fact is a fact which is used in more than one fact table. Conformed fact is a fact table which can be used across multiple data marts in combination with multiple fact tables.**

**Question12:- What is Aggregate tables?**

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

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

**Question1:- What is Summary Information?**

**Summary Information is the area in a Data Warehouse where predefined aggregations are kept.**

**Question14:- What is ETL?**

**ETL stands for Extract-> Transform -> Load.**

**It is the process of using a software to extract the desired data from various sources, then transform that data by using rules and lookup tables to meet you requirement, and then loading it into a target data warehouse.**

**SOME ELT TOOLS:-**

**Informatica Power Center**

**Talend Studio \*\*\*\***

**DataStage**

**Oracle Warehosue Builder**

**SAP Data Services**

**SQL Server Integration Services(SSIS) \*\*\***

**Question16:- What is Data MART?**

**Data mart is a smaller version of the Data Warehouse which deals with a single subject.**

**Data marts are focused on one are. Hence, they draw data from a limited number of sources.**

**Time taken to build Data Marts is very less compared to the time taken to build a Data Warehouse.**

**Question17:- What is Data Meta?**

**Metadata is defined as data about data.**

**Metadata is a DWH defines the source data i.e. Flat file, Relational Database and other objects.**

**Metadata is used to define which table is source and target, and which concept is used to build business logic called transformation to the actual output.**

**Question18:- What is Data Mining?**

**Data mining is the process of analyzing data in different dimensions & summarizing it into useful info.**

**Data warehousing is about storing analytical data in a structure suitable for data mining. This analytical data is extracted from operational systems usually on a daily basis.**

**Question19:- What is OLAP servers?**

**Multi-Dimensional OLAP:-**

**MOLAP is a form of OLAP that processes and stores the data directly into a multidimensional database. Benefit is, it can perform complex calculations but, only limited data can be handled.**

**Relational OLAP:-**

**ROLAP is a form of OLAP that performs analysis of multidimensional data stored in a relational database rather than a multidimensional database. Greater amount of data can be processed but it requires more processing time/disk space.**

**Hybrid OLAP:-**

**HOLAP is a combination of the advantages of MOLAP and ROLAP. HOLAP can "drill through" from the cube into underlying relational data.**

**Question20:- What is OLAP is faster?**

**MOLAP.**

**Question21:- What are the operations that can be performed using OLAP cube?**

**functions which can be performed by OLAP are:-**

**ROLL - UP**

**DRILL-DOWN**

**SLICE = only 1 dimension**

**DICE = two or more dimension**

**PIVOT**

**Question22:- What is Normalization? Benefit of Normalization?**

**Normalization is the process of splitting up data into multiple tables.**

**Different normalized forms of data are 1NF,2NF, 3NF.**

**The process converting data in 1NF TO 2NF TO 3NF is called normalizing data.**

**The benefit with Normalization is that, it helps in reducing data redundancy.**

**Question23:- What is Dimensional modelling?**

**Dimensional modelling is a concept which can be used by a data warehouse designers to build their own data warehouse . This model can be stored in two types of tables - Facts and Dimensional table. Fact table has facts and measurements of the business and dimension table contains the context of measurements.**

**Types of dimensional modelling are:-**

**Conceptual modelling**

**Logical Modelling**

**Physical modelling**

**Question24:- What is SCD ?**

**SCD is short for Slowly Changing Dimensions, and it applies to the cases where records change over time.**

**3 types of SCDs :-**

**SC1 1- the new record replaces the original record**

**SCD2 - the new record is added to the existing customer dimension table**

**SCD3- the original data is modified to include new data**

**Question25:- What Schemas do data warehouse implement?**

**Star schema.**

**Snowflake schema.**

**Fact constellation schema.**

**Question26:- What is real-time Data Warehousing?**

**Real-time data warehousing captures the business data whenever it occurs.**

**When ever a business activity gets completed, the data will be available in the flow and become available for use instantly.**

**Question27:- What language is used for defining Schema Definition?**

**Data Mining Query Language(DMQL) is used for Schema Definiton.**