

---

**Trainer: Jagan Byri, Course : Informatica.**

*LearnersParadise Confidential*

*Sign up for a free Informatica trial class at [learnersparadise.com](http://learnersparadise.com)*

---

### Data Warehouse Interview Questions

Informatica Group URL for Real Time Problems: <http://groups.yahoo.com/group/informaticadevelopment/>  
<http://groups.yahoo.com/group/informaticadevelopment/>

Kimbel Ralph URL's:-

<http://www.dbmsmag.com/9612d05.html>

<http://www.dbmsmag.com/9701d05.html>

STAR - SCHEMA :-

<http://www.starlab.vub.ac.be/staff/robert/Information%20Systems/Halpin%203rd%20ed/Infosys%20Ch1.pdf>

ODS Design:-

<http://www.compaq.nl/products/servers/alphaserver/pdf/SPD-ODS%20Service-V1.0.pdf>

[http://www.intelligententerprise.com/010613/warehouse1\\_1.shtml?database](http://www.intelligententerprise.com/010613/warehouse1_1.shtml?database)

1.Can 2 Fact Tables share same dimensions Tables? How many Dimension tables are associated with one Fact Table ur project?

**Ans:** Yes

2.What is ROLAP, MOLAP, and DOLAP...?

**Ans:** **ROLAP** (Relational OLAP), **MOLAP** (Multidimensional OLAP), and **DOLAP** (Desktop OLAP). In these three OLAP architectures, the interface to the analytic layer is typically the same; what is quite different is how the data is physically stored.

**In MOLAP**, the premise is that online analytical processing is best implemented by storing the data multidimensionally; that is, data must be stored multidimensionally in order to be viewed in a multidimensional manner.

**In ROLAP**, architects believe to store the data in the relational model; for instance, OLAP capabilities are best provided against the relational database.

**DOLAP**, is a variation that exists to provide portability for the OLAP user. It creates multidimensional datasets that can be transferred from server to desktop, requiring only the DOLAP software to exist on the target system. This provides significant advantages to portable computer users, such as salespeople who are frequently on the road and do not have direct access to their office server.

3.What is an MDDB? and What is the difference between MDDBs and RDBMSs?

**Ans:** **Multidimensional Database** There are two primary technologies that are used for storing the data used in OLAP applications.

These two technologies are multidimensional databases (MDDB) and relational databases (RDBMS). The major difference between MDDBs and RDBMSs is in how they store data. **Relational databases** store their data in a series of tables and columns. **Multidimensional databases**, on the other hand, store their data in a large multidimensional arrays.

For **example**, in an MDDB world, you might refer to a sales figure as Sales with Date, Product, and Location coordinates of 12-1-2001, Car, and south, respectively.

#### **Advantages of MDDB:**

##### **Retrieval is very fast because**

- The data corresponding to any combination of dimension members can be retrieved with a single I/O.
- Data is clustered compactly in a multidimensional array.
- Values are calculated ahead of time.
- The index is small and can therefore usually reside completely in memory.

##### **Storage is very efficient because**

- The blocks contain only data.
- A single index locates the block corresponding to a combination of sparse dimension numbers.

---

[LearnersParadise.com](http://LearnersParadise.com) – Online Training with ‘Live’ trainer.  
2900 Gordon Avenue, Suite# 100-24, Santa Clara, CA USA 95051

---

**Trainer: Jagan Byri, Course : Informatica.**

*LearnersParadise Confidential*

*Sign up for a free Informatica trial class at [learnersparadise.com](http://learnersparadise.com)*

---

4. What is MDB modeling and RDB Modeling?

**Ans:**

5. What is Maplet and how do u create Maplet?

**Ans:** A maplet is a reusable object that represents a set of transformations. It allows you to reuse transformation logic and can contain as many transformations as you need.

Create a maplet when you want to use a standardized set of transformation logic in several mappings. For **example**, if you have a several fact tables that require a series of dimension keys, you can create a maplet containing a series of Lookup transformations to find each dimension key. You can then use the maplet in each fact table mapping, rather than recreate the same lookup logic in each mapping.

**To create a new maplet:**

1. In the Maplet Designer, choose Maplets-Create Maplet.
2. Enter a descriptive maplet name.  
The recommended naming convention for maplets is mpltMapletName.
3. Click OK.  
The Mapping Designer creates a new maplet in the Maplet Designer.
4. Choose Repository-Save.

6. What for is the transformations are used?

**Ans:** Transformations are the manipulation of data from how it appears in the source system(s) into another form in the data warehouse or mart in a way that enhances or simplifies its meaning. In short, u transform data into information.

This includes **Datamerging, Cleansing, Aggregation:** -

**Datamerging:** Process of standardizing data types and fields. Suppose one source system calls integer type data as smallint where as another calls similar data as decimal. The data from the two source systems needs to be rationalized when moved into the oracle data format called number.

**Cleansing:** This involves identifying any changing inconsistencies or inaccuracies.

- Eliminating inconsistencies in the data from multiple sources.
- Converting data from different systems into single consistent data set suitable for analysis.
- Meets a standard for establishing data elements, codes, domains, formats and naming conventions.
- Correct data errors and fills in for missing data values.

**Aggregation:** The process where by multiple detailed values are combined into a single summary value typically summation numbers representing dollars spend or units sold.

- Generate summarized data for use in aggregate fact and dimension tables.

**Data Transformation** is an interesting concept in that some transformation can occur during the "extract," some during the "transformation," or even – in limited cases--- during "load" portion of the ETL process. The type of transformation function u need will most often determine where it should be performed. Some transformation functions could even be performed in more than one place. B'z'e many of the transformations u will want to perform already exist in some form or another in more than one of the three environments (source database or application, ETL tool, or the target db).

7. What is the difference between OLTP & OLAP?

**Ans:** **OLTP** stand for Online Transaction Processing. This is standard, normalized database structure. OLTP is designed for Transactions, which means that inserts, updates, and deletes must be fast. Imagine a call center that takes orders. Call takers are continually taking calls and entering orders that may contain numerous items. Each order and each item must be inserted

---

**Trainer: Jagan Byri, Course : Informatica.**

*LearnersParadise Confidential*

*Sign up for a free Informatica trial class at [learnersparadise.com](http://learnersparadise.com)*

---

into a database. Since the performance of database is critical, we want to maximize the speed of inserts (and updates and deletes). **To maximize performance, we typically try to hold as few records in the database as possible.**

**OLAP** stands for Online Analytical Processing. OLAP is a term that means many things to many people. Here, we will use the term OLAP and Star Schema pretty much interchangeably. We will assume that **star schema database is an OLAP system.** (This is not the same thing that **Microsoft** calls OLAP; they extend OLAP to mean the cube structures built using their product, **OLAP Services**). Here, we will assume that any system of **read-only, historical, aggregated data** is an OLAP system.

A data warehouse(or mart) is way of storing data for later retrieval. This retrieval is almost always used to support decision-making in the organization. That is why many data warehouses are considered to be **DSS (Decision-Support Systems)**.

Both a data warehouse and a data mart are storage mechanisms for **read-only, historical, aggregated data**.

By **read-only**, we mean that the person looking at the data won't be changing it. If a user wants at the sales yesterday for a certain product, they should not have the ability to change that number.

The "**historical**" part may just be a few minutes old, but usually it is at least a day old. A data warehouse usually holds data that goes back a certain period in time, such as five years. In contrast, standard OLTP systems usually only hold data as long as it is "current" or active. An order table, for example, may move orders to an archive table once they have been completed, shipped, and received by the customer.

When we say that data warehouses and data marts hold **aggregated data**, we need to stress that there are many levels of **aggregation** in a typical data warehouse.

8. If data source is in the form of Excel Spread sheet then how do use?

**Ans:** PowerMart and PowerCenter treat a Microsoft Excel source as a relational database, not a flat file. Like relational sources, the Designer uses ODBC to import a Microsoft Excel source. You do not need database permissions to import Microsoft Excel sources.

**To import an Excel source definition, you need to complete the following tasks:**

- Install the Microsoft Excel ODBC driver on your system.
- Create a Microsoft Excel ODBC data source for each source file in the ODBC 32-bit Administrator.
- Prepare Microsoft Excel spreadsheets by defining ranges and formatting columns of numeric data.
- Import the source definitions in the Designer.

Once you define ranges and format cells, you can import the ranges in the Designer. Ranges display as source definitions when you import the source.

9. Which db is RDBMS and which is MDDB can u name them?

**Ans: MDDB ex.** Oracle Express Server(OES), Essbase by Hyperion Software, Powerplay by Cognos and

**RDBMS ex.** Oracle , SQL Server ...etc.

10. What are the modules/tools in Business Objects? Explain their purpose briefly?

**Ans: BO Designer**, Business Query for Excel, **BO Reporter**, Infoview, Explorer, WEBI, **BO Publisher**, and Broadcast Agent, **BO ZABO**).

**InfoView:** IT portal entry into WebIntelligence & Business Objects.

Base module required for all options to view and refresh reports.

**Reporter:** Upgrade to create/modify reports on LAN or Web.

**Explorer:** Upgrade to perform OLAP processing on LAN or Web.

---

[LearnersParadise.com](http://LearnersParadise.com) – Online Training with 'Live' trainer.  
2900 Gordon Avenue, Suite# 100-24, Santa Clara, CA USA 95051

**Trainer: Jagan Byri, Course : Informatica.**

*LearnersParadise Confidential*

*Sign up for a free Informatica trial class at [learnersparadise.com](http://learnersparadise.com)*

---

**Designer:** Creates semantic layer between user and database.

**Supervisor:** Administer and control access for group of users.

**WebIntelligence:** Integrated query, reporting, and OLAP analysis over the Web.

**Broadcast Agent:** Used to schedule, run, publish, push, and broadcast pre-built reports and spreadsheets, including event notification and response capabilities, event filtering, and calendar based notification, over the LAN, e-mail, pager, Fax, Personal Digital Assistant( PDA), Short Messaging Service(SMS), etc.

**Set Analyzer** - Applies set-based analysis to perform functions such as exclusion, intersections, unions, and overlaps visually.

**Developer Suite** – Build packaged, analytical, or customized apps.

11. What are the Ad hoc queries, Canned Queries/Reports? and How do you create them?

(Plz check this page.....C:\BObjects\Queries\Data Warehouse - About Queries.htm)

**Ans:** The data warehouse will contain two types of query. There will be **fixed queries** that are clearly defined and well understood, such as **regular reports, canned queries** (standard reports) and **common aggregations**. There will also be **ad hoc queries** that are unpredictable, both in quantity and frequency.

**Ad Hoc Query:** Ad hoc queries are the starting point for *any* analysis into a database. Any business analyst wants to know what is inside the database. He then proceeds by calculating totals, averages, maximum and minimum values for most attributes within the database. These are **unpredictable** element of a data warehouse. It is exactly that ability to run any query when desired and expect a reasonable response that makes the data warehouse worthwhile, and makes the design such a significant challenge.

The end-user access tools are capable of automatically generating the database query that answers any Question posed by the user. The user will typically pose questions in terms that they are familiar with (for example, sales by store last week); this is converted into the database query by the access tool, which is aware of the structure of information within the data warehouse.

**Canned queries:** Canned queries are predefined queries. In most instances, canned queries contain prompts that allow you to customize the query for your specific needs. For **example**, a prompt may ask you for a School, department, term, or section ID. In this instance you would enter the name of the School, department or term, and the query will retrieve the specified data from the Warehouse. You can measure resource requirements of these queries, and the results can be used for capacity planning and for database design.

The main reason for using a canned query or report rather than creating your own is that your chances of misinterpreting data or getting the wrong answer are reduced. You are assured of getting the right data and the right answer.

12. How many Fact tables and how many dimension tables did you? Which table precedes what?

**Ans:** <http://www.ciobriefings.com/whitepapers/StarSchema.asp>

13. What is the difference between STAR SCHEMA & SNOW FLAKE SCHEMA?

**Ans:** <http://www.ciobriefings.com/whitepapers/StarSchema.asp>

14. Why did you choose STAR SCHEMA only? What are the benefits of STAR SCHEMA?

**Ans:** Because it's denormalized structure, i.e., Dimension Tables are denormalized. Why to denormalize means the first (and often only) answer is : speed. OLTP structure is designed for data inserts, updates, and deletes, but not data retrieval. Therefore, we can often squeeze some speed out of it by denormalizing some of the tables and having queries go against fewer tables. These queries are faster because they perform fewer joins to retrieve the same recordset. Joins are also confusing to many End users. By denormalizing, we can present the user with a view of the data that is far easier for them to understand.

**Benefits of STAR SCHEMA:**

- Far fewer Tables.
- Designed for analysis across time.
- Simplifies joins.
- Less database space.
- Supports “drilling” in reports.
- Flexibility to meet business and technical needs.

15. How do u load the data using Informatica?

**Ans:** Using **session**.

16. (i) What is FTP? (ii) How do u connect to remote? (iii) Is there another way to use FTP without a special utility?

**Ans: (i):** The **FTP** (File Transfer Protocol) utility program is commonly used for copying files to and from other computers. These computers may be at the same site or at different sites thousands of miles apart. FTP is general protocol that works on UNIX systems as well as other non- UNIX systems.

**(ii): Remote connect commands:**

**ftp** machinename

**ex:** **ftp 129.82.45.181** or **ftp iesg**

If the remote machine has been reached successfully, FTP responds by asking for a *loginname* and *password*. When u enter ur own loginname and password for the remote machine, it returns the prompt like below

ftp>

and permits u access to ur own home directory on the remote machine. U should be able to move around in ur own directory and to copy files to and from ur local machine using the FTP interface commands.

**Note:** U can set the mode of file transfer to **ASCII** ( default and transmits seven bits per character).

Use the **ASCII** mode with any of the following:

- Raw Data (e.g. \*.dat or \*.txt, codebooks, or other plain text documents)
- SPSS Portable files.
- HTML files.

If u set mode of file transfer to **Binary** (the binary mode transmits all eight bits per byte and thus provides less chance of a transmission error and must be used to transmit files other than ASCII files).

For **example** use **binary mode** for the following types of files:

- SPSS System files
- SAS Dataset
- Graphic files (eg., \*.gif, \*.jpg, \*.bmp, etc.)
- Microsoft Office documents (\*.doc, \*.xls, etc.)

**(iii):** Yes. If u r using Windows, u can access a text-based FTP utility from a DOS prompt.

**To do this, perform the following steps:**

1. From the Start → Programs →MS-Dos Prompt
2. Enter “ftp [ftp.geocities.com](http://ftp.geocities.com).” A prompt will appear  
(or)

Enter **ftp** to get ftp prompt → ftp> →**open** hostname **ex. ftp>open ftp.geocities.com** (It connect to the specified host).

3. Enter ur yahoo! GeoCities member name.
4. enter your yahoo! GeoCities pwd.

You can now use standard FTP commands to manage the files in your Yahoo! GeoCities directory.

---

---

**Trainer: Jagan Byri, Course : Informatica.**

*LearnersParadise Confidential*

*Sign up for a free Informatica trial class at [learnersparadise.com](http://learnersparadise.com)*

---

17. What cmd is used to transfer multiple files at a time using FTP?

**Ans:** mget ==> To copy multiple files from the remote machine to the local machine. You will be prompted for a y/n answer before transferring each file **mget \*** (copies all files in the current remote directory to ur current local directory, using the same file names).

mput ==> To copy multiple files from the local machine to the remote machine.

18. What is a Filter Transformation? or what options u have in Filter Transformation?

**Ans:** The Filter transformation provides the means for **filtering records in a mapping**. You pass all the rows from a source transformation through the Filter transformation, then enter a filter condition for the transformation. All ports in a Filter transformation are input/output, and **only records that meet the condition pass through the Filter transformation**.

**Note: Discarded rows do not appear in the session log or reject files**

To maximize session performance, include the Filter transformation as close to the sources in the mapping as possible. Rather than passing records you plan to discard through the mapping, you then filter out unwanted data early in the flow of data from sources to targets.

You cannot concatenate ports from more than one transformation into the Filter transformation; the input ports for the filter must come from a single transformation. Filter transformations exist within the flow of the mapping and cannot be unconnected. **The Filter transformation does not allow setting output default values.**

19. What are default sources which will be supported by Informatica Powermart ?

**Ans :**

- Relational tables, views, and synonyms.
- Fixed-width and delimited flat files that do not contain binary data.
- COBOL files.

20. When do u create the Source Definition ? Can I use this Source Defn to any Transformation?

**Ans:** When **working with a file that contains fixed-width binary data**, you must **create the source definition**.

The Designer displays the source definition as a table, consisting of names, datatypes, and constraints. **To use a source definition in a mapping**, connect a source definition to a **Source Qualifier** or **Normalizer transformation**. The Informatica Server uses these transformations to read the source data.

21. What is Active & Passive Transformation ?

**Ans: Active and Passive Transformations**

Transformations can be active or passive. An **active transformation** can change the number of records passed through it. A **passive transformation** never changes the record count. For **example**, the Filter transformation removes rows that do not meet the filter condition defined in the transformation.

**Active transformations that might change the record count include the following:**

- Advanced External Procedure
- Aggregator
- Filter
- Joiner
- Normalizer
- Rank
- Source Qualifier

---

**Trainer: Jagan Byri, Course : Informatica.**

*LearnersParadise Confidential*

*Sign up for a free Informatica trial class at [learnersparadise.com](http://learnersparadise.com)*

---

**Note:** If you use PowerConnect to access ERP sources, the ERP Source Qualifier is also an active transformation.

/\*

**You can connect only one of these active transformations to the same transformation or target**, since the Informatica Server cannot determine how to concatenate data from different sets of records with different numbers of rows.

\*/

**Passive transformations that never change the record count include the following:**

- Lookup
- Expression
- External Procedure
- Sequence Generator
- Stored Procedure
- Update Strategy

**You can connect any number of these passive transformations, or connect one active transformation with any number of passive transformations, to the same transformation or target.**

22. What is staging Area and Work Area?

**Ans: Staging Area : -**

- Holding Tables on DW Server.
- Loaded from Extract Process
- Input for Integration/Transformation
- May function as Work Areas
- Output to a work area or Fact Table

**Work Area: -**

- Temporary Tables
- Memory

23. What is Metadata? (plz refer DATA WHING IN THE REAL WORLD BOOK page # 125)

**Ans: Defn:** "Data About Data"

Metadata contains descriptive data for end users. In a data warehouse the term **metadata** is used in a number of different situations.

**Metadata is used for:**

- Data transformation and load
- Data management
- Query management

**Data transformation and load:**

Metadata may be used during data transformation and load to describe the source data and any changes that need to be made. The advantage of storing metadata about the data being transformed is that as source data changes the changes can be captured in the metadata, and transformation programs automatically regenerated.

**For each source data field the following information is reqd:**

**Source Field:**

- Unique identifier (to avoid any confusion occurring betn 2 fields of the same anme from different sources).

---

**Trainer: Jagan Byri, Course : Informatica.**

*LearnersParadise Confidential*

*Sign up for a free Informatica trial class at [learnersparadise.com](http://learnersparadise.com)*

---

- Name (Local field name).
- Type (storage type of data, like character, integer, floating point...and so on).
- Location
  - system ( system it comes from ex. Accounting system).
  - object ( object that contains it ex. Account Table).

**The destination field needs to be described in a similar way to the source:**

**Destination:**

- Unique identifier
- Name
- Type (database data type, such as Char, Varchar, Number and so on).
- Tablename (Name of the table th field will be part of).

**The other information that needs to be stored is the transformation or transformations that need to be applied to turn the source data into the destination data:**

**Transformation:**

- Transformation (s)
  - Name
  - Language (name of the language that transformation is written in).
    - module name
    - syntax

The **Name** is the unique identifier that differentiates this from any other similar transformations.

The **Language** attribute contains the name of the language that the transformation is written in.

The other attributes are module **name** and **syntax**. Generally these will be mutually exclusive, with only one being defined. For **simple transformations** such as simple SQL functions the **syntax** will be stored. For **complex transformations** the name of the module that contains the code is stored instead.

**Data management:**

Metadata is reqd to describe the data as it resides in the data warehouse. This is needed by the warehouse manager to allow it to track and control all data movements. Every object in the database needs to be described.

**Metadata is needed for all the following:**

- Tables
  - Columns
    - name
    - type
- Indexes
  - Columns
    - name
    - type
- Views
  - Columns
    - name
    - type
- Constraints
  - name
  - type
  - table
    - columns

Aggregations, Partition information also need to be stored in Metadata( for details refer page # 30)

**Query Generation:**

Metadata is also required by the query manger to enable it to generate queries. The same metadata can be used by the Whouse manager to describe the data in the data warehouse is also reqd by the query manager.

The query mangaer will also generate metadata about the queries it has run. This metadata can be used to build a history of all quries run and generate a query profile for each user, group of users and the data warehouse as a whole.

**The metadata that is reqd for each query is:**

- query
- tables accessed
  - columns accessed
    - name
    - refence identifier
- restrictions applied
  - column name
  - table name
  - reference identifier
  - restriction
- join Criteria applied
  - .....
  - .....
- aggregate functions used
  - .....
  - .....
- group by criteria
  - .....
  - .....
- sort criteria
  - .....
  - .....
- syntax
- execution plan
- resources
  - .....
  - .....

24. What kind of Unix flavoures u r experienced?

**Ans:** Solaris 2.5 SunOs 5.5 (Operating System)

Solaris 2.6 SunOs 5.6 (Operating System)

Solaris 2.8 SunOs 5.8 (Operating System)

**AIX 4.0.3**

5.5.1 2.5.1 **May 96** sun4c, sun4m, sun4d, sun4u, x86, ppc

5.6 2.6 **Aug. 97** sun4c, sun4m, sun4d, sun4u, x86

5.7 7 **Oct. 98** sun4c, sun4m, sun4d, sun4u, x86

5.8 8 **2000** sun4m, sun4d, sun4u, x86

25. What are the tasks that are done by Informatica Server?

**Ans:**The Informatica Server performs the following tasks:

---

---

**Trainer: Jagan Byri, Course : Informatica.**

*LearnersParadise Confidential*

*Sign up for a free Informatica trial class at [learnersparadise.com](http://learnersparadise.com)*

---

- Manages the scheduling and execution of sessions and batches
- Executes sessions and batches
- Verifies permissions and privileges
- Interacts with the Server Manager and pmcmd.

The Informatica Server moves data from sources to targets based on metadata stored in a repository. For instructions on how to move and transform data, the Informatica Server reads a mapping (a type of metadata that includes transformations and source and target definitions). Each mapping uses a session to define additional information and to optionally override mapping-level options. You can group multiple sessions to run as a single unit, known as a batch.

26. What are the two programs that communicate with the Informatica Server?

**Ans:** Informatica provides **Server Manager** and **pmcmd** programs to communicate with the Informatica Server:

**Server Manager.** A client application used to create and manage sessions and batches, and to monitor and stop the Informatica Server. You can use information provided through the Server Manager to troubleshoot sessions and improve session performance.

**pmcmd.** A command-line program that allows you to start and stop sessions and batches, stop the Informatica Server, and verify if the Informatica Server is running.

27. When do you reinitialize Aggregate Cache?

**Ans:** Reinitializing the aggregate cache overwrites historical aggregate data with new aggregate data. When you reinitialize the aggregate cache, instead of using the captured changes in source tables, you typically need to use the entire source table.

**For example,** you can reinitialize the aggregate cache if the source for a session changes incrementally every day and completely changes once a month. When you receive the new monthly source, you might configure the session to reinitialize the aggregate cache, truncate the existing target, and use the new source table during the session.

**/? Note: To be clarified when server manager works for following ?/**

To reinitialize the aggregate cache:

1. In the Server Manager, open the session property sheet.
2. Click the Transformations tab.
3. Check Reinitialize Aggregate Cache.
4. Click OK three times to save your changes.
5. Run the session.

The Informatica Server creates a new aggregate cache, overwriting the existing aggregate cache.

**/? To be check for step 6 & step 7 after successful run of session... ?/**

**6. After running the session, open the property sheet again.**

**7. Click the Data tab.**

8. Clear Reinitialize Aggregate Cache.

9. Click OK.

28. (i) What is Target Load Order in Designer?

**Ans: Target Load Order:** - In the Designer, you can set the order in which the Informatica Server sends records to various target definitions in a mapping. This feature is crucial if you want to maintain referential integrity when inserting, deleting, or updating records in tables that have the primary key and foreign key constraints applied to them. The Informatica Server writes data to all the targets connected to the same Source Qualifier or Normalizer simultaneously, to maximize performance.

**Trainer: Jagan Byri, Course : Informatica.**

*LearnersParadise Confidential*

*Sign up for a free Informatica trial class at [learnersparadise.com](http://learnersparadise.com)*

28. (ii) What are the minimum conditions that you need to have so as to use Target Load Order Option in Designer?

**Ans:** You need to have **Multiple Source Qualifier transformations**.

To specify the order in which the Informatica Server sends data to targets, create one Source Qualifier or Normalizer transformation for each target within a mapping. To set the target load order, you then determine the order in which each Source Qualifier sends data to connected targets in the mapping.

When a mapping includes a **Joiner transformation**, the Informatica Server sends all records to targets connected to that **Joiner at the same time, regardless of the target load order**.

28(iii). How do you set the Target load order?

**Ans: To set the target load order:**

1. Create a mapping that contains multiple Source Qualifier transformations.
2. After you complete the mapping, choose Mappings-Target Load Plan.  
A dialog box lists all Source Qualifier transformations in the mapping, as well as the targets that receive data from each Source Qualifier.
3. Select a Source Qualifier from the list.
4. Click the Up and Down buttons to move the Source Qualifier within the load order.
5. Repeat steps 3 and 4 for any other Source Qualifiers you wish to reorder.
6. Click OK and Choose Repository-Save.

29. What can you do with Repository Manager?

**Ans:** We can do the following tasks using **Repository Manager** :-

→ **To create usernames**, you must have one of the following sets of privileges:

- Administer Repository privilege
- Super User privilege

→ **To create a user group**, you must have one of the following privileges :

- Administer Repository privilege
- Super User privilege

→ **To assign or revoke privileges**, you must have one of the following privileges..

- Administer Repository privilege
- Super User privilege

**Note:** You cannot change the privileges of the default user groups or the default repository users.

30. What can you do with Designer ?

**Ans:** The Designer client application provides **five tools to help you create mappings**:

**Source Analyzer.** Use to import or create source definitions for flat file, Cobol, ERP, and relational sources.

**Warehouse Designer.** Use to import or create target definitions.

**Transformation Developer.** Use to create reusable transformations.

**Mapplet Designer.** Use to create mapplets.

**Mapping Designer.** Use to create mappings.

**Note:** The Designer allows you to work with multiple tools at one time. You can also work in multiple folders and repositories

31. What are the different types of Tracing Levels you have in Transformations?

**Ans:** Tracing Levels in Transformations :-

<u>Level</u>	<u>Description</u>
Terse	Indicates when the Informatica Server initializes the session and its components. Summarizes session results, but not at the level of individual records.

Normal	Includes initialization information as well as error messages and notification of rejected data.
Verbose initialization	Includes all information provided with the Normal setting plus more extensive information about initializing transformations in the session.
Verbose data	Includes all information provided with the Verbose initialization setting.

**Note:** By default, the tracing level for every transformation is Normal.

To add a slight performance boost, you can also set the tracing level to Terse, writing the minimum of detail to the session log when running a session containing the transformation.

31(i). What the difference is between a database, a data warehouse and a data mart?

**Ans:** -- A **database** is an organized collection of information.

-- A **data warehouse** is a very large database with special sets of tools to extract and cleanse data from operational systems and to analyze data.

-- A **data mart** is a focused subset of a data warehouse that deals with a single area of data and is organized for quick analysis.

32. What is Data Mart, Data Warehouse and Decision Support System explain briefly?

**Ans: Data Mart:**

A **data mart** is a repository of data gathered from operational data and other sources that is designed to serve a particular community of knowledge workers. In scope, the data may derive from an enterprise-wide database or data warehouse or be more specialized. **The emphasis of a data mart** is on meeting the specific demands of a particular group of knowledge users in terms of analysis, content, presentation, and ease-of-use. Users of a data mart can expect to have data presented in terms that are familiar.

In practice, the terms data mart and data warehouse each tend to imply the presence of the other in some form. However, most writers using the term seem to agree that the **design of a data mart tends to start from an analysis of user needs** and that a **data warehouse tends to start from an analysis of what data already exists and how it can be collected in such a way that the data can later be used**. A data warehouse is a central aggregation of data (which can be distributed physically); a data mart is a data repository that may derive from a data warehouse or not and that emphasizes ease of access and usability for a particular designed purpose. In general, a data warehouse tends to be a strategic but somewhat unfinished concept; a data mart tends to be tactical and aimed at meeting an immediate need.

**Data Warehouse:**

A **data warehouse** is a central repository for all or significant parts of the data that an enterprise's various business systems collect. The term was coined by W. H. Inmon. **IBM sometimes uses the term "information warehouse."**

Typically, a **data warehouse** is housed on an enterprise mainframe server. Data from various online transaction processing (OLTP) applications and other sources is selectively extracted and organized on the data warehouse database for use by analytical applications and user queries. **Data warehousing emphasizes** the capture of data from diverse sources for useful analysis and access, but does not generally start from the point-of-view of the end user or knowledge worker who may need access to specialized, sometimes local databases. The latter idea is known as the data mart.

**data mining, Web mining, and a decision support system (DSS)** are three kinds of applications that can make use of a data warehouse.

**Decision Support System:**

A decision support system (DSS) is a computer program application that analyzes business data and presents it so that users can make business decisions more easily. It is an "informational application" (in distinction to an "operational application" that collects the data in the course of normal business operation).

**Trainer: Jagan Byri, Course : Informatica.**

*LearnersParadise Confidential*

*Sign up for a free Informatica trial class at [learnersparadise.com](http://learnersparadise.com)*

---

**Typical information that a decision support application might gather and present would be:**

Comparative sales figures between one week and the next

Projected revenue figures based on new product sales assumptions

The consequences of different decision alternatives, given past experience in a context that is described

A decision support system may present information graphically and may include an expert system or artificial intelligence (AI). It may be aimed at business executives or some other group of knowledge workers.

33. What r the differences between Heterogeneous and Homogeneous?

**Ans: Heterogeneous**

Stored in different Schemas

Stored in different file or db types

Spread across in several countries

Different platform n H/W config.

**Homogeneous**

Common structure

Same database type

Same data center

Same platform and H/Ware configuration.

34. How do you use DDL commands in PL/SQL block **ex.** Accept table name from user and drop it, if available else display msg?

**Ans:** To invoke DDL commands in PL/SQL blocks we have to use **Dynamic SQL**, the **Package** used is **DBMS\_SQL**.

35. What r the steps to work with Dynamic SQL?

**Ans:** Open a Dynamic cursor, Parse SQL stmt, Bind i/p variables (if any), Execute SQL stmt of Dynamic Cursor and Close the Cursor.

36. Which package, procedure is used to find/check free space available for db objects like table/procedures/views/synonyms...etc?

**Ans:** The Package → is **DBMS\_SPACE**

The Procedure → is **UNUSED\_SPACE**

The Table → is **DBA\_OBJECTS**

**Note:** See the script to find free space @ c:\informatica\tbl\_free\_space

37. Does informatica allow if Empld is PKey in Target tbl and source data is 2 rows with same Empld?If u use lookup for the same situation does it allow to load 2 rows or only 1?

**Ans:** => No, it will not it generates pkey constraint violation. (it loads 1 row)  
=> Even then no if Empld is Pkey.

38. If Ename varchar2(40) from 1 source(siebel), Ename char(100) from another source (oracle) and the target is having Name varchar2(50) then how does informatica handles this situation? **How Informatica handles string and numbers datatypes sources?**

39. How do u debug mappings? I mean where do u attack?

40. How do u qry the Metadata tables for Informatica?

41(i). When do u use connected lookup n when do u use unconnected lookup?

**Ans:**

**Trainer: Jagan Byri, Course : Informatica.**

*LearnersParadise Confidential*

*Sign up for a free Informatica trial class at [learnersparadise.com](http://learnersparadise.com)*

**Connected Lookups :** -

A connected Lookup transformation is part of the mapping data flow. With connected lookups, you can have multiple return values. That is, you can pass multiple values from the same row in the lookup table out of the Lookup transformation.

**Common uses for connected lookups include:**

- => Finding a name based on a number **ex.** Finding a Dname based on deptno
- => Finding a value based on a range of dates
- => Finding a value based on multiple conditions

**Unconnected Lookups :** -

An unconnected Lookup transformation exists separate from the data flow in the mapping. You write an expression using the :LKP reference qualifier to call the lookup within another transformation.

**Some common uses for unconnected lookups include:**

- => Testing the results of a lookup in an expression
- => Filtering records based on the lookup results
- => Marking records for update based on the result of a lookup (for example, updating slowly changing dimension tables)
- => Calling the same lookup multiple times in one mapping

41(ii). What r the differences between Connected lookups and Unconnected lookups?

**Ans: Although both types of lookups perform the same basic task, there are some important differences:**

Connected Lookup	Unconnected Lookup
Part of the mapping data flow.	Separate from the mapping data flow.
Can return multiple values from the same row.	Returns one value from each row.
You link the lookup/output ports to another transformation.	You designate the return value with the Return port (R).
Supports default values.	Does not support default values.
If there's no match for the lookup condition, the server returns the <b>default value</b> for all output ports.	If there's no match for the lookup condition, the server returns <b>NULL</b> .
<b>More visible.</b> Shows the data passing in and out of the lookup.	<b>Less visible.</b> You write an expression using :LKP to tell the server when to perform the lookup.
Cache includes all lookup columns used in the mapping (that is, lookup table columns included in the lookup condition and lookup table columns linked as output ports to other transformations).	Cache includes lookup/output ports in the Lookup condition and lookup/return port.

42. What u need concentrate after getting explain plan?

**Ans:** The 3 most significant columns in the plan table are named OPERATION,OPTIONS, and OBJECT\_NAME.For each step, these tell u which operation is going to be performed and which object is the target of that operation.

Ex:-

```
*****
TO USE EXPLAIN PLAN FOR A QRY...
*****
SQL> EXPLAIN PLAN
2 SET STATEMENT_ID = 'PKAR02'
```

```
3 FOR
4 SELECT JOB,MAX(SAL)
5 FROM EMP
6 GROUP BY JOB
7 HAVING MAX(SAL) >= 5000;
```

Explained.

\*\*\*\*\*

TO QUERY THE PLAN TABLE :-

\*\*\*\*\*

```
SQL> SELECT RTRIM(ID)||' '||
2   LPAD(' ', 2*(LEVEL-1))||OPERATION
3   ||' '||OPTIONS
4   ||' '||OBJECT_NAME STEP_DESCRIPTION
5 FROM PLAN_TABLE
6 START WITH ID = 0 AND STATEMENT_ID = 'PKAR02'
7 CONNECT BY PRIOR ID = PARENT_ID
8 AND STATEMENT_ID = 'PKAR02'
9 ORDER BY ID;
```

STEP\_DESCRIPTION

```
-----
0 SELECT STATEMENT
1 FILTER
2 SORT GROUP BY
3 TABLE ACCESS FULL EMP
```

43. How components are interfaced in Psoft?

**Ans:**

44. How do u do the analysis of an ETL?

**Ans:**

45. What is Standard, Reusable Transformation and Mapplet?

**Ans:** Mappings contain two types of transformations, **standard** and **reusable**. Standard transformations exist within a single mapping. You cannot reuse a standard transformation you created in another mapping, nor can you create a shortcut to that transformation. However, often you want to create transformations that perform common tasks, such as calculating the average salary in a department. Since a standard transformation cannot be used by more than one mapping, you have to set up the same transformation each time you want to calculate the average salary in a department.

**Trainer: Jagan Byri, Course : Informatica.**

*LearnersParadise Confidential*

*Sign up for a free Informatica trial class at [learnersparadise.com](http://learnersparadise.com)*

**Mapplet:** A mapplet is a reusable object that represents a set of transformations. It allows you to reuse transformation logic and can contain as many transformations as you need. A mapplet can contain transformations, reusable transformations, and shortcuts to transformations.

46. How do you copy Mapping, Repository, Sessions?

**Ans:** To copy an object (such as a mapping or reusable transformation) from a shared folder, press the Ctrl key and drag and drop the mapping into the destination folder.

To **copy a mapping** from a non-shared folder, drag and drop the mapping into the destination folder.

In both cases, the destination folder must be open with the related tool active.

For **example**, to copy a mapping, the Mapping Designer must be active. To copy a Source Definition, the Source Analyzer must be active.

**Copying Mapping:**

- To copy the mapping, open a workbook.
- In the Navigator, click and drag the mapping slightly to the right, not dragging it to the workbook.
- When asked if you want to make a copy, click Yes, then enter a new name and click OK.
- Choose Repository-Save.

**Repository Copying:** You can copy a repository from one database to another. You use this feature before upgrading, to preserve the original repository. Copying repositories provides a quick way to copy all metadata you want to use as a basis for a new repository.

If the database into which you plan to copy the repository contains an existing repository, the Repository Manager deletes the existing repository. If you want to preserve the old repository, cancel the copy. Then back up the existing repository before copying the new repository.

To **copy a repository**, you must have one of the following privileges:

- Administer Repository privilege
- Super User privilege

**To copy a repository:**

1. In the Repository Manager, choose Repository-Copy Repository.
2. Select a repository you wish to copy, then enter the following information:

Copy Repository Field	Required/ Optional	Description
Repository	Required	Name for the repository copy. Each repository name must be unique within the domain and should be easily distinguished from all other repositories.
Database Username	Required	Username required to connect to the database. This login must have the appropriate database permissions to create the repository.
Database Password	Required	Password associated with the database username. Must be in US-ASCII.
ODBC Data Source	Required	Data source used to connect to the database.
Native Connect String	Required	Connect string identifying the location of the database.
Code Page	Required	Character set associated with the repository. Must be a superset of the code page of the repository you want to copy.

**If you are not connected to the repository you want to copy, the Repository Manager asks you to log in.**

[LearnersParadise.com](http://LearnersParadise.com) – Online Training with ‘Live’ trainer.  
2900 Gordon Avenue, Suite# 100-24, Santa Clara, CA USA 95051

---

**Trainer: Jagan Byri, Course : Informatica.**

*LearnersParadise Confidential*

*Sign up for a free Informatica trial class at [learnersparadise.com](http://learnersparadise.com)*

---

3. Click OK.
5. If asked whether you want to delete an existing repository data in the second repository, click OK to delete it. Click Cancel to preserve the existing repository.

**Copying Sessions:**

In the Server Manager, you can copy stand-alone sessions within a folder, or copy sessions in and out of batches.

**To copy a session, you must have one of the following:**

- Create Sessions and Batches privilege with read and write permission
- Super User privilege

**To copy a session:**

1. In the Server Manager, select the session you wish to copy.
2. Click the Copy Session button or choose Operations-Copy Session.

The Server Manager makes a copy of the session. The Informatica Server names the copy after the original session, appending a number, such as session\_name1.

47. What are shortcuts, and what is advantage?

**Ans:** Shortcuts allow you to use metadata across folders without making copies, ensuring uniform metadata. A shortcut inherits all properties of the object to which it points. Once you create a shortcut, you can configure the shortcut name and description.

When the object the shortcut references changes, the shortcut inherits those changes. By using a shortcut instead of a copy, you ensure each use of the shortcut exactly matches the original object. **For example**, if you have a shortcut to a target definition, and you add a column to the definition, the shortcut automatically inherits the additional column.

Shortcuts allow you to reuse an object without creating multiple objects in the repository. **For example**, you use a source definition in ten mappings in ten different folders. Instead of creating 10 copies of the same source definition, one in each folder, you can create 10 shortcuts to the original source definition.

You can create shortcuts to objects in shared folders. If you try to create a shortcut to a non-shared folder, the Designer creates a copy of the object instead.

**You can create shortcuts to the following repository objects:**

- Source definitions
- Reusable transformations
- Mapplets
- Mappings
- Target definitions
- Business components

**You can create two types of shortcuts:**

**Local shortcut.** A shortcut created in the same repository as the original object.

**Global shortcut.** A shortcut created in a local repository that references an object in a global repository.

**Advantages:** One of the primary advantages of using a shortcut is maintenance. If you need to change all instances of an object, you can edit the original repository object. All shortcuts accessing the object automatically inherit the changes.

**Shortcuts have the following advantages over copied repository objects:**

- You can maintain a common repository object in a single location. If you need to edit the object, all shortcuts immediately inherit the changes you make.

---

**Trainer: Jagan Byri, Course : Informatica.**

*LearnersParadise Confidential*

*Sign up for a free Informatica trial class at [learnersparadise.com](http://learnersparadise.com)*

---

- You can restrict repository users to a set of predefined metadata by asking users to incorporate the shortcuts into their work instead of developing repository objects independently.
- You can develop complex mappings, maplets, or reusable transformations, then reuse them easily in other folders.
- You can save space in your repository by keeping a single repository object and using shortcuts to that object, instead of creating copies of the object in multiple folders or multiple repositories.

48. What are Pre-session and Post-session Options?

(Plzz refer Help Using Shell Commands n Post-Session Commands and Email)

**Ans:** The Informatica Server can perform one or more shell commands before or after the session runs. Shell commands are operating system commands. You can use pre- or post- session shell commands, for **example**, to delete a reject file or session log, or to archive target files before the session begins.

**The status of the shell command, whether it completed successfully or failed, appears in the session log file.**

**To call a pre- or post-session shell command you must:**

1. Use any valid UNIX command or shell script for UNIX servers, or any valid DOS or batch file for Windows NT servers.
2. Configure the session to execute the pre- or post-session shell commands.

You can configure a session to stop if the Informatica Server encounters an error while executing pre-session shell commands.

**For example**, you might use a shell command to copy a file from one directory to another. For a **Windows NT server** you would use the following shell command to copy the SALES\_ADJ file from the target directory, L, to the source, H:  
copy L:\sales\sales\_adj H:\marketing\

For a **UNIX server**, you would use the following command line to perform a similar operation:  
cp sales/sales\_adj marketing/

**Tip:** Each shell command runs in the same environment (UNIX or Windows NT) as the Informatica Server. Environment settings in one shell command script do not carry over to other scripts. To run all shell commands in the same environment, call a single shell script that in turn invokes other scripts.

49. What are Folder Versions?

**Ans:** In the **Repository Manager**, you can create different versions within a folder to help you archive work in development. You can copy versions to other folders as well. When you save a version, you save all metadata at a particular point in development. Later versions contain new or modified metadata, reflecting work that you have completed since the last version.

Maintaining different versions lets you revert to earlier work when needed. By archiving the contents of a folder into a version each time you reach a development landmark, you can access those versions if later edits prove unsuccessful.

You create a folder version after completing a version of a difficult mapping, then continue working on the mapping. If you are unhappy with the results of subsequent work, you can revert to the previous version, then create a new version to continue development. Thus you keep the landmark version intact, but available for regression.

**Note:** You can only work within one version of a folder at a time.

50. How do automate/schedule sessions/batches n did u use any tool for automating Sessions/batch?

---

[LearnersParadise.com](http://LearnersParadise.com) – Online Training with ‘Live’ trainer.  
2900 Gordon Avenue, Suite# 100-24, Santa Clara, CA USA 95051



**Trainer: Jagan Byri, Course : Informatica.**

*LearnersParadise Confidential*

*Sign up for a free Informatica trial class at [learnersparadise.com](http://learnersparadise.com)*

**Ans:** We scheduled our sessions/batches using Server Manager.

You can either schedule a session to run at a given time or interval, or you can manually start the session.

U needto hv create sessions n batches with Read n Execute permissions or super user privilege.

**If you configure a batch to run only on demand, you cannot schedule it.**

**Note:** We did not use any tool for automation process.

51. What are the differences between 4.7 and 5.1 versions?

**Ans:** New Transformations added like XML Transformation and MQ Series Transformation, and PowerMart and PowerCenter both are same from 5.1version.

52. What r the procedure that u need to undergo before moving Mappings/sessions from Testing/Development to Production?

**Ans:**

53. How many values it (informatica server) returns when it passes thru Connected Lookup n Unconncted Lookup?

**Ans:** Connected Lookup can return multiple values where as Unconnected Lookup will return only one values that is Return Value.

54. What is the difference between PowerMart and PowerCenter in 4.7.2?

**Ans: If You Are Using PowerCenter**

PowerCenter allows you to register and run multiple Informatica Servers against the same repository. Because you can run these servers at the same time, you can distribute the repository session load across available servers to improve overall performance.

With **PowerCenter**, you receive all product functionality, including distributed metadata, the ability to organize repositories into a data mart domain and share metadata across repositories.

A PowerCenter license lets you create a single repository that you can configure as a global repository, the core component of a data warehouse.

**If You Are Using PowerMart**

This version of PowerMart includes all features except distributed metadata and multiple registered servers. Also, the various options available with PowerCenter (such as PowerCenter Integration Server for BW, PowerConnect for IBM DB2, PowerConnect for SAP R/3, and PowerConnect for PeopleSoft) are not available with PowerMart.

55. What kind of modifications u can do/perform with each Transformation?

**Ans: Using transformations, you can modify data in the following ways:**

<b>Task</b>	<b>Transformation</b>
Calculate a value	Expression
Perform an aggregate calculations	Aggregator
Modify text	Expression
Filter records	Filter, Source Qualifier
Order records queried by the Informatica Server	Source Qualifier
Call a stored procedure	Stored Procedure
Call a procedure in a shared library or in the COM layer of Windows NT	External Procedure

[LearnersParadise.com](http://LearnersParadise.com) – Online Training with ‘Live’ trainer.  
2900 Gordon Avenue, Suite# 100-24, Santa Clara, CA USA 95051



**Trainer: Jagan Byri, Course : Informatica.**

*LearnersParadise Confidential*

*Sign up for a free Informatica trial class at [learnersparadise.com](http://learnersparadise.com)*

Generate primary keys	Sequence Generator
Limit records to a top or bottom range	Rank
Normalize records, including those read from COBOL sources	Normalizer
Look up values	Lookup
Determine whether to insert, delete, update, or reject records	Update Strategy
Join records from different databases or flat file systems	Joiner

56. Expressions in Transformations, Explain briefly how do u use?

**Ans:** Expressions in Transformations

To transform data passing through a transformation, you can write an expression. The most obvious examples of these are the **Expression** and **Aggregator transformations**, which perform calculations on either single values or an entire range of values within a port. Transformations that use expressions include the following:

Transformation	How It Uses Expressions
Expression	Calculates the result of an expression for each row passing through the transformation, using values from one or more ports.
Aggregator	Calculates the result of an aggregate expression, such as a sum or average, based on all data passing through a port or on groups within that data.
Filter	Filters records based on a condition you enter using an expression.
Rank	Filters the top or bottom range of records, based on a condition you enter using an expression.
Update Strategy	Assigns a numeric code to each record based on an expression, indicating whether the Informatica Server should use the information in the record to insert, delete, or update the target.

In each transformation, you use the Expression Editor to enter the expression. The Expression Editor supports the transformation language for building expressions. The transformation language uses SQL-like functions, operators, and other components to build the expression. For **example**, as in SQL, the transformation language includes the functions COUNT and SUM. However, the PowerMart/PowerCenter transformation language includes additional functions not found in SQL.

When you enter the expression, you can use values available through ports. For **example**, if the transformation has two input ports representing a price and sales tax rate, you can calculate the final sales tax using these two values. The ports used in the expression can appear in the same transformation, or you can use output ports in other transformations.

57. In case of Flat files (which comes thru FTP as source) has not arrived then what happens?Where do u set this option?

**Ans:** U get an fatal error which cause server to **fail/stop the session**.

U can set Event-Based Scheduling Option in Session Properties under General tab-->Advanced options..

Event-Based	Required/ Optional	Description
Indicator File to Wait For	Optional	Required to use <b>event-based scheduling</b> . Enter the indicator file (or directory and file) <b>whose arrival schedules the session</b> . If you do not enter a directory, the Informatica Server assumes the file appears

[LearnersParadise.com](http://LearnersParadise.com) – Online Training with ‘Live’ trainer.  
2900 Gordon Avenue, Suite# 100-24, Santa Clara, CA USA 95051

---

**Trainer: Jagan Byri, Course : Informatica.**

*LearnersParadise Confidential*

*Sign up for a free Informatica trial class at [learnersparadise.com](http://learnersparadise.com)*

---

in the server variable directory \$PMRootDir.

58. What is the Test Load Option and when you use in Server Manager?

**Ans:** When **testing sessions in development**, you may not need to process the entire source. If this is true, use the Test Load Option (Session Properties → General Tab → Target Options → Choose Target Load options as Normal (option button), with Test Load checked (Check box) and No. of rows to test ex. 2000 (Text box with Scrolls)). You can also click the Start button.

---

59. SCD Type 2 and SGT difference?

60. Differences between 4.7 and 5.1?

61. Tuning Informatica Server for improving performance? Performance Issues?

**Ans:** See /\* C:\pkar\Informatica\Performance Issues.doc \*/

62. What is Override Option? Which is better?

63. What will happen if u increase buffer size?

64. what will happen if u increase commit Intervals? and also decrease commit Intervals?

65. What kind of Complex mapping u did? And what sort of problems u faced?

66. If u have 10 mappings designed and u need to implement some changes (may be in existing mapping or new mapping need to be designed) then how much time it takes from easier to complex?

67. Can u refresh Repository in 4.7 and 5.1? and also can u refresh pieces (partially) of repository in 4.7 and 5.1?

68. What is BI?

**Ans:** <http://www.visionnet.com/bi/index.shtml>

69. Benefits of BI?

**Ans:** <http://www.visionnet.com/bi/bi-benefits.shtml>

70. BI Faq

**Ans:** <http://www.visionnet.com/bi/bi-faq.shtml>

71. What is difference between data scrubbing and data cleansing?

**Ans:** Scrubbing data is the process of cleaning up the junk in legacy data and making it accurate and useful for the next generations of automated systems. This is perhaps the most difficult of all conversion activities. Very often, this is made more difficult when the customer wants to make good data out of bad data. This is the dog work. It is also the most important and can not be done without the active participation of the user.

**DATA CLEANING** - a two step process including **DETECTION** and then **CORRECTION** of errors in a data set

72. What is Metadata and Repository?

**Ans:**

**Metadata.** "Data about data" .

---

[LearnersParadise.com](http://LearnersParadise.com) – Online Training with 'Live' trainer.  
2900 Gordon Avenue, Suite# 100-24, Santa Clara, CA USA 95051



---

**Trainer: Jagan Byri, Course : Informatica.**

*LearnersParadise Confidential*

*Sign up for a free Informatica trial class at [learnersparadise.com](http://learnersparadise.com)*

---

It contains descriptive data for end users.  
Contains data that controls the ETL processing.  
Contains data about the current state of the data warehouse.  
ETL updates metadata, to provide the most current state.

**Repository.** The place where you store the metadata is called a repository. The more sophisticated your repository, the more complex and detailed metadata you can store in it. PowerMart and PowerCenter use a relational database as the repository.

73. SQL \* LOADER?

**Ans:** [http://download-west.oracle.com/otndoc/oracle9i/901\\_doc/server.901/a90192/ch03.htm#1004678](http://download-west.oracle.com/otndoc/oracle9i/901_doc/server.901/a90192/ch03.htm#1004678)

74. Debugger in Mapping?

75. Parameters passing in 5.1 version exposure?

76. What is the filename which u need to configure in Unix while Installing Informatica?

77. How do u select duplicate rows using Informatica i.e., how do u use Max(Rowid)/Min(Rowid) in Informatica?

---

[LearnersParadise.com](http://LearnersParadise.com) – Online Training with ‘Live’ trainer.  
2900 Gordon Avenue, Suite# 100-24, Santa Clara, CA USA 95051

Disclaimer: LearnersParadise is only distributor of the document. The content of the document is solely responsibility of the author.