What is Data warehouse?

> According to Bill Inmon, known as father of Data warehousing. "A Data warehouse is a subject oriented, integrated ,time variant, non volatile collection of data in support of management's decision making process".

http://questionsdoor.com/informatica-interview-questions/

What are the types of data warehouses?

➤ There are three types of data warehouses
➤ Enterprise Data Warehouse
➤ ODS (operational data store)
➤ Data Mart

What is Data mart?

➢ A data mart is a subset of data warehouse that is designed for a particular line of business, such as sales, marketing, or finance. In a dependent data mart, data can be derived from an enterprise wide data warehouse. In an independent data mart can be collected directly from sources.

What is star schema?

➤ A star schema is the simplest form of data warehouse schema that consists of one or more dimensional and fact tables.

What is snow flake schema?

> A Snowflake schema is nothing but one Fact table which is connected to a number of dimension tables, The snowflake and star schema are methods of storing data which are multidimensional in nature.

What are ETL Tools?

➢ ETL Tools are stands for Extraction, Transformation, and Loading the data into the data warehouse for decision making. ETL refers to the methods involved in accessing and manipulating source data and loading it into target database.

What are Dimensional table?

> Dimension tables contain attributes that describe fact records in the fact table.

What is data Modelling?

- ➤ Data Modeling is representing the real world set of data structures or entities and their relationship in their of data models, required for a database.Data Modelling consists of various types like :
- ➤ Conceptual data modeling
- ➤ Logical data modeling
- ➤ Physical data modeling
- ➤ Enterprise data modeling
- ➤ Relation data modeling
- ➤ Dimensional data modeling.

What is Surrogate key?

➢ Surrogate key is a substitution for the natural primary key. It is just a unique identifier or number of each row that can be used for the primary key to the table.

What is Data Mining?

➢ A Data Mining is the process of analyzing data from different perpectives and summarizing it into useful information.

What is Operational Data Store?

➤ A ODS is an operational data store which comes as a second layer in a datawarehouse architecture. It has got the characteristics of both OLTP and DSS systems.

What is the Difference between OLTP and OLAP?

➢ OLTP is nothing but OnLine Transaction Processing which contains a normalised tables .
But OLAP(Online Analtical Programming) contains the history of OLTP data which is non-volatile acts as a Decisions Support System.

How many types of dimensions are available in Informatica?

> There are three types of dimensions available are :
> Junk dimension
> Degenerative Dimension
> Conformed Dimension

What are the designer tools for creating tranformations?

➢ Mapping designer
➢ Transformation developer
➢ Mapplet designer

What is the maplet?

➢ Maplet is a set of transformations that you build in the maplet designer and you can use in multiple mapings.

What is Session and Batches?

> **Session:** A session is a set of commands that describes the server to move data to the target.
> **Batch:** A Batch is set of tasks that may include one or more numbar of tasks (sessions, ewent wait, email, command, etc).

What are slowly changing dimensions?

- Dimensions that change overtime are called Slowly Changing Dimensions(SCD).
- Slowly Changing Dimension-Type1 : Which has only current records.
- Slowly Changing Dimension-Type2 : Which has current records + historical records.
- Slowly Changing Dimension-Type3 : Which has current records + one previous records.

What are 2 modes of data movement in Informatica Server?

➢ There are two modes of data movement are:

Normal Mode in which for every record a separate DML stmt will be prepared and executed.

Bulk Mode in which for multiple records DML stmt will be preapred and executed thus improves performance.

What is the difference between Active and Passive transformation?

➢ **Active Transformation:**An active transformation can change the number of rows that pass through it from source to target i.e it eliminates rows that do not meet the condition in transformation. **Passive Transformation:**A passive transformation does not change the number of rows that pass through it i.e it passes all rows through the transformation.

Difference between connected and unconnected transformation?

➤ **Connected Transformation:**Connected transformation is connected to other transformations or directly to target table in the mapping.
UnConnected Transformation: An unconnected transformation is not connected to other transformations in the mapping. It is called within another transformation, and returns a value to that transformation.

What are various types of Aggregation?

➤ Various types of aggregation are
➤ SUM
➤ AVG
➤ vCOUNT
➤ MAX
➤ MIN
➤ FIRST
➤ LAST
➤ MEDIAN
➤ PERCENTILE
➤ STDDEV
➤ VARIANCE

What are Aggregator Transformation?

> **Aggregator transformation** is an Active and Connected transformation. This transformation is useful to perform calculations such as averages and sums (mainly to perform calculations on multiple rows or groups).

What are Expression transformation?

➢ **Expression transformation** is a Passive and Connected transformation. This can be used to calculate values in a single row before writing to the target.

What are Filter transformation?

➢ **Filter transformation** is an Active and Connected transformation. This can be used to filter rows in a mapping that do not meet the condition.

What are Joiner transformation?

> **Joiner Transformation** is an Active and Connected transformation. This can be used to join two sources coming from two different locations or from same location.

Why we use lookup transformations?

> Lookup Transformations can access data from relational tables that are not sources in mapping.

What are Normalizer transformation?

➢ **Normalizer Transformation** is an Active and Connected transformation. It is used mainly with COBOL sources where most of the time data is stored in denormalized format. Also, Normalizer transformation can be used to create multiple rows from a single row of data.

What are Rank transformation?

➤ **Rank transformation** is an Active and Connected transformation. It is used to select the top or bottom rank of data.

What are Router transformation?

➢ **Router transformation** is an Active and Connected transformation. It is similar to filter transformation. The only difference is, filter transformation drops the data that do not meet the condition whereas router has an option to capture the data that do not meet the condition. It is useful to test multiple conditions.

What are Sorter transformation?

➤ **Sorter transformation** is a Connected and an Active transformation. It allows to sort data either in ascending or descending order according to a specified field.

Name four output files that informations server creates during session running?

- ➢ Session Log
- ➢ Workflow Log
- ➢ Errors Log
- ➢ Badfile

What is the difference between lookup cache and lookup index?

> Look up Cache contains Index cache and data cache.
> Index cache : Contains columns used in condition.

What are the difference between static cache and dynamic cache?

➤ **Dynamic cache** decreases the performance in comparision to static cache.
Static cache do not see such things just insert data as many times as it is coming

Define maping and sessions?

➢ **Maping**: It is a set of source and target definitions linked by transformation objects that define the rules for transformation. **Session** : It is a set of instructions that describe how and when to move data from source to targets.

What is a command that used to run a batch?

➢ pmcmd is used to start a batch.

What is Datadriven?

> The informatica server follows instructions coded into update strategy transformations with in the session maping determine how to flag records for insert, update, delete or reject.

What is power center repository?

➤ The PowerCenter repository allows you to share metadata across repositories to create a data mart domain.

What is parameter file?

➢ A parameter file is a file created by text editor such as word pad or notepad. U can define the following values in parameter file.
Maping parameters
Maping variables
session parameters.

What are the types of lookup caches?

- Static cache
- Dynamic cache
- Persistent cache
- Shared cache
- Recache.

What are Stored Procedure transformation?

➢ **Stored Procedure transformation** is an Passive & Connected or UnConnected transformation. It is useful to automate time-consuming tasks and it is also used in error handling, to drop and recreate indexes and to determine the space in database, a specialized calculation.

What is fact table?

> The centralized table in a star schema is called as fact table. Fact tables are three types

> additive

> non-additive

> semi additive

What transformations are used for variable port?

➤ Expression Transformation
➤ Aggregated transformation
➤ Rank Transformation

What is Data Transformation Manager(DTM)?

➢ After the load manager performs validations for the session, it creates the DTM process. The DTM process is the second process associated with the session run.

How can you define a transformation?

➤ A transformation is a repository object that generates, modifies, or passes data. The Designer provides a set of transformations that perform specific functions.

What are Lookup transformation?

➤ **Lookup transformation** is Passive and it can be both Connected and UnConnected as well. It is used to look up data in a relational table, view, or synonym. Lookup definition can be imported either from source or from target tables.

What are Source Qualifier transformation?

➢ **Source Qualifier transformation** is an Active and Connected transformation. When adding a relational or a flat file source definition to a mapping, it is must to connect it to a Source Qualifier transformation. The Source Qualifier performs the various tasks such as overriding default SQL query, filtering records; join data from two or more tables etc.

What is difference between maplet and reusable transformation?

➤ **Maplet** consists of set of transformations that is reusable.
A reusable transformation is a single transformation that can be reusable.

What are Update Strategy transformation?

➤ **Update strategy transformation** is an active and connected transformation. It is used to update data in target table, either to maintain history of data or recent changes. You can specify how to treat source rows in table, insert, update, delete or data driven.

How many types of dimensions are available in informatica?

➤ There are three types of dimensions.
➤ Star schema
➤ Snowflake schema
➤ Glaxy schema

What is difference between maplet and reusable transformation?

> **Maplet** : one or more transformations.
> set of transformations that are reusable.
> **Reusable transformation**: only one transformation.
> Single transformation which is reusable.

What are different types of parsing?

➤ Quick parsing
➤ Thorough parsing

What are Lookup and Fact Tables?

➤ A lookup (Dimension) table contains information about the entities. In general the Dimension and details objects are derived from lookup tables. A fact table contains the statistical information about transactions.

What is Designer?

> Designer is the Business objects product that is intended to develop the universes. These universe is the semantic - layer of the database structure that isolates from technical issues.
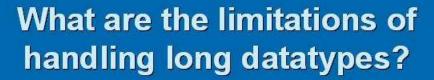
What is Surrogate Key?

➤ Surrogate keys are keys that are maintained within the data warehouse instead of keys taken from source data systems.

What are the pitfalls of DWH?

- Limited value of data (Historical data not current data)
- DW solutions complicate business processes
- DW solutions may have too long a learning curve
- Costs of cleaning, capturing and delivering data

How do you handle large datasets?

➢ By Using Bulk utility mode at the session level and if possible by disabling constraints after consulting with DBA; Using Bulk utility mode would mean that no writing is taking place in Roll Back Segment so loading is faster. However the pitfall is that recovery is not possible.

What are the types of OLAP?

- **ROLAP (Relational OLAP)** - Users see their data organized in cubes and dimensions but the data is really stored in RDBMS. The performance is slow. A storage mode that uses tables in a relational database to store multidimensional structures.
  MOLAP (Multidimensional OLAP) - Users see their data organized in cubes and dimensions but the data is really stored in MDBMS. Query performance is fast.
  HOLAP (Hybrid OLAP) - It is a combination of ROLAP and HOLAP. EG: HOLOs. In this one will find data queries on aggregated data as well as detailed data.

What is the difference between data mart and data warehouse?

➤ Data mart used on a business division/department level where as data warehouse is used on enterprise level.

What is Meta data?

➢ Data about the data, contains the location and description of data warehouse system components such as name, definitions and end user views.

How does the recovery mode work in informatica?

➤ In case of load failure an entry is made in OPB_SERV_ENTRY(?) table from where the extent of loading can be determined.

What is Aggregate Awareness?

> Aggregate awareness is a feature of DESIGNER that makes use of aggregate tables in a database. These are tables that contain pre-calculated data. The purpose of these tables is to enhance the performance of SQL transactions; they are thus used to speed up the execution of queries.

What is a difference between OLTP and OLAP?

- **OLTP**
- It focus on day to day transaction.
- Data Stability
- Dynamic
- Highly normalized.
- Access Frequency High.
- **OLAP**
- It focus on future predictions and decisions
- Static until refreshed
- Demoralized and replicated data
- Medium to low.

When should you use a star schema and when a snowflake schema?

➢ A star schema is a simplest data warehouse schema. Snowflake schema is similar to the star schema. It normalizes dimension table to save data storage space. It can be used to represent hierarchies of information.

What parameters can be tweaked to get better performance from a session?

➤ DTM shared memory, Index cache memory, Data cache memory, by indexing, using persistent cache, increasing commit interval etc.
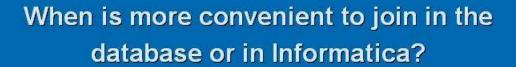
What are the benefits of DWH?

- Immediate information delivery
- Data Integration from across, even outside the organization
- Future vision of historical trends
- Tools for looking at data in new ways
- Enhanced customer service.

Is It Possible to invoke Informatica batch or session outside Informatica UI?

➢ PMCMD.

Why we are going for surrogate keys?

➤ Data tables in various source systems may use different keys for the same entity.

➤ Keys may change or be reused in the source data systems.

➤ Changes in organizational structures may move keys in the hierarchy.

When is more convenient to join in the database or in Informatica?

- Definitely at the database level
- at the source Qualifier query itself
- rather than using Joiner transformation

How do you measure session performance?

➢ By checking Collect performance Data check box.

What is Dimension Table?

> It contains data used to reference data stored in the fact table.

> Fewer rows

> Primarily character data

> One primary key (dimensional key)

> Updatable data

What is a database connection?

- A connection is a set of parameters that provides access to an RDBMS. These parameters include system information such as the data account, user identification, and the path to the database. Designer provides three types of connections: secured, shared, and personal.

What are the parts of Informatica Server?

➢ Load Manager
➢ Data Transformation manager
➢ Reader
➢ Writer

What is the functionality of repository manager?

➤ Repository manager is used to navigate through multiple folders and perform the basic repository tasks.

What is parameter file?

➢ Parameter file is to define the values for parameters and variables used in a session. A parameter file is a file created by text editor such as word pad or notepad.

What is the default source option for update strategy transformation?

➢ Data driven.

What is the difference between sequential batch and concurrent batch?

> In sequential batch one session will end then the other will begin. Where as in the concurrent. Batch the sessions will run simultaneously depending on the CPU availability.

What are the types of tracing level?

➢ Normal
➢ Verbose
➢ Verbose init
➢ Verbose data

What is update strategy transformation?

➤ This transformation is used to maintain the history data or just most recent changes in to target table.

What is power center repository?

> The Power center repository allows you to share metadata across repositories to create a data mart domain. In a data mart domain, You can create a single global repository to store metadata used across an enterprise, and a number of local repositories to share the global metadata as needed.

What are designer objects?

- Source
- Target
- Transformation
- Mapping
- Mapplet

Can you generate reports in Informatcia?

> By using Metadata reporter we can generate reports in informatica.

What is a reusable transformation?

➢ Reusable transformation: Developed in transformation developer. It can be invoked from more than one mapping or mapplet.

What are different Types of facts?

- There are three types of facts :
- Additive : Additive facts are facts that can be summed up through all of the dimensions in the fact table.
- Semi–Additive : Semi–additive facts are facts that can be summed up for some of the dimensions in the fact table, but not the others.
- Non–Additive: Non–additive facts are facts that cannot be summed up for any of the dimensions present in the fact table.

What are the types of designer tools?

➢ The Designer provides the following tools are :

➢ Source Analyzer

➢ Warehouse Designer

➢ Transformation Developer

➢ Mapplet Designer

➢ Mapping Designer

What is the subquery?

➢ Subquery is a query whose return values are used in filtering conditions of the main query.

What is tracing level?

➢ Tracing level represents the amount of information that informatcia server writes in a log file.

Why you use repository connectivity?

➤ When you edit,schedule the sesion each time,informatica server directly communicates the repository to check whether or not the session and users are valid.All the metadata of sessions and mappings will be stored in repository.

What are the options in the target session of update strategy transformation?

- Insert
- Delete
- Update
- Update as update
- Update as insert
- Update else insert
- Truncate table

What is difference between ER Modeling and Dimensional Modeling?

➢ ER Modeling is used for normalizing the OLTP database design.

➢ Dimesional modeling is used for de-normalizing the ROLAP / MOLAP design.

What is the advantage of Lookup persistent cache?

➢ When Lookup cache is saved in Look up Transformation It is called persistent cache. The first time session runs it is saved on the disk and utilized in subsequent running of the Session. It is used when the look up table is Static i.e. doesn't change frequently.

What are all the types of dimensions?

- Informational Dimension
- Structural Dimension
- Categorical Dimension
- Partitioning Dimension

Which transformation should u need while using the cobol sources as source defintions?

➢ Normalizer transformaiton which is used to normalize the data.Since cobol sources are oftenly consists of Denormailzed data.

How many ways you create ports?

➤ Drag the port from another transforamtion.
➤ Click the add buttion on the ports tab.

Types of transformations available in Informatica?

- Aggregator
- Application Source Qualifier
- Custom
- Expression
- External Procedure
- Filter
- Input
- Joiner
- Lookup
- Normalizer
- Output
- Rank
- Router
- Sequence Generator
- Sorter
- Source Qualifier
- Stored Procedure
- Transaction Control
- Union
- Update Strategy
- XML Generator
- XML Parser
- XML Source Qualifier

How can you improve session performance in aggregator transformation?

➢ Use sorted input.

Why we use stored procedure transformation?

➢ A stored procedure transformation is an important tool for populating and maintaing databases.

http://questionsdoor.com/informatica-interview-questions/