Informatica Interview Question Collections

Description:

    BISP is committed to provide BEST learning material to the beginners and advance learners. In the same series, we have prepared a list of Informatica Interview Questions. Join our professional training program and learn from experts.

History:

Version	Description Change	Author	Publish Date
0.1	Initial Draft	Kuldeep	12th Aug 2011
0.1	Review#1	Amit Sharma	18th Aug 2011

1. What are the components of Informatica? And what is the purpose of each?

Ans: Informatica Designer, Server Manager & Repository Manager. Designer for Creating Source & Target definitions, Creating Mapplets and Mappings etc. Server Manager for creating sessions & batches, Scheduling the sessions & batches, Monitoring the triggered sessions and batches, giving post and pre session commands, creating database connections to various instances etc. Repository Manage for Creating and Adding repositories, Creating & editing folders within a repository, Establishing users, groups, privileges & folder permissions, Copy, delete, backup a repository, Viewing the history of sessions, Viewing the locks on various objects and removing those locks etc.

2. What is a repository? And how to add it in an informatica client?

Ans: It's a location where all the mappings and sessions related information is stored. Basically it's a database where the metadata resides. We can add a repository through the Repository manager.

3. Name at least 5 different types of transformations used in mapping design and state the use of each.

Ans: Source Qualifier – Source Qualifier represents all data queries from the source, Expression – Expression performs simple calculations,
Filter – Filter serves as a conditional filter,
Lookup – Lookup looks up values and passes to other objects,
Aggregator - Aggregator performs aggregate calculations.

4. How can a transformation be made reusable?

Ans: In the edit properties of any transformation there is a check box to make it reusable, by checking that it becomes reusable. You can even create reusable transformations in Transformation developer.

5. How are the sources and targets definitions imported in informatica designer? How to create Target definition for flat files?

Ans: When you are in source analyzer there is a option in main menu to Import the source from Database, Flat File, Cobol File & XML file, by selecting any one of them you can import a source definition. When you are in Warehouse Designer there is an option in main menu to import the target from Database, XML from File and XML from sources you can select any one of these.
There is no way to import target definition as file in Informatica designer. So while creating the target definition for a file in the warehouse designer it is created considering it as a table, and then in the session properties of that mapping it is specified as file.

6. Explain what is sql override for a source table in a mapping.
Ans: The Source Qualifier provides the SQL Query option to override the default query. You can enter any SQL statement supported by your source database. You might enter your own SELECT statement, or have the database perform aggregate calculations, or call a stored procedure or stored function to read the data and perform some tasks.
7. What is lookup override?

Ans: This feature is similar to entering a custom query in a Source Qualifier transformation. When entering a Lookup SQL Override, you can enter the entire override, or generate and edit the default SQL statement.
The lookup query override can include WHERE clause.

8. What are mapplets? How is it different from a Reusable Transformation?
Ans: 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. You create mapplets in the Mapplet Designer.
Its different than a reusable transformation as it may contain a set of transformations, while a reusable transformation is a single one.
9. How to use an oracle sequence generator in a mapping?

Ans: We have to write a stored procedure, which can take the sequence name as input and dynamically generates a nextval from that sequence. Then in the mapping we can use that stored procedure through a procedure transformation.

10. What is a session and how to create it?
Ans: A session is a set of instructions that tells the Informatica Server how and when to move data from sources to targets. You create and maintain sessions in the Server Manager.

11. How to create the source and target database connections in server manager?

Ans: In the main menu of server manager there is menu "Server Configuration", in that there is the menu "Database connections". From here you can create the Source and Target database connections.

12. Where are the source flat files kept before running the session?

Ans: The source flat files can be kept in some folder on the Informatica server or any other machine, which is in its domain.

13. What are the oracle DML commands possible through an update strategy?

Ans: dd_insert, dd_update, dd_delete & dd_reject.

14. How to update or delete the rows in a target, which do not have key fields?

Ans: To Update a table that does not have any Keys we can do a SQL Override of the Target Transformation by specifying the WHERE conditions explicitly. Delete cannot be done this way. In this case you have to specifically mention the Key for Target table definition on the Target transformation in the Warehouse Designer and   delete the row using the Update Strategy transformation.

15. What is option by which we can run all the sessions in a batch simultaneously?
Ans: In the batch edit box there is an option called concurrent. By checking that all the sessions in that Batch will run concurrently.

16. Informatica settings are available in which file?
Ans: Informatica settings are available in a file **pmdesign.ini**   in Windows folder.

17. How can we join the records from two heterogeneous sources in a mapping?
Ans: By using a joiner.

18. Difference between Connected & Unconnected look-up.

Ans: An unconnected Lookup transformation exists separate from the pipeline in the mapping. You write an expression using the :LKP reference qualifier to call the lookup within another transformation. While the connected lookup forms a part of the whole flow of mapping.

19. Difference between Lookup Transformation & Unconnected Stored Procedure Transformation – Which one is faster ?

20. Compare Router Vs Filter & Source Qualifier Vs Joiner.

Ans: A Router transformation has input ports and output ports. Input ports reside in the input group, and output ports reside in the output groups. Here you can test data based on one or more group filter conditions.
But in filter you can filter data based on one or more conditions before writing it to targets.

A source qualifier can join data coming from same source database. While a joiner is used to combine data from heterogeneous sources. It can even join data from two tables from same database.

A source qualifier can join more than two sources. But a joiner can join only two sources.

21. **How to Join 2 tables connected to a Source Qualifier w/o having any relationship defined ?**

Ans: By writing an sql override.

22. **In a mapping there are 2 targets to load header and detail, how to ensure that header loads first then detail table.**

Ans: Constraint Based Loading (if no relationship at oracle level) OR Target Load Plan (if only 1 source qualifier for both tables) OR select first the header target table and then the detail table while dragging them in mapping.

23. **A mapping just take 10 seconds to run, it takes a source file and insert into target, but before that there is a Stored Procedure transformation which takes around 5 minutes to run and gives output 'Y' or 'N'. If Y then continue feed or else stop the feed. (Hint: since SP transformation takes more time compared to the mapping, it shouldn't run row wise).**

Ans: There is an option to run the stored procedure before starting to load the rows.

Data warehousing concepts

1.What is difference between view and materialized view?

Views contains query whenever execute views it has read from base table
Where as M views loading or replicated takes place only once, which gives you better query performance

Refresh m views 1.on commit and 2. on demand
(Complete, never, fast, force)

2.What is bitmap index why it's used for DWH?

A bitmap for each key value replaces a list of rowids. Bitmap index more efficient for data warehousing because low cardinality, low updates, very efficient for where class

3.What is star schema? And what is snowflake schema?

The center of the star consists of a large fact table and the points of the star are the dimension tables.

Snowflake schemas normalized dimension tables to eliminate redundancy. That is, the
Dimension data has been grouped into multiple tables instead of one large table.

Star schema contains demoralized dimension tables and fact table, each primary key values in dimension table associated with foreign key of fact tables.
Here a fact table contains all business measures (normally numeric data) and foreign key values, and dimension tables has details about the subject area.

Snowflake schema basically a normalized dimension tables to reduce redundancy in the dimension tables

4.Why need staging area database for DWH?

Staging area needs to clean operational data before loading into data warehouse.
Cleaning in the sense your merging data which comes from different source

5.What are the steps to create a database in manually?

create os service and create init file and start data base no mount stage then give create data base command.

6.Difference between OLTP and DWH?

OLTP system is basically application orientation (eg, purchase order it is functionality of an application)
Where as in DWH concern is subject orient (subject in the sense custorer, product, item, time)

OLTP

· Application Oriented
· Used to run business
· Detailed data
· Current up to date
· Isolated Data
· Repetitive access
· Clerical User
· Performance Sensitive
· Few Records accessed at a time (tens)
· Read/Update Access
· No data redundancy
· Database Size 100MB-100 GB

DWH

· Subject Oriented
· Used to analyze business
· Summarized and refined
· Snapshot data
· Integrated Data
· Ad-hoc access
· Knowledge User
· Performance relaxed
· Large volumes accessed at a time(millions)
· Mostly Read (Batch Update)
· Redundancy present
· Database Size 100 GB - few terabytes

7.Why need data warehouse?

A single, complete and consistent store of data obtained from a variety of different sources made available to end users in a what they can understand and use in a business context.

A process of transforming data into information and making it available to users in a timely enough manner to make a difference Information

Technique for assembling and managing data from various sources for the purpose of answering business questions. Thus making decisions that were not previous possible

8.What is difference between data mart and data warehouse?

A data mart designed for a particular line of business, such as sales, marketing, or finance.

Where as data warehouse is enterprise-wide/organizational

The data flow of data warehouse depending on the approach

9.What is the significance of surrogate key?

Surrogate key used in slowly changing dimension table to track old and new values and it's derived from primary key.

10.What is slowly changing dimension. What kind of scd used in your project?

Dimension attribute values may change constantly over the time. (Say for example customer dimension has customer_id,name, and address) customer address may change over time.
How will you handle this situation?
There are 3 types, one is we can overwrite the existing record, second one is create additional new record at the time of change with the new attribute values.
Third one is create new field to keep new values in the original dimension table.

11.What is difference between primary key and unique key constraints?

Primary key maintains uniqueness and not null values
Where as unique constrains maintain unique values and null values

12.What are the types of index? And is the type of index used in your project?

Bitmap index, B-tree index, Function based index, reverse key and composite index.
We used Bitmap index in our project for better performance.

13.How is your DWH data modeling(Details about star schema)?

14.A table have 3 partitions but I want to update in 3rd partitions how will you do?

Specify partition name in the update statement. Say for example
Update employee partition(name) a, set a.empno=10 where ename='Ashok'

15.When you give an update statement how memory flow will happen and how oracles allocate memory for that?

Oracle first checks in Shared sql area whether same Sql statement is available if it is there it uses. Otherwise allocate memory in shared sql area and then create run time memory in Private sql area to create parse tree and execution plan. Once it completed stored in the shared sql area wherein previously allocated memory

16.Write a query to find out 5th max salary? In Oracle, DB2, SQL Server

Select (list the columns you want) from (select salary from employee order by salary)
Where rownum<5

17.When you give an update statement how undo/rollback segment will work/what are the steps?

Oracle keep old values in undo segment and new values in redo entries. When you say rollback it replace old values from undo segment. When you say commit erase the undo segment values and keep new vales in permanent.

Informatica Administration

18.What is DTM? How will you configure it?

DTM transform data received from reader buffer and its moves transformation to transformation on row by row basis and it uses transformation caches when necessary.

19.You transfer 100000 rows to target but some rows get discard how will you trace them? And where its get loaded?

Rejected records are loaded into bad files. It has record indicator and column indicator.

Record indicator identified by (0-insert,1-update,2-delete,3-reject) and column indicator identified by (D-valid,O-overflow,N-null,T-truncated).

Normally data may get rejected in different reason due to transformation logic

20.What are the different uses of a repository manager?

Repository manager used to create repository which contains metadata the informatica uses to transform data from source to target. And also it use to create informatica user's and folders and copy, backup and restore the repository

21.How do you take care of security using a repository manager?

Using repository privileges, folder permission and locking.

Repository privileges(Session operator, Use designer, Browse repository, Create session and batches, Administer repository, administer server, super user)

Folder permission(owner, groups, users)

Locking(Read, Write, Execute, Fetch, Save)

22.What is a folder?

Folder contains repository objects such as sources, targets, mappings, transformation which are helps logically organize our data warehouse.

23.Can you create a folder within designer?

Not possible

24.What are shortcuts? Where it can be used? What are the advantages?

There are 2 shortcuts(Local and global) Local used in local repository and global used in global repository. The advantage is reuse an object without creating multiple objects. Say for example a source definition want to use in 10 mappings in 10 different folder without creating 10 multiple source you create 10 shotcuts.

25.How do you increase the performance of mappings?

Use single pass read(use one source qualifier instead of multiple SQ for same table)
Minimize data type conversion (Integer to Decimal again back to Integer)
Optimize transformation(when you use Lookup, aggregator, filter, rank and joiner)
Use caches for lookup
Aggregator use presorted port, increase cache size, minimize input/out port as much as possible
Use Filter wherever possible to avoid unnecessary data flow

26.Explain Informatica Architecture?

Informatica consist of client and server. Client tools such as Repository manager, Designer, Server manager. Repository data base contains metadata it read by informatica server used read data from source, transforming and loading into target.

27.How will you do sessions partitions?

It's not available in power part 4.7

Transformation

28.What are the constants used in update strategy?

DD_INSERT, DD_UPDATE, DD_DELETE, DD_REJECT

29.What is difference between connected and unconnected lookup transformation?

Connected lookup return multiple values to other transformation
Where as unconnected lookup return one values
If lookup condition matches Connected lookup return user defined default values
Where as unconnected lookup return null values
Connected supports dynamic caches where as unconnected supports static

30.What you will do in session level for update strategy transformation?

In session property sheet set Treat rows as "Data Driven"

31.What are the port available for update strategy , sequence generator, Lookup, stored procedure transformation?

Transformations   Port
Update strategy   Input, Output
Sequence Generator   Output only
Lookup   Input, Output, Lookup, Return
Stored Procedure   Input, Output

32.Why did you used connected stored procedure why don't use unconnected stored procedure?

33.What is active and passive transformations?

Active transformation change the no. of records when passing to targe(example filter)
where as passive transformation will not change the transformation(example expression)

34.What are the tracing level?

Normal – It contains only session initialization details and transformation details no. records rejected, applied
Terse - Only initialization details will be there
Verbose Initialization – Normal setting information plus detailed information about the transformation.
Verbose data – Verbose init. Settings and all information about the session

35.How will you make records in groups?

Using group by port in aggregator

36.Need to store value like 145 into target when you use aggregator, how will you do that?

Use Round() function

37.How will you move mappings from development to production database?

    Copy all the mapping from development repository and paste production repository while paste it will promt whether you want replace/rename. If say replace informatica replace all the source tables with repository database.

38.What is difference between aggregator and expression?

Aggregator is active transformation and expression is passive transformation
Aggregator transformation used to perform aggregate calculation on group of records   really
Where as expression used perform calculation with single record

39.Can you use mapping without source qualifier?

Not possible, If source RDBMS/DBMS/Flat file use SQ or use normalizer if the source cobol feed

40.When do you use a normalizer?

    Normalized can be used in Relational to denormilize data.

41.What are stored procedure transformations. Purpose of sp transformation. How did you go about using your project?

Connected and unconnected stored procudure.
Unconnected stored procedure used for data base level activities such as pre and post load

Connected stored procedure used in informatica level for example passing one parameter as input and capturing return value from the stored procedure.

Normal   - row wise check
Pre-Load Source - (Capture source incremental data for incremental aggregation)
Post-Load Source - (Delete Temporary tables)
Pre-Load Target - (Check disk space available)
Post-Load Target – (Drop and recreate index)

42.What is lookup and difference between types of lookup. What exactly happens when a lookup is cached. How does a dynamic lookup cache work.

Lookup transformation used for check values in the source and target tables(primary key values).
There are 2 type connected and unconnected transformation
Connected lookup returns multiple values if condition true
Where as unconnected return a single values through return port.
Connected lookup return default user value if the condition does not mach
Where as unconnected return null values
Lookup cache does:
Read the source/target table and stored in the lookup cache

43.What is a joiner transformation?

Used for heterogeneous sources(A relational source and a flat file)
Type of joins:
Assume 2 tables has values(Master - 1, 2, 3 and Detail - 1, 3, 4)

Normal(If the condition mach both master and detail tables then the records will be displaced. Result set 1, 3)
Master Outer(It takes all the rows from detail table and maching rows from master table. Result set 1, 3, 4)
Detail Outer(It takes all the values from master source and maching values from detail table. Result set 1, 2, 3)
Full Outer(It takes all values from both tables)

44.What is aggregator transformation how will you use in your project?

Used perform aggregate calculation on group of records and we can use conditional clause to filter data

45.Can you use one mapping to populate two tables in different schemas?

Yes we can use

46.Explain lookup cache, various caches?

Lookup transformation used for check values in the source and target tables(primary key values).

Various Caches:

Persistent cache (we can save the lookup cache files and reuse them the next time process the lookup transformation)

Re-cache from database (If the persistent cache not synchronized with lookup table you can configure the lookup transformation to rebuild the lookup cache)

Static cache (When the lookup condition is true, Informatica server return a value from lookup cache and it's does not update the cache while it processes the lookup transformation)

Dynamic cache (Informatica server dynamically inserts new rows or update existing rows in the cache and the target. Suppose if we want lookup a target table we can use dynamic cache)

Shared cache (we can share lookup transformation between multiple transformations in a mapping. 2 lookup in a mapping can share single lookup cache)

47.Which path will the cache be created?

User specified directory. If we say c:\ all the cache files created in this directory.

48.Where do you specify all the parameters for lookup caches?

Lookup property sheet/tab.

49.How do you remove the cache files after the transformation?

After session complete, DTM remove cache memory and deletes caches files.
In case using persistent cache and Incremental aggregation then caches files will be saved.

50.What is the use of aggregator transformation?

To perform Aggregate calculation

Use conditional clause to filter data in the expression Sum(commission, Commission >2000)

Use non-aggregate function iif (max(quantity) > 0, Max(quantitiy), 0))

51.What are the contents of index and cache files?

Index caches files hold unique group values as determined by group by port in the transformation.

Data caches files hold row data until it performs necessary calculation.

52.How do you call a store procedure within a transformation?

In the expression transformation create new out port in the expression write :sp.stored procedure name(arguments)

53.Is there any performance issue in connected & unconnected lookup? If yes, How?

Yes

Unconnected lookup much more faster than connected lookup why because in unconnected not connected to any other transformation we are calling it from other transformation so it minimize lookup cache value

Where as connected transformation connected to other transformation so it keeps values in the lookup cache.

54.What is dynamic lookup?

When we use target lookup table, Informatica server dynamically insert new values or it updates if the values exist and passes to target table.

55.How Informatica read data if source have one relational and flat file?

Use joiner transformation after source qualifier before other transformation.

56.How you will load unique record into target flat file from source flat files has duplicate data?

There are 2 we can do this either we can use Rank transformation or oracle external table
In rank transformation using group by port (Group the records) and then set no. of rank 1. Rank transformation return one value from the group. That the values will be a unique one.

57.Can you use flat file for repository?

No, We cant

58.Can you use flat file for lookup table?

No, We cant

59.Without Source Qualifier and joiner how will you join tables?

In session level we have option user defined join. Where we can write join condition.

60.Update strategy set DD_Update but in session level have insert. What will happens?

Insert take place. Because this option override the mapping level option

Sessions and batches

61.What are the commit intervals?

Source based commit (Based on the no. of active source records(Source qualifier) reads. Commit interval set 10000 rows and source qualifier reads 10000 but due to transformation logic 3000 rows get rejected when 7000 reach target commit will fire, so writer buffer does not rows held the buffer)

Target based commit (Based on the rows in the buffer and commit interval. Target based commit set 10000 but writer buffer fills every 7500, next time buffer fills 15000 now commit statement will fire then 22500 like go on.)

62.When we use router transformation?

When we want perform multiple condition to filter out data then we go for router. (Say for example source records 50 filter condition mach 10 records remaining 40 records get filter out but still we want perform few more filter condition to filter remaining 40 records.)

63.How did you schedule sessions in your project?

Run once (set 2 parameter date and time when session should start)

Run Every (Informatica server run session at regular interval as we configured, parameter Days, hour, minutes, end on, end after, forever)

Customized repeat (Repeat every 2 days, daily frequency hr, min, every week, every month)

Run only on demand(Manually run) this not session scheduling.

64.How do you use the pre-sessions and post-sessions in sessions wizard, what for they used?

Post-session used for email option when the session success/failure send email. For that we should configure
Step1. Should have a informatica startup account and create outlook profile for that user
Step2. Configure Microsoft exchange server in mail box applet(control panel)
Step3. Configure informatica server miscellaneous tab have one option called MS exchange profile where we have specify the outlook profile name.

Pre-session used for even scheduling (Say for example we don't know whether source file available or not in particular directory. For that we write one DOS command to move file directory to destination and set event based scheduling option in session property sheet Indicator file wait for).

65.What are different types of batches. What are the advantages and dis-advantages of a concurrent batch?

Sequential(Run the sessions one by one)

Concurrent (Run the sessions simultaneously)

Advantage of concurrent batch:

It's takes informatica server resource and reduce time it takes run session separately.
Use this feature when we have multiple sources that process large amount of data in one session. Split sessions and put into one concurrent batches to complete quickly.

Disadvantage

Require more shared memory otherwise session may get failed

66.How do you handle a session if some of the records fail. How do you stop the session in case of errors. Can it be achieved in mapping level or session level?

It can be achieved in session level only. In session property sheet, log files tab one option is the error handling Stop on ------ errors. Based on the error we set informatica server stop the session.

67.How you do improve the performance of session.

If we use Aggregator transformation use sorted port, Increase aggregate cache size, Use filter before aggregation so that it minimize unnecessary aggregation.

Lookup transformation use lookup caches

Increase DTM shared memory allocation

Eliminating transformation errors using lower tracing level(Say for example a mapping has 50 transformation when transformation error occur informatica server has to write in session log file it affect session performance)

68.Explain incremental aggregation. Will that increase the performance? How?

Incremental aggregation capture whatever changes made in source used for aggregate calculation in a session, rather than processing the entire source and recalculating the same calculation each time session run. Therefore it improve session performance.

Only use incremental aggregation following situation:

Mapping have aggregate calculation
Source table changes incrementally
Filtering source incremental data by time stamp

Before Aggregation have to do following steps:

Use filter transformation to remove pre-existing records

Reinitialize aggregate cache when source table completely changes for example incremental changes happing daily and complete changes happenings monthly once. So when the source table completely change we have reinitialize the aggregate cache and truncate target table use new source table. Choose Reinitialize cache in the aggregation behavior in transformation tab

69.Concurrent batches have 3 sessions and set each session run if previous complete but 2nd fail then what will happen the batch?

Batch will fail

General Project

70. How many mapping, dimension tables, Fact tables and any complex mapping you did? And what is your database size, how frequently loading to DWH?

I did 22 Mapping, 4 dimension table and one fact table. One complex mapping I did for slowly changing dimension table. Database size is 9GB. Loading data every day
71. What are the different transformations used in your project?

Aggregator, Expression, Filter, Sequence generator, Update Strategy, Lookup, Stored Procedure, Joiner, Rank, Source Qualifier.

72. How did you populate the dimensions tables?

73. What are the sources you worked on?
Oracle

74. How many mappings have you developed on your whole dwh project?

45 mappings

75. What is OS used your project?

Windows NT

76. Explain your project (Fact table, dimensions, and database size)

Fact table contains all business measures (numeric values) and foreign key values, Dimension table contains details about subject area like customer, product

77. What is difference between Informatica power mart and power center?

Using power center we can create global repository
Power mart used to create local repository
Global repository configure multiple server to balance session load
Local repository configure only single server

78. Have you done any complex mapping?

Developed one mapping to handle slowly changing dimension table.

79. Explain details about DTM?

Once we session start, load manager start DTM and it allocate session shared memory and contains reader and writer. Reader will read source data from source qualifier using SQL statement and move data to DTM then DTM transform data to transformation to transformation and row by row basis finally move data to writer then writer write data into target using SQL statement.

I-Flex Interview (14th May 2003)

80. What are the key you used other than primary key and foreign key?

Used surrogate key to maintain uniqueness to overcome duplicate value in the primary key.

81. Data flow of your Data warehouse(Architecture)

DWH is a basic architecture (OLTP to Data warehouse from DWH OLAP analytical and report building.

82. Difference between Power part and power center?

Using power center we can create global repository
Power mart used to create local repository
Global repository configure multiple server to balance session load
Local repository configure only single server

83. What are the batches and it's details?

Sequential(Run the sessions one by one)

Concurrent (Run the sessions simultaneously)

Advantage of concurrent batch:

It's takes informatica server resource and reduce time it takes run session separately.
Use this feature when we have multiple sources that process large amount of data in one session. Split sessions and put into one concurrent batches to complete quickly.

Disadvantage

Require more shared memory otherwise session may get failed

84. What is external table in oracle. How oracle read the flat file

Used for read flat file. Oracle internally write SQL loader script with control file.

85. What are the index you used? Bitmap join index?

Bitmap index used in data warehouse environment to increase query response time, since DWH has low cardinality, low updates, very efficient for where clause.

Bitmap join index used to join dimension and fact table instead reading 2 different index.

86.What are the partitions in 8i/9i? Where you will use hash partition?

In oracle8i there are 3 partition (Range, Hash, Composite)
In Oracle9i List partition is additional one

Range (Used for Dates values for example in DWH (Date values are Quarter 1, Quarter 2, Quarter 3, Quater4)

Hash (Used for unpredictable values say for example we cant able predict which value to allocate which partition then we go for hash partition. If we set partition 5 for a column oracle allocate values into 5 partition accordingly).

List (Used for literal values say for example a country have 24 states create 24 partition for 24 states each)

Composite (Combination of range and hash)

91.What is main difference mapplets and mapping?

Reuse the transformation in several mappings, where as mapping not like that.

If any changes made in mapplets it automatically inherited in all other instance mapplets.

92. What is difference between the source qualifier filter and filter transformation?
Source qualifier filter only used for relation source where as Filter used any kind of source.

Source qualifier filter data while reading where as filter before loading into target.

93. What is the maximum no. of return value when we use unconnected transformation?

Only one.

94. What are the environments in which informatica server can run on?

Informatica client runs on Windows 95 / 98 / NT, Unix Solaris, Unix AIX(IBM)

Informatica Server runs on Windows NT / Unix

Minimum Hardware requirements

Informatica Client Hard disk 40MB, RAM 64MB

Informatica Server Hard Disk 60MB, RAM 64MB

95. Can unconnected lookup do everything a connected lookup transformation can do?

No, We cant call connected lookup in other transformation. Rest of things it's possible

96. In 5.x can we copy part of mapping and paste it in other mapping?

I think its possible

97. What option do you select for a sessions in batch, so that the sessions run one after the other?

We have select an option called "Run if previous completed"

98. How do you really know that paging to disk is happening while you are using a lookup transformation? Assume you have access to server?

We have collect performance data first then see the counters parameter lookup_readtodisk if it's greater than 0 then it's read from disk

Step1. Choose the option "Collect Performance data" in the general tab session property

sheet.

Step2. Monitor server then click server-request → session performance details

Step3. Locate the performance details file named called session_name.perf file in the session log file directory

Step4. Find out counter parameter **lookup_readtodisk** if it's greater than 0 then informatica read lookup table values from the disk. Find out how many rows in the cache see **Lookup_rowsincache**

99. List three option available in informatica to tune aggregator transformation?

Use Sorted Input to sort data before aggregation
Use Filter transformation before aggregator

Increase Aggregator cache size

100.Assume there is text file as source having a binary field to, to source qualifier What native data type informatica will convert this binary field to in source qualifier?

Binary data type for relational source for flat file ?

101.Variable v1 has values set as 5 in designer(default), 10 in parameter file, 15 in repository. While running session which value informatica will read?

Informatica read value 15 from repository

102. Joiner transformation is joining two tables s1 and s2. s1 has 10,000 rows and s2 has 1000 rows . Which table you will set master for better performance of joiner transformation? Why?

Set table S2 as Master table because informatica server has to keep master table in the cache so if it is 1000 in cache will get performance instead of having 10000 rows in cache

103. Source table has 5 rows. Rank in rank transformation is set to 10. How many rows the rank transformation will output?

5 Rank

104. How to capture performance statistics of individual transformation in the mapping and explain some important statistics that can be captured?

Use tracing level Verbose data

105. Give a way in which you can implement a real time scenario where data in a table is changing and you need to look up data from it. How will you configure the lookup transformation for this purpose?

In slowly changing dimension table use type 2 and model 1

106. What is DTM process? How many threads it creates to process data, explain each thread in brief?

DTM receive data from reader and move data to transformation to transformation on row by row basis. It's create 2 thread one is reader and another one is writer

107. Suppose session is configured with commit interval of 10,000 rows and source has 50,000 rows explain the commit points for source based commit & target based commit. Assume appropriate value wherever required?

Target Based commit (First time Buffer size full 7500 next time 15000)

Commit Every 15000, 22500, 30000, 40000, 50000

Source Based commit(Does not affect rows held in buffer)

Commit Every 10000, 20000, 30000, 40000, 50000

108.What does first column of bad file (rejected rows) indicates?

First Column - Row indicator (0, 1, 2, 3)

Second Column – Column Indicator (D, O, N, T)

109. What is the formula for calculation rank data caches? And also Aggregator, data, index caches?

Index cache size = Total no. of rows * size of the column in the lookup condition (50 * 4)

Aggregator/Rank transformation Data Cache size = (Total no. of rows * size of the column in the lookup condition) + (Total no. of rows * size of the connected output ports)

110. Can unconnected lookup return more than 1 value? No

INFORMATICA TRANSFORMATIONS

- Aggregator
- Expression
- External Procedure
- Advanced External Procedure
- Filter
- Joiner
- Lookup
- Normalizer
- Rank
- Router
- Sequence Generator
- Stored Procedure
- Source Qualifier
- Update Strategy
- XML source qualifier

Expression Transformation

- You can use ET to calculate values in a single row before you write to the target
- You can use ET, to perform any non-aggregate calculation
- To perform calculations involving multiple rows, such as sums of averages, use the Aggregator. Unlike ET the Aggregator Transformation allow you to group and sort data

Calculation

To use the Expression Transformation to calculate values for a single row, you must include the following ports.
- Input port for each value used in the calculation
- Output port for the expression

NOTE

You can enter multiple expressions in a single ET. As long as you enter only one expression for each port, you can create any number of output ports in the Expression Transformation. In this way, you can use one expression transformation rather than creating separate transformations for each calculation that requires the same set of data.

Sequence Generator Transformation

- Create keys
- Replace missing values
- This contains two output ports that you can connect to one or more transformations. The server generates a value each time a row enters a connected transformation, even if that value is not used.
- There are two parameters NEXTVAL, CURRVAL
- The SGT can be reusable
- You can not edit any default ports (NEXTVAL, CURRVAL)

SGT Properties
- Start value
- Increment By
- End value
- Current value
- Cycle   (If selected, server cycles through sequence range. Otherwise, Stops with configured end value)
- Reset
- No of cached values

NOTE
- Reset is disabled for Reusable SGT
- Unlike other transformations, you cannot override SGT properties at session level. This protects the integrity of sequence values generated.

Aggregator Transformation

Difference between Aggregator and Expression Transformation

We can use Aggregator to perform calculations on groups. Where as the Expression transformation permits you to calculations on row-by-row basis only.

The server performs aggregate calculations as it reads and stores necessary data group and row data in an aggregator cache.

When Incremental aggregation occurs, the server passes new source data through the mapping and uses historical cache data to perform new calculation incrementally.

Components
- Aggregate Expression
- Group by port

-        Aggregate cache

When a session is being run using aggregator transformation, the server creates Index and data caches in memory to process the transformation. If the server requires more space, it stores overflow values in cache files.

NOTE

The performance of aggregator transformation can be improved by using "Sorted Input option". When this is selected, the server assumes all data is sorted by group.

Incremental Aggregation

-        Using this, you apply captured changes in the source to aggregate calculation in a session. If the source changes only incrementally and you can capture changes, you can configure the session to process only those changes
-        This allows the sever to update the target incrementally, rather than forcing it to process the entire source and recalculate the same calculations each time you run the session.

Steps:
-        The first time you run a session with incremental aggregation enabled, the server process the entire source.
-        At the end of the session, the server stores aggregate data from that session ran in two files, the index file and data file. The server creates the file in local directory.
-        The second time you run the session, use only changes in the source as source data for the session. The server then performs the following actions:
    (1)   For each input record, the session checks the historical information in the index file for a corresponding group, then:
                If it finds a corresponding group –
                        The server performs the aggregate operation incrementally, using the aggregate data for that group, and saves the incremental changes.
                Else
                        Server create a new group and saves the record data
    (2)   When writing to the target, the server applies the changes to the existing target.
        o   Updates modified aggregate groups in the target
        o   Inserts new aggregate data
        o   Delete removed aggregate data
        o   Ignores unchanged aggregate data
        o   Saves modified aggregate data in Index/Data files to be used as historical data the next time you run the session.
Each Subsequent time you run the session with incremental aggregation, you use only the incremental source changes in the session.
If the source changes significantly, and you want the server to continue saving the aggregate data for the future incremental changes, configure the server to overwrite existing aggregate data with new aggregate data.

Use Incremental Aggregator Transformation Only IF:
-        Mapping includes an aggregate function
-        Source changes only incrementally
-        You can capture incremental changes. You might do this by filtering source data by timestamp.
External Procedure Transformation
-        When Informatica's transformation does not provide the exact functionality we need, we can develop complex functions with in a dynamic link library or Unix shared library.
-        To obtain this kind of extensibility, we can use Transformation Exchange (TX) dynamic invocation interface built into Power mart/Power Center.
-        Using TX, you can create an External Procedure Transformation and bind it to an External Procedure that you have developed.
-        Two types of External Procedures are available
COM External Procedure (Only for WIN NT/2000)
Informatica External Procedure (available for WINNT, Solaris, HPUX etc)
Components of TX:

(a)   External Procedure

This exists separately from Informatica Server. It consists of C++, VB code written by developer. The code is compiled and linked to a DLL or Shared memory, which is loaded by the Informatica Server at runtime.

(b)   External Procedure Transformation

This is created in Designer and it is an object that resides in the Informatica Repository. This serves in many ways

o   This contains metadata describing External procedure

o   This allows an External procedure to be references in a mappingby adding an instance of an External Procedure transformation.

All External Procedure Transformations must be defined as reusable transformations.

Therefore you cannot create External Procedure transformation in designer. You can create only with in the transformation developer of designer and add instances of the transformation to mapping.

Difference Between Advanced External Procedure And External Procedure Transformation

Advanced External Procedure Transformation

- The Input and Output functions occur separately
- The output function is a separate callback function provided by Informatica that can be called from Advanced External Procedure Library.
- The Output callback function is used to pass all the output port values from the Advanced External Procedure library to the informatica Server.
- Multiple Outputs (Multiple row Input and Multiple rows output)
- Supports Informatica procedure only
- Active Transformation
- Connected only

External Procedure Transformation

- In the External Procedure Transformation, an External Procedure function does both input and output, and it's parameters consists of all the ports of the transformation.
- Single return value (One row input and one row output)
- Supports COM and Informatica Procedures
- Passive transformation
- Connected or Unconnected

By Default, The Advanced External Procedure Transformation is an active transformation. However, we can configure this to be a passive by clearing "IS ACTIVE" option on the properties tab

LOOKUP Transformation

- We are using this for lookup data in a related table, view or synonym
- You can use multiple lookup transformations in a mapping
- The server queries the Lookup table based in the Lookup ports in the transformation. It compares lookup port values to lookup table column values, bases on lookup condition.

Types:

(a) Connected (or) unconnected.

(b) Cached (or) uncached .

If you cache the lkp table , you can choose to use a dynamic or static cache . by default ,the LKP cache remains static and doesn't change during the session .with dynamic cache ,the server inserts rows into the cache during the session ,information recommends that you cache the target table as Lookup .this enables you to lookup values in the target and insert them if they don't exist..

You can configure a connected LKP to receive input directly from the mapping pipeline .(or) you can configure an unconnected LKP to receive input from the result of an expression in another transformation.

Differences Between Connected and Unconnected Lookup:

connected

o   Receives input values directly from the pipeline.

o   uses Dynamic or static cache

o   Returns multiple values

o   supports user defined default values.

Unconnected

o   Recieves input values from the result of LKP expression in another transformation

o   Use static cache only.

o   Returns only one value.

o   Doesn't supports user-defined default values.

NOTES

o   Common use of unconnected LKP is to update slowly changing dimension tables.

o   Lookup components are

(a)   Lookup table. B) Ports c) Properties d) condition.

Lookup tables: This can be a single table, or you can join multiple tables in the same Database using a Lookup query override.You can improve Lookup initialization time by adding an index to the Lookup table.
Lookup ports: There are 3 ports in connected LKP transformation (I/P,O/P,LKP) and 4 ports unconnected LKP(I/P,O/P,LKP and return ports).
- if you've certain that a mapping doesn't use a Lookup ,port ,you delete it from the transformation. This reduces the amount of memory.

Lookup Properties: you can configure properties such as SQL override .for the Lookup,the Lookup table name ,and tracing level for the transformation.
Lookup condition: you can enter the conditions ,you want the server to use to determine whether input data qualifies values in the Lookup or cache .

when you configure a LKP condition for the transformation, you compare transformation input values with values in the Lookup table or cache ,which represented by LKP ports .when you run session ,the server queries the LKP table or cache for all incoming values based on the condition.

NOTE
-       If you configure a LKP to use static cache ,you can following operators =,>,<,>=,<=,!=.
but if you use an dynamic cache only =can be used .
-       when you don't configure the LKP for caching ,the server queries the LKP table for each input row .the result will be same, regardless of using cache
However using a Lookup cache can increase session performance, by Lookup table, when the source table is large.

Performance tips:
-       Add an index to the columns used in a Lookup condition.
-       Place conditions with an equality opertor (=) first.
-       Cache small Lookup tables .
-       Don't use an ORDER BY clause in SQL override.
-       Call unconnected Lookups with :LKP reference qualifier.

Normalizer Transformation
Normalization is the process of organizing data.
In database terms ,this includes creating normalized tables and establishing relationships between those tables. According to rules designed to both protect the data, and make the database more flexible by eliminating redundancy and inconsistent dependencies.
NT normalizes records from COBOL and relational sources ,allowing you to organizet the data according to you own needs.
A NT can appear anywhere is a data flow when you normalize a relational source.
Use a normalizer transformation, instead of source qualifier transformation when you normalize a COBOL source.
The occurs statement is a COBOL file nests multiple records of information in a single record.
Using the NT ,you breakout repeated data with in a record is to separate record into separate records.For each new record it creates, the NT generates an unique identifier. You can use this key value to join the normalized records.

Stored Procedure Transformation
- DBA creates stored procedures to automate time consuming tasks that are too complicated for standard SQL statements.
- A stored procedure is a precompiled collection of transact SQL statements and optional flow control statements, similar to an executable script.
- Stored procedures are stored and run with in the database. You can run a stored procedure with EXECUTE SQL statement in a database client tool, just as SQL statements. But unlike standard procedures allow user defined variables, conditional statements and programming features.

Usages of Stored Procedure
- Drop and recreate indexes.
- Check the status of target database before moving records into it.
- Determine database space.
- Perform a specialized calculation.

NOTE
- The Stored Procedure must exist in the database before creating a Stored Procedure Transformation, and the Stored procedure can exist in a source, target or any database with a valid connection to the server.

TYPES
- Connected Stored Procedure Transformation (Connected directly to the mapping)
- Unconnected Stored Procedure Transformation (Not connected directly to the flow of the mapping. Can be called from an Expression Transformation or other transformations)

Running a Stored Procedure
The options for running a Stored Procedure Transformation:
- Normal , Pre load of the source, Post load of the source, Pre load of the target, Post load of the target
You can run several stored procedure transformation in different modes in the same mapping.
Stored Procedure Transformations are created as normal type by default, which means that they run during the mapping, not before or after the session. They are also not created as reusable transformations.

If you want to:	Use below mode
Run a SP before/after the session	Unconnected
Run a SP once during a session	Unconnected
Run a SP for each row in data flow	Unconnected/Connected
Pass parameters to SP and receive a single return value	Connected

A normal connected SP will have an I/P and O/P port and return port also an output port, which is marked as 'R'.

Error Handling
- This can be configured in server manager (Log & Error handling)
- By default, the server stops the session

Rank Transformation

- This allows you to select only the top or bottom rank of data. You can get returned the largest or smallest numeric value in a port or group.
- You can also use Rank Transformation to return the strings at the top or the bottom of a session sort order. During the session, the server caches input data until it can perform the rank calculations.
- Rank Transformation differs from MAX and MIN functions, where they allows to select a group of top/bottom values, not just one value.
- As an **active** transformation, Rank transformation might change the number of rows passed through it.

Rank Transformation Properties

- Cache directory
- Top or Bottom rank
- Input/Output ports that contain values used to determine the rank.

Different ports in Rank Transformation

I	-	Input
O	-	Output
V	-	Variable
R	-	Rank

Rank Index

The designer automatically creates a RANKINDEX port for each rank transformation. The server uses this Index port to store the ranking position for each row in a group.

The RANKINDEX is an output port only. You can pass the RANKINDEX to another transformation in the mapping or directly to a target.

Filter Transformation
- As an active transformation, the Filter Transformation may change the no of rows passed through it.
- A filter condition returns TRUE/FALSE for each row that passes through the transformation, depending on whether a row meets the specified condition.
- Only rows that return TRUE pass through this filter and discarded rows do not appear in the session log/reject files.
- To maximize the session performance, include the Filter Transformation as close to the source in the mapping as possible.
- The filter transformation does not allow setting output default values.
- To filter out row with NULL values, use the ISNULL and IS_SPACES functions.

Joiner Transformation
Source Qualifier: can join data origination from a common source database
Joiner Transformation: Join tow related heterogeneous sources residing in different locations or File systems.
To join more than two sources, we can add additional joiner transformations.

<div align="center"><u>**SESSION LOGS**</u></div>

Information that reside in a session log:
- Allocation of system shared memory
- Execution of Pre-session commands/ Post-session commands
- Session Initialization
- Creation of SQL commands for reader/writer threads
- Start/End timings for target loading
- Error encountered during session
- Load summary of Reader/Writer/ DTM statistics

Other Information
- By default, the server generates log files based on the server code page.

Thread Identifier
Ex: CMN_1039
Reader and Writer thread codes have 3 digit and Transformation codes have 4 digits.
The number following a thread name indicate the following:
(a) Target load order group number
(b) Source pipeline number
(c) Partition number
(d) Aggregate/ Rank boundary number

Log File Codes
Error Codes Description

BR -   Related to reader process, including ERP, relational and flat file.
CMN -   Related to database, memory allocation
DBGR -   Related to debugger
EP-   External Procedure
LM -   Load Manager
TM -   DTM
REP -   Repository
WRT -   Writer

Load Summary
(a) Inserted
(b) Updated
(c) Deleted
(d) Rejected

Statistics details
(a) Requested rows shows the no of rows the writer actually received for the specified operation
(b) Applied rows shows the number of rows the writer successfully applied to the target (Without Error)
(c) Rejected rows show the no of rows the writer could not apply to the target
(d) Affected rows shows the no of rows affected by the specified operation

Detailed transformation statistics
The server reports the following details for each transformation in the mapping
(a) Name of Transformation
(b) No of I/P rows and name of the Input source
(c) No of O/P rows and name of the output target
(d) No of rows dropped
Tracing Levels

Normal   - Initialization and status information, Errors encountered, Transformation errors, rows skipped, summarize session details (Not at the level of individual rows)

Terse   - Initialization information as well as error messages, and notification of rejected data

Verbose Init   - Addition to normal tracing, Names of Index, Data files used and detailed transformation statistics.

Verbose Data   - Addition to Verbose Init, Each row that passes in to mapping detailed transformation statistics.

NOTE

When you enter tracing level in the session property sheet, you override tracing levels configured for transformations in the mapping.

MULTIPLE SERVERS

With Power Center, we can register and run multiple servers against a local or global repository. Hence you can distribute the repository session load across available servers to improve overall performance. (You can use only one Power Mart server in a local repository)

Issues in Server Organization
- Moving target database into the appropriate server machine may improve efficiency
- All Sessions/Batches using data from other sessions/batches need to use the same server and be incorporated into the same batch.
- Server with different speed/sizes can be used for handling most complicated sessions.

Session/Batch Behavior
- By default, every session/batch run on its associated Informatica server. That is selected in property sheet.
- In batches, that contain sessions with various servers, the property goes to the servers, that's of outer most batch.

Session Failures and Recovering Sessions

Two types of errors occurs in the server
- Non-Fatal
- Fatal

(a) Non-Fatal Errors

It is an error that does not force the session to stop on its first occurrence. Establish the error threshold in the session property sheet with the stop on option. When you enable this option, the server counts Non-Fatal errors that occur in the reader, writer and transformations.

Reader errors can include alignment errors while running a session in Unicode mode.

Writer errors can include key constraint violations, loading NULL into the NOT-NULL field and database errors.

Transformation errors can include conversion errors and any condition set up as an ERROR,. Such as NULL Input.

(b) Fatal Errors

This occurs when the server can not access the source, target or repository. This can include loss of connection or target database errors, such as lack of database space to load data.

If the session uses normalizer (or) sequence generator transformations, the server can not update the sequence values in the repository, and a fatal error occurs.

© Others

Usages of ABORT function in mapping logic, to abort a session when the server encounters a transformation error.

Stopping the server using *pmcmd* (or) Server Manager

Performing Recovery
- When the server starts a recovery session, it reads the OPB_SRVR_RECOVERY table and notes the rowid of the last row commited to the target database. The server then reads all sources again and starts processing from the next rowid.
- By default, perform recovery is disabled in setup. Hence it won't make entries in OPB_SRVR_RECOVERY table.
- The recovery session moves through the states of normal session schedule, waiting to run, Initializing, running, completed and failed. If the initial recovery fails, you can run recovery as many times.
- The normal reject loading process can also be done in session recovery process.
- The performance of recovery might be low, if
  o Mapping contain mapping variables
  o Commit interval is high

Un recoverable Sessions

Under certain circumstances, when a session does not complete, you need to truncate the target and run the session from the beginning.

Commit Intervals

A commit interval is the interval at which the server commits data to relational targets during a session.

(a) Target based commit
- Server commits data based on the no of target rows and the key constraints on the target table. The commit point also depends on the buffer block size and the commit pinterval.
- During a session, the server continues to fill the writer buffer, after it reaches the commit interval. When the buffer block is full, the Informatica server issues a commit command. As a result, the amount of data committed at the commit point generally exceeds the commit interval.
- The server commits data to each target based on primary –foreign key constraints.

(b) Source based commit
- Server commits data based on the number of source rows. The commit point is the commit interval you configure in the session properties.
- During a session, the server commits data to the target based on the number of rows from an active source in a single pipeline. The rows are referred to as source rows.

- A pipeline consists of a source qualifier and all the transformations and targets that receive data from source qualifier.
- Although the Filter, Router and Update Strategy transformations are active transformations, the server does not use them as active sources in a source based commit session.
- When a server runs a session, it identifies the active source for each pipeline in the mapping. The server generates a commit row from the active source at every commit interval.
- When each target in the pipeline receives the commit rows the server performs the commit.

Reject Loading

During a session, the server creates a reject file for each target instance in the mapping. If the writer of the target rejects data, the server writers the rejected row into the reject file.

You can correct those rejected data and re-load them to relational targets, using the reject loading utility. (You cannot load rejected data into a flat file target)

Each time, you run a session, the server appends a rejected data to the reject file.

Locating the BadFiles

$PMBadFileDir
Filename.bad

When you run a partitioned session, the server creates a separate reject file for each partition.

Reading Rejected data

Ex: 3,D,1,D,D,0,D,1094345609,D,0,0.00

To help us in finding the reason for rejecting, there are two main things.

(a) Row indicator

Row indicator tells the writer, what to do with the row of wrong data.

Row indicator	Meaning	Rejected By
0	Insert	Writer or target
1	Update	Writer or target
2	Delete	Writer or target
3	Reject	Writer

If a row indicator is 3, the writer rejected the row because an update strategy expression marked it for reject.

(b) Column indicator

Column indicator is followed by the first column of data, and another column indicator. They appears after every column of data and define the type of data preceding it

Column Indicator	Meaning	Writer Treats as
D	Valid Data	Good Data. The target accepts it unless a database error occurs, such as finding duplicate key.
O	Overflow	Bad Data.
N	Null	Bad Data.
T	Truncated	Bad Data

NOTE

NULL columns appear in the reject file with commas marking their column.

Correcting Reject File

Use the reject file and the session log to determine the cause for rejected data.

Keep in mind that correcting the reject file does not necessarily correct the source of the reject.

Correct the mapping and target database to eliminate some of the rejected data when you run the session again.

Trying to correct target rejected rows before correcting writer rejected rows is not recommended since they may contain misleading column indicator.

For example, a series of "N" indicator might lead you to believe the target database does not accept NULL values, so you decide to change those NULL values to Zero.

However, if those rows also had a 3 in row indicator. Column, the row was rejected b the writer because of an update strategy expression, not because of a target database restriction.

If you try to load the corrected file to target, the writer will again reject those rows, and they will contain inaccurate 0 values, in place of NULL values.

Why writer can reject ?
- Data overflowed column constraints
- An update strategy expression

Why target database can Reject ?
- Data contains a NULL column
- Database errors, such as key violations

Steps for loading reject file:
-       After correcting the rejected data, rename the rejected file to reject_file.in
-       The rejloader used the data movement mode configured for the server. It also used the code page of server/OS. Hence do not change the above, in middle of the reject loading
-       Use the reject loader utility
  Pmrejldr pmserver.cfg [folder name] [session name]

Other points

The server does not perform the following option, when using reject loader
- (a)   Source base commit
- (b)   Constraint based loading
- (c)   Truncated target table
- (d)   FTP targets
- (e)   External Loading

Multiple reject loaders

You can run the session several times and correct rejected data from the several session at once. You can correct and load all of the reject files at once, or work on one or two reject files, load then and work on the other at a later time.

External Loading

You can configure a session to use Sybase IQ, Teradata and Oracle external loaders to load session target files into the respective databases.

The External Loader option can increase session performance since these databases can load information directly from files faster than they can the SQL commands to insert the same data into the database.

Method:

When a session used External loader, the session creates a control file and target flat file. The control file contains information about the target flat file, such as data format and loading instruction for the External Loader. The control file has an extension of "*.ctl " and you can view the file in $PmtargetFilesDir.

For using an External Loader:

The following must be done:
-       configure an external loader connection in the server manager
-       Configure the session to write to a target flat file local to the server.
-       Choose an external loader connection for each target file in session property sheet.

Issues with External Loader:
-       Disable constraints
-       Performance issues
  - o   Increase commit intervals
  - o   Turn off database logging
-       Code page requirements
-       The server can use multiple External Loader within one session (Ex: you are having a session with the two target files. One with Oracle External Loader and another with Sybase External Loader)

Other Information:
-       The External Loader performance depends upon the platform of the server
-       The server loads data at different stages of the session
-       The serve writes External Loader initialization and completing messaging in the session log. However, details about EL performance, it is generated at EL log, which is getting stored as same target directory.
-       If the session contains errors, the server continues the EL process. If the session fails, the server loads partial target data using EL.
-       The EL creates a reject file for data rejected by the database. The reject file has an extension of "*.ldr" reject.

- The EL saves the reject file in the target file directory
- You can load corrected data from the file, using database reject loader, and not through Informatica reject load utility (For EL reject file only)

Configuring EL in session
- In the server manager, open the session property sheet
- Select File target, and then click flat file options

Caches

- server creates index and data caches in memory for aggregator ,rank ,joiner and Lookup transformation in a mapping.
- Server stores key values in index caches and output values in data caches : if the server requires more memory ,it stores overflow values in cache files .
- When the session completes, the server releases caches memory, and in most circumstances, it deletes the caches files .
- Caches Storage overflow :
- releases caches memory, and in most circumstances, it deletes the caches files .

Caches Storage overflow :

Transformation	index cache	data cache
Aggregator	stores group values As configured in the Group-by ports.	stores calculations based on Group-by ports
Rank	stores group values as Configured in the Group-by	stores ranking information based on Group-by ports .
Joiner	stores index values for The master source table As configured in Joiner condition.	stores master source rows .
Lookup	stores Lookup condition Information.	stores lookup data that's Not stored in the index cache.

Determining cache requirements

To calculate the cache size, you need to consider column and row requirements as well as processing overhead.
- server requires processing overhead to cache data and index information.
  Column overhead includes a null indicator, and row overhead can include row to key information.

Steps:
- first, add the total column size in the cache to the row overhead.
- Multiply the result by the no of groups (or) rows in the cache this gives the minimum cache requirements .
- For maximum requirements, multiply min requirements by 2.

Location:
-by default , the server stores the index and data files in the directory $PMCacheDir.
-the server names the index files PMAGG*.idx and data files PMAGG*.dat. if the size exceeds 2GB,you may find multiple index and data files in the directory .The server appends a number to the end of filename(PMAGG*.id*1,id*2,etc).

Aggregator Caches

- when server runs a session with an aggregator transformation, it stores data in memory until it completes the aggregation.
- when you partition a source, the server creates one memory cache and one disk cache and one and disk cache for each partition .It routes data from one partition to another based on group key values of the transformation.
- server uses memory to process an aggregator transformation with sort ports. It doesn't use cache memory .you don't need to configure the cache memory, that use sorted ports.

Index cache:
#Groups (($\sum$ column size) + 7)
Aggregate data cache:
#Groups (($\sum$ column size) + 7)

Rank Cache

- when the server runs a session with a Rank transformation, it compares an input row with rows with rows in data cache. If the input row out-ranks a stored row,the Informatica server replaces the stored row with the input row.
- If the rank transformation is configured to rank across multiple groups, the server ranks incrementally for each group it finds .

Index Cache :
#Groups (($\sum$ column size) + 7)
Rank Data Cache:
#Group [(#Ranks * ($\sum$ column size + 10)) + 20]

Joiner Cache:
- When server runs a session with joiner transformation, it reads all rows from the master source and builds memory caches based on the master rows.
- After building these caches, the server reads rows from the detail source and performs the joins
- Server creates the Index cache as it reads the master source into the data cache. The server uses the Index cache to test the join condition. When it finds a match, it retrieves rows values from the data cache.
- To improve joiner performance, the server aligns all data for joiner cache or an eight byte boundary.

Index Cache :

$\quad$ #Master rows [(∑ column size) + 16)

Joiner Data Cache:

$\quad$ **#**Master row [(∑ column size) + 8]

Lookup cache:
- When server runs a lookup transformation, the server builds a cache in memory, when it process the first row of data in the transformation.
- Server builds the cache and queries it for the each row that enters the transformation.
- If you partition the source pipeline, the server allocates the configured amount of memory for each partition. If two lookup transformations share the cache, the server does not allocate additional memory for the second lookup transformation.
- The server creates index and data cache files in the lookup cache drectory and used the server code page to create the files.

Index Cache :

$\quad$ #Rows in lookup table [(∑ column size) + 16)

Lookup Data Cache:

$\quad$ **#**Rows in lookup table [(∑ column size) + 8]

Transformations

A transformation is a repository object that generates, modifies or passes data.

(a) <u>Active Transformation:</u>
    a. Can change the number of rows, that passes through it (Filter, Normalizer, Rank ..)

(b) <u>Passive Transformation:</u>
    a. Does not change the no of rows that passes through it (Expression, lookup ..)

<u>NOTE:</u>

- Transformations can be connected to the data flow or they can be unconnected
- An unconnected transformation is not connected to other transformation in the mapping. It is called with in another transformation and returns a value to that transformation

Reusable Transformations:

When you are using reusable transformation to a mapping, the definition of transformation exists outside the mapping while an instance appears with mapping.

All the changes you are making in transformation will immediately reflect in instances.

You can create reusable transformation by two methods:

(a) Designing in transformation developer
(b) Promoting a standard transformation

Change that reflects in mappings are like expressions. If port name etc. are changes they won't reflect.

Handling High-Precision Data:

- Server process decimal values as doubles or decimals.
- When you create a session, you choose to enable the decimal data type or let the server process the data as double (Precision of 15)

<u>Example:</u>

- You may have a mapping with decimal (20,0) that passes through. The value may be 40012030304957666903.

  If you enable decimal arithmetic, the server passes the number as it is. If you do not enable decimal arithmetic, the server passes $4.00120303049577 \times 10^{19}$.

  If you want to process a decimal value with a precision greater than 28 digits, the server automatically treats as a double value.

Mapplets

When the server runs a session using a mapplets, it expands the mapplets. The server then runs the session as it would any other sessions, passing data through each transformations in the mapplet.

If you use a reusable transformation in a mapplet, changes to these can invalidate the mapplet and every mapping using the mapplet.

You can create a non-reusable instance of a reusable transformation.

Mapplet Objects:
- (a)   Input transformation
- (b)   Source qualifier
- (c)   Transformations, as you need
- (d)   Output transformation

Mapplet Won't Support:
-       Joiner
-       Normalizer
-       Pre/Post session stored procedure
-       Target definitions
-       XML source definitions

Types of Mapplets:
- (a)   Active Mapplets   -   Contains one or more active transformations
- (b)   Passive Mapplets   -   Contains only passive transformation

Copied mapplets are not an instance of original mapplets. If you make changes to the original, the copy does not inherit your changes

You can use a single mapplet, even more than once on a mapping.

Ports

Default value for I/P port   -   NULL
Default value for O/P port -   ERROR
Default value for variables -   Does not support default values

Session Parameters

This parameter represent values you might want to change between sessions, such as DB Connection or source file.

We can use session parameter in a session property sheet, then define the parameters in a session parameter file.

The user defined session parameter are:
- (a)   DB Connection
- (b)   Source File directory
- (c)   Target file directory
- (d)   Reject file directory

Description:

Use session parameter to make sessions more flexible. For example, you have the same type of transactional data written to two different databases, and you use the database connections TransDB1 and TransDB2 to connect to the databases. You want to use the same mapping for both tables.

Instead of creating two sessions for the same mapping, you can create a database connection parameter, like $DBConnectionSource, and use it as the source database connection for the session.

When you create a parameter file for the session, you set $DBConnectionSource to TransDB1 and run the session. After it completes set the value to TransDB2 and run the session again.

NOTE:

You can use several parameter together to make session management easier.

Session parameters do not have default value, when the server can not find a value for a session parameter, it fails to initialize the session.

Session Parameter File

-       A parameter file is created by text editor.
-       In that, we can specify the folder and session name, then list the parameters and variables used in the session and assign each value.
-       Save the parameter file in any directory, load to the server
-       We can define following values in a parameter
  - o   Mapping parameter
  - o   Mapping variables
  - o   Session parameters
-       You can include parameter and variable information for more than one session in a single parameter file by creating separate sections, for each session with in the parameter file.
-       You can override the parameter file for sessions contained in a batch by using a batch parameter file. A batch parameter file has the same format as a session parameter file

Locale

Informatica server can transform character data in two modes

(a) **ASCII**
      a. Default one
      b. Passes 7 byte, US-ASCII character data
(b) **UNICODE**
      a. Passes 8 bytes, multi byte character data
      b. It uses 2 bytes for each character to move data and performs additional checks at session level, to ensure data integrity.

Code pages contains the encoding to specify characters in a set of one or more languages. We can select a code page, based on the type of character data in the mappings.
Compatibility between code pages is essential for accurate data movement.
The various code page components are

-       Operating system Locale settings
-       Operating system code page
-       Informatica server data movement mode
-       Informatica server code page
-       Informatica repository code page

Locale
(a) System Locale -   System Default
(b) User locale   -   setting for date, time, display
©   Input locale

Mapping Parameter and Variables

These represent values in mappings/mapplets.
If we declare mapping parameters and variables in a mapping, you can reuse a mapping by altering the parameter and variable values of the mappings in the session.
This can reduce the overhead of creating multiple mappings when only certain attributes of mapping needs to be changed.
When you want to use the same value for a mapping parameter each time you run the session.
Unlike a mapping parameter, a mapping variable represent a value that can change through the session. The server saves the value of a mapping variable to the repository at the end of each successful run and used that value the next time you run the session.

Mapping objects:
Source, Target, Transformation, Cubes, Dimension

Debugger

We can run the Debugger in two situations
(a)   Before Session: After saving mapping, we can run some initial tests.
(b)   After Session: real Debugging process

MEadata Reporter:
-       Web based application that allows to run reports against repository metadata
-       Reports including executed sessions, lookup table dependencies, mappings and source/target schemas.

Repository

Types of Repository
(a) Global Repository
      a. This is the hub of the domain use the GR to store common objects that multiple developers can use through shortcuts. These may include operational or application source definitions, reusable transformations, mapplets and mappings
(b) Local Repository
      a. A Local Repository is with in a domain that is not the global repository. Use4 the Local Repository for development.

©   Standard Repository
      a. A repository that functions individually, unrelated and unconnected to other repository

NOTE:
-       Once you create a global repository, you can not change it to a local repository
-       However, you can promote the local to global repository

Batches

-       Provide a way to group sessions for either serial or parallel execution by server
-       Batches
      o Sequential (Runs session one after another)
      o Concurrent (Runs sessions at same time)

Nesting Batches
Each batch can contain any number of session/batches. We can nest batches several levels deep, defining batches within batches

---

Nested batches are useful when you want to control a complex series of sessions that must run sequentially or concurrently

Scheduling

When you place sessions in a batch, the batch schedule override that session schedule by default. However, we can configure a batched session to run on its own schedule by selecting the "Use Absolute Time Session" Option.

Server Behavior

Server configured to run a batch overrides the server configuration to run sessions within the batch. If you have multiple servers, all sessions within a batch run on the Informatica server that runs the batch.

The server marks a batch as failed if one of its sessions is configured to run if "Previous completes" and that previous session fails.

Sequential Batch

If you have sessions with dependent source/target relationship, you can place them in a sequential batch, so that Informatica server can run them is consecutive order.

They are two ways of running sessions, under this category

        (a) Run the session, only if the previous completes successfully
        (b) Always run the session (this is default)

Concurrent Batch

In this mode, the server starts all of the sessions within the batch, at same time

Concurrent batches take advantage of the resource of the Informatica server, reducing the time it takes to run the session separately or in a sequential batch.

Concurrent batch in a Sequential batch

If you have concurrent batches with source-target dependencies that benefit from running those batches in a particular order, just like sessions, place them into a sequential batch.

Server Concepts

The Informatica server used three system resources

    (a)   CPU
    (b)   Shared Memory
    (c)   Buffer Memory

Informatica server uses shared memory, buffer memory and cache memory for session information and to move data between session threads.

LM Shared Memory

Load Manager uses both process and shared memory. The LM keeps the information server list of sessions and batches, and the schedule queue in process memory.

Once a session starts, the LM uses shared memory to store session details for the duration of the session run or session schedule. This shared memory appears as the configurable parameter (LMSharedMemory) and the server allots 2,000,000 bytes as default.

This allows you to schedule or run approximately 10 sessions at one time.

DTM Buffer Memory

The DTM process allocates buffer memory to the session based on the DTM buffer poll size settings, in session properties. By default, it allocates 12,000,000 bytes of memory to the session.

DTM divides memory into buffer blocks as configured in the buffer block size settings. (Default: 64,000 bytes per block)

Running a Session

The following tasks are being done during a session

1.	LM locks the session and read session properties
2.	LM reads parameter file
3.	LM expands server/session variables and parameters
4.	LM verifies permission and privileges
5.	LM validates source and target code page
6.	LM creates session log file
7.	LM creates DTM process
8.	DTM process allocates DTM process memory
9.	DTM initializes the session and fetches mapping
10.	DTM executes pre-session commands and procedures
11.	DTM creates reader, writer, transformation threads for each pipeline
12.	DTM executes post-session commands and procedures
13.	DTM writes historical incremental aggregation/lookup to repository
14.	LM sends post-session emails

Stopping and aborting a session

- If the session you want to stop is a part of batch, you must stop the batch
- If the batch is part of nested batch, stop the outermost batch
- When you issue the stop command, the server stops reading data. It continues processing and writing data and committing data to targets
- If the server cannot finish processing and committing data, you can issue the ABORT command. It is similar to stop command, except it has a 60 second timeout. If the server cannot finish processing and committing data within 60 seconds, it kills the DTM process and terminates the session.

Recovery:

- After a session being stopped/aborted, the session results can be recovered. When the recovery is performed, the session continues from the point at which it stopped.
- If you do not recover the session, the server runs the entire session the next time.
- Hence, after stopping/aborting, you may need to manually delete targets before the session runs again.

NOTE:
ABORT command and ABORT function, both are different.

When can a Session Fail

- Server cannot allocate enough system resources
- Session exceeds the maximum no of sessions the server can run concurrently
- Server cannot obtain an execute lock for the session (the session is already locked)
- Server unable to execute post-session shell commands or post-load stored procedures
- Server encounters database errors
- Server encounter Transformation row errors (Ex: NULL value in non-null fields)
- Network related errors

When Pre/Post Shell Commands are useful

- To delete a reject file
- To archive target files before session begins

Session Performance

- Minimum log (Terse)
- Partitioning source data.
- Performing ETL for each partition, in parallel. (For this, multiple CPUs are needed)
- Adding indexes.
- Changing commit Level.
- Using Filter trans to remove unwanted data movement.
- Increasing buffer memory, when large volume of data.
- Multiple lookups can reduce the performance. Verify the largest lookup table and tune the expressions.
- In session level, the causes are small cache size, low buffer memory and small commit interval.
- At system level,
  - o   WIN NT/2000-U the task manager.
  - o   UNIX: VMSTART, IOSTART.

Hierarchy of optimization

- Target.
- Source.
- Mapping
- Session.
- System.

Optimizing Target Databases:

- Drop indexes /constraints
- Increase checkpoint intervals.
- Use bulk loading /external loading.
- Turn off recovery.
- Increase database network packet size.

Source   level

- Optimize the query (using group by, group by).
- Use conditional filters.
- Connect to RDBMS using IPC protocol.

Mapping

- Optimize data type conversions.
- Eliminate transformation errors.
- Optimize transformations/ expressions.

Session:

- concurrent batches.
- Partition sessions.

- Reduce error tracing.
- Remove staging area.
- Tune session parameters.

System:
- improve network speed.
- Use multiple preservers on separate systems.
- Reduce paging.

Session Process

Info server uses both process memory and system shared memory to perform ETL process.
It runs as a daemon on UNIX and as a service on WIN NT.
The following processes are used to run a session:
(a)   LOAD manager process: - starts a session
    • creates DTM process, which creates the session.
(b)   DTM process: -   creates threads to initialize the session
                        -      read, write and transform data.
                        -      handle pre/post session opertions.

Load manager processes:
- manages session/batch scheduling.
- Locks session.
- Reads parameter file.
- Expands server/session variables, parameters .
- Verifies permissions/privileges.
- Creates session log file.

DTM process:
The primary purpose of the DTM is to create and manage threads that carry out the session tasks.
The DTM allocates process memory for the session and divides it into buffers. This is known as   buffer memory.
The default memory allocation is 12,000,000 bytes .it creates the main thread, which is called master thread .this manages all other threads.

Various threads	functions
Master thread-	handles stop and abort requests from load manager.
Mapping thread-	one thread for each session.
	Fetches session and mapping information.
	Compiles mapping.
	Cleans up after execution.
Reader thread-	one thread for each partition.
	Relational sources uses relational threads and
	Flat files use file threads.
Writer thread-	one thread for each partition writes to target.
Transformation thread-	One or more transformation for each partition.

Note:
When you run a session, the   threads for a partitioned source execute concurrently. The threads use buffers to move/transform data.

1. **Explain about your projects**
- Architecture
- Dimension and Fact tables
- Sources and Targets
- Transformations used
- Frequency of populating data
- Database size

2. What is dimension modeling?
Unlike ER model the dimensional model is very asymmetric
with one large central table called as fact table connected to multiple
dimension tables .It is also called star schema.

3. What are mapplets?
Mapplets are reusable objects that represents collection of transformations
Transformations not to be included in mapplets are
Cobol source definitions
Joiner transformations

Normalizer Transformations
Non-reusable sequence generator transformations
Pre or post session procedures
Target definitions
XML Source definitions
IBM MQ source definitions
Power mart 3.5 style Lookup functions

4. What are the transformations that use cache for performance?
Aggregator, Lookups, Joiner and Ranker

5. What the active and passive transformations?

An active transformation changes the number of rows that pass through the
mapping.
1. Source Qualifier
2. Filter transformation
3. Router transformation
4. Ranker
5. Update strategy
6. Aggregator
7. Advanced External procedure
8. Normalizer
9. Joiner

Passive transformations do not change the number of rows that pass through
the mapping.
1. Expressions
2. Lookup
3. Stored procedure
4. External procedure
5. Sequence generator
6. XML Source qualifier

6. What is a lookup transformation?

Used to look up data in a relational table, views, or synonym, The
informatica server queries the lookup table based on the lookup ports in the
transformation. It compares lookup transformation port values to lookup
table column values based on the lookup condition. The result is passed to
other transformations and the target.

Used to :
Get related value
Perform a calculation
Update slowly changing dimension tables.
Diff between connected and unconnected lookups. Which is better?
Connected :
Received input values directly from the pipeline
Can use Dynamic or static cache.
Cache includes all lookup columns used in the mapping
Can return multiple columns from the same row
If there is no match , can return default values
Default values can be specified.

Un connected :

Receive input values from the result of a LKP expression in another transformation.

Only static cache can be used.

Cache includes all lookup/output ports in the lookup condition and lookup or return port.

Can return only one column from each row.

If there is no match it returns null.

Default values cannot be specified.

Explain various caches :

Static:

Caches the lookup table before executing the transformation.

Rows are not added dynamically.

Dynamic:

Caches the rows as and when it is passed.

Unshared:

Within the mapping if the lookup table is used in more than
one transformation then the cache built for the first lookup can be used for
the others. It cannot be used across mappings.

Shared:

If the lookup table is used in more than one
transformation/mapping then the cache built for the first lookup can be used
for the others. It can be used across mappings.

Persistent :

If the cache generated for a Lookup needs to be preserved
for subsequent use then persistent cache is used. It will not delete the
index and data files. It is useful only if the lookup table remains
constant.

What are the uses of index and data caches?

The conditions are stored in index cache and records from
the lookup are stored in data cache

7. Explain aggregate transformation?

The aggregate transformation allows you to perform aggregate calculations,
such as averages, sum, max, min etc. The aggregate transformation is unlike
the Expression transformation, in that you can use the aggregator
transformation to perform calculations in groups. The expression
transformation permits you to perform calculations on a row-by-row basis
only.

Performance issues ?

The Informatica server performs calculations as it reads and stores
necessary data group and row data in an aggregate cache.
 Create Sorted input ports and pass the input records to aggregator in
sorted forms by groups then by port

Incremental aggregation?

In the Session property tag there is an option for
performing incremental aggregation. When the Informatica server performs
incremental aggregation , it passes new source data through the mapping and
uses historical cache (index and data cache) data to perform new aggregation
calculations incrementally.

What are the uses of index and data cache?
The group data is stored in index files and Row data stored
in data files.

8. Explain update strategy?
Update strategy defines the sources to be flagged for insert, update,
delete, and reject at the targets.
What are update strategy constants?
DD_INSERT,0   DD_UPDATE,1   DD_DELETE,2
DD_REJECT,3

If DD_UPDATE is defined in update strategy and Treat source
rows as INSERT in Session . What happens?
Hints: If in Session anything other than DATA DRIVEN is
mentions then Update strategy in the mapping is ignored.

**What are the three areas where the rows can be flagged for
particular treatment?**
In mapping, In Session treat Source Rows and In Session
Target Options.

What is the use of Forward/Reject rows in Mapping?

9. Explain the expression transformation ?
Expression transformation is used to calculate values in a single row before
writing to the target.
What are the default values for variables?
Hints: Straing = Null, Number = 0, Date = 1/1/1753

10. Difference between Router and filter transformation?
In filter transformation the records are filtered based on the condition and
rejected rows are discarded. In Router the multiple conditions are placed
and the rejected rows can be assigned to a port.

How many ways you can filter the records?
1. Source Qualifier
2. Filter transformation
3. Router transformation
4. Ranker
5. Update strategy
.

11. How do you call stored procedure and external procedure transformation ?
External Procedure can be called in the Pre-session and post session tag in
the Session property sheet.
Store procedures are to be called in the mapping designer by three methods
1. Select the icon and add a Stored procedure transformation
2. Select transformation - Import Stored Procedure
3. Select Transformation - Create and then select stored procedure.

12. Explain Joiner transformation and where it is used?
While a Source qualifier transformation can join data originating from a
common source database, the joiner transformation joins two related
heterogeneous sources residing in different locations or file systems.
Two relational tables existing in separate databases

Two flat files in different file systems.
Two different ODBC sources
In one transformation how many sources can be coupled?
Two sources can be couples. If more than two is to be couples add another
Joiner in the hierarchy.
What are join options?
Normal (Default)
Master Outer
Detail Outer
Full Outer

13. Explain Normalizer transformation?
The normaliser transformation normalises records from COBOL and relational
sources, allowing you to organise the data according to your own needs. A
Normaliser transformation can appear anywhere in a data flow when you
normalize a relational source. Use a Normaliser transformation instead of
the Source Qualifier transformation when you normalize COBOL source. When
you drag a COBOL source into the Mapping Designer Workspace, the Normaliser
transformation appears, creating input and output ports for every columns in
the source.

14. What is Source qualifier transformation?
When you add relational or flat file source definition to a mapping , you
need to connect to a source Qualifier transformation. The source qualifier
represents the records that the informatica server reads when it runs a
session.
Join Data originating from the same source database.
Filter records when the Informatica server reads the source data.
Specify an outer join rather than the default inner join.
Specify sorted ports
Select only distinct values from the source
Create a custom query to issue a special SELECT statement for the
Informatica server to read the source data.

15. What is Ranker transformation?
Filters the required number of records from the top or from the bottom.

16. What is target load option?
It defines the order in which informatica server loads the data into the
targets.
This is to avoid integrity constraint violations

17. How do you identify the bottlenecks in Mappings?
Bottlenecks can occur in
1. Targets
    The most common performance bottleneck occurs when the
informatica server writes to a target
    database. You can identify target bottleneck by
configuring the session to write to a flat file target.
    If the session performance increases significantly when
you write to a flat file, you have a target
    bottleneck.

Solution :
Drop or Disable index or constraints
Perform bulk load (Ignores Database log)
Increase commit interval (Recovery is compromised)
Tune the database for RBS, Dynamic Extension etc.,

2. Sources
Set a filter transformation after each SQ and see the records are not through.
If the time taken is same then there is a problem.
You can also identify the Source problem by
Read Test Session - where we copy the mapping with sources, SQ and remove all transformations
and connect to file target. If the performance is same then there is a Source bottleneck.
Using database query - Copy the read query directly from the log. Execute the query against the
source database with a query tool. If the time it takes to execute the query and the time to fetch
the first row are significantly different, then the query can be modified using optimizer hints.
Solutions:
Optimize Queries using hints.
Use indexes wherever possible.

3. Mapping
If both Source and target are OK then problem could be in mapping.
Add a filter transformation before target and if the time is the same then there is a problem.
(OR) Look for the performance monitor in the Sessions property sheet and view the counters.
Solutions:
If High error rows and rows in lookup cache indicate a mapping bottleneck.
Optimize Single Pass Reading:
Optimize Lookup transformation :
    1. Caching the lookup table:
        When caching is enabled the informatica server caches the lookup table and queries the
        cache during the session. When this option is not enabled the server queries the lookup
        table on a row-by row basis.
        Static, Dynamic, Shared, Un-shared and Persistent cache
    2. Optimizing the lookup condition
        Whenever multiple conditions are placed, the condition with equality sign should take
        precedence.
    3. Indexing the lookup table
        The cached lookup table should be indexed on order by columns. The session log contains
        the ORDER BY statement
        The un-cached lookup since the server issues a SELECT statement for each row passing

into lookup transformation, it is better to index the lookup table on the columns in the condition

Optimize Filter transformation:
You can improve the efficiency by filtering early in the data flow. Instead of using a filter transformation halfway through the mapping to remove a sizable amount of data.
Use a source qualifier filter to remove those same rows at the source,
If not possible to move the filter into SQ, move the filter transformation as close to the source
qualifier as possible to remove unnecessary data early in the data flow.
Optimize Aggregate transformation:
1. Group by simpler columns. Preferably numeric columns.
2. Use Sorted input. The sorted input decreases the use of aggregate caches. The server assumes all input data are sorted and as it reads it performs aggregate calculations.
3. Use incremental aggregation in session property sheet.
Optimize Seq. Generator transformation:
1. Try creating a reusable Seq. Generator transformation and use it in multiple mappings
2. The number of cached value property determines the number of values the informatica server caches at one time.
Optimize Expression transformation:
1. Factoring out common logic
2. Minimize aggregate function calls.
3. Replace common sub-expressions with local variables.
4. Use operators instead of functions.

4. Sessions
If you do not have a source, target, or mapping bottleneck, you may have a session bottleneck.
You can identify a session bottleneck by using the performance details. The informatica server
creates performance details when you enable Collect Performance Data on the General Tab of
the session properties.
Performance details display information about each Source Qualifier, target definitions, and
individual transformation. All transformations have some basic counters that indicate the
Number of input rows, output rows, and error rows.
Any value other than zero in the readfromdisk and writetodisk counters for Aggregate, Joiner,
or Rank transformations indicate a session bottleneck.
Low bufferInput_efficiency and BufferOutput_efficiency

counter also indicate a session
   bottleneck.
   Small cache size, low buffer memory, and small commit
intervals can cause session bottlenecks.
5. System (Networks)

18. How to improve the Session performance?
1 Run concurrent sessions
2 Partition session (Power center)
3. Tune Parameter - DTM buffer pool, Buffer block size, Index cache size,
data cache size, Commit Interval, Tracing level (Normal, Terse, Verbose
Init, Verbose Data)
 The session has memory to hold 83 sources and targets. If it is more, then
DTM can be increased.
 The informatica server uses the index and data caches for Aggregate, Rank,
Lookup and Joiner
 transformation. The server stores the transformed data from the above
transformation in the data
 cache before returning it to the data flow. It stores group information for
those transformations in
 index cache.
 If the allocated data or index cache is not large enough to store the date,
the server stores the data
 in a temporary disk file as it processes the session data. Each time the
server pages to the disk the
 performance slows. This can be seen from the counters .
 Since generally data cache is larger than the index cache, it has to be
more than the index.
4. Remove Staging area
5. Tune off Session recovery
6. Reduce error tracing

19. What are tracing levels?
Normal-default
Logs initialization and status information, errors
encountered, skipped rows due to transformation errors, summarizes session
results but not at the row level.
Terse
Log initialization, error messages, notification of rejected
data.
Verbose Init.
In addition to normal tracing levels, it also logs
additional initialization information, names of index and data files used
and detailed transformation statistics.
Verbose Data.
In addition to Verbose init, It records row level logs.

20. What is Slowly changing dimensions?
Slowly changing dimensions are dimension tables that have
slowly increasing data as well as updates to existing data.
21. What are mapping parameters and variables?
A mapping parameter is a user definable constant that takes up a value
before running a session. It can be used in SQ expressions, Expression
transformation etc.
Steps:

Define the parameter in the mapping designer - parameter & variables .
Use the parameter in the Expressions.
Define the values for the parameter in the parameter file.

A mapping variable is also defined similar to the parameter except that the
value of the variable is subjected to change.
It picks up the value in the following order.
1. From the Session parameter file
2. As stored in the repository object in the previous run.
3. As defined in the initial values in the designer.
4. Default values
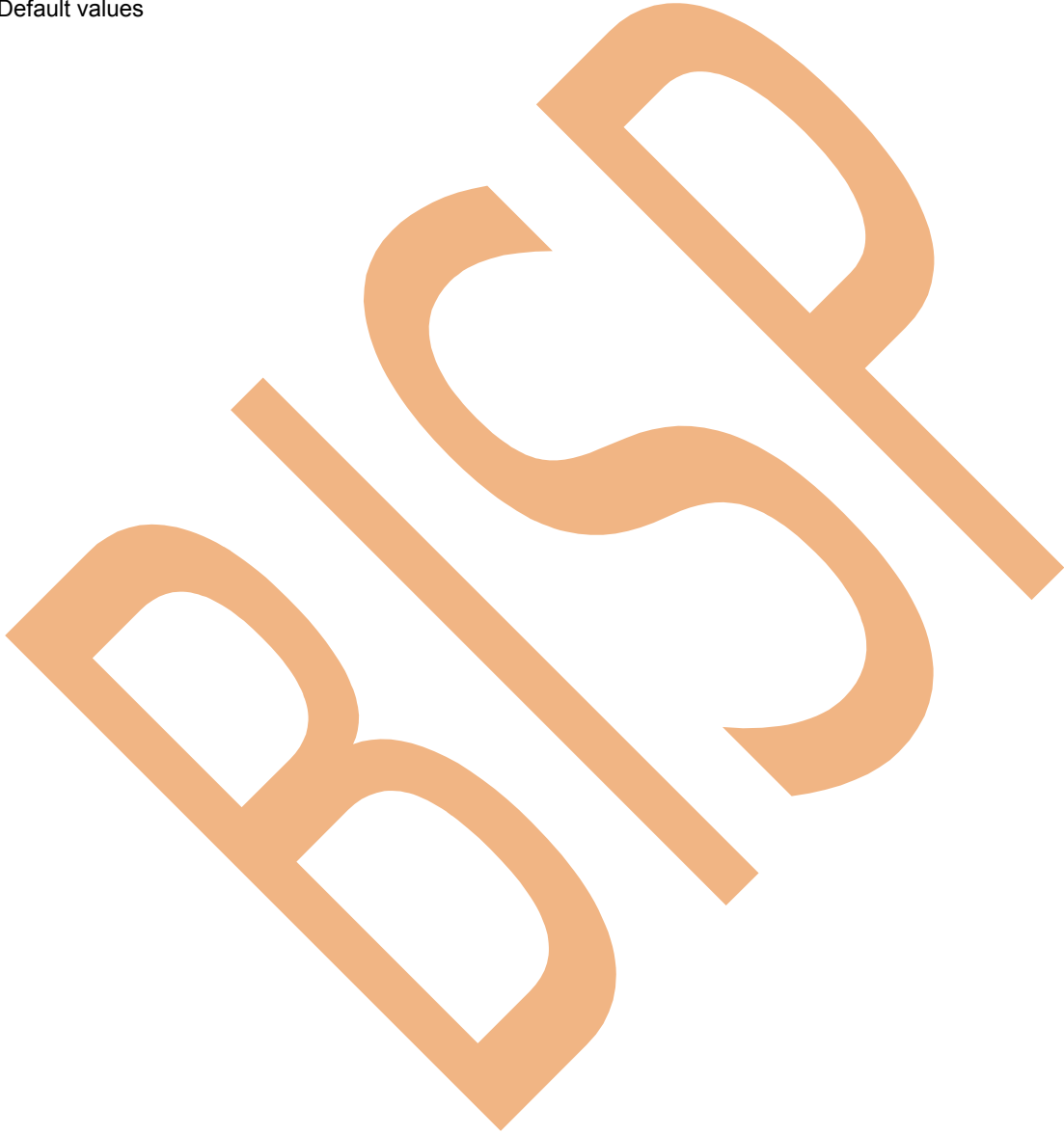
Q. How many types of Sql Statements are there in Oracle?

There are basically 6 types of sql statements. They are:

      a) Data Definition Language (DDL) The DDL statements define and maintain objects and drop objects.

      b) Data Manipulation Language (DML) The DML statements manipulate database data.

      c) Transaction Control Statements Manage change by DML

      d) Session Control -Used to control the properties of current session enabling and disabling roles and changing.
E.g. Alter Statements, Set Role

      e) System Control Statements-Change Properties of Oracle Instance. E.g. Alter System

      f) Embedded Sql- Incorporate DDL, DML and TCS in Programming Language. E.g. Using the Sql Statements in languages such as 'C', Open, Fetch, execute and close

Q. What is a Join?

A **join** is a query that combines rows from two or more tables, views, or materialized views ("snapshots"). Oracle performs a join whenever multiple tables appear in the queries FROM clause. The query's select list can select any columns from any of these tables. If any two of these tables have a column name in common, you must qualify all references to these columns throughout the query with table names to avoid ambiguity.

Q. What are join conditions?

Most join queries contain WHERE clause conditions that compare two columns, each from a different table. Such a condition is called a **join condition**. To execute a join, Oracle combines pairs of rows, each containing one row from each table, for which the join condition evaluates to TRUE. The columns in the join conditions need not also appear in the select list.

Q. What is an equijoin?

An **equijoin** is a join with a join condition containing an equality operator. An equijoin combines rows that have equivalent values for the specified columns.
Eg:
Select ename, job, dept.deptno, dname   From emp, dept   Where **emp.deptno = dept.deptno**;

Q. What are self joins?

A **self join** is a join of a table to itself. This table appears twice in the FROM clause and is followed by table aliases that qualify column names in the join condition.
Eg: SELECT e.ename || ' works for ' || e2.name "Employees and their Managers"
    FROM emp e1, emp e2 WHERE e1.mgr = e2.empno;

ENAME	EMPNO	MGR
BLAKE	12345	67890
KING	67890	22446

Result: BLAKE works for KING

Q. What is an Outer Join?

An outer join extends the result of a simple join. An outer join returns all rows that satisfy the join condition and those rows from one table for which no rows from the other satisfy the join condition. Such rows are not returned by a simple join. To write a query that performs an outer join of tables A and B and returns all rows from A, apply the outer join operator (+) to all columns of B in the join condition.
For all rows in A that have no matching rows in B, Oracle returns null for any select list expressions containing columns of B.

Outer join queries are subject to the following rules and restrictions:

❖ The (+) operator can appear only in the WHERE clause or, in the context of left correlation (that is, when specifying the TABLE clause) in the FROM clause, and can be applied only to a column of a table or view.
❖ If A and B are joined by multiple join conditions, you must use the (+) operator in all of these conditions. If you do not, Oracle will return only the rows resulting from a simple join, but without a warning or error to advise you that you do not have the results of an outer join.
❖ The (+) operator can be applied only to a column, not to an arbitrary expression. However, an arbitrary expression can contain a column marked with the (+) operator.
❖ A condition containing the (+) operator cannot be combined with another condition using the OR logical operator.
❖ A condition cannot use the IN comparison operator to compare a column marked with the (+) operator with an expression.
❖ A condition cannot compare any column marked with the (+) operator with a subquery.

If the WHERE clause contains a condition that compares a column from table B with a constant, the (+) operator must be applied to the column so that Oracle returns the rows from table A for which it has generated NULLs for this column. Otherwise Oracle will return only the results of a simple join.

In a query that performs outer joins of more than two pairs of tables, a single table can be the null-generated table for only one other table. For this reason, you cannot apply the (+) operator to columns of B in the join condition for A and B and the join condition for B and C.

Set Operators: UNION [ALL], INTERSECT, MINUS
Set operators combine the results of two component queries into a single result. Queries containing set operators are called compound queries.
The number and datatypes of the columns selected by each component query must be the same, but the column lengths can be different.

If you combine more than two queries with set operators, Oracle evaluates adjacent queries from left to right. You can use parentheses to specify a different order of evaluation.

Restrictions:

❖ These set operators are not valid on columns of type BLOB, CLOB, BFILE, varray, or nested table.
❖ The UNION, INTERSECT, and MINUS operators are not valid on LONG columns.
❖ To reference a column, you must use an alias to name the column.
❖ You cannot also specify the for_update_clause with these set operators.
❖ You cannot specify the order_by_clause in the subquery of these operators.

Table 3–9   Set Operators

Operator	Returns
UNION	All rows selected by either query.
UNION ALL	All rows selected by either query, including all duplicates.
INTERSECT	All distinct rows selected by both queries.
MINUS	All distinct rows selected by the first query but not the second.

All set operators have equal precedence. If a SQL statement contains multiple set operators, Oracle evaluates them from the left to right if no parentheses explicitly specify another order.

The corresponding expressions in the select lists of the component queries of a compound query must match in number and datatype. If component queries select character data, the datatype of the return values are determined as follows:

❖ If both queries select values of datatype CHAR, the returned values have datatype CHAR.
❖ If either or both of the queries select values of datatype VARCHAR2, the returned values have datatype VARCHAR2.

Q. What is a UNION?

The UNION operator eliminates duplicate records from the selected rows. We must match datatype (using the TO_DATE and TO_NUMBER functions) when columns do not exist in one or the other table.

Q. What is UNION ALL?
The UNION ALL operator does not eliminate duplicate selected rows.
Note: The UNION operator returns only distinct rows that appear in either result, while the UNION ALL operator returns all rows.

Q. What is an INTERSECT?
The INTERSECT operator returns only those rows returned by both queries. It shows only the distinct values from the rows returned by both queries.

Q. What is MINUS?
The MINUS operator returns only rows returned by the first query but not by the second. It also eliminates the duplicates from the first query.
Note: For compound queries (containing set operators UNION, INTERSECT, MINUS, or UNION ALL), the ORDER BY clause must use positions, rather than explicit expressions. Also, the ORDER BY clause can appear only in the last component query. The ORDER BY clause orders all rows returned by the entire compound query.

Q) What is a Transaction in Oracle?
A transaction is a Logical unit of work that compromises one or more SQL Statements executed by a single User. According to ANSI, a transaction begins with first executable statement and ends when it is explicitly committed or rolled back.
A transaction is an atomic unit.

Q. What are some of the Key Words Used in Oracle?
Some of the Key words that are used in Oracle are:
A) **Committing**: A transaction is said to be committed when the transaction makes permanent   changes resulting from the SQL statements.
b) **Rollback**: A transaction that retracts any of the changes resulting from SQL statements in Transaction.
c) **SavePoint**: For long transactions that contain many SQL statements, intermediate markers or savepoints are declared. Savepoints can be used to divide a transaction into smaller points.

We can declare intermediate markers called savepoints within the context of a transaction. Savepoints divide a long transaction into smaller parts. Using savepoints, we can arbitrarily mark our work at any point within a long transaction. We then have the option later of rolling back work performed before the current point in the transaction but after a declared savepoint within the transaction.
For example, we can use savepoints throughout a long complex series of updates so that if we make an error, we do not need to resubmit every statement.

d) **Rolling Forward**: Process of applying redo log during recovery is called rolling forward.
e) **Cursor**: A cursor is a handle (name or a pointer) for the memory associated with a specific statement. A cursor is basically an area allocated by Oracle for executing the Sql Statement. Oracle uses an implicit cursor statement for Single row query and Uses Explicit cursor for a multi row query.
f) **System Global Area (SGA)**: The SGA is a shared memory region allocated by the Oracle that contains Data and control information for one Oracle Instance. It consists of Database Buffer Cache and Redo log Buffer. (KPIT Infotech, Pune)
g) **Program Global Area (PGA)**: The PGA is a memory buffer that contains data and control information for server process.
g) **Database Buffer Cache**: Database Buffer of SGA stores the most recently used blocks of database data. The set of database buffers in an instance is called Database Buffer Cache.
h) **Redo log Buffer**: Redo log Buffer of SGA stores all the redo log entries.
i) **Redo Log Files**: Redo log files are set of files that protect altered database data in memory that has not been written to Data Files. They are basically used for backup when a database crashes.
j) **Process**: A Process is a 'thread of control' or mechanism in Operating System that executes series of steps.

Q. What are Procedure, functions and Packages?

Procedures and functions consist of set of PL/SQL statements that are grouped together as a unit to solve a specific problem or perform set of related tasks.

Procedures do not return values while Functions return one Value.

Packages: Packages provide a method of encapsulating and storing related procedures, functions, variables and other Package Contents

Q. What are Database Triggers and Stored Procedures?

Database Triggers: Database Triggers are Procedures that are automatically executed as a result of insert in, update to, or delete from table. Database triggers have the values old and new to denote the old value in the table before it is deleted and the new indicated the new value that will be used. DT is useful for implementing complex business rules which cannot be enforced using the integrity rules. We can have the trigger as Before trigger or After Trigger and at Statement or Row level.

e.g:: operations insert, update ,delete   3   before ,after   3*2   A total of 6 combinations

    At statement level(once for the trigger) or row level(for every execution)   6 * 2   A total of 12.

  Thus a total of 12 combinations are there and the restriction of usage of 12 triggers has been lifted from Oracle 7.3 Onwards.

Stored Procedures: Stored Procedures are Procedures that are stored in Compiled form in the database. The advantage of using the stored procedures is that many users can use the same procedure in compiled and ready to use format.

Q. How many Integrity Rules are there and what are they?

There are Three Integrity Rules. They are as follows:

      a) Entity Integrity Rule: The Entity Integrity Rule enforces that the Primary key cannot be Null

      b) Foreign Key Integrity Rule: The FKIR denotes that the relationship between the foreign key and the primary key has to be enforced. When there is data in Child Tables the Master tables cannot be deleted.

      c) Business Integrity Rules: The Third Integrity rule is about the complex business processes which cannot be implemented by the above 2 rules.

Q. What are the Various Master and Detail Relationships?

The various Master and Detail Relationship are

      a) No Isolated   : The Master cannot be deleted when a child is existing

      b) Isolated   : The Master can be deleted when the child is existing

      c) Cascading   : The child gets deleted when the Master is deleted.

Q. What are the Various Block Coordination Properties?

The various Block Coordination Properties are:

a) Immediate - Default Setting. The Detail records are shown when the Master Record are shown.

b) Deferred with Auto Query- Oracle Forms defer fetching the detail records until the operator navigates to the detail block.

c) Deferred with No Auto Query- The operator must navigate to the detail block and explicitly execute a query

Q. What are the Different Optimization Techniques?

The Various Optimization techniques are:

      a) Execute Plan: we can see the plan of the query and change it accordingly based on the indexes

      b) Optimizer_hint: set_item_property ('DeptBlock',OPTIMIZER_HINT,'FIRST_ROWS');

      Select /*+ First_Rows */ Deptno,Dname,Loc,Rowid from dept where (Deptno > 25)

      c) Optimize_Sql: By setting the Optimize_Sql = No, Oracle Forms assigns a single cursor for all SQL statements. This slow downs the processing because for every time the SQL must be parsed whenever they are executed.

      f45run module = my_firstform userid = scott/tiger optimize_sql = No

d) Optimize_Tp:
   By setting the Optimize_Tp= No, Oracle Forms assigns seperate cursor only for each query SELECT statement. All other SQL statements reuse the cursor.
f45run module = my_firstform userid = scott/tiger optimize_Tp = No

Q. How do u implement the If statement in the Select Statement?
We can implement the if statement in the select statement by using the Decode statement.
e.g select   DECODE (EMP_CAT,'1','First','2','Second', Null);

Q. How many types of Exceptions are there?
There are 2 types of exceptions. They are:
   a) System Exceptions
   e.g. When no_data_found, When too_many_rows
   b) User Defined Exceptions
   e.g. My_exception exception
   When My_exception then

Q. What are the inline and the precompiler directives?
The inline and precompiler directives detect the values directly.

Q. How do you use the same lov for 2 columns?
We can use the same lov for 2 columns by passing the return values in global values and using the global values in the code.

Q. How many minimum groups are required for a matrix report?
The minimum number of groups in matrix report is 4.

Q. What is the difference between static and dynamic lov?
The static lov contains the predetermined values while the dynamic lov contains values that come at run time.

Q. What are the OOPS concepts in Oracle?
Oracle does implement the OOPS concepts. The best example is the Property Classes. We can categorize the properties by setting the visual attributes and then attach the property classes for the objects. OOPS supports the concepts of objects and classes and we can consider the property classes as classes and the items as objects

Q. What is the difference between candidate key, unique key and primary key?
Candidate keys are the columns in the table that could be the primary keys and the primary key is the key that has been selected to identify the rows. Unique key is also useful for identifying the distinct rows in the table.

Q. What is concurrency?
Concurrency is allowing simultaneous access of same data by different users. Locks useful for accessing the database are:
a) Exclusive - The exclusive lock is useful for locking the row when an insert, update or delete is being done. This lock should not be applied when we do only select from the row.
b) Share lock - We can do the table as Share_Lock and as many share_locks can be put on the same resource.

Q. What are Privileges and Grants?

Privileges are the right to execute a particular type of SQL statements.
E.g. Right to Connect, Right to create, Right to resource
Grants are given to the objects so that the object might be accessed accordingly. The grant has to be given by the owner of the object.

Q. What are Table Space, Data Files, Parameter File and Control Files?
Table Space: The table space is useful for storing the data in the database.

When a database is created two table spaces are created.
>   a) System Table space: This data file stores all the tables related to the system and dba tables
>   b) User Table space: This data file stores all the user related tables

We should have separate table spaces for storing the tables and indexes so that the access is fast.

Data Files: Every Oracle Data Base has one or more physical data files. They store the data for the database. Every data file is associated with only one database. Once the Data file is created the size cannot change. To increase the size of the database to store more data we have to add data file.

Parameter Files: Parameter file is needed to start an instance.A parameter file contains the list of instance configuration parameters.

e.g. db_block_buffers = 500   db_name = ORA7   db_domain = u.s.acme lang

Control Files: Control files record the physical structure of the data files and redo log files
They contain the Db name, name and location of dbs, data files, redo log files and time stamp.

Q. Some of the terms related to Physical Storage of the Data.

The finest level of granularity of the data base is the data blocks.

Data Block   : One Data Block correspond to specific number of physical database space

Extent   : Extent is the number of specific number of contiguous data blocks.

Segments   : Set of Extents allocated for Extents. There are three types of Segments.
>   a) Data Segment: Non Clustered Table has data segment data of every table is stored in cluster data segment
>   b) Index Segment: Each Index has index segment that stores data
>   c) Roll Back Segment: Temporarily store 'undo' information

Q. What are the Pct Free and Pct Used?

Pct Free is used to denote the percentage of the free space that is to be left when creating a table. Similarly Pct Used is used to denote the percentage of the used space that is to be used when creating a table E.g. Pctfree 20, Pctused 40

Q. What is Row Chaining?
The data of a row in a table may not be able to fit the same data block. Data for row is stored in a chain of data blocks.

Q. What is a 2 Phase Commit?
Two Phase commit is used in distributed data base systems. This is useful to maintain the integrity of the database so that all the users see the same values. It contains DML statements or Remote Procedural calls that reference a remote object. There are basically 2 phases in a 2 phase commit.
>   a) Prepare Phase: Global coordinator asks participants to prepare
>   b) Commit Phase: Commit all participants to coordinator to Prepared, Read only or abort Reply

A two-phase commit mechanism guarantees that all database servers participating in a distributed transaction either all commit or all roll back the statements in the transaction. A two-phase commit mechanism also protects implicit DML operations performed by integrity constraints, remote procedure calls, and triggers.

Q. What is the difference between deleting and truncating of tables?
Deleting   a table will not remove the rows from the table but entry is there in the database dictionary and it can be retrieved But truncating a table deletes it completely and it cannot be retrieved.

Q. What are mutating tables?
When a table is in state of transition it is said to be mutating. E.g. If a row has been deleted then the table is said to be mutating and no operations can be done on the table except select.

Q. What are Codd Rules?
Codd Rules describe the ideal nature of a RDBMS. No RDBMS satisfies all the 12 codd rules and Oracle Satisfies 11 of the 12 rules and is the only RDBMS to satisfy the maximum number of rules.

Q. What is Normalization?
Normalization is the process of organizing the tables to remove the redundancy. There are mainly 5 Normalization rules.
1 Normal Form - A table is said to be in 1st Normal Form when the attributes are atomic
2 Normal Form - A table is said to be in 2nd Normal Form when all the candidate keys are dependant on the primary key
3rd Normal Form - A table is said to be third Normal form when it is not dependant transitively

Q. What is the Difference between a post query and a pre query?

A post query will fire for every row that is fetched but the pre query will fire only once.

Q. How can we delete the duplicate rows in the table?
We can delete the duplicate rows in the table by using the Rowid.

Delete emp where rowid=(select max(rowid) from emp group by empno)

Delete emp a where rownum=(select max(rownum) from emp g where a.empno=b.empno)

Q. Can U disable database trigger? How?
Yes. With respect to table   ALTER TABLE TABLE   [DISABLE all_trigger]

Q. What are pseudocolumns? Name them?
A pseudocolumn behaves like a table column, but is not actually stored in the table. You can select from pseudocolumns, but you cannot insert, update, or delete their values. This section describes these pseudocolumns:
  * CURRVAL   * NEXTVAL   * LEVEL   * ROWID   * ROWNUM

Q. How many columns can table have?
    The number of columns in a table can range from 1 to 254.

Q. Is space acquired in blocks or extents?
In extents.

Q. What is clustered index?
In an indexed cluster, rows are stored together based on their cluster key values. Can not be applied for HASH.

Q. What are the datatypes supported By oracle (INTERNAL)?
varchar2, Number, Char, MLSLABEL.

Q. What are attributes of cursor?
%FOUND , %NOTFOUND , %ISOPEN,%ROWCOUNT

Q. Can you use select in FROM clause of SQL select ? Yes.
Q. Describe the difference between a procedure, function and anonymous pl/sql block.
Candidate should mention use of DECLARE statement, a function must return a value while a procedure doesn't have to.

Q. What is a mutating table error and how can you get around it?
This happens with triggers. It occurs because the trigger is trying to modify a row it is currently using. The usual fix involves either use of views or temporary tables so the database is selecting from one while updating the other.

Q. Describe the use of %ROWTYPE and %TYPE in PL/SQL.
%ROWTYPE allows you to associate a variable with an entire table row. The %TYPE associates a variable with a single column type.

Q. What packages (if any) has Oracle provided for use by developers?

Oracle provides the DBMS_ series of packages. There are many which developers should be aware of such as DBMS_SQL, DBMS_PIPE, DBMS_TRANSACTION, DBMS_LOCK, DBMS_ALERT, DBMS_OUTPUT, DBMS_JOB, DBMS_UTILITY, DBMS_DDL, UTL_FILE. If they can mention a few of these and describe how they used them, even better. If they include the SQL routines provided by Oracle, great, but not really what was asked.

Q. Describe the use of PL/SQL tables.
PL/SQL tables are scalar arrays that can be referenced by a binary integer. They can be used to hold values for use in later queries or calculations. In Oracle 8 they will be able to be of the %ROWTYPE designation, or RECORD.

Q. When is a declare statement needed?
The DECLARE statement is used in PL/SQL anonymous blocks such as with stand alone, non-stored PL/SQL procedures. It must come first in a PL/SQL standalone file if it is used.

Q. In what order should a open/fetch/loop set of commands in a PL/SQL block be implemented if you use the %NOTFOUND cursor variable in the exit when statement? Why?
OPEN then FETCH then LOOP followed by the exit when. If not specified in this order will result in the final return being done twice because of the way the %NOTFOUND is handled by PL/SQL.

Q. What are SQLCODE and SQLERRM and why are they important for PL/SQL developers?
SQLCODE returns the value of the error number for the last error encountered. The SQLERRM returns the actual error message for the last error encountered. They can be used in exception handling to report, or, store in an error log table, the error that occurred in the code. These are especially useful for the WHEN OTHERS exception.

Q. How can you find within a PL/SQL block, if a cursor is open?
Use the %ISOPEN cursor status variable.

Q. How can you generate debugging output from PL/SQL?
Use the DBMS_OUTPUT package. Another possible method is to just use the SHOW ERROR command, but this only shows errors. The DBMS_OUTPUT package can be used to show intermediate results from loops and the status of variables as the procedure is executed. The new package UTL_FILE can also be used.

Q. What are the types of triggers?
There are 12 types of triggers in PL/SQL that consist of combinations of the BEFORE, AFTER, ROW, TABLE, INSERT, UPDATE, DELETE and ALL key words:
BEFORE ALL ROW INSERT
AFTER ALL ROW INSERT
BEFORE INSERT
AFTER INSERT

Q. How can variables be passed to a SQL routine?
By use of the & or double && symbol. For passing in variables numbers can be used (&1, &2,...,&8) to pass the values after the command into the SQLPLUS session. To be prompted for a specific variable, place the ampersanded variable in the code itself:
 "select * from dba_tables where owner=&owner_name;" . Use of double ampersands tells SQLPLUS to resubstitute the value for each subsequent use of the variable, a single ampersand will cause a reprompt for the value unless an ACCEPT statement is used to get the value from the user.

Q. You want to include a carriage return/linefeed in your output from a SQL script, how can you do this?
The best method is to use the CHR() function (CHR(10) is a return/linefeed) and the concatenation function "||". Another method, although it is hard to document and isn't always portable is to use the return/linefeed as a part of a quoted string.

Q. How can you call a PL/SQL procedure from SQL?
By use of the EXECUTE (short form EXEC) command. You can also wrap the call in a BEGIN END block and treat it as an anonymous PL/SQL block.

Q. How do you execute a host operating system command from within SQL?
By use of the exclamation point "!" (in UNIX and some other OS) or the HOST (HO) command.

Q. You want to use SQL to build SQL, what is this called and give an example?
This is called dynamic SQL. An example would be:
set lines 90 pages 0 termout off feedback off verify off
spool drop_all.sql
select 'drop user '||username||' cascade;' from dba_users

where username not in ("SYS','SYSTEM');
spool off

Essentially you are looking to see that they know to include a command (in this case DROP USER...CASCADE;) and that you need to concatenate using the '||' the values selected from the database.

Q. What SQLPlus command is used to format output from a select?
This is best done with the COLUMN command.

Q. You want to group the following set of select returns, what can you group on?
Max(sum_of_cost), min(sum_of_cost), count(item_no), item_no
The only column that can be grouped on is the "item_no" column, the rest have aggregate functions associated with them.

Q. What special Oracle feature allows you to specify how the cost based system treats a SQL statement?
The COST based system allows the use of HINTs to control the optimizer path selection. If they can give some example hints such as FIRST ROWS, ALL ROWS, USING INDEX, STAR, even better.

Q. You want to determine the location of identical rows in a table before attempting to place a unique index on the table, how can this be done?
Oracle tables always have one guaranteed unique column, the rowid column. If you use a min/max function against your rowid and then select against the proposed primary key you can squeeze out the rowids of the duplicate rows pretty quick. For example:

select rowid from emp e   where e.rowid > (select min(x.rowid)
          from emp x where x.emp_no = e.emp_no);
In the situation where multiple columns make up the proposed key, they must all be used in the where clause.

Q. What is a Cartesian product?
A Cartesian product is the result of an unrestricted join of two or more tables. The result set of a three table Cartesian product will have x * y * z number of rows where x, y, z correspond to the number of rows in each table involved in the join. This occurs if there are not at least n-1 joins where n is the number of tables in a SELECT.

Q. You are joining a local and a remote table, the network manager complains about the traffic involved, how can you reduce the network traffic?
Push the processing of the remote data to the remote instance by using a view to pre-select the information for the join. This will result in only the data required for the join being sent across.

Q. What is the default ordering of an ORDER BY clause in a SELECT statement? Ascending
Q. What is tkprof and how is it used?
The tkprof tool is a tuning tool used to determine cpu and execution times for SQL statements. You use it by first setting timed_statistics to true in the initialization file and then turning on tracing for either the entire database via the sql_trace parameter or for the session using the ALTER SESSION command. Once the trace file is generated you run the tkprof tool against the trace file and then look at the output from the tkprof tool. This can also be used to generate explain plan output.

Q. What is explain plan and how is it used?
The EXPLAIN PLAN command is a tool to tune SQL statements. To use it you must have an explain_table generated in the user you are running the explain plan for. This is created using the utlxplan.sql script. Once the explain plan table exists you run the explain plan command giving as its argument the SQL statement to be explained. The explain_plan table is then queried to see the execution plan of the statement. Explain plans can also be run using tkprof.

Q. How do you set the number of lines on a page of output? The width?
The SET command in SQLPLUS is used to control the number of lines generated per page and the width of those lines, for example SET PAGESIZE 60 LINESIZE 80 will generate reports that are 60 lines long with a line width of 80 characters. The PAGESIZE and LINESIZE options can be shortened to PAGES and LINES.

Q. How do you prevent output from coming to the screen?
The SET option TERMOUT controls output to the screen. Setting TERMOUT OFF turns off screen output. This option can be shortened to TERM.

Q. How do you prevent Oracle from giving you informational messages during and after a SQL statement execution?
The SET options FEEDBACK and VERIFY can be set to OFF.

Q. How do you generate file output from SQL? By use of the SPOOL command.
Data Modeler:

Q. Describe third normal form?
Expected answer: Something like: In third normal form all attributes in an entity are related to the primary key and only to the primary key

Q. Is the following statement true or false? Why or why not?
"All relational databases must be in third normal form"
False. While 3NF is good for logical design most databases, if they have more than just a few tables, will not perform well using full 3NF. Usually some entities will be denormalized in the logical to physical transfer process.

Q. What is an ERD?
An ERD is an Entity-Relationship-Diagram. It is used to show the entities and relationships for a database logical model.

Q. Why are recursive relationships bad? How do you resolve them?
A recursive relationship (one where a table relates to itself) is bad when it is a hard relationship (i.e. neither side is a "may" both are "must") as this can result in it not being possible to put in a top or perhaps a bottom of the table (for example in the EMPLOYEE table you couldn't put in the PRESIDENT of the company because he has no boss, or the junior janitor because he has no subordinates). These type of relationships are usually resolved by adding a small intersection entity.

Q. What does a hard one-to-one relationship mean (one where the relationship on both ends is "must")?
This means the two entities should probably be made into one entity.

Q. How should a many-to-many relationship be handled? By adding an intersection entity table
Q. What is an artificial (derived) primary key? When should an artificial (or derived) primary key be used?
A derived key comes from a sequence. Usually it is used when a concatenated key becomes too cumbersome to use as a foreign key.

Q. When should you consider denormalization?
Whenever performance analysis indicates it would be beneficial to do so without compromising data integrity.

Q. What is a Schema?
Associated with each database user is a schema. A schema is a collection of schema objects. Schema objects include tables, views, sequences, synonyms, indexes, clusters, database links, snapshots, procedures, functions, and packages.

Q. What do you mean by table?
Tables are the basic unit of data storage in an Oracle database. Data is stored in rows and columns.
A row is a collection of column information corresponding to a single record.

Q. Is there an alternative of dropping a column from a table? If yes, what?
Dropping a column in a large table takes a considerable amount of time. A quicker alternative is to mark a column as unused with the SET UNUSED clause of the ALTER TABLE statement. This makes the column data unavailable, although the data remains in each row of the table. After marking a column as unused, you can add another column that has the same name to the table. The unused column can then be dropped at a later time when you want to reclaim the space occupied by the column data.

Q. What is a rowid?
The rowid identifies each row piece by its location or address. Once assigned, a given row piece retains its rowid until the corresponding row is deleted, or exported and imported using the Export and Import utilities.

Q. What is a view? (KPIT Infotech, Pune)
A view is a tailored presentation of the data contained in one or more tables or other views. A view takes the output of a query and treats it as a table. Therefore, a view can be thought of as a stored query or a virtual table.

Unlike a table, a view is not allocated any storage space, nor does a view actually contain data. Rather, a view is defined by a query that extracts or derives data from the tables that the view references. These tables are called base tables. Base tables can in turn be actual tables or can be views themselves (including snapshots). Because a view is based on

other objects, a view requires no storage other than storage for the definition of the view (the stored query) in the data dictionary.

Q. What are the advantages of having a view?
The advantages of having a view are:
- ❖ To provide an additional level of table security by restricting access to a predetermined set of rows or columns of a table
- ❖ To hide data complexity
- ❖ To simplify statements for the user
- ❖ To present the data in a different perspective from that of the base table
- ❖ To isolate applications from changes in definitions of base tables
- ❖ To save complex queries
  For example, a query can perform extensive calculations with table information.
  By saving this query as a view, you can perform the calculations each time the view is queried.

Q. What is a Materialized View? (Honeywell, KPIT Infotech, Pune)
Materialized views, also called snapshots, are schema objects that can be used to summarize, precompute, replicate, and distribute data. They are suitable in various computing environments especially for data warehousing.
From a physical design point of view, Materialized Views resembles tables or partitioned tables and behave like indexes.

Q. What is the significance of Materialized Views in data warehousing?
In data warehouses, materialized views are used to precompute and store aggregated data such as sums and averages. Materialized views in these environments are typically referred to as summaries because they store summarized data. They can also be used to precompute joins with or without aggregations.

Cost-based optimization can use materialized views to improve query performance by automatically recognizing when a materialized view can and should be used to satisfy a request. The optimizer transparently rewrites the request to use the materialized view. Queries are then directed to the materialized view and not to the underlying detail tables or views.

Q. Differentiate between Views and Materialized Views? (KPIT Infotech, Pune)

Q. What is the major difference between an index and Materialized view?
Unlike indexes, materialized views can be accessed directly using a SELECT statement.

Q. What are the procedures for refreshing Materialized views?
Oracle maintains the data in materialized views by refreshing them after changes are made to their master tables.
The refresh method can be:
  a) incremental (fast refresh) or
  b) complete
For materialized views that use the fast refresh method, a materialized view log or direct loader log keeps a record of changes to the master tables.

Materialized views can be refreshed either on demand or at regular time intervals.
Alternatively, materialized views in the same database as their master tables can be refreshed whenever a transaction commits its changes to the master tables.

Q. What are materialized view logs?
A materialized view log is a schema object that records changes to a master table's data so that a materialized view defined on the master table can be refreshed incrementally. Another name for materialized view log is snapshot log. Each materialized view log is associated with a single master table. The materialized view log resides in the same database and schema as its master table.

Q. What is a synonym?
A synonym is an alias for any table, view, snapshot, sequence, procedure, function, or package. Because a synonym is simply an alias, it requires no storage other than its definition in the data dictionary.

Q. What are the advantages of having synonyms?
Synonyms are often used for security and convenience.
For example, they can do the following:
1. Mask the name and owner of an object
2. Provide location transparency for remote objects of a distributed database
3. Simplify SQL statements for database users

Q. What are the advantages of having an index? Or What is an index?
The purpose of an index is to provide pointers to the rows in a table that contain a given key value. In a regular index, this is achieved by storing a list of rowids for each key corresponding to the rows with that key value. Oracle stores each key value repeatedly with each stored rowid.

Q. What are the different types of indexes supported by Oracle?
The different types of indexes are:
    a.   B-tree indexes
    b.   B-tree cluster indexes
    c.   Hash cluster indexes
    d.   Reverse key indexes
    e.   Bitmap indexes

Q. Can we have function based indexes?
Yes, we can create indexes on functions and expressions that involve one or more columns in the table being indexed. A function-based index precomputes the value of the function or expression and stores it in the index.
You can create a function-based index as either a B-tree or a bitmap index.

Q. What are the restrictions on function based indexes?
The function used for building the index can be an arithmetic expression or an expression that contains a PL/SQL function, package function, C callout, or SQL function. The expression cannot contain any aggregate functions, and it must be DETERMINISTIC. For building an index on a column containing an object type, the function can be a method of that object, such as a map method. However, you cannot build a function-based index on a LOB column, REF, or nested table column, nor can you build a function-based index if the object type contains a LOB, REF, or nested table.

Q. What are the advantages of having a B-tree index?
The major advantages of having a B-tree index are:
1. B-trees provide excellent retrieval performance for a wide range of queries, including exact match and range searches.
2. Inserts, updates, and deletes are efficient, maintaining key order for fast retrieval.
3. B-tree performance is good for both small and large tables, and does not degrade as the size of a table grows.

Q. What is a bitmap index? (KPIT Infotech, Pune)
The purpose of an index is to provide pointers to the rows in a table that contain a given key value. In a regular index, this is achieved by storing a list of rowids for each key corresponding to the rows with that key value. Oracle stores each key value repeatedly with each stored rowid. In a bitmap index, a bitmap for each key value is used instead of a list of rowids. Each bit in the bitmap corresponds to a possible rowid. If the bit is set, then it means that the row with the corresponding rowid contains the key value. A mapping function converts the bit position to an actual rowid, so the bitmap index provides the same functionality as a regular index even though it uses a different representation internally. If the number of different key values is small, then bitmap indexes are very space efficient.
Bitmap indexing efficiently merges indexes that correspond to several conditions in a WHERE clause. Rows that satisfy some, but not all, conditions are filtered out before the table itself is accessed. This improves response time, often dramatically.

Q. What are the advantages of having bitmap index for data warehousing applications? (KPIT Infotech, Pune)
Bitmap indexing benefits data warehousing applications which have large amounts of data and ad hoc queries but a low level of concurrent transactions. For such applications, bitmap indexing provides:
1. Reduced response time for large classes of ad hoc queries
2. A substantial reduction of space usage compared to other indexing techniques
3. Dramatic performance gains even on very low end hardware

4. Very efficient parallel DML and loads

Q. What is the advantage of bitmap index over B-tree index?
Fully indexing a large table with a traditional B-tree index can be prohibitively expensive in terms of space since the index can be several times larger than the data in the table. Bitmap indexes are typically only a fraction of the size of the indexed data in the table.

Q. What is the limitation/drawback of a bitmap index?
Bitmap indexes are not suitable for OLTP applications with large numbers of concurrent transactions modifying the data. These indexes are primarily intended for decision support in data warehousing applications where users typically query the data rather than update it.
Bitmap indexes are not suitable for high-cardinality data.

Q. How do you choose between B-tree index and bitmap index?
The advantages of using bitmap indexes are greatest for low cardinality columns: that is, columns in which the number of distinct values is small compared to the number of rows in the table. If the values in a column are repeated more than a hundred times, then the column is a candidate for a bitmap index. Even columns with a lower number of repetitions and thus higher cardinality, can be candidates if they tend to be involved in complex conditions in the WHERE clauses of queries.

For example, on a table with one million rows, a column with 10,000 distinct values is a candidate for a bitmap index. A bitmap index on this column can out-perform a B-tree index, particularly when this column is often queried in conjunction with other columns.

B-tree indexes are most effective for high-cardinality data: that is, data with many possible values, such as CUSTOMER_NAME or PHONE_NUMBER. A regular Btree index can be several times larger than the indexed data. Used appropriately, bitmap indexes can be significantly smaller than a corresponding B-tree index.

Q. What are clusters?
Clusters are an optional method of storing table data. A cluster is a group of tables that share the same data blocks because they share common columns and are often used together.
For example, the EMP and DEPT table share the DEPTNO column. When you cluster the EMP and DEPT tables, Oracle physically stores all rows for each department from both the EMP and DEPT tables in the same data blocks.

Q. What is partitioning? (KPIT Infotech, Pune)
Partitioning addresses the key problem of supporting very large tables and indexes by allowing you to decompose them into smaller and more manageable pieces called partitions. Once partitions are defined, SQL statements can access and manipulate the partitions rather than entire tables or indexes. Partitions are especially useful in data warehouse applications, which commonly store and analyze large amounts of historical data.

Q. What are the different partitioning methods?
Two primary methods of partitioning are available:
1. range partitioning, which partitions the data in a table or index according to a range of values, and
2. hash partitioning, which partitions the data according to a hash function.

Another method, composite partitioning, partitions the data by range and further subdivides the data into sub partitions using a hash function.

Q. What is the necessity to have table partitions?
The need to partition large tables is driven by:
- Data Warehouse and Business Intelligence demands for ad hoc analysis on great quantities of historical data
- Cheaper disk storage
- Application performance failure due to use of traditional techniques

Q. What are the advantages of storing each partition in a separate tablespace?
The major advantages are:

1. You can contain the impact of data corruption.
2. You can back up and recover each partition or subpartition independently.
3. You can map partitions or subpartitions to disk drives to balance the I/O load.

Q. What are the advantages of partitioning?
Partitioning is useful for:
1. Very Large Databases (VLDBs)
2. Reducing Downtime for Scheduled Maintenance
3. Reducing Downtime Due to Data Failures
4. DSS Performance
5. I/O Performance
6. Disk Striping: Performance versus Availability
7. Partition Transparency

Q. What is Range Partitioning? (KPIT Infotech, Pune)
Range partitioning maps rows to partitions based on ranges of column values. Range partitioning is defined by the partitioning specification for a table or index:
PARTITION BY RANGE (column_list) and by the partitioning specifications for each individual partition:
VALUES LESS THAN (value_list)

Q. What is Hash Partitioning?
Hash partitioning uses a hash function on the partitioning columns to stripe data into partitions. Hash partitioning allows data that does not lend itself to range partitioning to be easily partitioned for performance reasons such as parallel DML, partition pruning, and partition-wise joins.

Q. What are the advantages of Hash partitioning over Range Partitioning?
Hash partitioning is a better choice than range partitioning when:
a) You do not know beforehand how much data will map into a given range
b) Sizes of range partitions would differ quite substantially
c) Partition pruning and partition-wise joins on a partitioning key are important

Q. What are the rules for partitioning a table?
A table can be partitioned if:
    – It is not part of a cluster
    – It does not contain LONG or LONG RAW datatypes

Q. What is a global partitioned index?
In a global partitioned index, the keys in a particular index partition may refer to rows stored in more than one underlying table partition or subpartition. A global index can only be range-partitioned, but it can be defined on any type of partitioned table.

Q. What is a local index?
In a local index, all keys in a particular index partition refer only to rows stored in a single underlying table partition. A local index is created by specifying the LOCAL attribute.

Q. What are CLOB and NCLOB datatypes? (Mascot)
The CLOB and NCLOB datatypes store up to four gigabytes of character data in the database. CLOBs store single-byte character set data and NCLOBs store fixed-width and varying-width multibyte national character set data (NCHAR data).

Q. What is PL/SQL?
PL/SQL is Oracle's procedural language extension to SQL. PL/SQL enables you to mix SQL statements with procedural constructs. With PL/SQL, you can define and execute PL/SQL program units such as procedures, functions, and packages.
PL/SQL program units generally are categorized as anonymous blocks and stored procedures.

Q. What is an anonymous block?
An anonymous block is a PL/SQL block that appears within your application and it is not named or stored in the database.

Q. What is a Stored Procedure?

A stored procedure is a PL/SQL block that Oracle stores in the database and can be called by name from an application. When you create a stored procedure, Oracle parses the procedure and stores its parsed representation in the database.

Q. What is a distributed transaction?

A distributed transaction is a transaction that includes one or more statements that update data on two or more distinct nodes of a distributed database.

Q. What are packages? (KPIT Infotech, Pune)

A package is a group of related procedures and functions, together with the cursors and variables they use, stored together in the database for continued use as a unit.

While packages allow the administrator or application developer the ability to organize such routines, they also offer increased functionality (for example, global package variables can be declared and used by any procedure in the package) and performance (for example, all objects of the package are parsed, compiled, and loaded into memory once).

Q. What are procedures and functions? (KPIT Infotech, Pune)

A procedure or function is a schema object that consists of a set of SQL statements and other PL/SQL constructs, grouped together, stored in the database, and executed as a unit to solve a specific problem or perform a set of related tasks. Procedures and functions permit the caller to provide parameters that can be input only, output only, or input and output values.

Q. What is the difference between Procedure and Function?

Procedures and functions are identical except that functions always return a single value to the caller, while procedures do not return values to the caller.

Q. What is a DML and what do they do?

Data manipulation language (DML) statements query or manipulate data in existing schema objects. They enable you to:

1. Retrieve data from one or more tables or views (SELECT)
2. Add new rows of data into a table or view (INSERT)
3. Change column values in existing rows of a table or view (UPDATE)
4. Remove rows from tables or views (DELETE)
5. See the execution plan for a SQL statement (EXPLAIN PLAN)
6. Lock a table or view, temporarily limiting other users' access (LOCK TABLE)

Q. What is a DDL and what do they do?

Data definition language (DDL) statements define, alter the structure of, and drop schema objects. DDL statements enable you to:

1. Create, alter, and drop schema objects and other database structures, including the database itself and database users (CREATE, ALTER, DROP)
2. Change the names of schema objects (RENAME)
3. Delete all the data in schema objects without removing the objects' structure (TRUNCATE)
4. Gather statistics about schema objects, validate object structure, and list chained rows within objects (ANALYZE)
5. Grant and revoke privileges and roles (GRANT, REVOKE)
6. Turn auditing options on and off (AUDIT, NOAUDIT)
7. Add a comment to the data dictionary (COMMENT)

Q. What are shared sql's?

Oracle automatically notices when applications send identical SQL statements to the database. The SQL area used to process the first occurrence of the statement is shared—that is, used for processing subsequent occurrences of that same statement. Therefore, only one shared SQL area exists for a unique statement. Since shared SQL areas are shared memory areas, any Oracle process can use a shared SQL area. The sharing of SQL areas reduces memory usage on the database server, thereby increasing system throughput.

Q. What are triggers?

Oracle allows to define procedures called triggers that execute implicitly when an INSERT, UPDATE, or DELETE statement is issued against the associated table or, in some cases, against a view, or when database system actions occur. These procedures can be written in PL/SQL or Java and stored in the database, or they can be written as C callouts.

Q. What is Cost-based Optimization?
Using the cost-based approach, the optimizer determines which execution plan is most efficient by considering available access paths and factoring in information based on statistics for the schema objects (tables or indexes) accessed by the SQL statement.

Q. What is Rule-Based Optimization?
Using the rule-based approach, the optimizer chooses an execution plan based on the access paths available and the ranks of these access paths.

Q. What is meant by degree of parallelism?
The number of parallel execution servers associated with a single operation is known as the degree of parallelism.

Q. What is meant by data consistency?
Data consistency means that each user sees a consistent view of the data, including visible changes made by the user's own transactions and transactions of other users.

Q. What are Locks?
Locks are mechanisms that prevent destructive interaction between transactions accessing the same resource—either user objects such as tables and rows or system objects not visible to users, such as shared data structures in memory and data dictionary rows.

Q. What are the locking modes used in Oracle?
Oracle uses two modes of locking in a multiuser database:
Exclusive lock mode: Prevents the associates resource from being shared. This lock mode is obtained to modify data. The first transaction to lock a resource exclusively is the only transaction that can alter the resource until the exclusive lock is released.

Share lock mode: Allows the associated resource to be shared, depending on the operations involved. Multiple users reading data can share the data, holding share locks to prevent concurrent access by a writer (who needs an exclusive lock). Several transactions can acquire share locks on the same resource.

Q. What is a deadlock?
A deadlock can occur when two or more users are waiting for data locked by each other.

Q. How can you avoid deadlocks?
Multitable deadlocks can usually be avoided if transactions accessing the same tables lock those tables in the same order, either through implicit or explicit locks.
For example, all application developers might follow the rule that when both a master and detail table are updated, the master table is locked first and then the detail table. If such rules are properly designed and then followed in all applications, deadlocks are very unlikely to occur.

Q. What is redo log?
The redo log, present for every Oracle database, records all changes made in an Oracle database. The redo log of a database consists of at least two redo log files that are separate from the datafiles (which actually store a database's data). As part of database recovery from an instance or media failure, Oracle applies the appropriate changes in the database's redo log to the datafiles, which updates database data to the instant that the failure occurred.
A database's redo log can consist of two parts: the online redo log and the archived redo log.

Q. What are Rollback Segments?
Rollback segments are used for a number of functions in the operation of an Oracle database. In general, the rollback segments of a database store the old values of data changed by ongoing transactions for uncommitted transactions. Among other things, the information in a rollback segment is used during database recovery to undo any uncommitted changes applied from the redo log to the datafiles. Therefore, if database recovery is necessary, then the data is in a consistent state after the rollback segments are used to remove all uncommitted data from the datafiles.

Q. What is SGA?

The System Global Area (SGA) is a shared memory region that contains data and control information for one Oracle instance. An SGA and the Oracle background processes constitute an Oracle instance.

Oracle allocates the system global area when an instance starts and deallocates it when the instance shuts down. Each instance has its own system global area.

Users currently connected to an Oracle server share the data in the system global area. For optimal performance, the entire system global area should be as large as possible (while still fitting in real memory) to store as much data in memory as possible and minimize disk I/O.

The information stored within the system global area is divided into several types of memory structures, including the database buffers, redo log buffer, and the shared pool. These areas have fixed sizes and are created during instance startup.

Q. What is PCTFREE?

The PCTFREE parameter sets the minimum percentage of a data block to be reserved as free space for possible updates to rows that already exist in that block.

Q. What is PCTUSED?

The PCTUSED parameter sets the minimum percentage of a block that can be used for row data plus overhead before new rows will be added to the block. After a data block is filled to the limit determined by PCTFREE, Oracle considers the block unavailable for the insertion of new rows until the percentage of that block falls below the parameter PCTUSED. Until this value is achieved, Oracle uses the free space of the data block only for updates to rows already contained in the data block.

Notes:

Nulls are stored in the database if they fall between columns with data values. In these cases they require one byte to store the length of the column (zero).

Trailing nulls in a row require no storage because a new row header signals that the remaining columns in the previous row are null. For example, if the last three columns of a table are null, no information is stored for those columns. In tables with many columns, the columns more likely to contain nulls should be defined last to conserve disk space.

Two rows can both contain all nulls without violating a unique index.

NULL values in indexes are considered to be distinct except when all the non-NULL values in two or more rows of an index are identical, in which case the rows are considered to be identical. Therefore, UNIQUE indexes prevent rows containing NULL values from being treated as identical.

Bitmap indexes include rows that have NULL values, unlike most other types of indexes. Indexing of nulls can be useful for some types of SQL statements, such as queries with the aggregate function COUNT.

Bitmap indexes on partitioned tables must be local indexes.

PL/SQL is Oracle's procedural language extension to SQL. PL/SQL combines the ease and flexibility of SQL with the procedural functionality of a structured programming language, such as IF ... THEN, WHILE, and LOOP.

When designing a database application, a developer should consider the advantages of using stored PL/SQL:

❑Because PL/SQL code can be stored centrally in a database, network traffic between applications and the database is reduced, so application and system performance increases.

❑Data access can be controlled by stored PL/SQL code. In this case, the users of PL/SQL can access data only as intended by the application developer (unless another access route is granted).

❑PL/SQL blocks can be sent by an application to a database, executing complex

operations without excessive network traffic.

Even when PL/SQL is not stored in the database, applications can send blocks of
PL/SQL to the database rather than individual SQL statements, thereby again
reducing network traffic.

The following sections describe the different program units that can be defined and
stored centrally in a database.

Committing and Rolling Back Transactions

The changes made by the SQL statements that constitute a transaction can be either committed or rolled back. After a
transaction is committed or rolled back, the next transaction begins with the next SQL statement.

Committing a transaction makes permanent the changes resulting from all SQL statements in the transaction. The
changes made by the SQL statements of a transaction become visible to other user sessions' transactions that start only
after the transaction is committed.

Rolling back a transaction retracts any of the changes resulting from the SQL statements in the transaction. After a
transaction is rolled back, the affected data is left unchanged as if the SQL statements in the transaction were never
executed.

Introduction to the Data Dictionary

One of the most important parts of an Oracle database is its *data dictionary*, which is
a **read-only** set of tables that provides information about its associated database. A
data dictionary contains:

❏The definitions of all schema objects in the database (tables, views, indexes,
clusters, synonyms, sequences, procedures, functions, packages, triggers,
and so on)

❏How much space has been allocated for, and is currently used by, the
schema objects

❏Default values for columns

❏Integrity constraint information

❏The names of Oracle users

❏Privileges and roles each user has been granted

❏Auditing information, such as who has accessed or updated various
schema objects

❏Other general database information

The data dictionary is structured in tables and views, just like other database data.
All the data dictionary tables and views for a given database are stored in that
database's SYSTEM tablespace.

Not only is the data dictionary central to every Oracle database, it is an important
tool for all users, from end users to application designers and database
administrators. To access the data dictionary, you use SQL statements. Because the
data dictionary is read-only, you can issue only queries (SELECT statements)
against the tables and views of the data dictionary.

Q. What is the function of DUMMY table?

The table named DUAL is a small table in the data dictionary that Oracle and user written programs can reference to
guarantee a known result. This table has one column called DUMMY and one row containing the value "X".

Databases, tablespaces, and datafiels are closely related, but they have important differences:

Databases and tablespaces: An Oracle database consists of one or more logical storage units called tablespaces, which
collectively store all of the database's data.

Tablespaces and datafiles: Each table in an Oracle database consists of one or more files called datafiles, which are
physical structures that conform with the operating system in which Oracle is running.

databases and datafiles:
A database's data is collectively stored in the datafiles that
constitute each tablespace of the database. For example, the
simplest Oracle database would have one tablespace and one
datafile. Another database might have three tablespaces, each
consisting of two datafiles (for a total of six datafiles).

Nulls
A *null* is the absence of a value in a column of a row. Nulls indicate missing,
unknown, or inapplicable data. A null should not be used to imply any other value,
such as zero. A column allows nulls unless a NOT NULL or PRIMARY KEY
integrity constraint has been defined for the column, in which case no row can be
inserted without a value for that column.
Nulls are stored in the database if they fall between columns with data values. In
these cases they require one byte to store the length of the column (zero).
Trailing nulls in a row require no storage because a new row header signals that the
remaining columns in the previous row are null. For example, if the last three
columns of a table are null, no information is stored for those columns. In tables
with many columns, the columns more likely to contain nulls should be defined last
to conserve disk space.

Most comparisons between nulls and other values are by definition neither true nor
false, but unknown. To identify nulls in SQL, use the IS NULL predicate. Use the
SQL function NVL to convert nulls to non-null values.
Nulls are not indexed, except when the cluster key column value is null or the index
is a bitmap index.

What are different types of locks?

Q. Master table and Child table performances and comparisons in Oracle?

Q. What are the different types of Cursors? Explain. (Honeywell)

Q. What are the different types of Deletes?

Q. Can a View be updated?

Interview Questions from Honeywell

1. What is pragma?
2. Can you write commit in triggers?
3. Can you call user defined functions in select statements
4. Can you call insert/update/delete in select statements. If yes how? If no what is the other way?
5. After update how do you know, how many records got updated
6. Select statement does not retrieve any records. What exception is raised?

Interview Questions from Shreesoft
1. How many columns can a PLSQL table have
Interview Questions from mascot
   1. What is Load balancing & what u have used to do this? (SQL Loader)
   2. What r Routers?

   PL/SQL

1. What are different types of joins?
2. Difference between Packages and Procedures

3. Difference between Function and Procedures
4. How many types of triggers are there? When do you use Triggers
5. Can you write DDL statements in Triggers? (No)
6. What is Hint?
7. How do you tune a SQL query?

Interview Questions from KPIT Infotech, Pune

1. Package body
2. What is molar query?
3. What is row level security

General:
Why ORACLE is the best database for Datawarehousing
For data loading in Oracle, what are conventional loading and direct-path loading ?
7. If you use oracle SQL*Loader, how do you transform data with it during loading ? Example.
Three ways SQL*Loader could doad data, what are those three types ?
What are the contents of "bad files" and "discard files" when using SQL*Loader ?
How do you use commit frequencies ? how does it affect loading performance ?
What are the other factors of the database on which the loading performance depend ?
* WHAT IS PARALLELISM ?
* WHAT IS A PARALLEL QUERY ?
* WHAT ARE DIFFERENT WAYS OF LOADING DATA TO DATAWAREHOUSE USING ORACLE?
* WHAT IS TABLE PARTITIONING? HOW IT IS USEFUL TO WAREHOUSE DATABASE?
* WHAT ARE DIFFERENT TYPES OF PARTITIONING IN ORACLE?
* WHAT IS A MATERIALIZED VIEW? HOW IT IS DIFFERENT FROM NORMAL AND INLINE VIEWS?
* WHAT IS INDEXING? WHAT ARE DIFFERENT TYPES OF INDEXES SUPPORTED BY ORACLE?
* WHAT ARE DIFFERENT STORAGE OPTIONS SUPPORTED BY ORACLE?
* WHAT IS QUERY OPTIMIZER? WHAT ARE DIFFERENT TYPES OF OPTIMIZERS SUPPORTED BY ORACLE?
* EXPLAIN ROLLUP,CUBE,RANK AND DENSE_RANK FUNCTIONS OF ORACLE 8i.

The advantages of using bitmap indexes are greatest for low cardinality columns: that is, columns in which the number of distinct values is small compared to the number of rows in the table. A gender column, which only has two distinct values (male and female), is ideal for a bitmap index. However, data warehouse administrators will also choose to build bitmap indexes on columns with much higher cardinalities.

Local vs global: A B-tree index on a partitioned table can be local or global. Global indexes must be fully rebuilt after a direct load, which can be very costly when loading a relatively small number of rows into a large table. For this reason, it is strongly recommended that indexes on partitioned tables should be defined as local indexes unless there is a well-justified performance requirement for a global index. Bitmap indexes on partitioned tables are always local.

Why Constraints are Useful in a Data Warehouse
Constraints provide a mechanism for ensuring that data conforms to guidelines specified by the database administrator. The most common types of constraints include unique constraints (ensuring that a given column is unique), not-null constraints, and foreign-key constraints (which ensure that two keys share a primary key-foreign key relationship).

Materialized Views for Data Warehouses
In data warehouses, materialized views can be used to precompute and store aggregated data such as the sum of sales. Materialized views in these environments are typically referred to as summaries, because they store summarized data. They can also be used to precompute joins with or without aggregations. A materialized view eliminates the overhead associated with expensive joins or aggregations for a

large or important class of queries.

The Need for Materialized Views
Materialized views are used in data warehouses to increase the speed of queries on
very large databases. Queries to large databases often involve joins between tables
or aggregations such as SUM, or both. These operations are very expensive in terms
of time and processing power.

How does MV's work?
The query optimizer can use materialized views by
automatically recognizing when an existing materialized view can and should be
used to satisfy a request. It then transparently rewrites the request to use the
materialized view. Queries are then directed to the materialized view and not to the
underlying detail tables. In general, rewriting queries to use materialized views
rather than detail tables results in a significant performance gain.

If a materialized view is to be used by query rewrite, it must be stored in the same
database as its fact or detail tables. A materialized view can be partitioned, and you
can define a materialized view on a partitioned table and one or more indexes on
the materialized view.

The types of materialized views are:
☐ Materialized Views with Joins and Aggregates
☐ Single-Table Aggregate Materialized Views
☐ Materialized Views Containing Only Joins

Some Useful system tables:
user_tab_partitions
user_tab_columns

Doc3

Repository related Questions

Q. What is the difference between PowerCenter and PowerMart?
With PowerCenter, you receive all product functionality, including the ability to register multiple servers, share metadata
across repositories, and partition data.
A PowerCenter license lets you create a single repository that you can configure as a global repository, the core
component of a data warehouse.
PowerMart includes all features except distributed metadata, multiple registered servers, and data partitioning. Also, the
various options available with PowerCenter (such as PowerCenter Integration Server for BW, PowerConnect for IBM DB2,
PowerConnect for IBM MQSeries, PowerConnect for SAP R/3, PowerConnect for Siebel, and PowerConnect for
PeopleSoft) are not available with PowerMart.

Q. What are the new features and enhancements in PowerCenter 5.1?
The major features and enhancements to PowerCenter 5.1 are:
   a) Performance Enhancements

- **High precision decimal arithmetic.** The Informatica Server optimizes data throughput to increase performance
  of sessions using the Enable Decimal Arithmetic option.
- **To_Decimal and Aggregate functions.** The Informatica Server uses improved algorithms to increase
  performance of To_Decimal and all aggregate functions such as percentile, median, and average.
- **Cache management.** The Informatica Server uses better cache management to increase performance of
  Aggregator, Joiner, Lookup, and Rank transformations.
- **Partition sessions with sorted aggregation.** You can partition sessions with Aggregator transformation that
  use sorted input. This improves memory usage and increases performance of sessions that have sorted data.

When enabled, the Informatica Client and Informatica Server lift code page selection and validation restrictions. You can select any supported code page for source, target, lookup, and stored procedure data.

c) Designer Features and Enhancements

- **Debug mapplets.** You can debug a mapplet within a mapping in the Mapping Designer. You can set breakpoints in transformations in the mapplet.
- **Support for slash character (/) in table and field names.** You can use the Designer to import source and target definitions with table and field names containing the slash character (/). This allows you to import SAP BW source definitions by connecting directly to the underlying database tables.

d) Server Manager Features and Enhancements

- **Continuous sessions.** You can schedule a session to run continuously. A continuous session starts automatically when the Load Manager starts. When the session stops, it restarts immediately without rescheduling. Use continuous sessions when reading real time sources, such as IBM MQSeries.
- **Partition sessions with sorted aggregators.** You can partition sessions with sorted aggregators in a mapping.
- **Register multiple servers against a local repository.** You can register multiple PowerCenter Servers against a local repository.

Q. What is a repository?

The Informatica repository is a relational database that stores information, or *metadata*, used by the Informatica Server and Client tools. The repository also stores administrative information such as usernames and passwords, permissions and privileges, and product version.

We create and maintain the repository with the Repository Manager client tool. With the Repository Manager, we can also create folders to organize metadata and groups to organize users.

Q. What are different kinds of repository objects? And what it will contain?

Repository objects displayed in the Navigator can include sources, targets, transformations, mappings, mapplets, shortcuts, sessions, batches, and session logs.

Q. What is a metadata?

Designing a data mart involves writing and storing a complex set of instructions. You need to know where to get data (sources), how to change it, and where to write the information (targets). PowerMart and PowerCenter call this set of instructions metadata. Each piece of metadata (for example, the description of a source table in an operational database) can contain comments about it.

In summary, Metadata can include information such as mappings describing how to transform source data, sessions indicating when you want the Informatica Server to perform the transformations, and connect strings for sources and targets.

Q. What are folders?
Folders let you organize your work in the repository, providing a way to separate different types of metadata or different projects into easily identifiable areas.

Q. What is a Shared Folder?
A shared folder is one, whose contents are available to all other folders in the same repository. If we plan on using the same piece of metadata in several projects (for example, a description of the CUSTOMERS table that provides data for a variety of purposes), you might put that metadata in the shared folder.

Q. What are mappings?
A mapping specifies how to move and transform data from sources to targets. Mappings include source and target definitions and transformations. Transformations describe how the Informatica Server transforms data. Mappings can also include shortcuts, reusable transformations, and mapplets. Use the Mapping Designer tool in the Designer to create mappings.

Q. What are mapplets?

You can design a mapplet to contain sets of transformation logic to be reused in multiple mappings within a folder, a repository, or a domain. Rather than recreate the same set of transformations each time, you can create a mapplet containing the transformations, then add instances of the mapplet to individual mappings. Use the Mapplet Designer tool in the Designer to create mapplets.

Q. What are Transformations?

A transformation generates, modifies, or passes data through ports that you connect in a mapping or mapplet. When you build a mapping, you add transformations and configure them to handle data according to your business purpose. Use the Transformation Developer tool in the Designer to create transformations.

Q. What are Reusable transformations?

You can design a transformation to be reused in multiple mappings within a folder, a repository, or a domain. Rather than recreate the same transformation each time, you can make the transformation reusable, then add instances of the transformation to individual mappings. Use the Transformation Developer tool in the Designer to create reusable transformations.

Q. What are Sessions and Batches?

Sessions and batches store information about how and when the Informatica Server moves data through mappings. You create a session for each mapping you want to run. You can group several sessions together in a batch. Use the Server Manager to create sessions and batches.

Q. What are Shortcuts?

We can create shortcuts to objects in shared folders. Shortcuts provide the easiest way to reuse objects. We use a shortcut as if it were the actual object, and when we make a change to the original object, all shortcuts inherit the change.

Shortcuts to folders in the same repository are known as local shortcuts. Shortcuts to the global repository are called global shortcuts.

We use the Designer to create shortcuts.

Q. What are Source definitions?

Detailed descriptions of database objects (tables, views, synonyms), flat files, XML files, or Cobol files that provide source data. For example, a source definition might be the complete structure of the EMPLOYEES table, including the table name, column names and datatypes, and any constraints applied to these columns, such as NOT NULL or PRIMARY KEY. Use the Source Analyzer tool in the Designer to import and create source definitions.

Q. What are Target definitions?

Detailed descriptions for database objects, flat files, Cobol files, or XML files to receive transformed data. During a session, the Informatica Server writes the resulting data to session targets. Use the Warehouse Designer tool in the Designer to import or create target definitions.

Q. What is Dynamic Data Store?
The need to share data is just as pressing as the need to share metadata. Often, several data marts in the same organization need the same information. For example, several data marts may need to read the same product data from operational sources, perform the same profitability calculations, and format this information to make it easy to review.
If each data mart reads, transforms, and writes this product data separately, the throughput for the entire organization is lower than it could be. A more efficient approach would be to read, transform, and write the data to one central data store shared by all data marts. Transformation is a processing-intensive task, so performing the profitability calculations once saves time.
Therefore, this kind of dynamic data store (DDS) improves throughput at the level of the entire organization, including all data marts. To improve performance further, you might want to capture incremental changes to sources. For example, rather than reading all the product data each time you update the DDS, you can improve performance by capturing only the inserts, deletes, and updates that have occurred in the PRODUCTS table since the last time you updated the DDS. The DDS has one additional advantage beyond performance: when you move data into the DDS, you can format it in a standard fashion. For example, you can prune sensitive employee data that should not be stored in any data mart. Or you can display date and time values in a standard format. You can perform these and other data cleansing tasks when you move data into the DDS instead of performing them repeatedly in separate data marts.

Q. When should you create the dynamic data store? Do you need a DDS at all?

To decide whether you should create a dynamic data store (DDS), consider the following issues:

- **How much data do you need to store in the DDS?** The one principal advantage of data marts is the selectivity of information included in it. Instead of a copy of everything potentially relevant from the OLTP database and flat files, data marts contain only the information needed to answer specific questions for a specific audience (for example, sales performance data used by the sales division). A dynamic data store is a hybrid of the galactic warehouse and the individual data mart, since it includes all the data needed for all the data marts it supplies. If the dynamic data store contains nearly as much information as the OLTP source, you might not need the intermediate step of the dynamic data store. However, if the dynamic data store includes substantially less than all the data in the source databases and flat files, you should consider creating a DDS staging area.
- **What kind of standards do you need to enforce in your data marts?** Creating a DDS is an important technique in enforcing standards. If data marts depend on the DDS for information, you can provide that data in the range and format you want everyone to use. For example, if you want all data marts to include the same information on customers, you can put all the data needed for this standard customer profile in the DDS. Any data mart that reads customer data from the DDS should include all the information in this profile.
- **How often do you update the contents of the DDS?** If you plan to frequently update data in data marts, you need to update the contents of the DDS at least as often as you update the individual data marts that the DDS feeds. You may find it easier to read data directly from source databases and flat file systems if it becomes burdensome to update the DDS fast enough to keep up with the needs of individual data marts. Or, if particular data marts need updates significantly faster than others, you can bypass the DDS for these fast update data marts.
- **Is the data in the DDS simply a copy of data from source systems, or do you plan to reformat this information before storing it in the DDS?** One advantage of the dynamic data store is that, if you plan on reformatting information in the same fashion for several data marts, you only need to format it once for the dynamic data store. Part of this question is whether you keep the data normalized when you copy it to the DDS.
- **How often do you need to join data from different systems?** On occasion, you may need to join records queried from different databases or read from different flat file systems. The more frequently you need to perform this type of heterogeneous join, the more advantageous it would be to perform all such joins within the DDS, then make the results available to all data marts that use the DDS as a source.

Q. What is a Global repository?

The centralized repository in a domain, a group of connected repositories. Each domain can contain one global repository. The global repository can contain common objects to be shared throughout the domain through global shortcuts. Once created, you cannot change a global repository to a local repository. You can promote an existing local repository to a global repository.

Q. What is Local Repository?

Each local repository in the domain can connect to the global repository and use objects in its shared folders. A folder in a local repository can be copied to other local repositories while keeping all local and global shortcuts intact.

Q. What are the different types of locks?

There are five kinds of locks on repository objects:

- **Read lock.** Created when you open a repository object in a folder for which you do not have write permission. Also created when you open an object with an existing write lock.
- **Write lock.** Created when you create or edit a repository object in a folder for which you have write permission.
- **Execute lock.** Created when you start a session or batch, or when the Informatica Server starts a scheduled session or batch.
- **Fetch lock.** Created when the repository reads information about repository objects from the database.
- **Save lock.** Created when you save information to the repository.

Q. After creating users and user groups, and granting different sets of privileges, I find that none of the repository users can perform certain tasks, even the Administrator.

Repository privileges are limited by the database privileges granted to the database user who created the repository. If the database user (one of the default users created in the Administrators group) does not have full database privileges in the repository database, you need to edit the database user to allow all privileges in the database.

Q. I created a new group and removed the Browse Repository privilege from the group. Why does every user in the group still have that privilege?

Privileges granted to individual users take precedence over any group restrictions. Browse Repository is a default privilege granted to all new users and groups. Therefore, to remove the privilege from users in a group, you must remove the privilege from the group, and every user in the group.

Q. I do not want a user group to create or edit sessions and batches, but I need them to access the Server Manager to stop the Informatica Server.

To permit a user to access the Server Manager to stop the Informatica Server, you must grant them both the Create Sessions and Batches, and Administer Server privileges. To restrict the user from creating or editing sessions and batches, you must restrict the user's write permissions on a folder level.

Alternatively, the user can use *pmcmd* to stop the Informatica Server with the Administer Server privilege alone.

Q. How does read permission affect the use of the command line program, *pmcmd*?

To use *pmcmd*, you do not need to view a folder before starting a session or batch within the folder. Therefore, you do not need read permission to start sessions or batches with *pmcmd*. You must, however, know the exact name of the session or batch and the folder in which it exists.

With *pmcmd*, you can start any session or batch in the repository if you have the Session Operator privilege or execute permission on the folder.

Q. My privileges indicate I should be able to edit objects in the repository, but I cannot edit any metadata.

You may be working in a folder with restrictive permissions. Check the folder permissions to see if you belong to a group whose privileges are restricted by the folder owner.

Q. I have the Administer Repository Privilege, but I cannot access a repository using the Repository Manager.

To perform administration tasks in the Repository Manager with the Administer Repository privilege, you must also have the default privilege Browse Repository. You can assign Browse Repository directly to a user login, or you can inherit Browse Repository from a group.

Questions related to Server Manager

Q. What is Event-Based Scheduling?

When you use event-based scheduling, the Informatica Server starts a session when it locates the specified indicator file. To use event-based scheduling, you need a shell command, script, or batch file to create an indicator file when all sources are available. The file must be created or sent to a directory local to the Informatica Server. The file can be of any format recognized by the Informatica Server operating system. The Informatica Server deletes the indicator file once the session starts.

Use the following syntax to ping the Informatica Server on a UNIX system:

pmcmd ping [{user_name | %user_env_var} {password | %password_env_var}] [hostname:]portno

Use the following syntax to start a session or batch on a UNIX system:

pmcmd start {user_name | %user_env_var} {password | %password_env_var} [hostname:]portno [folder_name:] {session_name | batch_name} [:pf=param_file] session_flag wait_flag

Use the following syntax to stop a session or batch on a UNIX system:

pmcmd stop {user_name | %user_env_var} {password | %password_env_var} [hostname:]portno[folder_name:] {session_name | batch_name} session_flag

Use the following syntax to stop the Informatica Server on a UNIX system:

pmcmd stopserver {user_name | %user_env_var} {password | %password_env_var} [hostname:]portno

Q. What are the different types of Commit intervals?

The different commit intervals are:

- **Target-based commit.** The Informatica Server commits data based on the number of target rows and the key constraints on the target table. The commit point also depends on the buffer block size and the commit interval.
- **Source-based commit.** The Informatica Server commits data based on the number of source rows. The commit point is the commit interval you configure in the session properties.

Designer Questions

Q. What are the tools provided by Designer?

The Designer provides the following tools:

- **Source Analyzer.** Use to import or create source definitions for flat file, XML, 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.

Q. What is a transformation?

A transformation is a repository object that generates, modifies, or passes data. You configure logic in a transformation that the Informatica Server uses to transform data. The Designer provides a set of transformations that perform specific functions. For example, an Aggregator transformation performs calculations on groups of data.

Each transformation has rules for configuring and connecting in a mapping. For more information about working with a specific transformation, refer to the chapter in this book that discusses that particular transformation.

You can create transformations to use once in a mapping, or you can create reusable transformations to use in multiple mappings.

Q. What are the different types of Transformations? (Mascot)

a) **Aggregator transformation:** The Aggregator transformation allows you to perform aggregate calculations, such as averages and sums. The Aggregator transformation is unlike the Expression transformation, in that you can use the Aggregator transformation to perform calculations on groups. The Expression transformation permits you to perform calculations on a row-by-row basis only. (Mascot)

b**) Expression transformation:** You can use the Expression transformations to calculate values in a single row before you write to the target. For example, you might need to adjust employee salaries, concatenate first and last names, or convert strings to numbers. You can use the Expression transformation to perform any non-aggregate calculations. You can also use the Expression transformation to test conditional statements before you output the results to target tables or other transformations.

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

d) **Joiner transformation:** While a Source Qualifier transformation can join data originating from a common source database, the Joiner transformation joins two related heterogeneous sources residing in different locations or file systems.

e) **Lookup transformation:** Use a Lookup transformation in your mapping to look up data in a relational table, view, or synonym. Import a lookup definition from any relational database to which both the Informatica Client and Server can connect. You can use multiple Lookup transformations in a mapping.

The Informatica Server queries the lookup table based on the lookup ports in the transformation. It compares Lookup transformation port values to lookup table column values based on the lookup condition. Use the result of the lookup to pass to other transformations and the target.

Q. What is the difference between Aggregate and Expression Transformation? (Mascot)

Q. What is Update Strategy?

When we design our data warehouse, we need to decide what type of information to store in targets. As part of our target table design, we need to determine whether to maintain all the historic data or just the most recent changes.
The model we choose constitutes our *update strategy*, how to handle changes to existing records.

Update strategy flags a record for update, insert, delete, or reject. We use this transformation when we want to exert fine control over updates to a target, based on some condition we apply. For example, we might use the Update Strategy transformation to flag all customer records for update when the mailing address has changed, or flag all employee records for reject for people no longer working for the company.

Q. Where do you define update strategy?
We can set the Update strategy at two different levels:
- **Within a session.** When you configure a session, you can instruct the Informatica Server to either treat all records in the same way (for example, treat all records as inserts), or use instructions coded into the session mapping to flag records for different database operations.
- **Within a mapping.** Within a mapping, you use the Update Strategy transformation to flag records for insert, delete, update, or reject.

Q. What are the advantages of having the Update strategy at Session Level?

Q. What is a lookup table? (KPIT Infotech, Pune)
The lookup table can be a single table, or we can join multiple tables in the same database using a lookup query override. The Informatica Server queries the lookup table or an in-memory cache of the table for all incoming rows into the Lookup transformation.

If your mapping includes heterogeneous joins, we can use any of the mapping sources or mapping targets as the lookup table.

Q. What is a Lookup transformation and what are its uses?
We use a Lookup transformation in our mapping to look up data in a relational table, view or synonym.

We can use the Lookup transformation for the following purposes:

- ❖ Get a related value. For example, if our source table includes employee ID, but we want to include the employee name in our target table to make our summary data easier to read.
- ❖ Perform a calculation. Many normalized tables include values used in a calculation, such as gross sales per invoice or sales tax, but not the calculated value (such as net sales).
- ❖ Update slowly changing dimension tables. We can use a Lookup transformation to determine whether records already exist in the target.

Q. What are connected and unconnected Lookup transformations?

We can configure a connected Lookup transformation to receive input directly from the mapping pipeline, or we can configure an unconnected Lookup transformation to receive input from the result of an expression in another transformation.

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

A common use for unconnected Lookup transformations is to update slowly changing dimension tables.

Q. What is the difference between connected lookup and unconnected lookup?

Differences between Connected and Unconnected Lookups:

Connected Lookup	Unconnected Lookup
Receives input values directly from the pipeline.	Receives input values from the result of a :LKP expression in another transformation.
We can use a dynamic or static cache	We can use a static cache
Supports user-defined default values	Does not support user-defined default values

Q. What is Sequence Generator Transformation? (Mascot)

The Sequence Generator transformation generates numeric values. We can use the Sequence Generator to create unique primary key values, replace missing primary keys, or cycle through a sequential range of numbers.

The Sequence Generation transformation is a connected transformation. It contains two output ports that we can connect to one or more transformations.

Q. What are the uses of a Sequence Generator transformation?

We can perform the following tasks with a Sequence Generator transformation:
- o Create keys
- o Replace missing values
- o Cycle through a sequential range of numbers

Q. What are the advantages of Sequence generator? Is it necessary, if so why?

We can make a Sequence Generator reusable, and use it in multiple mappings. We might reuse a Sequence Generator when we perform multiple loads to a single target.

For example, if we have a large input file that we separate into three sessions running in parallel, we can use a Sequence Generator to generate primary key values. If we use different Sequence Generators, the Informatica Server might accidentally generate duplicate key values. Instead, we can use the same reusable Sequence Generator for all three sessions to provide a unique value for each target row.

Q. How is the Sequence Generator transformation different from other transformations?

The Sequence Generator is unique among all transformations because we cannot add, edit, or delete its default ports (NEXTVAL and CURRVAL).

Unlike other transformations we cannot override the Sequence Generator transformation properties at the session level. This protecxts the integrity of the sequence values generated.

Q. What does Informatica do? How it is useful?
Q. What is the difference between Informatica version 1.7.2 and 1.7.3?
Q. What are the complex filters used till now in your applications?
Q. Feartures of Informatica
Q. Have you used Informatica? which version?
Q. How do you set up a schedule for data loading from scratch? describe step-by-step.
Q. How do you use mapplet?
Q. What are the different data source types you have used with Informatica?
Q. Is it possible to run one loading session with one particular target and multiple types of data sources?
This section describes new features and enhancements to PowerCenter 6.0 and PowerMart 6.0.

Designer

- **Compare objects.** The Designer allows you to compare two repository objects of the same type to identify differences between them. You can compare sources, targets, transformations, mapplets, mappings, instances, or mapping/mapplet dependencies in detail. You can compare objects across open folders and repositories.
- Copying objects. In each Designer tool, you can use the copy and paste functions to copy objects from one workspace to another. For example, you can select a group of transformations in a mapping and copy them to a new mapping.
- **Custom tools.** The Designer allows you to add custom tools to the Tools menu. This allows you to start programs you use frequently from within the Designer.
- **Flat file targets.** You can create flat file target definitions in the Designer to output data to flat files. You can create both fixed-width and delimited flat file target definitions.
- **Heterogeneous targets.** You can create a mapping that outputs data to multiple database types and target types. When you run a session with heterogeneous targets, you can specify a database connection for each relational target. You can also specify a file name for each flat file or XML target.
- Link paths. When working with mappings and mapplets, you can view link paths. Link paths display the flow of data from a column in a source, through ports in transformations, to a column in the target.
- Linking ports. You can now specify a prefix or suffix when automatically linking ports between transformations based on port names.
- **Lookup cache.** You can use a dynamic lookup cache in a Lookup transformation to insert and update data in the cache and target when you run a session.
- **Mapping parameter and variable support in lookup SQL override.** You can use mapping parameters and variables when you enter a lookup SQL override.
- **Mapplet enhancements.** Several mapplet restrictions are removed. You can now include multiple Source Qualifier transformations in a mapplet, as well as Joiner transformations and Application Source Qualifier transformations for IBM MQSeries. You can also include both source definitions and Input transformations in one mapplet. When you work with a mapplet in a mapping, you can expand the mapplet to view all transformations in the mapplet.
- **Metadata extensions.** You can extend the metadata stored in the repository by creating metadata extensions for repository objects. The Designer allows you to create metadata extensions for source definitions, target definitions, transformations, mappings, and mapplets.
- **Numeric and datetime formats.** You can define formats for numeric and datetime values in flat file sources and targets. When you define a format for a numeric or datetime value, the Informatica Server uses the format to read from the file source or to write to the file target.
- Pre- and post-session SQL. You can specify pre- and post-session SQL in a Source Qualifier transformation and in a mapping target instance when you create a mapping in the Designer. The Informatica Server issues pre-SQL commands to the database once before it runs the session. Use pre-session SQL to issue commands to the database such as dropping indexes before extracting data. The Informatica Server issues post-session SQL commands to the database once after it runs the session. Use post-session SQL to issue commands to a database such as re-creating indexes.
- Renaming ports. If you rename a port in a connected transformation, the Designer propagates the name change to expressions in the transformation.
- **Sorter transformation.** The Sorter transformation is an active transformation that allows you to sort data from relational or file sources in ascending or descending order according to a sort key. You can increase session performance when you use the Sorter transformation to pass data to an Aggregator transformation configured for sorted input in a mapping.

- Tips. When you start the Designer, it displays a tip of the day. These tips help you use the Designer more efficiently. You can display or hide the tips by choosing Help-Tip of the Day.
- Tool tips for port names. Tool tips now display for port names. To view the full contents of the column, position the mouse over the cell until the tool tip appears.
- View dependencies. In each Designer tool, you can view a list of objects that depend on a source, source qualifier, transformation, or target. Right-click an object and select the View Dependencies option.
- Working with multiple ports or columns. In each Designer tool, you can move multiple ports or columns at the same time.

Informatica Server

- **Add timestamp to workflow logs.** You can configure the Informatica Server to add a timestamp to messages written to the workflow log.
- Expanded pmcmd capability. You can use pmcmd to issue a number of commands to the Informatica Server. You can use pmcmd in either an interactive or command line mode. The interactive mode prompts you to enter information when you omit parameters or enter invalid commands. In both modes, you can enter a command followed by its command options in any order. In addition to commands for starting and stopping workflows and tasks, pmcmd now has new commands for working in the interactive mode and getting details on servers, sessions, and workflows.
- **Error handling.** The Informatica Server handles the abort command like the stop command, except it has a timeout period. You can specify when and how you want the Informatica Server to stop or abort a workflow by using the Control task in the workflow. After you start a workflow, you can stop or abort it through the Workflow Monitor or *pmcmd*.
- **Export session log to external library.** You can configure the Informatica Server to write the session log to an external library.
- **Flat files.** You can specify the precision and field length for columns when the Informatica Server writes to a flat file based on a flat file target definition, and when it reads from a flat file source. You can also specify the format for datetime columns that the Informatica Server reads from flat file sources and writes to flat file targets.
- **Write Informatica Windows Server log to a file.** You can now configure the Informatica Server on Windows to write the Informatica Server log to a file.

Metadata Reporter

- **List reports for jobs, sessions, workflows, and worklets.** You can run a list report that lists all jobs, sessions, workflows, or worklets in a selected repository.
- **Details reports for sessions, workflows, and worklets. You can run a details report** to view details about each session, workflow, or worklet in a selected repository.
- **Completed session, workflow, or worklet detail reports.** You can run a completion details report, which displays details about how and when a session, workflow, or worklet ran, and whether it ran successfully.
- **Installation on WebLogic**. You can now install the Metadata Reporter on WebLogic and run it as a web application.

Repository Manager

- **Metadata extensions.** You can extend the metadata stored in the repository by creating metadata extensions for repository objects. The Repository Manager allows you to create metadata extensions for source definitions, target definitions, transformations, mappings, mapplets, sessions, workflows, and worklets.
- *pmrep* **security commands.** You can use *pmrep* to create or delete repository users and groups. You can also use *pmrep* to modify repository privileges assigned to users and groups.
- Tips. When you start the Repository Manager, it displays a tip of the day. These tips help you use the Repository Manager more efficiently. You can display or hide the tips by choosing Help-Tip of the Day.

Repository Server

The Informatica Client tools and the Informatica Server now connect to the repository database over the network through the Repository Server.

- **Repository Server.** The Repository Server manages the metadata in the repository database. It accepts and manages all repository client connections and ensures repository consistency by employing object locking. The Repository Server can manage multiple repositories on different machines on the network.

- **Repository connectivity changes.** When you connect to the repository, you must specify the host name of the machine hosting the Repository Server and the port number the Repository Server uses to listen for connections. You no longer have to create an ODBC data source to connect a repository client application to the repository.

Transformation Language

- **New functions.** The transformation language includes two new functions, ReplaceChr and ReplaceStr. You can use these functions to replace or remove characters or strings in text data.

- **SETVARIABLE.** The SETVARIABLE function now executes for rows marked as insert or update.

Workflow Manager

The Workflow Manager and Workflow Monitor replace the Server Manager. Instead of creating a session, you now create a process called a workflow in the Workflow Manager. A workflow is a set of instructions on how to execute tasks such as sessions, emails, and shell commands. A session is now one of the many tasks you can execute in the Workflow Manager.

The Workflow Manager provides other tasks such as Assignment, Decision, and Event-Wait tasks. You can also create branches with conditional links. In addition, you can batch workflows by creating worklets in the Workflow Manager.

- **DB2 external loader.** You can use the DB2 EE external loader to load data to a DB2 EE database. You can use the DB2 EEE external loader to load data to a DB2 EEE database. The DB2 external loaders can insert data, replace data, restart load operations, or terminate load operations.

- **Environment SQL.** For relational databases, you may need to execute some SQL commands in the database environment when you connect to the database. For example, you might want to set isolation levels on the source and target systems to avoid deadlocks. You configure environment SQL in the database connection. You can use environment SQL for source, target, lookup, and stored procedure connections.

- **Email.** You can create email tasks in the Workflow Manager to send emails when you run a workflow. You can configure a workflow to send an email anywhere in the workflow logic, including after a session completes or after a session fails. You can also configure a workflow to send an email when the workflow suspends on error.

- **Flat file targets.** In the Workflow Manager, you can output data to a flat file from either a flat file target definition or a relational target definition.

- **Heterogeneous targets.** You can output data to different database types and target types in the same session. When you run a session with heterogeneous targets, you can specify a database connection for each relational target. You can also specify a file name for each flat file or XML target.

- **Metadata extensions.** You can extend the metadata stored in the repository by creating metadata extensions for repository objects. The Workflow Manager allows you to create metadata extensions for sessions, workflows, and worklets.

- **Oracle 8 direct path load support.** You can load data directly to Oracle 8i in bulk mode without using an external loader. You can load data directly to an Oracle client database version 8.1.7.2 or higher.

- **Partitioning enhancements.** To improve session performance, you can set partition points at multiple transformations in a pipeline. You can also specify different partition types at each partition point.

- **Server variables.** You can use new server variables to define the workflow log directory and workflow log count.

- Teradata TPump external loader. You can use the Teradata TPump external loader to load data to a Teradata database. You can use TPump in sessions that contain multiple partitions.

- Tips. When you start the Workflow Manager, it displays a tip of the day. These tips help you use the Workflow Manager more efficiently. You can display or hide the tips by choosing Help-Tip of the Day.

- **Workflow log.** In addition to session logs, you can configure the Informatica Server to create a workflow log to record details about workflow runs.

- **Workflow Monitor.** You use a tool called the Workflow Monitor to monitor workflows, worklets, and tasks. The Workflow Monitor displays information about workflow runs in two views: Gantt Chart view or Task view. You can run, stop, abort, and resume workflows from the Workflow Monitor.

Q: How do I connect job streams/sessions or batches across folders? (30 October 2000)

For quite a while there's been a deceptive problem with sessions in the Informatica repository. For management and maintenance reasons, we've always wanted to separate mappings, sources, targets, in to subject areas or functional areas of the business. This makes sense until we try to run the entire Informatica job stream. Understanding of course that only the folder in which the map has been defined can house the session. This makes it difficult to run jobs / sessions across folders - particularly when there are necessary job dependancies which must be defined. The purpose of this article is to introduce an alternative solution to this problem. It requires the use of shortcuts.

The basics are like this: Keep the map creations, sources, and targets subject oriented. This allows maintenance to be easier (by subect area). Then once the maps are done, change the folders to allow shortcuts (done from the repository manager). Create a folder called: "MY_JOBS" or something like that. Go in to designer, open "MY_JOBS", expand the source folders, and create shortcuts to the mappings in the source folders.

Go to the session manager, and create sessions for each of the short-cut mappings in MY_JOBS. Then batch them as you see fit. This will allow a single folder for running jobs and sessions housed anywhere in any folder across your repository.

Q: How do I get maximum speed out of my database connection? (12 September 2000)

In Sybase or MS-SQL Server, go to the Database Connection in the Server Manager. Increase the packet size. Recommended sizing depends on distance traveled from PMServer to Database - 20k Is usually acceptable on the same

subnet. Also, have the DBA increase the "maximum allowed" packet size setting on the Database itself. Following this change, the DBA will need to restart the DBMS. Changing the Packet Size doesn't mean all connections will connect at this size, it just means that anyone specifying a larger packet size for their connection may be able to use it. It should increase speed, and decrease network traffic. Default IP Packets are between 1200 bytes and 1500 bytes.

In Oracle: there are two methods. For connection to a local database, setup the protocol as IPC (between PMServer and a DBMS Server that are hosted on the same machine). IPC is not a protocol that can be utilized across networks (apparently). IPC stands for Inter Process Communication, and utilizes memory piping (RAM) instead of client context, through the IP listner. For remote connections there is a better way: Listner.ORA and TNSNames.ORA need to be modified to include SDU and TDU settings. SDU = Service Layer Data Buffer, and TDU = Transport Layer Data Buffer. Both of which specify packet sizing in Oracle connections over IP. Default for Oracle is 1500 bytes. Also note: these settings can be used in IPC connections as well, to control the IPC Buffer sizes passed between two local programs (PMServer and Oracle Server)

Both the Server and the Client need to be modified. The server will allow packets up to the max size set - but unless the client specifies a larger packet size, the server will default to the smallest setting (1500 bytes). Both SDU and TDU should be set the same. See the example below:

TNSNAMES.ORA

LOC=(DESCRIPTION= (SDU = 20480) (TDU=20480)

LISTENER.ORA

LISTENER=....(SID_DESC= (SDU = 20480) (TDU=20480) (SID_NAME = beqlocal) ....

Q: How do I get a Sequence Generator to "pick up" where another "left off"? (8 June 2000)

- To perform this mighty trick, one can use an unconnected lookup on the Sequence ID of the target table. Set the properties to "LAST VALUE", input port is an ID. the condition is: **SEQ_ID >= input_ID**. Then in an expression set up a variable port: connect a NEW self-resetting sequence generator to a new input port in the expression. The variable port's expression should read: **IIF(v_seq = 0 OR ISNULL(v_seq) = true, :LKP.lkp_sequence(1), v_seq)**. Then, set up an output port. Change the output port's expression to read: **v_seq + input_seq** (from the resetting sequence generator). Thus you have just completed an "append" without a break in sequence numbers.

Q: How do I query the repository to see which sessions are set in TEST MODE? (8 June 2000)

- Runthefollowing   select:
select   *   from   opb_load_session   where   bit_option   =   13;
It's actually BIT # 2 in this bit_option setting, so if you have a mask, or a bit-level function you can then AND it with a mask of 2, if this is greater than zero, it's been set for test load.

Q: How do I "validate" all my mappings at once? (31 March 2000)

- Issue   the   following   command   WITH   CARE. UPDATE OPB_MAPPING SET IS_VALID = 1;
  - Then disconnect from the database, and re-connect. In session manager, and designer as well.

Q: How do I validate my entire repository? (12 September 2000)

- To add the menu option, change this registry entry on your client. HKEY_CURRENT_USER/Software/Informatica/PowerMart Client Tools/4.7/Repository Manager Options .   Add the following string Name: EnableCheckReposit Data.

Validate Repository forces Informatica to run through the repository, and check the repo for errors

Q: How do I work around a bug in 4.7? I can't change the execution order of my stored procedures that I've imported? (31 March 2000)

- Issue the following statements WITH CARE:
select widget_id from OPB_WIDGET where WIDGET_NAME = <widget name>
(write down the WIDGET ID)
- select * from OPB_WIDGET_ATTR where WIDGET_ID = <widget_id>
- update OPB_WIDGET_ATTR set attr_value = <execution order> where WIDGET_ID = <widget_id> and attr_id = 5
- COMMIT;
The <execution order> is the number of the order in which you want the stored proc to execute. Again, disconnect from both designer and session manager repositories, and re-connect to "re-read" the local cache.

Q: How do I keep the session manager from "Quitting" when I try to open a session? (23 March 2000)

- Informatica Tech Support has said: if you are using a flat file as a source, and your "file name" in the "Source Options" dialog is longer than 80 characters, it will "kill" the Session Manager tool when you try to re-open it. You can fix the session by: logging in to the repository via SQLPLUS, or ISQL, and finding the table called: OPB_LOAD_SESSION, find the Session ID associated with the session name - write it down. Then *select FNAME from OPB_LOAD_FILES where Session_ID = <session_id>*. Change / **update OPB_LOAD_FILES**

set FNAME= <new file name> column, change the length back to less than 80 characters, and commit the changes. Now the session has been repaired. Try to keep the directory to that source file in the DIRECTORY entry box above the file name box. Try to keep all the source files together in the same source directory if possible.

Q: *How do I repair a "damaged" repository? (16 March 2000)*

- There really isn't a good repair tool, nor is there a "great" method for repairing the repository. However, I have some suggestions which might help. If you're running in to a session which causes the session manager to "quit" on you when you try to open it, or you have a map that appears to have "bad sources", there may be something you can do. There are varying degrees of damage to the repository - mostly caused because the sequence generator that PM/PC relies on is buried in a table in the repository - and they generate their own sequence numbers. If this table becomes "corrupted" or generates the wrong sequences, you can get repository errors all over the place. It can spread quickly. Try the following steps to repair a repository: (USE AT YOUR OWN RISK) The recommended path is to backup the repository, send it to Technical Support - and tell them it's damaged.

1. Delete the session, disconnect, re-connect, then re-create the session, then attempt to edit the new session again. If the new session won't open up (srvr mgr quits), then there are more problems - PM/PC is not successfully attaching sources and targets to the session (SEE: OPB_LOAD_SESSION table (SRC_ID, TARGET_ID) columns - they will be zero, when they should contain an ID.
2. Delete the session, then open the map. Delete the source and targets from the MAP. Save the map and invalidate it - forcing an update to the repository and it's links. Drag the sources and targets back in to the map and re-connect them. Validate and Save. Then try re-building the session (back to step one). If there is still a failure, then there are more problems.
3. Delete the session and the map entirely. Save the repository changes - thus requesting a delete in the repository. While the "delete" may occur - some of the tables in the repository may not be "cleansed". There may still be some sources, targets, and transformation objects (reusable) left in the repository. Rebuild the map from scratch - then save it again... This will create a new MAP ID in the OPB_MAPPING table, and force PM/PC to create new ID links to existing Source and Target objects (as well as all the other objects in the map).
4. If that didn't work - you may have to delete the sources, reusable objects, and targets, as well as the session and the map. Then save the repository - again, trying to "remove" the objects from the repository itself. Then re-create them. This forces PM/PC to assign new ID's to ALL the objects in the map, the map, and the session - hopefully creating a "good" picture of all that was rebuilt.

- Or try this method:
1. Create a NEW repository -> call it REPO_A (for reference only).
2. Copy any of the MAPPINGS that don't have "problems" opening in their respective sessions, and copy the mappings (using designer) from the old repository (REPO_B) to the new repository (REPO_A). This will create NEW ID's for all the mappings, CAUTION: You will lose your sessions.
3. DELETE the old repository (REPO_B).
4. Create a new repository in the OLD Repository Space (REPO_B)..
5. Copy the maps back in to the original repository (Recreated Repository) From REPO_A to REPO_B.
6. Rebuild the sessions, then re-create all of the objects you originally had trouble with.

- You can apply this to FOLDER level and Repository Manager Copying, but you need to make sure that none of the objects within a folder have <u>any</u> problems.

- What this does: creates new ID's, resets the sequence generator, re-establishes all the links to the objects in the tables, and drop's out (by process of elimination) any objects you've got problems with.

- Bottom line: PM/PC client tools have trouble when the links between ID's get broken. It's fairly rare that this occurs, but when it does - it can cause heartburn.

Q: *How do I clear the locks that are left in the repository? (3 March 2000)*

Clearing locks is typically a task for the repository manager. Generally it's done from within the Repository Manager: Edit Menu -> Show Locks. Select the locks, then press "remove". Typically locks are left on objects when a client is rebooted without properly exiting Informatica. These locks can keep others from editing the objects. They can also keep scheduled executions from occurring. It's not uncommon to want to clear the locks automatically - on a prescheduled time table, or at a specified time. This can be done safely only if no-one has an object out for editing at the time of deletion of the lock. The suggested method is to log in to the database from an automated script, and issue a "delete from OPB_OBJECT_LOCKS" table.

Q: *How do I turn on the option for Check Repository? (3 March 2000)*

According to Technical Support, it's only available by adjusting the registry entries on the client. PM/PC need to be told it's in Admin mode to work. Below are the steps to turn on the Administration Mode on the client. Be aware - this may be a security risk, anyone using that terminal will have access to these features.

1)start   repository   manager
2)   repository   menu   go   to   check   repository
3)   if   the   option   is   not   there   you   need   to   edit   your   registry   using   regedit
go to: HKEY_CURRENT_USER>>SOFTWARE>>INFORMATICA>>PowerMart Client Tools>>Repository Manager Options

go to your specific version 4.5 or 4.6 and then go to Repository Manager. In there add two strings:
1) EnableAdminMode 1
2) EnableCheckReposit 1
- both should be spelled as shown the value for both is 1

Q: *How do I generate an Audit Trail for my repository (ORACLE / Sybase) ?*
Download one of two *USE AT YOUR OWN RISK* zip files. The first is available now for PowerMart 4.6.x and PowerCenter 1.6x. It's a 7k zip file: Informatica Audit Trail v0.1a The other file (for 4.5.x is coming...). Please note: this is FREE **software that plugs in to ORACLE 7x, and ORACLE 8x, and Oracle 8i**. It has NOT been built for Sybase, Informix, or DB2. If someone would care to adapt it, and send it back to me, I'll be happy to post these also. It has **limited support** - has **not** been fully **tested in a multi-user environment**, any feedback would be appreciated. **NOTE: SYBASE VERSION IS ON IT'S WAY.**

Q: *How do I "tune" a repository? My repository is slowing down after a lot of use, how can I make it faster?*
In Oracle: Schedule a nightly job to ANALYZE TABLE for ALL INDEXES, creating histograms for the tables - keep the cost based optimizer up to date with the statistics. In SYBASE: schedule a nightly job to UPDATE STATISTICS against the tables and indexes. In Informix, DB2, and RDB, see your owners manuals about maintaining SQL query optimizer statistics.

Q: *How do I achieve "best performance" from the Informatica tool set?*
By balancing what Informatica is good at with what the databases are built for. There are reasons for placing some code at the database level - particularly views, and staging tables for data. Informatica is extremely good at reading/writing and manipulating data at very high rates of throughput. However - to achieve optimum performance (in the Gigabyte to Terabyte range) there needs to be a balance of Tuning in Oracle, utilizing staging tables, views for joining source to target data, and throughput of manipulation in Informatica. For instance: Informatica will never achieve the speeds of "append" or straight inserts that Oracle SQL*Loader, or Sybase BCP achieve. This is because these two tools are written internally - specifically for the purposes of loading data (direct to tables / disk structures). The API that Oracle / Sybase provide Informatica with is not nearly as equipped to allow this kind of direct access (to eliminate breakage when Oracle/Sybase upgrade internally). The basics of Informatica are: 1) Keep maps as simple as possible 2) break complexity up in to multiple maps if possible 3) rule of thumb: one MAP per TARGET table 4) Use staging tables for LARGE sets of data 5) utilize SQL for it's power of sorts, aggregations, parallel queries, temp spaces, etc... (setup views in the database, tune indexes on staging tables) 6) Tune the database - partition tables, move them to physical disk areas, etc... separate the logic.

Q: *How do I get an Oracle Sequence Generator to operate faster?*
The first item is: use a function to call it, not a stored procedure. Then, make sure the sequence generator and the function are local to the SOURCE or TARGET database, DO NOT use synonyms to place either the sequence or function in a remote instance (synonyms to a separate schema/database on the same instance may be only a slight performance hit). This should help - possibly double the throughput of generating sequences in your map. The other item is: see slide presentations on performance tuning for your sessions / maps for a "best" way to utilize an Oracle sequence generator. Believe it or not - the write throughput shown in the session manager per target table is directly affected by calling an external function/procedure which is generating sequence numbers. It does NOT appear to affect the read throughput numbers. This is a difficult problem to solve when you have low "write throughput" on any or all of your targets. Start with the sequence number generator (if you can), and try to optimize the map for this.

Q: *I have a mapping that runs for hours, but it's not doing that much. It takes 5 input tables, uses 3 joiner transformations, a few lookups, a couple expressions and a filter before writing to the target. We're running PowerMart 4.6 on an NT 4 box. What tuning options do I have?*
Without knowing the complete environment, it's difficult to say what the problem is, but here's a few solutions with which you can experiment. If the NT box is not dedicated to PowerMart (PM) during its operation, identify what it contends with and try rescheduling things such that PM runs alone. PM needs all the resources it can get. If it's a dedicated box, it's a well known fact that PM consumes resources at a rapid clip, so if you have room for more memory, get it, particularly since you mentioned use of the joiner transformation. Also toy with the caching parameters, but remember that each joiner grabs the full complement of memory that you allocate. So if you give it 50Mb, the 3 joiners will really want 150Mb. You can also try breaking up the session into parallel sessions and put them into a batch, but again, you'll have to manage memory carefully because of the joiners. Parallel sessions is a good option if you have a multiple-processor CPU, so if you have vacant CPU slots, consider adding more CPU's. If a lookup table is relatively big (more than a few thousand rows), try turning the cache flag off in the session and see what happens. So if you're trying to look up a "transaction ID" or something similar out of a few million rows, don't load the table into memory. Just look it up, but be sure the table has appropriate indexes. And last, if the sources live on a pretty powerful box, consider creating a view on the source system that essentially does the same thing as the joiner transformations and possibly some of the lookups. Take advantage of the source system's hardware to do a lot of the work before handing down the result to the resource constrained NT box.

Q: *Is there a "best way" to load tables?*
Yes - If all that is occurring is inserts (to a single target table) - then the BEST method of loading that target is to configure and utilize the bulk loading tools. For Sybase it's BCP, for Oracle it's SQL*Loader. With multiple targets, break the maps apart (see slides), one for INSERTS only, and remove the update strategies from the insert only maps (along with

unnecessary lookups) - then watch the throughput fly. We've achieved 400+ rows per second per table in to 5 target Oracle tables (Sun Sparc E4500, 4 CPU's, Raid 5, 2 GIG RAM, Oracle 8.1.5) without using SQL*Loader. On an NT 366 mhz P3, 128 MB RAM, single disk, single target table, using SQL*Loader we've loaded 1 million rows (150 MB) in 9 minutes total - all the map had was one expression to left and right trim the ports (12 ports, each row was 150 bytes in length). 3 minutes for SQL*Loader to load the flat file - DIRECT, Non-Recoverable.

Q: *How do I guage that the performance of my map is acceptable?*

If you have a small file (under 6MB) and you have pmserver on a Sun Sparc 4000, Solaris 5.6, 2 cpu's, 2 gigs RAM, (baseline configuration - if your's is similar you'll be ok). For NT: 450 MHZ PII 128 MB RAM (under 3 MB file size), then it's nothing to worry about unless your write throughput is sitting at 1 to 5 rows per second. If you are in this range, then your map is too complex, or your tables have not been optimized. On a baseline defined machine (as stated above), expected read throughput will vary - depending on the source, write throughput for relational tables (tables in the database) should be upwards of 150 to 450+ rows per second. To calculate the total write throughput, add all of the rows per second for each target together, run the map several times, and average the throughput. If your map is running "slow" by these standards, then see the slide presentations to implement a different methodology for tuning. The suggestion here is: break the map up - 1 map per target table, place common logic in to maplets.

Q: *How do I create a "state variable"?*

Create a variable port in an expression (v_MYVAR), set the data type to Integer (for this example), set the expression to: IIF((ISNULL(v_MYVAR) = true or v_MYVAR = 0) *[and <your condition>]*, 1, v_MYVAR).> What happens here, is that upon initialization Informatica may set the v_MYVAR to NULL, or zero.> The first time this code is executed it is set to "1".> Of course – you can set the variable to any value you wish – and carry that through the transformations.> Also – you can add your own AND condition (as indicated in italics), and only set the variable when a specific condition has been met.> The variable port will hold it's value for the rest of the transformations.> This is a good technique to use for lookup values when a single lookup value is necessary based on a condition being met (such as a key for an "unknown" value).> You can change the data type to character, and use the same examination – simply remove the "or v_MYVAR = 0" from the expression – character values will be first set to NULL.

Q: *How do I pass a variable in to a session?*

There is no direct method of passing variables in to maps or sessions.> In order to get a map/session to respond to data driven (variables) – a data source must be provided.> If working with flat files – it can be another flat file, if working with relational data sources it can be with another relational table.> Typically a relational table works best, because SQL joins can then be employed to filter the data sets, additional maps and source qualifiers can utilize the data to modify or alter the parameters during run-time.

Q: *How can I create one map, one session, and utilize multiple source files of the same format?*

In UNIX it's very easy: create a link to the source file desired, place the link in the SrcFiles directory, run the session.> Once the session has completed successfully, change the link in the SrcFiles directory to point to the next available source file.> Caution: the only downfall is that you cannot run multiple source files (of the same structure) in to the database simultaneously.> In other words – it forces the same session to be run serially, but if that outweighs the maintenance and speed is not a major issue, feel free to implement it this way.> On NT you would have to physically move the files in and out of the SrcFiles directory.   Note: the difference between creating a link to an individual file, and changing SrcFiles directory to link to a specific directory is this: changing a link to an individual file allows multiple sessions to link to all different types of sources, changing SrcFiles to be a link itself is restrictive – also creates Unix Sys Admin pressures for directory rights to PowerCenter (one level up).

Q: *How can I move my Informatica Logs / BadFiles directories to other disks without changing anything in my sessions?*

Use the UNIX Link command – ask the SA to create the link and grant read/write permissions – have the "real" directory placed on any other disk you wish to have it on.

Q: *How do I handle duplicate rows coming in from a flat file?*

If you don't care about "reporting" duplicates, use an aggregator. Set the Group By Ports to group by the primary key in the parent target table. Keep in mind that using an aggregator causes the following: The last duplicate row in the file is pushed through as the one and only row, loss of ability to detect which rows are duplicates, caching of the data before processing in the map continues. If you wish to report duplicates, then follow the suggestions in the presentation slides (available on this web site) to institute a staging table. See the pro's and cons' of staging tables, and what they can do for you.

Q: *Where can I find a history / metrics of the load sessions that have occurred in Informatica? (8 June 2000)*

The tables which house this information are OPB_LOAD_SESSION, OPB_SESSION_LOG, and OPB_SESS_TARG_LOG. OPB_LOAD_SESSION contains the single session entries, OPB_SESSION_LOG contains a historical log of all session runs that have taken place. OPB_SESS_TARG_LOG keeps track of the errors, and the target tables which have been loaded.   Keep in mind these tables are tied together by Session_ID. If a session is deleted from OPB_LOAD_SESSION, it's history is not necessarily deleted from OPB_SESSION_LOG, nor from OPB_SESS_TARG_LOG. Unfortunately - this leaves un-identified session ID's in these tables. However, when you can join them together, you can get the start and complete times from each session. I would suggest using a view to get the data out (beyond the MX views) - and record it in another metrics table for historical reasons. It could even be done by putting a TRIGGER on these tables (possibly the best solution)...

Q: *Where can I find more information on what the Informatica Repository Tables are?*

On this web-site. We have published an unsupported view of what we believe to be housed in specific tables in the Informatica Repository. Check it out - we'll be adding to this section as we go. Right now it's just a belief of what we see in the tables. Repository Table Meta-Data Definitions

Q: *Where can I find / change the settings regarding font's, colors, and layouts for the designer?*
You can find all the font's, colors, layouts, and controls in the registry of the individual client. All this information is kept at: HKEY_CURRENT_USER\Software\Informatica\PowerMart Client Tools\<ver>. Below here, you'll find the different folders which allow changes to be made. Be careful, deleting items in the registry could hamper the software from working properly.

Q: *Where can I find tuning help above and beyond the manuals?*
Right here. There are slide presentations, either available now, or soon which will cover tuning of Informatica maps and sessions - it does mean that the architectural solution proposed here be put in place.

Q: *Where can I find the map's used in generating performance statistics?*
A windows ZIP file will soon be posted, which houses a repository backup, as well as a simple PERL program that generates the source file, and a SQL script which creates the tables in Oracle. You'll be able to download this, and utilize this for your own benefit.

Q: *Why doesn't constraint based load order work with a maplet? (08 May 2000)*
If your maplet has a sequence generator (reusable) that's mapped with data straight to an "OUTPUT" designation, and then the map splits the output to two tables: parent/child - and your session is marked with "Constraint Based Load Ordering" you may have experienced a load problem - where the constraints do not appear to be met?? Well - the problem is in the perception of what an "OUTPUT" designation is. The OUTPUT component is NOT an "object" that collects a "row" as a row, before pushing it downstream. An OUTPUT component is merely a pass-through structural object - as indicated, there are no data types on the INPUT or OUTPUT components of a maplet - thus indicating merely structure. To make the constraint based load order work properly, move all the ports through a single expression, then through the OUTPUT component - this will force a single row to be "put together" and passed along to the receiving maplet. Otherwise - the sequence generator generates 1 new sequence ID for each split target on the other side of the OUTPUT component.

Q: *Why doesn't 4.7 allow me to set the Stored Procedure connection information in the Session Manager -> Transformations Tab? (31 March 2000)*
This functionality used to exist in an older version of PowerMart/PowerCenter. It was a good feature - as we could control when the procedure was executed (ie: source pre-load), but execute it in a target database connection. It appears to be a removed piece of functionality. We are asking Informatica to put it back in.

Q: *Why doesn't it work when I wrap a sequence generator in a view, with a lookup object?*
First - to wrap a sequence generator in a view, you must create an Oracle stored function, then call the function in the select statement in a view. Second, Oracle dis-allows an order by clause on a column returned from a user function (It will cut your connection - and report an oracle error). I think this is a bug that needs to be reported to Oracle. An Informatica lookup object automatically places an "order by" clause on the return ports / output ports in the order they appear in the object. This includes any "function" return. The minute it executes a non-cached SQL lookup statement with an order by clause on the function return (sequence number) - Oracle cuts the connection. Thus keeping this solution from working (which would be slightly faster than binding an external procedure/function).

Q: *Why doesn't a running session QUIT when Oracle or Sybase return fatal errors?*
The session will only QUIT when it's threshold is set: "Stop on 1 errors". Otherwise the session will continue to run.

Q: *Why doesn't a running session return a non-successful error code to the command line when Oracle or Sybase return any error?*
If the session is not bounded by it's threshold: set "Stop on 1 errors" the session will run to completion - and the server will consider the session to have completed successfully - even if Oracle runs out of Rollback or Temp Log space, even if Sybase has a similar error. To correct this - set the session to stop on 1 error, then the command line: pmcmd will return a non-zero (it failed) type of error code. - as will the session manager see that the session failed.

Q: *Why doesn't the session work when I pass a text date field in to the to_date function?*
In order to make to_date(xxxx,<format>) work properly, we suggest surrounding your expression with the following: IIF(is_date(<date>,<format>) = true, to_date(<date>,<format>), NULL) This will prevent session errors with "transformation error" in the port. If you pass a non-date to a to_date function it will cause the session to bomb out. By testing it first, you ensure 1) that you have a real date, and 2) your format matches the date input. The format should match the expected date input directly - spaces, no spaces, and everything in between. For example, if your date is: 1999103022:31:23 then you want a format to be: YYYYMMDDHH24:MI:SS with no spaces.

Q: *Why doesn't the session control an update to a table (I have no update strategy in the map for this target)?*
In order to process ANY update to any target table, you <u>must</u> put an update strategy in the map, process a DD_UPDATE command, change the session to "data driven". There is a second method: without utilizing an update strategy, set the SESSION properties to "UPDATE" instead of "DATA DRIVEN", but be warned ALL targets will be updated in place - with failure if the rows don't exist. Then you can set the update flags in the mapping's sessions to control updates to the target. Simply setting the "update flags" in a session is not enough to force the update to complete - even though the log may show an update SQL statement, the log will also show: cannot insert (duplicate key) errors.

Q: *Who is the Informatica Sales Team in the Denver Region?*
Christine Connor (Sales), and Alan Schwab (Technical Engineer).

Q: *Who is the contact for Informatica consulting across the country?*

Q: *What happens when I don't connect input ports to a maplet? (14 June 2000)*

Potentially Hazardous values are generated in the maplet itself. Particularly for numerics. If you didn't connect ALL the ports to an input on a maplet, chances are you'll see sporadic values inside the maplet - thus sporadic results. Such as ZERO in certain decimal cases where NULL is desired. This is because both the INPUT and OUTPUT objects of a maplet are nothing more than an interface, which defines the structure of a data row - they are NOT like an expression that actually "receives" or "puts together" a row image. This can cause a misunderstanding of how the maplet works - if you're not careful, you'll end up with unexpected results.

Q: *What is the Local Object Cache? (3 March 2000)*

The local object cache is a cache of the Informatica objects which are retrieved from the repository when a connection is established to a repository. The cache is not readily accessed because it's housed within the PM/PC client tool. When the client is shut-down, the cache is released. Apparently the refresh cycle of this local cache requires a full disconnect/reconnect to the repository which has been updated. This cache will house two different images of the same object. For instance: a shared object, or a shortcut to another folder. If the actual source object is updated (source shared, source shortcut), updates can only be seen in the current open folder if a disconnect/reconnect is performed against that repository. There is no apparent command to refresh the cache from the repository. This may cause some confusion when updating objects then switching back to the mapping where you'd expect to see the newly updated object appear.

Q: *What is the best way to "version control"?*

It seems the general developer community agrees on this one, the Informatica Versioning leaves a lot to be desired. We suggest not utilizing the versioning provided. For two reasons: one, it's extremely unwieldy (you lose all your sessions), and the repository grows exponentially because Informatica copies objects to increase the version number. We suggest two different approaches; 1) utilizing a backup of the repository - synchronize Informatica repository backups (as opposed to DBMS repo backups) with all the developers. Make your backup consistently and frequently. Then - if you need to back out a piece, restore the whole repository. 2) Build on this with a second "scratch" repository, save and restore to the "scratch" repository ONE version of the folders. Drag and drop the folders to and from the "scratch" development repository. Then - if you need to VIEW a much older version, restore that backup to the scratch area, and view the folders. In this manner - you can check in the whole repository backup binary to an outside version control system like PVCS, CCS, SCM, etc... Then restore the whole backup in to acceptance - use the backup as a "VERSION" or snapshot of everything in the repository - this way items don't get lost, and disconnected versions do not get migrated up in to production.

Q: *What is the best way to handle multiple developer environments?*

The school of thought is still out on this one. As with any - there are many many ways to handle this. One idea is presented here (which seems to work well, and be comfortable to those who already worked in shared Source Code environments). The idea is this: All developers use shared folders, shared objects, and global repositories. In development - it's all about communication between team members - so that the items being modified are assigned to individuals for work. With this methodology - all maps can use common mapplets, shared sources, targets, and other items. The one problem with this is that the developers MUST communicate about what they are working on. This is a common and familiar method to working on shared source code - most development teams feel comfortable with this, as do managers. The problem with another commonly utilized method (one folder per developer), is that you end up with run-away development environments. Code re-use, and shared object use nearly always drop to zero percent (caveat: unless you are following SEI / CMM / KPA Level 5 - and you have a dedicated CM (Change Management) person in the works. Communication is still of utmost importance, however now you have the added problem of "checking in" what looks like different source tables from different developers, but the objects are named the same... Among other problems that arise.

Q: *What is the web address to submit new enhancement requests?*

- Informatica's enhancement request web address is: mailto:featurerequest@informatica.com

Q: *What is the execution order of the ports in an expression?*

All ports are executed TOP TO BOTTOM in a serial fashion, but they are done in the following groups: All input ports are pushed values first. Then all variables are executed (top to bottom physical ordering in the expression). Last - all output expressions are executed to push values to output ports - again, top to bottom in physical ordering. You can utilize this to your advantage, by placing lookups in to variables, then using the variables "later" in the execution cycle.

Q: *What is a suggested method for validating fields / marking them with errors?*

One of the successful methods is to create an expression object, which contains variables.> One variable per port that is to be checked.> Set the error "flag" for that field, then at the bottom of the expression trap each of the error fields.> From this port you can choose to set flags based on each individual error which occurred, or feed them out as a combination of concatenated field names – to be inserted in to the database as an error row in an error tracking table.

Q: *What does the error "Broken Pipe" mean in the PMSERVER.ERR log on Unix?*

One of the known causes for this error message is: when someone in the client User Interface queries the server, then presses the "cancel" button that appears briefly in the lower left corner.> It is harmless – and poses no threat.

Q: *What is the best way to create a readable "DEBUG" log?*

Create a table in a relational database which resembles your flat file source (assuming you have a flat file source).> Load the data in to the relational table.> Then – create your map from top to bottom and turn on VERBOSE DATA log at the

session level.> Go back to the map, over-ride the SQL in the SQL Qualifier to only pull one to three rows through the map, then run the session.> In this manner, the DEBUG log will be readable, errors will be much easier to identify – and once the logic is fixed, the whole data set can be run through the map with NORMAL logging.> Otherwise you may end up with a huge (Megabyte) log.> The other two ways to create debugging logs are: 1) switch the session to TEST LOAD, set it to 3 rows, and run… The problem with this is that the reader will read ALL of the source data.> 2) change the output to a flat file…. The problem with this is that your log ends up huge (depends on the number of source rows you have).

Q: *What is the best methodology for utilizing Informatica's Strengths?*

It depends on the purpose. However – there is a basic definition of how well the tool will perform with throughput and data handling, if followed in general principal – you will have a winning situation.> 1) break all complex maps down in to small manageable chunks.> Break up any logic you can in to steps.> Informatica does much better with smaller more maintainable maps. 2) Break up complex logic within an expression in to several different expressions.> Be wary though: the more expressions the slower the throughput – only break up the logic if it's too difficult to maintain.> 3) Follow the guides for table structures and data warehouse structures which are available on this web site.> For reference: load flat files to staging tables, load staging tables in to operational data stores / reference stores / data warehousing sources, load data warehousing sources in to star schemas or snowflakes, load star schemas or snowflakes in to highly de-normalized reporting tables.> By breaking apart the logic you will see the fastest throughput.

Q: *When is it right to use SQL*Loader / BCP as a piped session versus a tail process?*

SQL*Loader / BCP as a piped session should be used when no intermediate file is necessary, or the source data is too large to stage to an intermediate file, there is not enough disk or time to place all the source data in to an intermediate file.> The downfalls currently are this: as a piped process (for PowerCenter 1.5.2 and 1.6 / PowerMart v4.52. and 4.6)> the core does NOT stop when either BCP or SQL*Loader "quit" or terminate.> The core will only stop after reading all of the source data in to the data reader thread.> This is dangerous if you have a huge file you wish to process – and it's scheduled as a monitored process.> Which means: a 5 hour load (in which SQL*Loader / BCP stopped within the first 5 minutes) will only stop and signal a page after 5 hours of reading source data.

Q: *What happens when Informatica causes DR Watson's on NT? (30 October 2000)*

This is just my theory for now, but here's the best explanation I can come up with. Typically this occurs when there is not enough physical RAM available to perform the operation. Usually this only happens when SQLServer is installed on the same machine as the PMServer - however if this is not your case, some of this may still apply. PMServer starts up child threads just like Unix. The threads share the global shared memory area - and rely on NT's Thread capabilities.   The DR Watson seems to appear when a thread attempts to deallocate, or allocate real memory. There's none left (mostly because of SQLServer). The memory manager appears to return an error, or asks the thread to wait while it reorganizes virtual RAM to make way for the physical request. Unfortunately the thread code doesn't pay attention to this request, resulting in a memory violation. The other theory is the thread attempts to free memory that's been swapped to virtual, or has been "garbage collected" and cleared already - thus resulting again in a protected memory mode access violation - thus a DR Watson. Typically the DR Watson can cause the session to "freeze up". The only way to clear this is to stop and restart the PMSERVER service - in some cases it requires a full machine reboot. The only other possibility is when PMServer is attempting to free or shut down a thread - maybe there's an error in the code which causes the DR Watson. In any case, the only real fix is to increase the physical RAM on the machine, or to decrease the number of concurrent sessions running at any given point, or to decrease the amount of RAM that each concurrent session is using.

Q: *What happens when Informatica CORE DUMPS on Unix? (12 April 2000)*

Many things can cause a core dump, but the question is: how do you go about "finding out" what cuased it, how do you work to solve it, and is there a simple fix? This case was found to be frequent (according to tech support) among setups of New Unix Hardware - causing unnecessary core dumps.   The IPC semaphore settings were set too low - causing X number of concurrent sessions to "die" with "writer process died" and "reader process died" etc... We are on a Unix Machine - Sun Solaris 5.7, anyone with this configuration might want to check the settings if they experience "Core Dumps" as well.

1. Run "sysdef", examine the IPC Semaphores section at the bottom of the output.
2. the folowing settings should be "increased"
3. SEMMNI - (semaphore identifiers), (7 x # of concurrent sessions to run in Informatica) + 10 for growth + DBMS setting (DBMS Setting: Oracle = 2 per user, Sybase = 40 (avg))
4. SEMMNU - (undo structures in system) = 0.80 x SEMMNI value
5. SEMUME - (max undo entries per process) = SEMMNU
6. SHMMNI - (shared memory identifiers) = SEMMNI + 10

- These settings must be changed by ROOT: etc/system file.
- About the CORE DUMP: To help Informatica figure out what's going wrong you can run a unix utility: "truss" in the following manner:

1. Shut down PMSERVER
2. login as "powermart" owner of pmserver - cd to the pmserver home directory.
3. Open Session Manager on another client - log in, and be ready to press "start" for the sessions/batches causing problems.
4. type: truss -f -o truss.out pmserver   <hit return>
5. On the client, press "start" for the sessions/batches having trouble.
6. When all the batches have completed or failed, press "stop server" from the Server Manager

- Your "truss.out" file will have been created - thus giving you a log of all the forked processes, and memory management /system calls that will help decipher what's happing. you can examine the "truss.out" file - look for: "killed" in the log.
- DONT FORGET: Following a CORE DUMP it's always a good idea to shut down the unix server, and bounce the box (restart the whole server).

Q: *What happens when Oracle or Sybase goes down in the middle of a transformation?*

It's up to the database to recover up to the last commit point.> If you're asking this question, you should be thinking about re-runnability of your processes.> Designing re-runability in to the processing/maps up front is the best preventative measure you can have.> Utilizing the recovery facility of PowerMart / PowerCenter appears to be sketchy at best – particularly in this area of recovery.> The transformation itself will eventually error out – stating that the database is no longer available (or something to that effect).

Q: *What happens when Oracle (or Sybase) is taken down for routine backup, but nothing is running in PMServer at the time?*

PMServer reports that the database is unavailable in the PMSERVER.err log.> When Oracle/Sybase comes back on line, PMServer will attempt to re-connect (if the repository is on the Oracle/Sybase instance that went down), and eventually it will succeed (when Oracle/Sybase becomes available again).> However – it is recommended that PMServer be scheduled to shutdown before Oracle/Sybase is taken off-line and scheduled to re-start after Oracle/Sybase is put back on-line.

Q: *What happens in a database when a cached LOOKUP object is created (during a session)?*

The session generates a select statement with an Order By clause. Any time this is issued, the databases like Oracle and Sybase will select (read) all the data from the table, in to the temporary database/space. Then the data will be sorted, and read in chunks back to Informatica server. This means, that hot-spot contention for a cached lookup will NOT be the table it just read from. It will be the TEMP area in the database, particularly if the TEMP area is being utilized for other things. Also - once the cache is created, it is not re-read until the next running session re-creates it.

Q: *Can you explain how "constraint based load ordering" works? (27 Jan 2000)*

Constraint based load ordering in PowerMart / PowerCenter works like this: it controls the order in which the target tables are committed to a relational database. It is of no use when sending information to a flat file. To construct the proper constraint order: links between the TARGET tables in Informatica need to be constructed. Simply turning on "constraint based load ordering" has no effect on the operation itself. Informatica does NOT read constraints from the database when this switch is turned on. Again, to take advantage of this switch, you must construct primary / foreign key relationships in the TARGET TABLES in the designer of Informatica. Creating primary / foreign key relationships is difficult - you are only allowed to link a single port (field) to a single table as a primary / foreign key.

Q: *It appears as if "constraint based load ordering" makes my session "hang" (it never completes). How do I fix this? (27 Jan 2000)*

We have a suggested method. The best known method for fixing this "hang" bug is to 1) open the map, 2) delete the target tables (parent / child pairs) 3) Save the map, 4) Drag in the targets again, Parent's FIRST 5) relink the ports, 6) Save the map, 7) refresh the session, and re-run it. What it does: Informatica places the "target load order" as the order in which the targets are created (in the map). It does this because the repository is Seuqence ID Based and the session derives it's "commit" order by the Sequence ID (unless constraint based load ordering is ON), then it tries to re-arrange the commit order based on the constraints in the Target Table definitions (in PowerMart/PowerCenter). Once done, this will solve the commit ordering problems, and the "constraint based" load ordering can even be turned off in the session. Informatica claims not to support this feature in a session that is not INSERT ONLY. However -we've gotten it to work successfully in DATA DRIVEN environments. The only known cause (according to Technical Support) is this: the writer is going to commit a child table (as defined by the key links in the targets). It checks to see if that particular parent row has been committed yet - but it finds nothing (because the reader filled up the memory cache with new rows). The memory that was holding the "committed" rows has been "dumped" and no longer exists. So - the writer waits, and waits, and waits - it never sees a "commit" for the parents, so it never "commits" the child rows. This only appears to happen with files larger than a certain number of rows (depending on your memory settings for the session). The only fix is this: Set "ThrottleReader=20" in the PMSERVER.CFG file. It apparently limits the Reader thread to a maximum of "20" blocks for each session - thus leaving the writer more room to cache the commit blocks. However - this too also hangs in certain situations. To fix this, Tech Support recommends moving to PowerMart 4.6.2 release (internal core apparently needs a fix). 4.6.2 appears to be "better" behaved but not perfect. The only other way to fix this is to turn off constraint based load ordering, choose a different architecture for your maps (see my presentations), and control one map/session per target table and their order of execution.

Q: *Is there a way to copy a session with a map, when copying a map from repository to repository? Say, copying from Development to Acceptance?*

Not that anyone is aware of. There is no direct straight forward method for copying a session. This is the one downside to attempting to version control by folder. You MUST re-create the session in Acceptance (UNLESS) you backup the Development repository, and RESTORE it in to acceptance. This is the only way to take all contents (and sessions) from one repository to another. In this fashion, you are versioning all of the repository at once. With the repository BINARY you can then check this whole binary in to PVCS or some other outside version control system. However, to recreate the session, the best method is to: bring up Development folder/repo, side by side with Acceptance folder/repo - then modify the settings in Acceptance as necessary.

Q: *Can I set Informatica up for Target flat file, and target relational database?*

Up through PowerMart 4.6.2, PowerCenter 1.6.2 this cannot be done in a single map. The best method for this is to stay relational with your first map, add a table to your database that looks exactly like the flat file (1 for 1 with the flat file), target the two relational tables. Then, construct another map which simply reads this "staging" table and dumps it to flat file. You can batch the maps together as sequential.

Q: *How can you optimize use of an Oracle Sequence Generator?*

In order to optimize the use of an Oracle Sequence Generator you must break up you map. The generic method for calling a sequence generator is to encapsulate it in a stored procedure. This is typically slow - and kills the performance. Your version of Informatica's tool should contain maplets to make this easier. Break the map up in to inserts only, and updates only. The suggested method is as follows: 1) Create a staging table - bring the data in straight from the flat file in to the staging table. 2) Create a maplet with the current logic in it. 3) create one INSERT map, and one Update map (separate inserts from updates) 4) create a SOURCE called: DUAL, containing the fields: DUMMY char(1), NEXTVAL NUMBER(15,0), CURRVAL number(15,0), 5) Copy the source in to your INSERT map, 6) delete the Source Qualifier for "dummy" 7) copy the "nextval" port in to the original source qualifier (the one that pulls data from the staging table) 8) Over-ride the SQL in the original source qualifier, (generate it, then change DUAL.NEXTVAL to the sequence name: SQ_TEST.NEXTVAL. 9) Feed the "nextval" port through the mapplet. 10) Change the where clause on the SQL over-ride to select only the data from the staging table that doesn't exist in the parent target (to be inserted. This is extremely fast, and will allow your inserts only map to operate at incredibly high throughput while using an Oracle Sequence Generator. Be sure to tune your indexes on the Oracle tables so that there is a high read throughput.

Q: *Why can't I over-ride the SQL in a lookup, and make the lookup non-cached?*

- Apparently Informatica hasn't made this feature available yet in their tool. It's a shame - it would simplify the method for pulling Oracle Sequence numbers from the database. For now - it's simply not implemented.

Q: *Does it make a difference if I push all my ports (fields) through an expression, or push only the ports which are used in the expression?*

- From the work that has been done - it doesn't make much of an impact on the overall speed of the map. If the paradigm is to push all ports through the expressions for readability then do so, however if it's easier to push the ports around the expression (not through it), then do so.

Q: *What is the affect of having multiple expression objects vs one expression object with all the expressions?*

- Less overall objects in the map make the map/session run faster. Consolodating expressions in to a single expression object is most helpful to throughput - but can increase the complexity (maintenance). Read the question/answer about execution cycles above for hints on how to setup a large expression like this.

Q.Am using a SP that returns a resultset. (ex : select * from cust where cust_id = @cust_id)I am supposed to load the contents of this into the target..As simple as it seems , I am not able to pass the the mapping parameters for cust_idAlso , I cannot have a mapping without SQ Tranf.

Ans: Here select * from cust where cust_id = @cust_id is wrong it should be like this: select * from cust where cust_id = '$$cust_id'

Q.My requirement is like this: Target table structure. Col1, col2, col3, filename

The source file structure will have col1, col2 and col3. All the 10 files have the same structure but different filenames. when i run my mapping thro' file list, i am able to load all the 10 files but the filename column is empty. Hence my requirement is that while reading from the file list, is there any way i can extract the filename and populate into my target table.what u have said is that it will populate into a separate table. But in no way i can find which record has come from which file. Pls help?

Ans: Here PMCMD command can be used with shell script to run the same session by changing the source file name dynamically in the parameter file.

Q.Hi all,i am fighting with this problem for a quiet a bit of time now.I need your help guys (plz)i am trying to load data from DB2 to Oracle.the column in DB2 is of LONGVARCHAR and the column in Oracle that i am mapping to is of CLOB data type.for this it is giving 'parameter binding error,illegal parameter value in LOB function'plz if anybody had faced this kind of problem,guide me.

      (log file give problem as follows:

      WRITER_1_*_1> WRT_8167 Start loading table [SHR_ASSOCIATION] at: Mon Jan 03 17:21:17 2005

      WRITER_1_*_1> Mon Jan 03 17:21:17 2005

      WRITER_1_*_1> WRT_8229 Database errors occurred:

      Database driver error...parameter binding failed

      ORA-24801: illegal parameter value in OCI lob function Database driver error...)

Ans: Informatica Powercenter below 6.2.1 doesn't supports CLOB/BLOB data types but this is supported in   7.0 onwards. So please upgrade to this version or change the data type of u r column to the suitable one.

Q.Hi We are doing production support, when I checked one mapping I found that for that mapping Source is Sybase and Target is Oracle table (in mapping) when I checked in the session for the same maping I found that In session properties they declared the target as Flat file Is it possible?? if so how....when it's possible?

Ans: I think they are loading the data from SYBASE source to Oracle Target using the External Loader.

Q.Is there *any* way to use a SQL statement as a source rather than a table or tables and join them in Informatica via aggregator's, Join's, etc... ?

Ans: SQL Override is there in the Source Qualifier Transformation.

Q.I have a data file in which each record may contain variable number of fields. I have to store these records in oracle table with one to one relationship between record in data file and record in table.

Ans: Question is not clear. But I think he should have the structure of all the records depending on its type. Then use a sequence transformation for getting an unique id for each record.