1. **What are fact tables?**

A fact table is a table that contains summarized numerical and historical data (facts) and a multipart index composed of foreign keys from the primary keys of related dimension tables.

1. **What are measures?**

Measures are numeric data based on columns in a fact table. They are the primary data which end users are interested in. E.g. a sales fact table may contain a profit measure which represents profit on each sale.

1. **What are aggregations?**

Aggregations are pre calculated numeric data. By calculating and storing the answers to a query before users ask for it, the query processing time can be reduced. This is key, in providing fast query performance in OLAP.

1. **what is aggregate table and aggregate fact table ... any examples of both?**

Aggregate table? contains summarized data. The materialized views are aggregated tables. for ex in sales we have only date transaction. if we want to create a report like sales by product per year. in such cases we aggregate the date? vales into week\_agg, month\_agg, quarter\_agg, year\_agg. to retrive date from this tables we use @aggrtegate function.

1. **What are non-additive facts in detail?**

A fact may be measure, metric or a dollar value. Measure and metric are non additive facts.  
  
Dollar value is additive fact. If we want to find out the amount for a particular place for a particular period of time, we can add the dollar amounts and come up with the total amount.  
  
A non additive fact, for eg measure height(s) for 'citizens by geographical location' , when we rollup 'city' data to 'state' level data we should not add heights of the citizens rather we may want to use it to derive 'count'

1. **Explain in detail about type 1, type 2(SCD), type 3?**

Type-1  
  
Most Recent Value  
  
Type-2(full History)  
  
i) Version Number  
  
ii) Flag  
  
iii) Date  
  
Type-3  
  
Current and one Perivies value

SCD'S slow change dimension  
  
there are three types  
scd1, scd2, scd3  
  
scd1:- suppose it the data got updated in the table then there r 2 methods one is to drop the table and upload the new one. but it’s a long process. here by using table compression we can update the particular data that has been modified. that is scd1  
  
scd2:- by using key generation we r going to generate the new rownum column if there r any update the next row will be updated one and row numbers will be incremented automatically by 1  
  
scd3:- i this one a extra column is added and updated information is stored in that column. if again table is update another column is added to it again...  
in this one scd2 is mostly used..

1. **Why Denormalization is promoted in Universe Designing?**

In a relational data model, for normalization purposes, some lookup tables are not merged as a single table. In a dimensional data modeling(star schema), these tables would be merged as a single table called DIMENSION table for performance and slicing data. Due to this merging of tables into one large Dimension table, it comes out of complex intermediate joins. Dimension tables are directly joined to Fact tables. Though, redundancy of data occurs in DIMENSION table, size of DIMENSION table is 15% only when compared to FACT table. So only Denormalization is promoted in Universe Designing.

1. **What is active data warehousing?**

An active data warehouse provides information that enables decision-makers within an organization to manage customer relationships nimbly, efficiently and proactively. Active data warehousing is all about integrating advanced decision support with day-to-day-even minute-to-minute-decision making in a way that increases quality of those customer touches which encourages customer loyalty and thus secure an organization's bottom line. The marketplace is coming of age as we progress from first-generation "passive" decision-support systems to current- and next-generation "active" data warehouse implementations

1. **Is OLAP databases are called decision support system ??? true/false?**

True, OLAP (online analytical processing) works by analyzing aggregated data to give final reports to top management to take action/decisions on business which is same as DSS.(Decision support system)

1. **What is snapshot?**

You can disconnect the report from the catalog to which it is attached by saving the report with a snapshot of the data. However, you must reconnect to the catalog if you want to refresh the data.

1. **What is** the difference between **Datawarehousing and BusinessIntelligence?**

Data warehousing deals with all aspects of managing the development, implementation and operation of a data warehouse or data mart including meta data management, data acquisition, data cleansing, data transformation, storage management, data distribution, data archiving, operational reporting, analytical reporting, security management, backup/recovery planning, etc.

Business intelligence, on the other hand, is a set of software tools that enable an organization to analyze measurable aspects of their business such as sales performance, profitability, operational efficiency, effectiveness of marketing campaigns, market penetration among certain customer groups, cost trends, anomalies and exceptions, etc. Typically, the term ? business intelligence? is used to encompass OLAP, data visualization, data mining and query/reporting tools. Think of the data warehouse as the back office and business intelligence as the entire business including the back office. The business needs the back office on which to function, but the back office without a business to support, makes no sense.

As explained, Data warehouse contains the data or data mart which final product of all the process like meta data management, acquire, data cleansing, transformation and load other process as mentioned above. Business intelligence is set of software which connects to data mart to do various reporting useful for the business or good running of the company. All the business decision is taken based on the data warehousing reporting. Hope this make sense.

1. **What is the difference between data warehouse and BI?**

Simply speaking, BI is the capability of analyzing the data of a data warehouse in advantage of that business. A BI tool analyzes the data of a data warehouse and to come into some business decision depending on the result of the analysis.

Business Intelligence is a collection of broad category of application programs and techniques used to querying, retrieving, reporting and analyzing the business information’s multi dimensionally.  
  
Business Intelligence is a collection of application specifications which allow the client applications to retrieve business information’s from the Data Ware House in order to make some business decisions.

1. **What is Dimensional Modeling?**

Dimensional Modeling is a design concept used by many data warehouse designers to build their data warehouse. In this design model all the data is stored in two types of tables - Facts table and Dimension table. Fact table contains the facts/measurements of the business and the dimension table contains the context of measurements ie, the dimensions on which the facts are calculated.

Dimensional Modeling is a design concept used by many data warehouse designers to build their data warehouse. In this design model all the data is stored in two types of tables - Facts table and Dimension table. Fact table contains the facts/measurements of the business and the dimension table contains the context of measurements ie, the dimensions on which the facts are calculated.

Dimensional Modeling is a design concept used by many data warehouse designers to build their data warehouse. In this design model all the data is stored in two types of tables - Facts table and Dimension table. Fact table contains the facts/measurements of the business and the dimension table contains the context of measurements ie, the dimensions on which the facts are calculated.

**I wanted to load only particular no of records suppose i have 100 records out of which i wanted to load first 10 records to target.**

Select only 10 records in your source qualifier query.  
You can use rownum function in oracle to select 10 records.

You can filter the records right in source qualifier via a query with condition as rownum<=10  
  
You can use a sequence generator and filter pair. Sequence generator would cycle the count from 1 every time and as soon as 10 reaches, filter transformation would reject the records.  
  
You can use a variable port for counting purpose and then use a filter to drop rows from flow as soon as count reaches to 10.

**If one wants to see the time dependent master data, what table is to be referred?**

Master file

**How to Add VB script to web pages?**

There are scripting languages like Javascript and Vbscript and they are designed as an extension to html language. The Web browsers like Microsoft Internet Explorer receive the scripts along with the rest of the web page document. It is the browser responsibility to parse and process the scripts. These scripts are widely used as a client side scripting languages.

**How Do I Make Remarks In My Code?**

Use the Rem keyword or a single apostrophe ('). If the Rem keyword follows code on a line, it must be separated from the code by a colon. However, when you use an apostrophe, the colon is not required. For example:  
Dim myArray() : Rem A colon is needed because it follows code on the same line  
For X = 2 to 8 ' No colon is needed for an apostrophe.  
Rem All text on this line will not be interpretted... IF Y=X Then

**What type of Indexing mechanism do we need to use for a typical datawarehouse?**

On the fact table it is best to use bitmap indexes. Dimension tables can use bitmap and/or the other types of clustered/non-clustered, unique/non-unique indexes.   
  
To my knowledge, SQLServer does not support bitmap indexes. Only Oracle supports bitmaps.

**Which columns go to the fact table and which columns go the dimension table?**

The Primary Key columns of the Tables(Entities) go to the Dimension Tables as Foreign Keys.  
  
The Primary Key columns of the Dimension Tables go to the Fact Tables as Foreign Keys.

**What is a level of Granularity of a fact table?**

Level of granularity means level of detail that you put into the fact table in a data warehouse. For example: Based on design you can decide to put the sales data in each transaction. Now, level of granularity would mean what detail are you willing to put for each transactional fact. Product sales with respect to each minute or you want to aggregate it up to minute and put that data.

**What does level of Granularity of a fact table signify?**

Granularity  
The first step in designing a fact table is to determine the granularity of the fact table. By  
granularity, we mean the lowest level of information that will be stored in the fact table. This  
constitutes two steps:   
  
Determine which dimensions will be included. Determine where along the hierarchy of each dimension the information will be kept. The determining factors usually go back to the  
requirements

**How are the Dimension tables designed?**

Most dimension tables are designed using Normalization principles upto 2NF. In some instances they are further normalized to 3NF.  
  
Find where data for this dimension are located.   
  
Figure out how to extract this data.   
  
Determine how to maintain changes to this dimension (see more on this in the next section).   
  
Change fact table and DW population routines.

**What are slowly changing dimensions?**

SCD(Slowly Changing Dimensions) stands for Slowly changing dimensions. Slowly changing dimensions are of three types  
  
SCD1: only maintained updated values.  
  
Ex: a customer address modified we update existing record with new address.  
  
SCD2: **adding a new row with a updated information (gsr)**

maintaining historical information and current information by using  
  
A) Effective Date  
  
B) Versions  
  
C) Flags  
  
or combination of these  
  
scd3: by **adding new columns** to target table we maintain historical information and current information

**What are non-additive facts?**

Non-Additive: Non-additive facts are facts that cannot be summed up for any of the dimensions present in the fact table.

**What are conformed dimensions?**

Conformed dimensions are dimensions which are common to the cubes. ( cubes are the schemas contains facts and dimension tables)   
  
Consider

Cube-1 contains F1, D1, D2, D3 and

Cube-2 contains F2, D1, D2, D4 are the Facts and Dimensions   
here D1, D2 are the Conformed Dimensions

**What is VLDB?**

VLDB stands for Very Large DataBase.  
  
It is an environment or storage space managed by a relational database management system (RDBMS) consisting of vast quantities of information.  
  
VLDB doesn’t refer to size of database or vast amount of information stored. It refers to the window of opportunity to take back up of the database.  
  
Window of opportunity refers to the time of interval and if the DBA was unable to take back up in the specified time then the database was considered as VLDB.

**What is SCD1, SCD2, SCD3?**

SCD Stands for Slowly changing dimensions.   
  
SCD1: only maintained updated values.  
  
Ex: a customer address modified we update existing record with new address.  
  
  
SCD2: maintaining historical information and current information by using  
  
A) Effective Date  
  
B) Versions  
  
C) Flags  
  
or combination of these  
  
SCD3: by adding new columns to target table we maintain historical information and current information.

**What are Semi-additive and factless facts and in which scenario will you use such kinds of fact tables?**

Snapshot facts are semi-additive, while we maintain aggregated facts we go for semi-additive.  
  
EX: Average daily balance   
  
a fact table without numeric fact columns is called fact less fact table.  
  
Ex: Promotion Facts  
  
While maintain the promotion values of the transaction (ex: product samples) because this table doesn’t contain any measures.

**What are conformed dimensions?**

Conformed dimensions mean the exact same thing with every possible fact table to which they are joined  
Ex:Date Dimensions is connected all facts like Sales facts,Inventory facts..etc

Conformed dimensions are the common elements in each datamarts, which, when combined into the data warehouse, from the overlapping glue.

Confirmed dimensions are the dimensions which are reusable and constant across the data marts within a data warehouse.

**Differences between star and snowflake schemas?**

Star schema - all dimensions will be linked directly with a fat table.  
Snow schema - dimensions maybe interlinked or may have one-to-many relationship with other tables.

star schema is less normalization, query complexity is less, transaction speed is high.  
snow flake is very good in normalization, query complexity is very high and the transaction speed is medium.

**How do you load the time dimension?**

Time dimensions are usually loaded by a program that loops through all possible dates that may appear in the data. It is not unusual for 100 years to be represented in a time dimension, with one row per day.

Time dimension are used to represent the datas or measures over a certain period of time. The server time dimension is the most widely used one by which we can represent the data’s in hierarchical manner such as quarter ->year->months->week wise representations.

**Why are OLTP database designs not generally a good idea for a Data Warehouse?**

Since in OLTP, tables are normalized and hence query response will be slow for end user and OLTP does not contain years of data and hence cannot be analyzed.

**Why should you put your data warehouse on a different system than your OLTP system?**

A OLTP system is basically " data oriented " (ER model) and not " Subject oriented "(Dimensional Model) .That is why we design a separate system that will have a subject oriented OLAP system...  
  
Moreover if a complex query is fired on an OLTP system will cause a heavy overhead on the OLTP server that will affect the day to day business directly.

The loading of a warehouse will likely consume a lot of machine resources. Additionally, users may create queries or reports that are very resource intensive because of the potentially large amount of data available. Such loads and resource needs will  
conflict with the needs of the OLTP systems for resources and will negatively impact those production systems.

**Explain the advantages of RAID 1, 1/0, and 5. What type of RAID setup would you put your TX logs.**

Transaction logs write sequentially and don't need to be read at all. The idea is to have each on RAID 1/0 because it has much better write performance than RAID 5.  
  
RAID 1 is also better for TX logs and costs less than 1/0 to implement. It has a tad less reliability and performance is a little worse generally speaking.  
  
RAID 5 is best for data generally because of cost and the fact it provides great read capability.