

The screenshot shows the homepage of the Interview Question Bank. At the top, there is a dark banner with the text "Crack Your Interview Today" in large, bold, yellow-orange letters. Below the banner is a green square icon containing a silhouette of two people at a desk. To the right of the icon, the text "Interview Question Bank" is displayed, followed by "India" in smaller letters. On the far right of the header, there are links for "Store", "Contact", and a shopping cart icon with "(0)".

Interview Question Bank

For Complete Study Material For Interview Preparations Click On Below Link

https://www.instamojo.com/interview_questions

[Click On Below Links For More Courses](#)

Informatica 650+ Interview Questions & Answers

https://www.instamojo.com/interview_questions/informatica-interview-questions-asked-in-top

SQL 500+ Interview Questions & Answers

https://www.instamojo.com/interview_questions/sql-interview-questions-asked-in-top-it-comp

Unix 500+ Interview Questions & Answers

https://www.instamojo.com/interview_questions/unix-interview-questions-asked-in-top-it-com

SAP BODS 250+ Interview Questions & Answers

https://www.instamojo.com/interview_questions/sap-bods-interview-questions-answers-asked-i

SAP BO 250+ Interview Questions & Answers

https://www.instamojo.com/interview_questions/sap-business-objects-interview-questions-ask

Python 500+ Interview Questions & Answers

https://www.instamojo.com/interview_questions/python-interview-questions-asked-in-top-it-c

HR 150+ Interview Questions & Answers

https://www.instamojo.com/interview_questions/hr-round-interview-questions-answers-asked-i

Data Warehousing and ETL Interview Questions & Answers

https://www.instamojo.com/interview_questions/data-warehousing-and-etl-interview-questions

SCD Types (0,1,2,3,4,5,6,7) & Explanation

https://www.instamojo.com/interview_questions/scd-5a66d

Campus Placement 2000+ Interview Questions & Answers For Freshers

https://www.instamojo.com/interview_questions/campus-placement-interview-questions-with-an

100+ Resume , CV & Cover Letters

https://www.instamojo.com/interview_questions/download-your-favorite-resume-template-and-c

Register Here For Job Referral Program

https://www.instamojo.com/interview_questions/apply-for-referral-program-and-get-an-opport

Informatica

Interview Questions & Answers

This Document Consist of all the Questions & Answers as below:

- *Questions asked in Top IT Companies in different Technical rounds*
 - *Technical Round 1 – For 0-1 Year Freshers/Experienced Candidates*
 - *Technical Round 2 – For 1-3 Years Experienced Candidates*
 - *Technical Round 3 – For 3-4 Years Experienced Candidates*
 - *Technical Round 4 – For 4-6 Years Experienced Candidates*
 - *Technical Round 5 – For 6+ Years Experienced Candidates*
- *Certification Questions*
- *Scenario Based Questions*
- *Most Frequently Asked Questions*
 - *Those questions have been repeated in the document to make sure you prepare well for the interview.*

What is ETL?

- **Extract:** This is the first step in the process of the ETL process in the Informatica tool which performs the process of obtaining data sets from the configured data source. This will then be loaded into the data warehouse. This process can also be called Capture.
- **Transform:** The second step in the process of ETL is Transform which will be done from one format to another format which will be supported by the structure of the data warehouse that will be stored into.
- **Load:** This is the third step in the ETL process which will perform the operations of storing the data into Database by creating indexes in the format of data warehouse structure after applying all the required transformations.

What is the need of an ETL tool?

The problem comes with traditional programming languages where you need to connect to multiple sources and you have to handle errors. For this you have to write complex code. ETL tools provide a ready-made solution for this. You dont need to worry about handling these things and can concentrate only on coding the requirement part.

Why it is used?

The conventional application for Informatica is:

- The company uses it to transfer from the current legacy system such as mainframe to the latest system of database. Consequently, the transfer of its existent data into the system could be carried out.
- Companies establishing their warehouses of data will need a tool of ETL to transfer the data to the warehouse from the Production system.
- The assimilation of information from several different systems such as numerous databases and system based on files could be completed utilizing Informatica.
- It could be utilized as a tool for cleansing data.

It is beneficial more than it provides a broad collection of product versions. Hence, consumers could choose a particular edition based on requisite. Informatica is consistently emphasized as a product of the integration of data product leader in the listing of Gartner Magic Quadrant.

Examples of Informatica

Their different applications in the area of Data Integration in the case of Informatica tool and below are the examples:

Example: The different steps in the process of ETL Informatica are as below in every case:

1. Create a Source Definition.
2. Create a Target Definition.
3. Design transformation rules.
4. Create all the mappings.
5. Create a session for each mapping.
6. Create a data connection.
7. Create workflow.
8. Start the process of a workflow.

The above process will always follow in the case of almost every ETL process which is a kind of generalized theme.

What are the different applications of Informatica ETL tool are?

1. It helps large organizations in the industry to store the data into many kinds of physical repositories such as data warehouses.
2. It used for reporting, analysis and pictorial representation of large data for easier presentation.
3. It has Data governance, data warehousing, data migration, data replication, and data synchronization applications.
4. It has Integrated Competency Centers and Matter Data Management facilities.
5. It acts as a Service Oriented Architecture (SOA) feature in terms of data exchange features.
6. It provides great services in Business Intelligence applications.
7. ETL has a vast number of applications in the area of Data Warehouse where it transforms the data into different kinds of formats in order to support the existing data warehouse design structures or data storage structures.
8. It provides integration of different heterogeneous systems such as multiple databases including file-based systems as well.
9. It can also be used as a cleansing tool in case of cleaning the data.
10. It also provides the migration of data from different technological or hardware streams.

What is Informatica PowerCenter?

Informatica is a data processing tool that is widely used for ETL to extract transform and load processing. This is an industry-leading software in the field of data processing and data governance. Informatica is mostly used in the data warehouse, business intelligence and data integration between the business applications domain. Informatica has built-in functionalities to connect with various source systems like databases, file systems or saas based applications using configurations, adapters and in-built connectors. Once the Source system is connected and the source data being captured, Informatica supports several out of the box transformation functionalities to transform the data, route, and process the configured and connected data to the target systems.

Explain different features of informatica?

The different features of what is Informatica are as below:

1. It allows the safe exchange of data in Business to Business and complete end to end visibility.

2. It extracts data from different kinds of databases with different types of data such as Structured, Unstructured and Semi-Structured data and transforms it into required form and stores into the target databases for the customer requirements.
3. It has the feature of parallel processing which is defined as the process of computing the execution of multiple processes simultaneously.
4. It provides the reusability of data, re-running the data and recovery of data.
5. It gives accurate deliverables and provides automation of deployments.
6. It provides high security to the data stored.
7. It provides recovery in the case of data disaster moments and customized access permissions.
8. It has a centralized cloud server that provides high security with easy accessibility and tracking data facility.
9. It provides high visibility in case of presentations for the visual data.
10. It provides easy data maintenance and data monitoring.

What are the important components of the informatica power center?

The important components of the informatica power center are listed below:

Domain: Domain is the primary unit for management and administration of services in Powercenter. The components of domain are one or more nodes, service manager an application services.

Node: Node is logical representation of machine in a domain. A domain can have multiple nodes. Master gateway node is the one that hosts the domain. You can configure nodes to run application services like integration service or repository service. All requests from other nodes go through the master gateway node.

Service Manager: Service manager is for supporting the domain and the application services. The Service Manager runs on each node in the domain. The Service Manager starts and runs the application services on a machine.

Application services: Group of services which represents the informatica server based functionality. Application services include PowerCenter repository service, integration service, Data integration service, Metadata manage service etc.

Powercenter Repository: The metadata is store in a relational database. The tables contain the instructions to extract, transform and load data.

Powercenter Repository service: Accepts requests from the client to create and modify the metadata in the repository. It also accepts requests from the integration service for metadata to run workflows.

Powercenter Integration Service: The integration service extracts data from the source, transforms the data as per the instructions coded in the workflow and loads the data into the targets.

Informatica Administrator: Web application used to administer the domain and powercenter security.

Metadata Manager Service: Runs the metadata manager web application. You can analyze the metadata from various metadata repositories. The important components of the informatica power center are listed below:

Domain: Domain is the primary unit for management and administration of services in Powercenter. The components of domain are one or more nodes, service manager and application services.

Node: Node is logical representation of machine in a domain. A domain can have multiple nodes. Master gateway node is the one that hosts the domain. You can configure nodes to run application services like integration service or repository service. All requests from other nodes go through the master gateway node.

Service Manager: Service manager is for supporting the domain and the application services. The Service Manager runs on each node in the domain. The Service Manager starts and runs the application services on a machine.

Application services: Group of services which represents the informatica server based functionality. Application services include powercenter repository service, integration service, Data integration service, Metadata manage service etc.

Powercenter Repository: The metadata is stored in a relational database. The tables contain the instructions to extract, transform and load data.

Powercenter Repository service: Accepts requests from the client to create and modify the metadata in the repository. It also accepts requests from the integration service for metadata to run workflows.

Powercenter Integration Service: The integration service extracts data from the source, transforms the data as per the instructions coded in the workflow and loads the data into the targets.

Informatica Administrator: Web application used to administer the domain and

powercenter security.

Metadata Manager Service: Runs the metadata manager web application. You can analyze the metadata from various metadata repositories.

What are the Advantages and Disadvantages of Informatica?

Below are the advantages and disadvantages of Informatica are mentioned below:

Below are the advantages:

- Effective interfaces of GUI for Session monitoring, Job Scheduling, ETL Design, Debugging, Administration, etc.
- Queued Message
- Third-party application data
- Mainframe and file-based data
- XML and unstructured data
- Accessibility to a broad variety of data sources of the company
- Relational data
- Could simply acclimate and merge with provided data handling utility through the vendor.
- Load stabilization and parallel processing
- The web-based sole point of authority for an application broad enterprise assuring a great quality of protection with diminished administration expense
- Grid choice gives an affordable answer to suffice the high demand processing with linear scalability and high accessibility.

Disadvantages

When you are browsing several webpages to locate the drawbacks of Informatics you might not be able to find any as there are very few drawbacks. This section comprises information that is never more shown on any other article. Below you might a few disadvantages:

- Workflow Monitor lack of sorting
- In the workflow monitor, there is no choice to separate out folders that are required or needed
- In the repository manager, it not possible to move an object from one folder to another
- Further, you cannot import XML export files
- Also, it includes Mappings and workflows development is also not possible

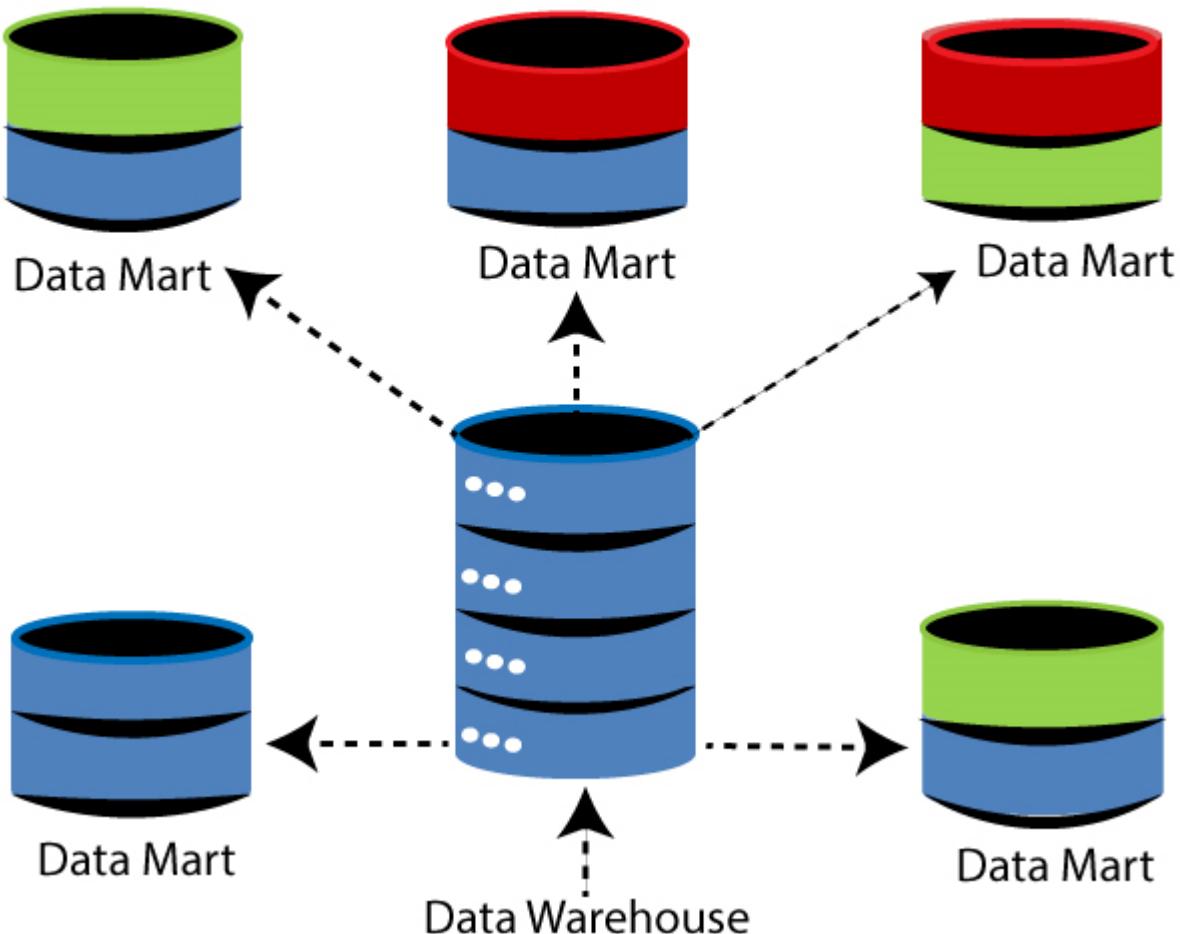
What are the different clients of PowerCenter?

Here is the list of PowerCenter clients:

- PowerCenter designer

- PowerCenter workflow monitor
- PowerCenter workflow manager
- PowerCenter repository manager

Explain the difference between a data warehouse and a data mart?



Data warehouse and Data mart are the structured repositories that store and manage the data. A data warehouse is used to store the data centrally for the entire business while data mart is used to store the specific data, not the entire business data. Querying the data from the data warehouse is a very tedious task, so data mart is used. The data mart is a collection of smaller sets of data which allows you to access the data faster and efficiently.

What is repository manager?

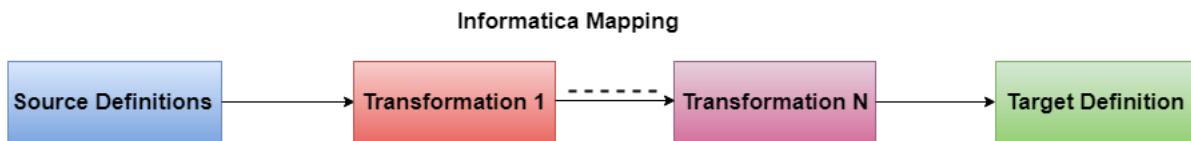
A repository is a place or a relational database used to store the information or metadata. Metadata can include various information such as mappings that describes how to transform the data, sessions describe when you want the Informatica server to perform the transformations, also stores the administrative information such username and password,

permissions and privileges, and product version. The repository is created and maintained by the Repository Manager client tool.

Repository Manager is a manager that manages and organizes the repository. Repository Manager can create the folders to organize the data and groups to handle multiple users.

What is mapping?

Mapping is a pipeline or structural flow of data that describes how data flows from source to the destination through transformations.



Mapping consists of the following components:

- **Source Definition:** Source Definition defines the structure and characteristics of the source such as data types, type of the data source, etc. You can create more than one source definitions by using the Informatica Source Analyzer.
- **Target Definition:** Target Definition defines the final destination or target where the data will be loaded.
- **Transformation:** Transformation defines how source data should be transformed, and various functions are applied during the transformation process.
- **Links:** Links define how data should flow from source definition to the target table by performing different transformations.

Which All Databases Powercenter Server On Windows Can Connect To?

PowerCenter Server on Windows can connect to following databases:

1. IBM DB2
2. Informix
3. Microsoft Access
4. Microsoft Excel
5. Microsoft SQL Server
6. Oracle
7. Sybase
8. Teradata

Which All Databases Powercenter Server On Unix Can Connect To?

PowerCenter Server on UNIX can connect to following databases:

1. IBM DB2

2. Informix
3. Oracle
4. Sybase
5. Teradata

What are the advantages of Informatica over other ETL tools?

Informatica is the world's most popular data integration tool. It interoperates with the widest range of different standards, systems, and applications; it's fast; and it is designed to adapt to the constant change in the field/market, the organization, and the system. Also, you can easily monitor jobs, and it's easy to identify the causes of failed jobs.

What is a transformation?

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

What are the types of transformation?

Types of Transformations:

1. Active Transformation
2. Passive Transformation
3. Active/Passive Transformation
4. Connected/Unconnected Transformation
5. Multigroup Transformation
6. Blocking Transformation

What is an active transformation?

Active Transformation:- An active transformation can perform any of the following actions:

- (a) **Change the number of rows that pass through the transformation**:- for eg, the Filter transformation is active because it removes rows that do not meet the filter condition.
- (b) **Change the transaction boundary**:- for eg, , the Transaction Control transformation is active because it defines a commit or roll back transaction based on an expression evaluated for each row.
- (c) **Change the row type**:- for eg, the Update Strategy transformation is active because it flags rows for insert, delete, update, or reject.

What is a passive transformation?

Passive Transformation:- An Passive transformation which will satisfy all below conditions:

- (a) **Do not Change the number of rows that pass through the transformation**
- (b) **Maintains the transaction boundary**
- (c) **Maintains the row type.**

Important Note:- The Designer does not allow you to connect multiple active transformations or an active and a passive transformation to the same downstream transformation or transformation input group because the Integration Service may not be able to concatenate the rows passed by active transformations. For example, one branch in a mapping contains an Update Strategy transformation that flags a row for delete. Another branch contains an Update Strategy transformation that flags a row for insert. If you connect these transformations to a single transformation input group, the Integration Service cannot combine the delete and insert operations for the row.

What is a connected transformation?

Connected Transformations:- Transformations which are connected to other transformation in the data flow are called Connected Transformations

What is an unconnected transformation?

Unconnected Transformations:- Transformations which are not connected to other transformation in the data flow are called as Unconnected Transformations. An unconnected transformation is called within another transformation, and returns a value to that transformation.

What are multi-group transformations?

Transformations having multiple input and output groups are called multi-group transformations.

Examples: Custom, HTTP, Joiner, Router, Union, Unstructured Data, XML source qualifier, XML Target definition, XML parser, XML generator

List out all the transformations which use cache?

Aggregator, Joiner, Lookup, Rank, Sorter

What is blocking transformation?

Transformation which blocks the input rows are called blocking transformation.

Example: Custom transformation, unsorted joiner

What is a reusable transformation?

A reusable transformation is the one which can be used in multiple mappings. Reusable

transformation is created in transformation developer.

List down all the transformation as per their category?

Active Transformations	Passive Transformations	Both Active & Passive Transformation
1. Aggregator	1. Data Masking	1. Custom
2. Application Source Qualifier &	2. Expression	2. Lookup
3. Source Qualifier	3. External Procedure	3. Unstructured Data
4. Sorter	4. Input	4. SQL
5. Union	5. Output	5. Java
6. Update Strategy	6. Sequence Generator	
7. Rank	7. Stored Procedure	
8. Router	8. HTTP	
9. Filter		
10. Normalizer		
11. Transaction Control		
12. Joiner		
13. XML Source Qualifier		
14. XML Generator		
15. XML Parser		

Connected and Unconnected Transformation

- 1. External Procedure
- 2. Stored Procedure
- 3. Lookup

Connected Transformation

Rest All transformations are connected.

Multigroup Transformations

- 1. Union
- 2. Router
- 3. Joiner
- 4. Custom
- 5. Unstructured Data
- 6. XML Source Qualifier
- 7. XML Generator
- 8. XML Parser

9. XML Target Definition

Blocking Transformations

1. Custom transformation with Input may block property enabled. 
2. Joiner transformation configured for unsorted Input. 

What Are The Methods For Creating Reusable Transformations?

There two methods for creating reusable transformations:

1. Using transformation developer tool.
2. Converting a non-reusable transformation into a reusable transformation in mapping.

How do you promote a non-reusable transformation to reusable transformation?

Edit the transformation and check the Make Reusable option

How to create a non-reusable instance of reusable transformations?

In the navigator, select an existing transformation and drag the transformation into the mapping workspace. Hold down the Ctrl key before you release the transformation.

Which transformation can be created only as reusable transformation but not as non-reusable transformation?

External procedure transformation.

Does an Informatica Transformation support only Aggregate expressions?

Apart from aggregate expressions Informatica Aggregator also supports non-aggregate expressions and conditional clauses.

At The Max How Many Transformations Can Be Us In A Mapping?

In a mapping we can use any number of transformations depending on the project, and the included transformations in the particular related transformations.

What is a union transformation?

A union transformation is used merge data from multiple sources similar to the UNION ALL SQL statement to combine the results from two or more SQL statements.

As union transformation gives UNION ALL output, how you will get the UNION output?

Pass the output of union transformation to a sorter transformation. In the properties of sorter transformation check the option select distinct. Alternatively you can pass the output of union transformation to aggregator transformation and in the aggregator transformation specify all ports as group by ports.

What are the guidelines to be followed while using union transformation?

The following rules and guidelines need to be taken care while working with union transformation:

- You can create multiple input groups, but only one output group.
- All input groups and the output group must have matching ports. The precision, datatype, and scale must be identical across all groups.
- The Union transformation does not remove duplicate rows. To remove duplicate rows, you must add another transformation such as a Router or Filter transformation.
- You cannot use a Sequence Generator or Update Strategy transformation upstream from a Union transformation.
- The Union transformation does not generate transactions.

Why union transformation is an active transformation?

Union is an active transformation because it combines two or more data streams into one. Though the total number of rows passing into the Union is the same as the total number of rows passing out of it, and the sequence of rows from any given input stream is preserved in the output, the positions of the rows are not preserved, i.e. row number 1 from input stream 1 might not be row number 1 in the output stream. Union does not even guarantee that the output is repeatable

What are the restrictions of Union Transformation?

- All input groups and the output group must have matching ports. The precision, datatype, and scale must be identical across all groups.
- We can create multiple input groups, but only one default output group.
- The Union transformation does not remove duplicate rows.
- We cannot use a Sequence Generator or Update Strategy transformation upstream from a Union transformation.
- The Union transformation does not generate transactions.

What is a transaction control transformation?

A transaction is a set of rows bound by a commit or rollback of rows. The transaction control transformation is used to commit or rollback a group of rows.

What is the commit type if you have a transaction control transformation in the mapping?

The commit type is "user-defined".

What are the different transaction levels available in transaction control transformation?

The following are the transaction levels or built-in variables:

- TC_CONTINUE_TRANSACTION: The Integration Service does not perform any transaction change for this row. This is the default value of the expression.
- TC_COMMIT_BEFORE: The Integration Service commits the transaction, begins a new transaction, and writes the current row to the target. The current row is in the new transaction.
- TC_COMMIT_AFTER: The Integration Service writes the current row to the target, commits the transaction, and begins a new transaction. The current row is in the committed transaction.
- TC_ROLLBACK_BEFORE: The Integration Service rolls back the current transaction, begins a new transaction, and writes the current row to the target. The current row is in the new transaction.
- TC_ROLLBACK_AFTER: The Integration Service writes the current row to the target, rolls back the transaction, and begins a new transaction. The current row is in the rolled back transaction.

What is a sorter transformation?

Sorter transformation is used to sort the data. You can sort the data either in ascending or descending order according to a specified sort key.

Why sorter is an active transformation?

It is an active transformation because it removes the duplicates from the key and consequently changes the number of rows..

How to improve the performance of a session using sorter transformation?

Sort the data using sorter transformation before passing in to aggregator or joiner transformation. As the data is sorted, the integration service uses the memory to do aggregate and join operations and does not use cache files to process the data.

How To Delete Duplicate Rows In Flat Files Source Is Any Option In Informatica?

Use a sorter transformation, in that you will have a "distinct" option make use of it .

What is an expression transformation?

An expression transformation is used to calculate values in a single row.

Example: salary+1000

How to generate sequence numbers using expression transformation?

Create a variable port in expression transformation and increment it by one for every row.

Assign this variable port to an output port.

What is SQL transformation?

SQL transformation process SQL queries midstream in a pipeline and you can insert, update, delete and retrieve rows from a database.

How do you configure a SQL transformation?

The following options are required to configure SQL transformation:

- Mode: Specifies the mode in which SQL transformation runs. SQL transformation supports two modes. They are script mode and query mode.
- Database type: The type of database that SQL transformation connects to.
- Connection type: Pass database connection to the SQL transformation at run time or specify a connection object.

What are the different modes in which a SQL transformation runs?

SQL transformation runs in two modes. They are:

- Script mode: The SQL transformation runs scripts that are externally located. You can pass a script name to the transformation with each input row. The SQL transformation outputs one row for each input row.
- Query mode: The SQL transformation executes a query that you define in a query editor. You can pass parameters to the query to define dynamic queries. You can output multiple rows when the query has a SELECT statement.

In which cases the SQL transformation becomes a passive transformation and active transformation?

If you run the SQL transformation in script mode, then it becomes passive transformation. If you run the SQL transformation in the query mode and the query has a SELECT statement, then it becomes an active transformation.

When you configure an SQL transformation to run in script mode, what are the ports that the designer adds to the SQL transformation?

The designer adds the following ports to the SQL transformation in script mode:

- ScriptName: This is an input port. ScriptName receives the name of the script to execute the current row.
- ScriptResult: This is an output port. ScriptResult returns PASSED if the script execution succeeds for the row. Otherwise it returns FAILED.
- ScriptError: This is an output port. ScriptError returns the errors that occur when a script fails for a row.

What are the types of SQL queries you can specify in the SQL transformation when you use it in query mode?

- Static SQL query: The query statement does not change, but you can use query parameters to change the data. The integration service prepares the query once and runs the query for all input rows.
- Dynamic SQL query: The query statement can be changed. The integration service prepares a query for each input row.

What are the types of connections to connect the SQL transformation to the database available?

- Static connection: Configure the connection object in the session. You must first create the connection object in workflow manager.
- Logical connection: Pass a connection name to the SQL transformation as input data at run time. You must first create the connection object in workflow manager.
- Full database connection: Pass the connect string, user name, password and other connection information to SQL transformation input ports at run time.

How do you find the number of rows inserted, updated or deleted in a table?

You can enable the NumRowsAffected output port to return the number of rows affected by the INSERT, UPDATE or DELETE query statements in each input row. This NumRowsAffected

option works in query mode.

What will be the output of NumRowsAffected port for a SELECT statement?

The NumRowsAffected output is zero for the SELECT statement.

When you enable the NumRowsAffected output port in script mode, what will be the output?

In script mode, the NumRowsAffected port always returns NULL.

How do you limit the number of rows returned by the select statement?

You can limit the number of rows by configuring the Max Output Row Count property. To configure unlimited output rows, set Max Output Row Count to zero.

What is a lookup transformation?

A lookup transformation is used to look up data in a flat file, relational table, view, and synonym.

What are the types of lookup transformations?

We can configure the Lookup transformation to one of the following types of lookups:

1. Connected or Unconnected
2. Relational or Flat File
3. Cached or Un Cached
4. Pipeline Lookup

1. Connected or Unconnected : In Connected Lookup port are connected to transformation in the pipeline. Unconnected Lookup ports are not connected to transformations in pipeline and result is send through as a result of :LKP expression.

2. Relational Lookup: When we create a Lookup transformation using a relational table as a lookup source then the resulting lookup is relational lookup. We can connect to the lookup source using ODBC and import the table definition as the structure for the Lookup transformation.

a) We can override the default SQL statement if we want to add a WHERE clause or query multiple tables.

b) We can use a dynamic lookup cache with relational lookups.

3. Flat File Lookup: When we use a flat file for a lookup source, we can use any flat file definition in the repository, or we can import it. When we import a flat file lookup source, the Designer invokes the Flat File Wizard.

4. Cached or Un-cached Lookup: We can check the option in Properties Tab to Cache to lookup or not. By default, lookup is cached.

5. Pipeline Lookup : When we use SAP application as lookup source in that case we use pipeline lookup.

Describe the differences between an SQL override and a lookup override?

When you want to limit the number of rows entering a mapping pipeline, you'd use an SQL override. When you want to limit the number of lookup rows to avoid scanning an entire table, you'd use the lookup override. Lookup override provides only one record even if multiple records for a condition exist. Also, SQL override doesn't use the "order by" clause—you have to manually enter it in the query.

What are the tasks of a lookup transformation?

Perform the following tasks with a Lookup transformation:

1. **Get a related value** : Retrieve a value from the lookup table based on a value in the source. For example, the source has an employee ID. Retrieve the employee name from the lookup table.
2. **Get multiple values** : Retrieve multiple rows from a lookup table. For example, return all employees in a department.
3. **Perform a calculation** : Retrieve a value from a lookup table and use it in a calculation. For example, retrieve a sales tax percentage, calculate a tax, and return the tax to a target.
4. **Update slowly changing dimension tables** : Determine whether rows exist in a target.

Explain in details Lookup ports?

Lookup Ports

Ports	Lookup Type	Number Needed	Description
I	Connected Unconnected	Minimum 1	Input port to Lookup. Usually ports used for Join condition are Input ports.
O	Connected Unconnected	Minimum 1	Ports going to another transformation from Lookup.
L	Connected Unconnected	Minimum 1	Lookup port. The Designer automatically Designates each column in the lookup source as a lookup (L) and output port (O).
R	Unconnected	1 Only	Return port. Use only in unconnected Lookup t/f only.

Explain in detail lookup properties?

Lookup Properties

Configure the lookup properties such as caching and multiple matches on the Lookup Properties tab. Configure the lookup condition or the SQL statements to query the lookup table. You can also change the Lookup table name.

Options	Lookup Type	Description
Lookup SQL Override	Relational	Overrides the default SQL statement to query the lookup table.
Lookup Table Name	Relational	Specifies the name of the table from which the transformation looks up and caches values.
Lookup Caching Enabled	Flat File, Relational	Indicates whether the Power Center Server caches lookup values during the session.
Lookup Policy on Multiple Match	Flat File, Relational	Determines what happens when the Lookup transformation finds multiple rows that match the lookup condition. Options: Use First Value or Use Last Value or Use Any Value or Report Error
Lookup Condition	Flat File, Relational	Displays the lookup condition you set in the Condition tab.
Connection Information	Relational	Specifies the database containing the lookup table.
Source Type	Flat File, Relational	Lookup is from a database or flat file.
Lookup Cache Directory Name	Flat File, Relational	Location where cache is build.
Lookup Cache Persistent	Flat File, Relational	Whether to use Persistent Cache or not.
Dynamic Lookup Cache	Flat File, Relational	Whether to use Dynamic Cache or not.
Recache From Lookup Source	Flat File, Relational	To rebuild cache if cache source changes and we are using Persistent Cache.
Insert Else Update	Relational	Use only with dynamic caching enabled. Applies to rows entering the Lookup transformation with the row type of insert.
Lookup Data Cache Size	Flat File, Relational	Data Cache Size
Lookup Index Cache Size	Flat File, Relational	Index Cache Size

Cache File Name Prefix	Flat File, Relational	Use only with persistent lookup cache. Specifies the file name prefix to use with persistent lookup cache files.
------------------------	-----------------------	--

How do you configure a lookup transformation?

Configure the lookup transformation to perform the following types of lookups:

- Relational or flat file lookup
- Pipeline lookup
- Connected or unconnected lookup
- Cached or Uncached lookup

What is a pipeline lookup transformation?

A pipeline lookup transformation is used to perform lookup on application sources such as JMS, MSMQ or SAP. A pipeline lookup transformation has a source qualifier as the lookups source.

What is connected and unconnected lookup transformation?

- A connected lookup transformation is connected to the transformations in the mapping pipeline. It receives source data, performs a lookup and returns data to the pipeline.
- An unconnected lookup transformation is not connected to the other transformations in the mapping pipeline. A transformation in the pipeline calls the unconnected lookup with a :LKP expression.

How many input parameters can exist in an unconnected lookup?

Any number of input parameters can exist. For instance, you can provide input parameters like column 1, column 2, column 3, and so on. But the return value would only be one.

What are the differences between connected and unconnected lookup transformation?

Difference Between Connected and Unconnected Lookup

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.
Cache includes all lookup columns used in the mapping.	Cache includes all lookup/output ports in the lookup condition and the lookup/return port.

If there is no match for the lookup condition, the Power Center Server returns the default value for all output ports.	If there is no match for the lookup condition, the Power Center Server returns NULL.
If there is a match for the lookup condition, the Power Center Server returns the result of the lookup condition for all lookup/output ports.	If there is a match for the lookup condition, the Power Center Server returns the result of the lookup condition into the return port.
Pass multiple output values to another transformation.	Pass one output value to another transformation.
Supports user-defined default values	Does not support user-defined default values.

Which Is Better Among Connected Lookup And Unconnected Lookup Transformations In Informatica Or Any Other Etl Tool?

If you are having defined source you can use connected, source is not well defined or from different database you can go for unconnected.

Connected and unconnected lookup depends on scenarios and performance If you are looking for a single value for look up and the value is like 1 in 1000 then you should go for unconnected lookup. Performance wise its better as we are not frequently using the transformation. If multiple columns are returned as lookup value then one should go for connected lookup.

How many input parameters can be present in an unconnected lookup?

The number of parameters that can combine in an unconnected lookup is numerous. However, no matter how many parameters are put, the return value would be only one. If we take an example like box 1, box 2, box 3, box 4 can be put in an unconnected lookup, but there is only one return value.

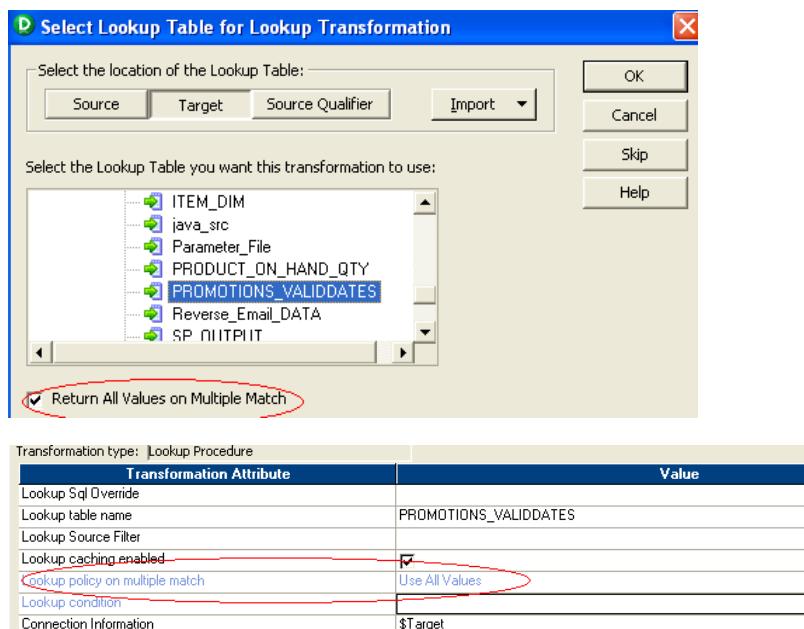
How Do You Create A Mapping Using Multiple Lookup Transformation?

Use unconnected lookup if same lookup repeats multiple times.

What happens if the Lookup transformation finds multiple matches?

# Report Error. The Integration Service reports an error and does not return a row.	
# Use First Value. Returns the first row that matches the lookup condition.	

#	Use Last Value. Return the last row that matches the lookup condition.
#	Use All Values. Return all matching rows.
#	Use Any Value. The Integration Service returns the first value that matches the lookup condition. It creates an index based on the key ports instead of all Lookup transformation ports.
#	Report Error. The Integration Service reports an error and does not return a row. If you do not enable the Output Old Value On Update option, the Lookup Policy On Multiple Match option is set to Report Error for dynamic lookups.



What is "Output Old Value on Update"?

This option is used when dynamic cache is enabled. When this option is enabled, the integration service outputs old values out of the lookup/output ports. When the Integration Service updates a row in the cache, it outputs the value that existed in the lookup cache before it updated the row based on the input data. When the Integration Service inserts a new row in the cache, it outputs null values. When you disable this property, the Integration

Service outputs the same values out of the lookup/output and input/output ports.

What is "Insert Else Update" and "Update Else Insert"?

These options are used when dynamic cache is enabled.

- Insert Else Update option applies to rows entering the lookup transformation with the row type of insert. When this option is enabled the integration service inserts new rows in the cache and updates existing rows when disabled, the Integration Service does not update existing rows.
- Update Else Insert option applies to rows entering the lookup transformation with the row type of update. When this option is enabled, the Integration Service updates existing rows, and inserts a new row if it is new. When disabled, the Integration Service does not insert new rows.

What are the options available to configure a lookup cache?

The following options can be used to configure a lookup cache:

- Persistent cache
- Recache from lookup source
- Static cache
- Dynamic cache
- Shared Cache
- Pre-build lookup cache

What is a cached lookup transformation and Uncached lookup transformation?

- Cached lookup transformation: The Integration Service builds a cache in memory when it processes the first row of data in a cached Lookup transformation. The Integration Service stores condition values in the index cache and output values in the data cache. The Integration Service queries the cache for each row that enters the transformation.
- Uncached lookup transformation: For each row that enters the lookup transformation, the Integration Service queries the lookup source and returns a value. The integration service does not build a cache.

How the integration service builds the caches for connected lookup transformation?

The Integration Service builds the lookup caches for connected lookup transformation in the following ways:

- Sequential cache: The Integration Service builds lookup caches sequentially. The Integration Service builds the cache in memory when it processes the first row of the data in a cached lookup transformation.
- Concurrent caches: The Integration Service builds lookup caches concurrently. It does not need to wait for data to reach the Lookup transformation.

How the integration service builds the caches for unconnected lookup transformation?

The Integration Service builds caches for unconnected Lookup transformations sequentially.

What is a dynamic cache?

The dynamic cache represents the data in the target. The Integration Service builds the cache when it processes the first lookup request. It queries the cache based on the lookup condition for each row that passes into the transformation. The Integration Service updates the lookup cache as it passes rows to the target. The integration service either inserts the row in the cache or updates the row in the cache or makes no change to the cache.

When you use a dynamic cache, do you need to associate each lookup port with the input port?

Yes. You need to associate each lookup/output port with the input/output port or a sequence ID. The Integration Service uses the data in the associated port to insert or update rows in the lookup cache.

What are the different values returned by NewLookupRow port?

The different values are

- 0 - Integration Service does not update or insert the row in the cache.
- 1 - Integration Service inserts the row into the cache.
- 2 - Integration Service updates the row in the cache.

Difference Between Static And Dynamic Cache And Explain With One Example?

Static - Once the data is cached , it will not change, example unconnected lookup uses static cache.

Dynamic - The cache is updated as to reflect the update in the table (or source) for which it is referring to.(ex. connected lookup).

What is a persistent cache?

If the lookup source does not change between session runs, then you can improve the performance by creating a persistent cache for the source. When a session runs for the first time, the integration service creates the cache files and saves them to disk instead of

deleting them. The next time when the session runs, the integration service builds the memory from the cache file.

What is a shared cache?

You can configure multiple Lookup transformations in a mapping to share a single lookup cache. The Integration Service builds the cache when it processes the first Lookup transformation. It uses the same cache to perform lookups for subsequent Lookup transformations that share the cache.

What is shared cache and its types?

Shared Cache :

- **Unnamed cache:** When Lookup transformations in a mapping have compatible caching structures, the IS shares the cache by default. You can only share static unnamed caches.
- **Named cache:** Use a persistent named cache when we want to share a cache file across mappings or share a dynamic and a static cache. The caching structures must match or be compatible with a named cache. You can share static and dynamic named caches.

How do you improve the performance of lookup transformation?

- Create an index on the columns used in the lookup condition
- Place conditions with equality operator first
- Cache small lookup tables.
- Join tables in the database: If the source and the lookup table are in the same database, join the tables in the database rather than using a lookup transformation.
- Use persistent cache for static lookups.
- Avoid ORDER BY on all columns in the lookup source. Specify explicitly the ORDER By clause on the required columns.
- For flat file lookups, provide Sorted files as lookup source.

What is an update strategy transformation?

Update strategy transformation is used to flag source rows for insert, update, delete or reject within a mapping. Based on this flagging each row will be either inserted or updated or deleted from the target. Alternatively the row can be rejected.

Why update strategy is an active transformation?

As update strategy transformation can reject rows, it is called as an active transformation.

What are the constants used in update strategy transformation for flagging the rows?

- DD_INSERT is used for inserting the rows. The numeric value is 0.
- DD_UPDATE is used for updating the rows. The numeric value is 1.
- DD_DELETE is used for deleting the rows. The numeric value is 2.
- DD_REJECT is used for rejecting the rows. The numeric value is 3.

If you place an aggregator after the update strategy transformation, how the output of aggregator will be affected?

The update strategy transformation flags the rows for insert, update and delete or reject before you perform aggregate calculation. How you flag a particular row determines how the aggregator transformation treats any values in that row used in the calculation. For example, if you flag a row for delete and then later use the row to calculate the sum, the integration service subtracts the value appearing in this row. If the row had been flagged for insert, the integration service would add its value to the sum.

How do you update the records with or without using Update Strategy?

We can use the session configurations to update the records. We can have several options for handling database operations such as insert, update, delete.

During session configuration, you can select a single database operation for all rows using the Treat Source Rows As setting from the ‘Properties’ tab of the session.

- **Insert:** – Treat all rows as inserts.
- **Delete:** – Treat all rows as deletes.
- **Update:** – Treat all rows as updates.
- **Data Driven :-** Integration Service follows instructions coded into Update Strategy flag rows for insert, delete, update, or reject.

Once determined how to treat all rows in the session, we can also set options for individual rows, which gives additional control over how each row behaves. We need to define these options in the Transformations view on mapping tab of the session properties.

- **Insert:** – Select this option to insert a row into a target table.
- **Delete:** – Select this option to delete a row from a table.
- **Update :** You have the following options in this situation:
 - Update as Update: – Update each row flagged for update if it exists in the target table.
 - Update as Insert: – Insert each row flagged for update.
 - Update else Insert: – Update the row if it exists. Otherwise, insert it.
- **Truncate Table:** – Select this option to truncate the target table before loading data.

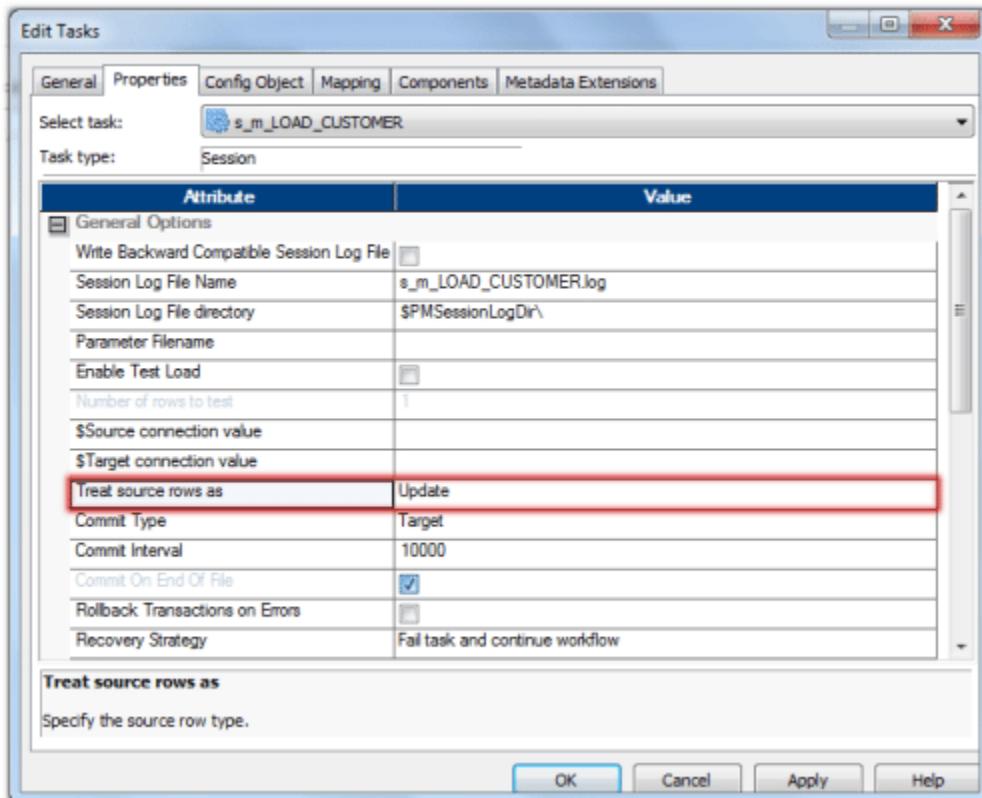
Steps:

1. Design the mapping just like an 'INSERT' only mapping, without Lookup, Update Strategy

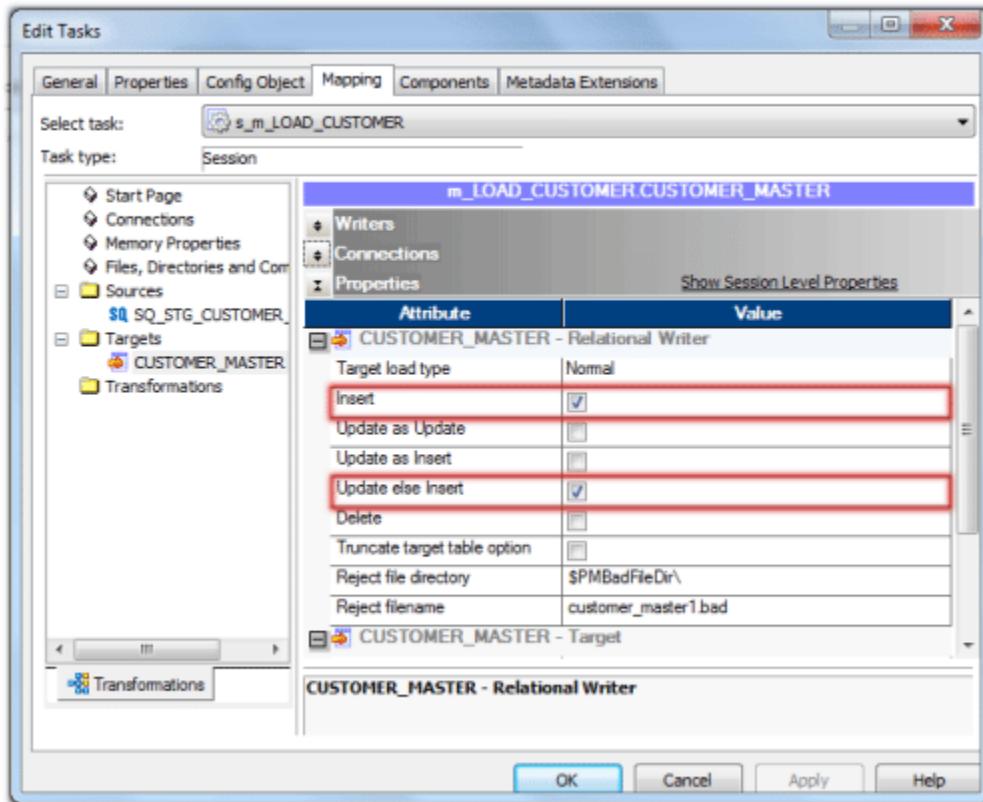


Transformation.

2. First set Treat Source Rows As property as shown in below image.



3. Next, set the properties for the target table as shown below. Choose the properties Insert and Update else Insert.



These options will make the session as Update and Insert records without using Update Strategy in Target Table.

When we need to update a huge table with few records and less inserts, we can use this solution to improve the session performance.

The solutions for such situations is not to use Lookup Transformation and Update Strategy to insert and update records.

The Lookup Transformation may not perform better as the lookup table size increases and it also degrades the performance.

Explain Use Of Update Strategy Transformation at mapping & session level?

To flag source records as INSERT, DELETE, UPDATE or REJECT for target database. Default flag is Insert. This is must for Incremental Data Loading.

This is the important transformation, is used to maintain the history data or just most recent changes into the target table.

We can set or flag the records by using these two levels.

1. Within a session :

When you configure the session, you can instruct the informatica server to either treat all the records in the same way.

2. Within a mapping :

within a mapping we use update strategy transformation to flag the records like insert,update,delete or reject.

If you have an update strategy transformation in the mapping, what should be the value selected for 'Treat Source Rows As' option in session properties?

The value selected for the option is 'Data Driven'. The integration service follows the instructions coded in the update strategy transformation.

What are data driven sessions?

When you configure a session using update strategy, the session property data driven instructs Informatica server to use the instructions coded in mapping to flag the rows for insert, update, delete or reject. This is done by mentioning DD_UPDATE or DD_INSERT or DD_DELETE in the update strategy transformation.

“Treat source rows as” property in session is set to “Data Driven” by default when using a update strategy transformation in a mapping.

If you have an update strategy transformation in the mapping and you did not selected the value 'Data Driven' for 'Treat Source Rows As' option in session, then how the session will behave?

If you do not choose Data Driven when a mapping contains an Update Strategy or Custom transformation, the Workflow Manager displays a warning. When you run the session, the Integration Service does not follow instructions in the Update Strategy transformation in the mapping to determine how to flag rows.

In which files the data rejected by update strategy transformation will be written?

If the update strategy transformation is configured to Forward Rejected Rows then the integration service forwards the rejected rows to next transformation and writes them to the session reject file. If you do not select the forward reject rows option, the integration service drops rejected rows and writes them to the session log file. If you enable row error handling, the Integration Service writes the rejected rows and the dropped rows to the row error logs. It does not generate a reject file.

What are Update Strategy's target table options?

Update as Update: Updates each row flagged for an update if it exists in the table.

Update as Insert: Inserts a new row for each update.

Update else Inserts Updates if the row exists, else inserts.

What is a stored procedure?

A stored procedure is a precompiled collection of database procedural statements. Stored procedures are stored and run within the database.

Give some examples where a stored procedure is used?

The stored procedure can be used to do the following tasks

- Check the status of a target database before loading data into it.
- Determine if enough space exists in a database.
- Perform a specialized calculation.
- Drop and recreate indexes.

What is a connected stored procedure transformation?

The stored procedure transformation is connected to the other transformations in the mapping pipeline.

In which scenarios a connected stored procedure transformation is used?

- Run a stored procedure every time a row passes through the mapping.
- Pass parameters to the stored procedure and receive multiple output parameters.

What is an unconnected stored procedure transformation?

The stored procedure transformation is not connected directly to the flow of the mapping. It either runs before or after the session or is called by an expression in another transformation in the mapping.

In which scenarios an unconnected stored procedure transformation is used?

- Run a stored procedure before or after a session
- Run a stored procedure once during a mapping, such as pre or post-session.

- Run a stored procedure based on data that passes through the mapping, such as when a specific port does not contain a null value.
- Run nested stored procedures.
- Call multiple times within a mapping.

What are the options available to specify when the stored procedure transformation needs to be run?

The following options describe when the stored procedure transformation runs:

- Normal: The stored procedure runs where the transformation exists in the mapping on a row-by-row basis. This is useful for calling the stored procedure for each row of data that passes through the mapping, such as running a calculation against an input port. Connected stored procedures run only in normal mode.
- Pre-load of the Source: Before the session retrieves data from the source, the stored procedure runs. This is useful for verifying the existence of tables or performing joins of data in a temporary table.
- Post-load of the Source: After the session retrieves data from the source, the stored procedure runs. This is useful for removing temporary tables.
- Pre-load of the Target: Before the session sends data to the target, the stored procedure runs. This is useful for verifying target tables or disk space on the target system.
- Post-load of the Target: After the session sends data to the target, the stored procedure runs. This is useful for re-creating indexes on the database.

A connected stored procedure transformation runs only in Normal mode. A unconnected stored procedure transformation runs in all the above modes.

What is execution order in stored procedure transformation?

The order in which the Integration Service calls the stored procedure used in the transformation, relative to any other stored procedures in the same mapping. Only used when the Stored Procedure Type is set to anything except Normal and more than one stored procedure exists.

What is PROC_RESULT in stored procedure transformation?

PROC_RESULT is a system variable, where the output of an unconnected stored procedure transformation is assigned by default.

What is Status Code in INFORMATICA?

Code provides an Error handling mechanism during each session. Status Code is issued by the stored procedure to recognize whether it is committed successfully or not and provides information to the INFORMATICA server to decide whether the session has to be stopped or continued.

What are the parameter types in a stored procedure?

There are three types of parameters exist in a stored procedure:

- IN: Input passed to the stored procedure
- OUT: Output returned from the stored procedure
- INOUT: Defines the parameter as both input and output. Only Oracle supports this parameter type.

What is a source qualifier transformation?

A source qualifier represents the rows that the integration service reads when it runs a session. Source qualifier is an active transformation.

Why you need a source qualifier transformation?

The source qualifier transformation converts the source data types into informatica native data types.

What are the different tasks a source qualifier can do?

- Join two or more tables originating from the same source (homogeneous sources) database.
- Filter the rows.
- Sort the data
- Selecting distinct values from the source
- Create custom query
- Specify a pre-sql and post-sql

What all task we can perform by Source Qualifier transformation?

Tasks can be performed as below:

1. **Join data originating from the same source database.** You can join two or more tables with primary key foreign key relationships by linking the sources to one Source Qualifier transformation.
2. **Filter rows** when the Integration Service reads source data. If you include a filter condition, the Integration Service adds a WHERE clause to the default query.
3. **Specify an outer join rather than the default inner join.** If you include a user-defined join, the Integration Service replaces the join information specified by the metadata in the SQL query.
4. **Specify sorted ports.** If you specify a number for sorted ports, the Integration Service adds an ORDER BY clause to the default SQL query.
5. **Select only distinct values from the source.** If you choose Select Distinct, the Integration Service adds a SELECT DISTINCT statement to the default SQL query.
6. **Create a custom query to issue a special SELECT statement** for the Integration Service to read source data. For example, you might use a custom query to perform aggregate calculations.

What is the default join in source qualifier transformation?

The source qualifier transformation joins the tables based on the primary key-foreign key relationship. "Inner equi join."

How to create a custom join in source qualifier transformation?

When there is no primary key-foreign key relationship between the tables, you can specify a custom join using the 'user-defined join' option in the properties tab of source qualifier.

What Are The Basic Needs To Join Two Sources In A Source Qualifier?

Basic need to join two sources using source qualifier:

1. Both sources should be in same database
2. The should have at least one column in common with same data types

How to join heterogeneous sources and flat files?

Use joiner transformation to join heterogeneous sources and flat files

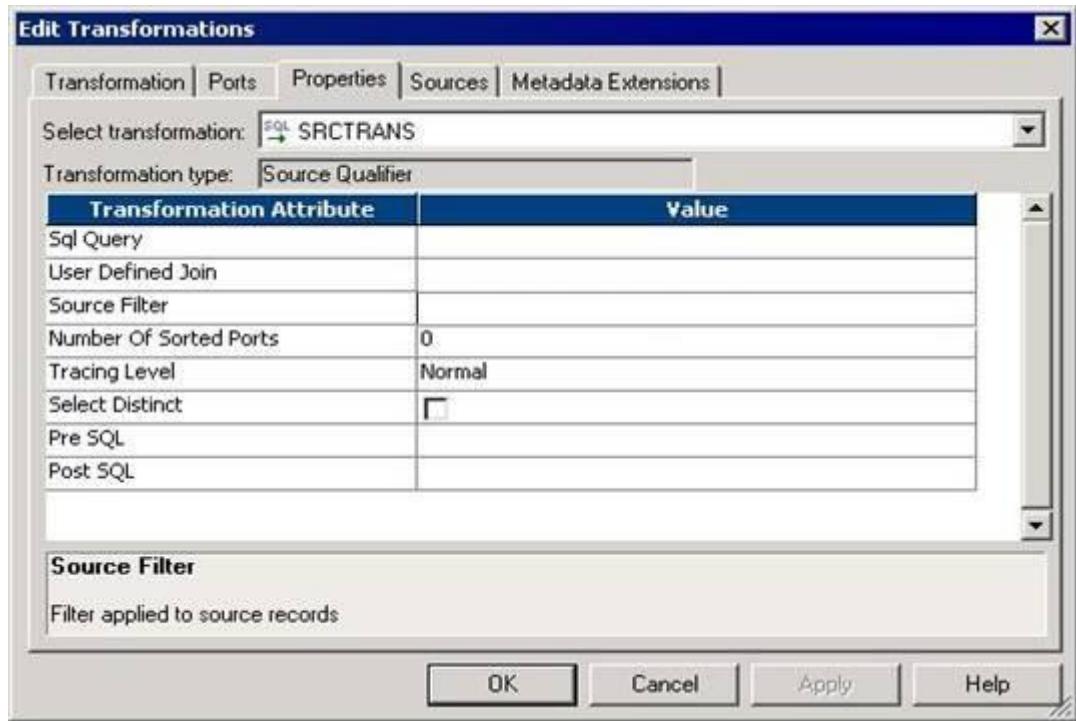
How do you configure a source qualifier transformation?

- SQL Query
- User-Defined Join
- Source Filter
- Number of Sorted Ports
- Select Distinct
- Pre-SQL

- Post-SQL

What are the properties of Source Qualifier transformation?

Source Qualifier Properties :-



1) SOURCE FILTER:

Specifies the filter condition the Integration Service applies when querying rows.

2) NUMBER OF SORTED PORTS:

Indicates the number of columns used when sorting rows queried from relational sources. If you select this option, the Integration Service adds an ORDER BY to the default query when it reads source rows. The ORDER BY includes the number of ports specified, starting from the top of the transformation.

When selected, the database sort order must match the session sort order.

By default it is 0. If we change it to 1, then the data will be sorted by column that is at the top in SQ.

Example:

```
SELECT
EMPNO,
EMP_NAME,
DEPTNO,
```

```
DEPT_NAME  
FROM EMPLOYEE
```

in above SQL, present in default query.

- If we want to sort as per EMP_NAME, then we need to change it to 2.
- If we change it to 2, then data will be sorted by top two columns. First it will be sorted by EMP_NAME then if a conflict occurs it will be sorted by EMPNO.

3) SELECT DISTINCT:

Specifies if you want to select only unique rows. The Integration Service includes a SELECT DISTINCT statement in default query if you choose this option.

- Just check the option in Properties tab to enable it.

4) PRE and POST SQL Commands :

- Pre-session SQL commands - are the commands that run against the source database before the Integration Service reads the source.
- Post-session SQL commands - are the commands that run against the source database after Integration Service writes to the target.
- Use a semi-colon (;) to separate multiple statements.

5) USER DEFINED JOINS :

Specifies the condition used to join data from multiple sources represented in the same Source Qualifier transformation

- We can specify equi join, left outer join and right outer join only. We Cannot specify full outer join. To use full outer join, we need to write SQL Query.

SQL Join syntax you can use

Join Type --> Syntax

Equi Join --> DEPT.DEPTNO=EMP.DEPTNO

Left Outer Join --> EMP LEFT OUTER JOIN DEPT ON DEPT.DEPTNO=EMP.DEPTNO

Right Outer Join --> EMP RIGHT OUTER JOIN DEPT ON DEPT.DEPTNO=EMP.DEPTNO

Informatica Join Syntax you can use

Equi Join --> { source1 INNER JOIN source2 on join condition } e.g. { CUSTOMER INNER JOIN PURCHASES on CUSTOMER.CUST_ID = PURCHASES.CUST_ID }

Left Outer Join --> { source1 LEFT OUTER JOIN source2 on join condition } e.g. { CUSTOMER LEFT OUTER JOIN PURCHASES on CUSTOMER.CUST_ID = PURCHASES.CUST_ID }

Right Outer Join --> { source1 RIGHT OUTER JOIN source2 on join condition } e.g. {

CUSTOMER RIGHT OUTER JOIN PURCHASES on CUSTOMER.CUST_ID = PURCHASES.CUST_ID }

}

Note: Brackets {} must be included while using infa syntax .

Both Syntax are valid and works in query override in Informatica.

6) SQL QUERY :

Defines a custom query that replaces the default query the Integration Service uses to read data from sources represented in this Source Qualifier transformation. A custom query overrides entries for a custom join or a source filter.

How to troubleshoot Source Qualifier Transformations?

Problem 1. I cannot perform a drag and drop operation, such as connecting ports.

Solution Review the error message on the status bar for details.

Problem 2. I cannot connect a source definition to a target definition.

Solution You cannot directly connect sources to targets. Instead, you need to connect them through a Source Qualifier transformation for relational and flat file sources, or through a Normalizer transformation for COBOL sources.

Problem 3. I cannot connect multiple sources to one target.

Solution The Designer does not allow you to connect multiple Source Qualifier transformations to a single target. There are two workarounds:

- (a) Reuse targets. Since target definitions are reusable, you can add the same target to the mapping multiple times. Then connect each Source Qualifier transformation to each target.
- (b) Join the sources in a Source Qualifier transformation. Then remove the WHERE clause from the SQL query.

What will happen if the SELECT list COLUMNS in the Custom override SQL Query and the OUTPUT PORTS order in SQ transformation do not match?

Mismatch or Changing the order of the list of selected columns to that of the connected transformation output ports may result in session failure.

Suppose we have a Source Qualifier transformation that populates two target tables. How do you ensure TGT2 is loaded after TGT1?

In the Workflow Manager, we can Configure Constraint-based load ordering for a session. The Integration Service orders the target load on a row-by-row basis. For every row generated by an active source, the Integration Service loads the corresponding transformed row first to the primary key table, then to the foreign key table.

Hence if we have one Source Qualifier transformation that provides data for multiple target tables having primary and foreign key relationships, we will go for Constraint-based load ordering.

Constraint-based loading

Revisiting Filter Transformation

What is a sequence generator transformation?

A Sequence generator transformation generates numeric values. Sequence generator transformation is a passive transformation.

What is the use of a sequence generator transformation?

A sequence generator is used to create unique primary key values, replace missing primary key values or cycle through a sequential range of numbers.

What are the ports in sequence generator transformation?

A sequence generator contains two output ports. They are CURRVAL and NEXTVAL.

What is the maximum number of sequence that a sequence generator can generate?

The maximum value is 9,223,372,036,854,775,807

When you connect both the NEXTVAL and CURRVAL ports to a target, what will be the output values of these ports?

The output values are

NEXTVAL CURRVAL

- 1 2
- 2 3
- 3 4
- 4 5
- 5 6

What will be the output value, if you connect only CURRVAL to the target without connecting NEXTVAL?

The integration service passes a constant value for each row.

What will be the value of CURRVAL in a sequence generator transformation?

CURRVAL is the sum of "NEXTVAL" and "Increment By" Value.

What is the number of cached values set to default for a sequence generator transformation?

For non-reusable sequence generators, the number of cached values is set to zero.

For reusable sequence generators, the number of cached values is set to 1000.

How do you configure a sequence generator transformation?

The following properties need to be configured for a sequence generator transformation:

- Start Value
- Increment By
- End Value
- Current Value
- Cycle
- Number of Cached Values

What is a router transformation?

A router is used to filter the rows in a mapping. Unlike filter transformation, you can specify one or more conditions in a router transformation. Router is an active transformation.

How to improve the performance of a session using router transformation?

Use router transformation in a mapping instead of creating multiple filter transformations to perform the same task. The router transformation is more efficient in this case. When you use a router transformation in a mapping, the integration service processes the incoming data only once. When you use multiple filter transformations, the integration service processes the incoming data for each transformation.

What are the different groups in router transformation?

The router transformation has the following types of groups:

- Input
- Output

How many types of output groups are there?

There are two types of output groups:

- User-defined group
- Default group

Where you specify the filter conditions in the router transformation?

You can create the group filter conditions in the groups tab using the expression editor.

Can you connect ports of two output groups from router transformation to a single

target?

No. You cannot connect more than one output group to one target or a single input group transformation.

What is rank transformation?

A rank transformation is used to select top or bottom rank of data. This means, it selects the largest or smallest numeric value in a port or group. Rank transformation also selects the strings at the top or bottom of a session sort order. Rank transformation is an active transformation.

What is rank cache?

The integration service compares input rows in the data cache, if the input row out-ranks a cached row, the integration service replaces the cached row with the input row. If you configure the rank transformation to rank across multiple groups, the integration service ranks incrementally for each group it finds. The integration service stores group information in index cache and row data in data cache.

What is Rank Index in Rank Transformation?

Rank Index is assigned by the task designer to each record. The rank index port is used to store ranking position for each row. Rank Transformation identifies each row from the top to bottom and then assigns Rank Index.

How do you specify the number of rows you want to rank in a rank transformation?

In the rank transformation properties, there is an option 'Number of Ranks' for specifying the number of rows you want to rank.

How to select either top or bottom ranking for a column?

In the rank transformation properties, there is an option 'Top/Bottom' for selecting the top or bottom ranking for a column.

How does Rank transformation handle string values?

Rank transformation can return the strings at the top or the bottom of a session sort order. When the Integration Service runs in Unicode mode, it sorts character data in the session using the selected sort order associated with the Code Page of IS which may be French, German, etc. When the Integration Service runs in ASCII mode, it ignores this setting and uses a binary sort order to sort character data

Can we specify ranking on more than one port?

No. We can specify to rank the data based on only one port. In the ports tab, you have to check the R option for designating the port as a rank port and this option can be checked only on one port.

What is normalizer transformation?

The normalizer transformation receives a row that contains multiple-occurring columns and returns a row for each instance of the multiple-occurring data. This means it converts column data in to row data. Normalizer is an active transformation.

Differences Between Normalizer And Normalizer Transformation?

Normalizer : It is a transformation mainly using for Cobol sources. It change the rows into columns and columns into rows.

Normalization : To remove the redundancy and inconsistency.

Normalizer Transformation : can be used to obtain multiple columns from a single row.

Which transformation is required to process the cobol sources?

Since the cobol sources contain denormalized data, normalizer transformation is used to normalize the cobol sources.

What Are Main Advantages And Purpose Of Using Normalizer Transformation In Informatica?

Normalizer Transformation is used mainly with COBOL sources where most of the time data is stored in de-normalized format. Also, Normalizer transformation can be used to create multiple rows from a single row of data.

1. Normalizer Transformation read the data from COBOL Sources.
2. It support Horizontal Pivot .It is a processing of single input into a multiple output

What is generated key and generated column id in a normalizer transformation?

- The integration service increments the generated key sequence number each time it process a source row. When the source row contains a multiple-occurring column or a multiple-occurring group of columns, the normalizer transformation returns a row for each occurrence. Each row contains the same generated key value.
- The normalizer transformation has a generated column ID (GCID) port for each multiple-occurring column. The GCID is an index for the instance of the multiple-occurring data. For example, if a column occurs 3 times in a source record, the normalizer returns a value of 1,2 or 3 in the generated column ID.

What is VSAM?

VSAM (Virtual Storage Access Method) is a file access method for an IBM mainframe operating system. VSAM organize records in indexed or sequential flat files.

What is VSAM normalizer transformation?

The VSAM normalizer transformation is the source qualifier transformation for a COBOL

source definition. A COBOL source is flat file that can contain multiple-occurring data and multiple types of records in the same file.

What is pipeline normalizer transformation?

Pipeline normalizer transformation processes multiple-occurring data from relational tables or flat files.

What is occurs clause and redefines clause in normalizer transformation?

- Occurs clause is specified when the source row has a multiple-occurring columns.
- A redefines clause is specified when the source has rows of multiple columns.

What is a joiner transformation?

A joiner transformation joins two heterogeneous sources. You can also join the data from the same source. The joiner transformation joins sources with at least one matching column. The joiner uses a condition that matches one or more joins of columns between the two sources.

How many joiner transformations are required to join n sources?

To join n sources n-1 joiner transformations are required.

What are Limitations on joiner transformation ?

- 1.Both pipelines begin with the same original data source.
- 2.Both input pipelines originate from the same Source Qualifier transformation.
- 3.Both input pipelines originate from the same Normalizer transformation.
- 4.Both input pipelines originate from the same Joiner transformation.
- 5.Either input pipelines contains an Update Strategy transformation.
- 6.Either input pipelines contains a Sequence Generator transformation.

What are the different types of joins?

- Normal join: In a normal join, the integration service discards all the rows from the master and detail source that do not match the join condition.
- Master outer join: A master outer join keeps all the rows of data from the detail source and the matching rows from the master source. It discards the unmatched rows from the master source.
- Detail outer join: A detail outer join keeps all the rows of data from the master source and the matching rows from the detail source. It discards the unmatched rows from the detail source.

- Full outer join: A full outer join keeps all rows of data from both the master and detail rows.

What Are The Joiner Caches?

Specifies the directory used to cache master records and the index to these records. By default, the cached files are created in a directory specified by the server variable \$PMCacheDir. If you override the directory, make sure the directory exists and contains enough disk space for the cache files. The directory can be a mapped or mounted drive. There are 2-types of cache in the joiner:

1. Data cache
2. Index Cache

How can you increase the performance in joiner transformation?

Below are the ways in which you can improve the performance of Joiner Transformation.

- Perform joins in a database when possible.
In some cases, this is not possible, such as joining tables from two different databases or flat file systems. To perform a join in a database, we can use the following options:
Create and Use a pre-session stored procedure to join the tables in a database.
Use the Source Qualifier transformation to perform the join.
- Join sorted data when possible
- For an unsorted Joiner transformation, designate the source with fewer rows as the master source.
- For a sorted Joiner transformation, designate the source with fewer duplicate key values as the master source.

In A Joiner Transformation, You Should Specify The Source With Fewer Rows As The Master Source. Why?

Joiner transformation compares each row of the master source against the detail source. The fewer unique rows in the master, the fewer iterations of the join comparison occur, which speeds the join process.

Joiner Transformation will cache Master table's data hence it is advised to define table with less #of rows as master.

Why joiner is a blocking transformation?

When the integration service processes an unsorted joiner transformation, it reads all master rows before it reads the detail rows. To ensure it reads all master rows before the detail rows, the integration service blocks all the details source while it caches rows from the master source. As it blocks the detail source, the unsorted joiner is called a blocking transformation.

Differentiate between Joiner and Lookup transformations?

Joiner	Lookup
It is not possible to override the query	It is possible to override the query
Only the '=' operator is available	All operators are available for use
Users cannot restrict the number of rows while reading relational tables	Users can restrict the number of rows while reading relational tables
It is possible to join tables with Joins	It behaves as Left Outer Join while connecting with the database

What are the settings used to configure the joiner transformation

- Master and detail source
- Type of join
- Join condition

What are the transformations that cannot be placed between the sort origin and the Joiner transformation so that we do not lose the input sort order?

The best option is to place the Joiner transformation directly after the sort origin to maintain sorted data. However, do not place any of the following transformations between the sort origin and the Joiner transformation:

Custom

UnsortedAggregator

Normalizer

Rank

Union transformation

XML Parser transformation

XML Generator transformation

Mapplet [if it contains any one of the above-mentioned transformations]

How To Join Two Tables Without Using The Joiner Transformation?

Its possible to join the two or more tables by using source qualifier. But provided the tables should have relationship.

When you drag and drop the tables you will get the source qualifier for each table. Delete all the source qualifiers. Add a common source qualifier for all. Right click on the source qualifier you will find EDIT click on it. Click on the properties tab, you will find sql query in that you can write your sql.

You can also do it using Session --- mapping---source there you have an option called User Defined Join there you can write your SQL.

How would you self-join in an Informatica mapping?

To self-join, place one transformation minimum between the source qualifier and the joiner in one branch minimum. You must pre-sort the data and then configure the joiner to accept sorted input.

What is a filter transformation?

A filter transformation is used to filter out the rows in mapping. The filter transformation allows the rows that meet the filter condition to pass through and drops the rows that do not meet the condition. Filter transformation is an active transformation.

Can we specify more than one filter condition in a filter transformation?

We can only specify one condition in the filter transformation. To specify more than one condition, we have to use router transformation?

In which case a filter transformation acts as passive transformation?

If the filter condition is set to TRUE, then it passes all the rows without filtering any data. In this case, the filter transformation acts as passive transformation.

Can we concatenate ports from more than one transformation into the filter transformation?

No. The input ports for the filter must come from a single transformation.

How to filter the null values and spaces?

Use the ISNULL and IS_SPACES functions

Example: IIF(ISNULL(commission),FALSE,TRUE)

How can we filter rows in Informatica?

There are two ways to filter rows in Informatica, they are as follows:

- **Source Qualifier Transformation:** It filters rows while reading data from a relational data source. It minimizes the number of rows while mapping to enhance performance. Also, Standard SQL is used by the filter condition for executing in the database.
- **Filter Transformation:** It filters rows within a mapped data from any source. It is added close to the source to filter out the unwanted data and maximize performance. It generates true or false values based on conditions.

How session performance can be improved by using filter transformation?

Keep the filter transformation as close as possible to the sources in the mapping. This allows the unwanted data to be discarded and the integration service processes only the required rows. If the source is relational source, use the source qualifier to filter the rows.

Difference Between Summary Filter And Details Filter?

Summary Filter - we can apply records group by that contain common values.

Detail Filter - we can apply to each and every record in a database.

What is the difference between Source qualifier and filter transformation?

Source qualifier transformation is used to represent rows that Integration service reads in a session. It is an active transformation. Using source qualifier the following tasks can be fulfilled:

1. When two tables from the same source database with primary key – foreign key transformation relationship is there then the sources can be linked to one source qualifier transformation.
2. Filtering rows when Integration service adds a where clause to the user's default query.
3. When a user wants an outer join instead of an inner join, then join information is replaced by metadata specified in SQL query.
4. When sorted ports are specified then the integration service uses order by clause to the default query.
5. If a user chooses to find a distinct value then integration service uses select distinct to the specified query.

When the data we need to filter is not a relational source then the user should use Filter transformation. It helps the user to meet the specified filter condition to let go or pass through. It will directly drop the rows that do not meet the condition and multiple conditions can be specified.

What Is The Difference Between Filter And Lookup Transformation?

1. Filter transformation is an Active transformation and Lookup is a Passive transformation.
2. Filter transformation is used to Filter rows based on condition and Lookup is used to look up data in a flat file or a relational table, view, or synonym.

What is aggregator transformation?

Aggregator transformation performs aggregate calculations like sum, average, count etc. It is an active transformation, changes the number of rows in the pipeline. Unlike expression transformation (performs calculations on a row-by-row basis), an aggregator transformation performs calculations on group of rows.

What is aggregate cache?

The integration service creates index and data cache in memory to process the aggregator transformation and stores the data group in index cache, row data in data cache. If the integration service requires more space, it stores the overflow values in cache files.

How can we improve performance of aggregate transformation?

- Use sorted input: Sort the data before passing into aggregator. The integration service uses memory to process the aggregator transformation and it does not use cache memory.
- Filter the unwanted data before aggregating.
- Limit the number of input/output or output ports to reduce the amount of data the aggregator transformation stores in the data cache.

What are the different types of aggregate functions?

Aggregate Function : - The transformation language includes the following Aggregate functions:

**AVG, COUNT, MAX, MIN, SUM
FIRST, LAST
MEDIAN, PERCENTILE, STDDEV, VARIANCE**

What are the different levels of aggregate functions?

There are three variations of aggregate function

- (i) Single Level Aggregate Function e.g. **SUM(QUANTITY)**
- (ii) Nested Aggregate Function e.g. **MAX(COUNT(ITEM))**
- (iii) Conditional Clause: - use conditional clauses in the aggregate expression to reduce the number of rows used in the aggregation. The conditional clause can be any clause that evaluates to TRUE or FALSE e.g. **SUM (COMMISSION, COMMISSION > 100)**

Do's & Don'ts While using Nested Aggregate function?

1. You can include multiple single level functions or multiple nested functions in output port of an aggregator.
2. You cannot have both types of functions together.
3. If you need to create both singe level and nested functions create separate aggregator transformation.

Why cannot you use both single level and nested aggregate functions in a single aggregate transformation?

The nested aggregate function returns only one output row, whereas the single level aggregate function returns more than one row. Since the number of rows returned are not same, you cannot use both single level and nested aggregate functions in the same transformation. If you include both the single level and nested functions in the same

aggregator, the designer marks the mapping or mapplet as invalid. So, you need to create separate aggregator transformations.

Up to how many levels, you can nest the aggregate functions?

We can nest up to two levels only.

Example: MAX(SUM(ITEM))

What are non aggregate functions?

Non-Aggregate Functions

We can also use non-aggregate functions in the aggregate expression.

IIF (MAX(QUANTITY) > 0, MAX(QUANTITY), 0))

What are Aggregator Index Cache & Data Cache?

Aggregator Index Cache & Data Cache:

- (a) The Power Center Server stores data in the aggregate cache until it completes Aggregate calculations.
- (b) It stores group values in an index cache and row data in the data cache. If the Power Center Server requires more space, it stores overflow values in cache files.

(A) Aggregator Index Cache: The index cache holds group information from the group by ports. If we are using Group By on DEPTNO, then this cache stores values 10, 20, 30 etc. All Group by Columns are in AGGREGATOR INDEX CACHE. E.g. DEPTNO

(B) Aggregator Data Cache: DATA CACHE is generally larger than the AGGREGATOR INDEX CACHE.

Data Cache contains:

- a) Variable ports if any
- b) Non group by input/output ports.
- c) Non group by input ports used in non-aggregate output expression.
- d) Port containing aggregate function

How Can I Get Distinct Values While Mapping In Informatica In Insertion?

You can add an aggregator before insert and group by the fields that need to be distinct.

How an Expression Transformation differs from Aggregator Transformation?

An Expression Transformation performs a calculation on a row-by-row basis. An Aggregator Transformation performs calculations on groups.

What is incremental aggregation?

When using incremental aggregation, we apply capture d changes in the source to aggregate calculations in a session. If the source changes only incrementally and we can capture changes, we can configure the session to process only those changes. This allows the Informatica Server to update our target incrementally, rather than forcing it to process the entire source and recalculate the same calculations each time you we the session.

Why cannot we use sorted input option for incremental aggregation?

In incremental aggregation, the aggregate calculations are stored in historical cache on the server. In this historical cache the data need not be in sorted order. If you give sorted input, the records come as presorted for that particular run but in the historical cache the data may not be in the sorted order. That is why this option is not allowed.

How the NULL values are handled in Aggregator?

You can configure the integration service to treat null values in aggregator functions as NULL or zero. By default the integration service treats null values as NULL in aggregate functions.

What are the performance considerations when working with Aggregator Transformation?

1. Filter the unnecessary data before aggregating it. Place a Filter transformation in the mapping before the Aggregator transformation to reduce unnecessary aggregation.
2. Improve performance by connecting only the necessary input/output ports to subsequent transformations, thereby reducing the size of the data cache.
3. Use Sorted input which reduces the amount of data cached and improves session performance.

What differs when we choose the Sorted Input for Aggregator Transformation?

Integration Service creates the index and data caches files in memory to process the Aggregator transformation. If the Integration Service requires more space as allocated for the index and data cache sizes in the transformation properties, it stores overflow values in cache files i.e. paging to disk. One way to increase session performance is to increase the index and data cache sizes in the transformation properties. But when we check Sorted Input the Integration Service uses memory to process an Aggregator transformation it does not use cache files.

Under what conditions selecting Sorted Input in aggregator will still not boost session performance?

1. Incremental Aggregation, session option is enabled.

2. The aggregate expression contains nested aggregate functions.
3. Source data is data-driven.

How can we access repository reports without SQL or other transformations?

We can access repository reports by using a metadata reporter. There is no need of using SQL or other transformation as it is a web app.

What is a mapplet?

A Mapplet is a reusable object that contains a set of transformations and enables to reuse that transformation logic in multiple mappings.

How to use mapplet input & output?

Mapplet Input:

Mapplet input can originate from a source definition and/or from an Input transformation in the mapplet. We can create multiple pipelines in a mapplet.

We use Mapplet Input transformation to give input to mapplet.

Use of Mapplet Input transformation is optional.

Mapplet Output:

The output of a mapplet is not connected to any target table.

We must use Mapplet Output transformation to store mapplet output.

A mapplet must contain at least one Output transformation with at least one connected port in the mapplet.

What Are The Unsupported Repository Objects For A Mapplet?

- COBOL source definition
- Joiner transformations
- Normalizer transformations
- Non reusable sequence generator transformations.
- Pre or post session stored procedures
- Target definitions
- Power mart 3.5 style Look Up functions
- XML source definitions
- IBM MQ source definitions.

How to implement Indirect Loading For Flat Files?

Suppose, you have 10 flat files of same structure. All the flat files have same number of columns and data type. Now we need to transfer all the 10 files to same target.

Names of files are say EMP1, EMP2 and so on.

Solution1:

1. Import one flat file definition and make the mapping as per need.
2. Now in session give the Source File name and Source File Directory location of one file.
3. Make workflow and run.
4. Now open session after workflow completes. Change the Filename and Directory to give information of second file. Run workflow again.
5. Do the above for all 10 files.

Solution2:

1. Import one flat file definition and make the mapping as per need.
2. Now in session give the Source Directory location of the files.
3. Now in Fieldname use \$InputFileName. This is a session parameter.
4. Now make a parameter file and give the value of \$InputFileName.
\$InputFileName=EMP1.txt
5. Run the workflow
6. Now edit parameter file and give value of second file. Run workflow again.
7. Do same for remaining files.

Solution3:

1. Import one flat file definition and make the mapping as per need.
2. Now make a notepad file that contains the location and name of each 10 flat files.

Sample:

D:\EMP1.txt

E:\EMP2.txt

E:\FILES\DW\EMP3.txt and so on

3. Now make a session and in Source file name and Source File Directory location fields, give the name and location of above created file.
4. In Source file type field, select Indirect.
5. Click Apply.
6. Validate Session
7. Make Workflow. Save it to repository and run.

How can we store previous session logs?

If you run the session in the time stamp mode then automatically session log out will not overwrite the current session log.

Go to Session Properties → Config Object → Log Options

Select the properties as follows:

Save session log by → Session Runs

Save session log for these runs → Change the number that you want to save the number of log files (Default is 0)

If you want to save all of the log files created by every run, and then select the option Save session log for these runs → Session Timestamp

You can find these properties in the session/workflow Properties.

What is a Data warehouse?

A Data warehouse is a subject-oriented, integrated, time-variant, non-volatile collection of data in support of management's decision-making process.

What is Data Mining?

Data Mining is the process of analysing data from different perspectives and summarizing it into useful information.

What is data cleaning?

Data cleaning is also known as data scrubbing.

Data cleansing is a process that ensures the set of data is correct and accurate. Data accuracy and consistency, data integration is checked during data cleaning. Data cleaning can be applied for a set of records or multiple sets of data that need to be merged.

What Is difference between Data Merging, Data Cleansing, Sampling?

Cleansing: TO identify and remove the redundancy and inconsistency.

sampling: just sample the data through send the data from source to target.

Data merging: It is a process of combining the data with similar structures in to a single output.

Data Cleansing: It is a process of identifying and rectifying the inconsistent and inaccurate data into consistent and accurate data.

Data Sampling: It is the process of sample by sending the data from source to target.

Compare Data Warehousing Top-down Approach With Bottom-up Approach?

Top down

ODS-->ETL-->Datawarehouse-->Datamart-->OLAP

Bottom up

ODS-->ETL-->Datamart-->Datawarehouse-->OLAP

What is Dimensional Modelling?

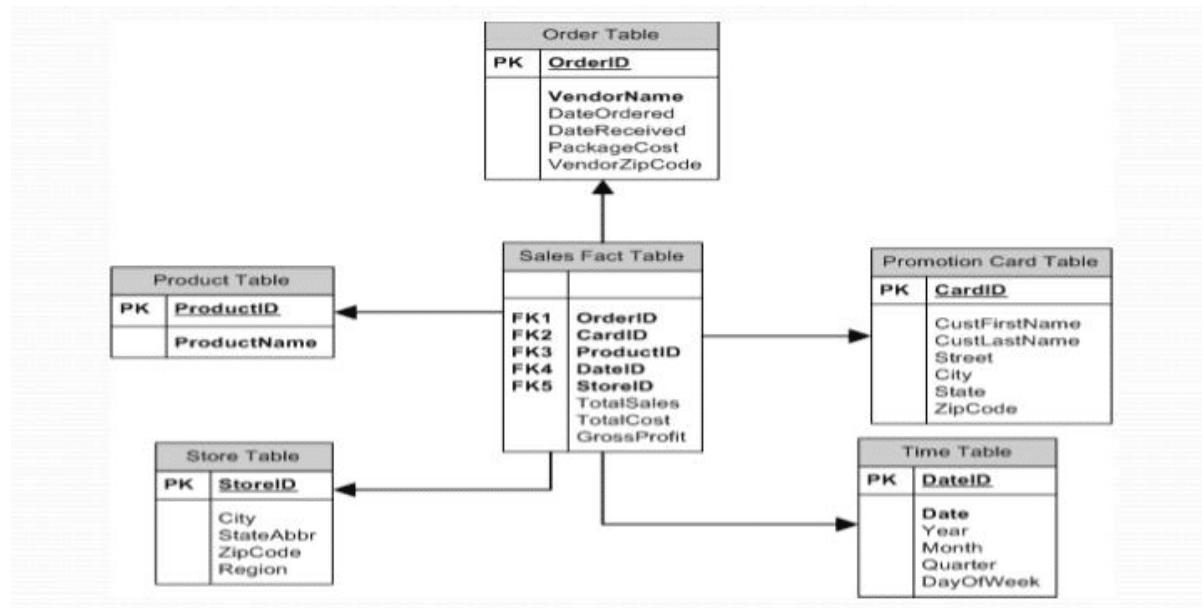
Dimensional data model concept involves two types of tables and it is different from the third normal form. This concept uses Facts table which contains the measurements of the business and Dimension table which contains the context (dimension of calculation) of the measurements.

1. Data Modelling:- It is a process of designing the database by fulfilling business requirements specifications.
2. A Data Modeler (or) Database Architect Designs the warehouse Database using a GUI based data modelling tool called “Erwin”.
3. Erwin is a deamidating tool from computer Associates (A).
4. A dimensional modelling consists of following types of schemas designed for Datawarehouse:
 - Star Schema.
 - Snowflake Schema.
 - Galaxy Schema.
5. A schema is a data model which consists of one or more tables.

Differentiate between various types of schemas in data warehousing?

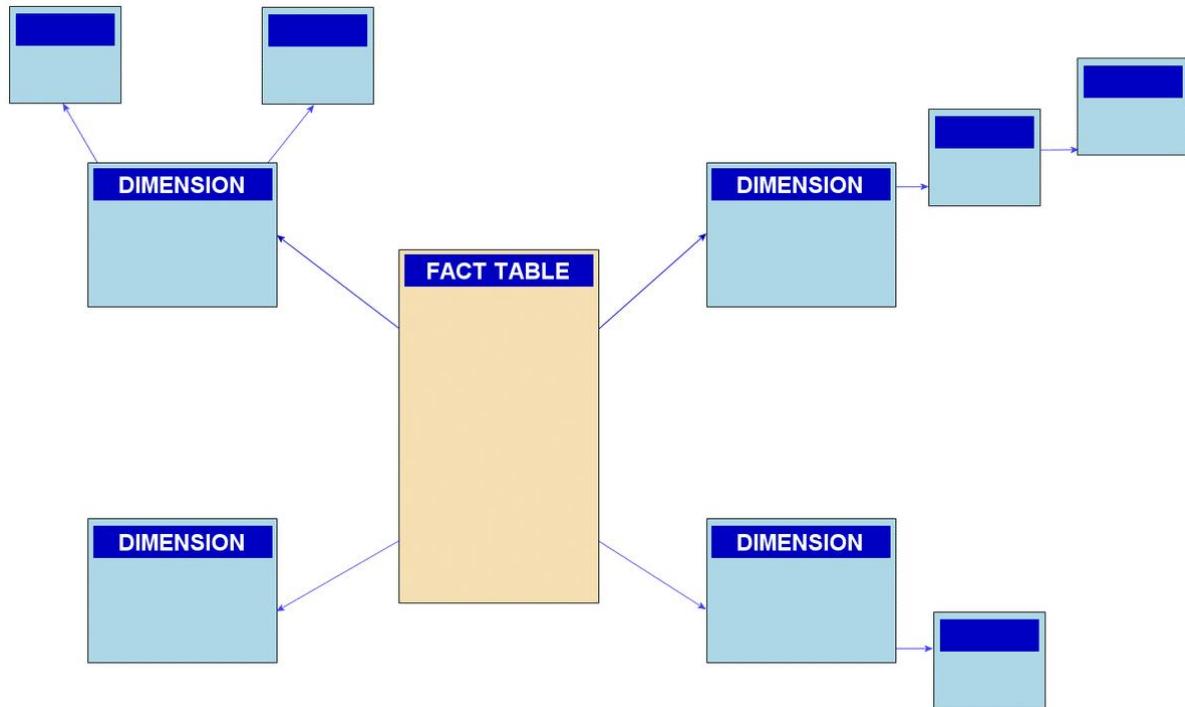
Star Schema

Star schema is the simplest style of data mart schema in computing. It is an approach which is most widely used to develop data warehouses and dimensional data marts. It features one or more fact tables referencing to numerous dimension tables.



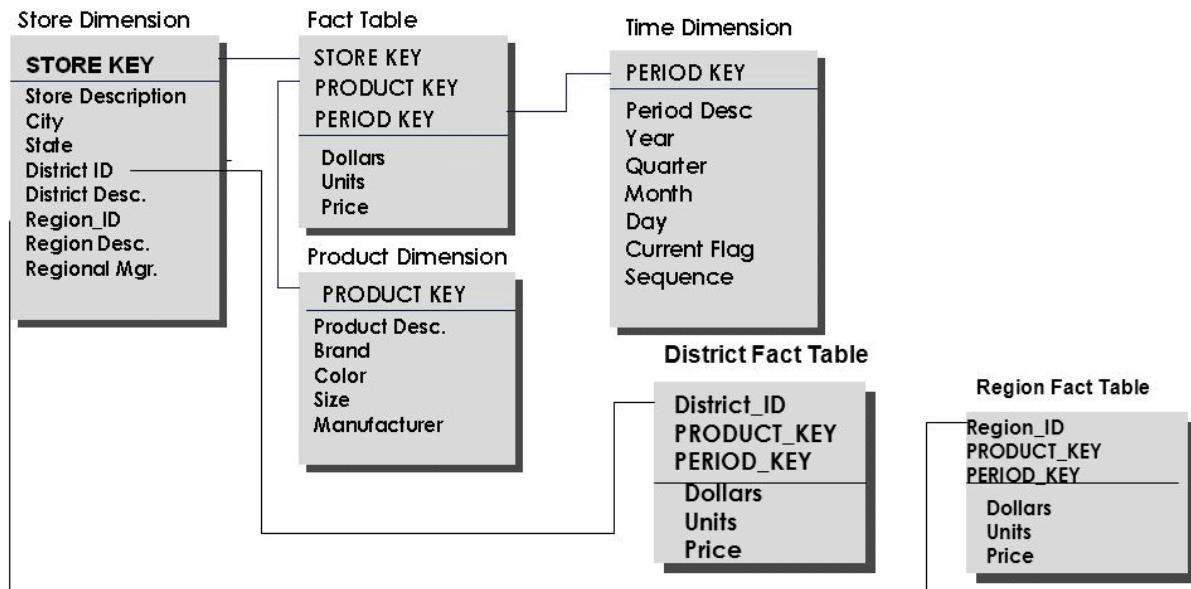
Snowflake Schema

A logical arrangement of tables in a multidimensional database, snowflake schema is represented by centralized fact tables which are connected to multidimensional tables. Dimensional tables in a star schema are normalized using snowflaking. Once normalized, the resultant structure resembles a snowflake with the fact table at the middle. Low-cardinality attributes are removed and separate tables are formed.



Fact Constellation Schema

Fact constellation schema is a measure of online analytical processing (OLAP), and OLAP happens to be a collection of multiple fact tables sharing dimension tables and viewed as a collection of stars. It can be seen as an extension of the star schema.



What are the Different methods of loading Dimension tables?

There are two different ways to load data in dimension tables.

Conventional (Slow) – All the constraints and keys are validated against the data before, it is loaded; this way data integrity is maintained.

Direct (Fast) – All the constraints and keys are disabled before the data is loaded. Once data is loaded, it is validated against all the constraints and keys. If data is found invalid or dirty it is not included in the index and all future processes are skipped on this data.

How Many Types Of Dimensions Are Available In Informatica?

The types of dimensions available are:

1. Junk dimension
2. Degenerative Dimension
3. Conformed Dimension
4. Role-playing dimension

What are Junk Dimensions?

A junk dimension is a collection of random transactional codes flags and/or text attributes that are unrelated to any particular dimension. The junk dimension is simply a structure that provides a convenient place to store the junk attributes.

Eg: Assume that we have a gender dimension and marital status dimension. In the fact table we need to maintain two keys referring to these dimensions. Instead of that create a junk dimension which has all the combinations of gender and marital status (cross join gender

and marital status table and create a junk table). Now we can maintain only one key in the fact table.

What are Degenerative Dimensions?

Degenerated Dimension:

A degenerate dimension is a dimension which is derived from the fact table and doesn't have its own dimension table.

Eg: A transactional code in a fact table..

What are conformed Dimensions?

Conformed Dimension:

Conformed dimensions mean the exact same thing with every possible fact table to which they are joined.

Eg: The date dimension table connected to the sales facts is identical to the date dimension connected to the inventory facts.

What do you understand by the term 'role-playing dimension'?

Dimensions which are often used for multiple purposes within the same database are called role-playing dimensions. For example, a date dimension can be used for "date of sale", as well as "date of delivery", or "date of hire".

Explain what staging area is and what is the purpose of a staging area?

Data staging is an area where you hold the data temporary on the data warehouse server.

Data staging includes the following steps

- Source data extraction and data transformation (restructuring)
- Data transformation (data cleansing, value transformation)
- Surrogate key assignments

Mention the types of metadata that are stored in repository?

The types of metadata, which is stored in repository, are Target definition, Source definition, Mapplet, Mappings, and Transformations.

What is a Star schema design?

A star schema is defined as a logical database design in which there will be a centrally located fact table that is surrounded by at least one or more dimension tables. This design is best suited for Data Warehouse or Data Mart.

What is the Snow Flake schema design?

In a Snow Flake design, the dimension table (de-normalized table) will be further divided into one or more dimensions (normalized tables) to organize the information in a better structural format. To design snowflake we should first design star schema design.

Discuss The Advantages & Disadvantages Of Star & Snowflake Schema?

In a STAR schema there is no relation between any two dimension tables, whereas in a SNOWFLAKE schema there is a possible relation between the dimension tables.

In star schema there is no relationship between two relational tables. All dimensions are de-normalized and query performance is degrades. In this snow flake schema dimensions are normalized. In this SF schema table space is increased. Maintenance cost is high. Query performance is increased.

How can we confirm all mappings in the repository simultaneously?

At a time, we can validate only one mapping. Hence, mapping cannot be validated simultaneously.

In Informatics' server, which files are created during the session RUMs?

The following types of files are created during the session RUMs:

- Errors log
- Bad file
- Workflow low
- Session log

What is Fact table?

Fact table contains measurements of business processes also fact table contains the foreign keys for the dimension tables. For example, if your business process is “paper production” then “average production of paper by one machine” or “weekly production of paper” would be considered as a measurement of the business process.

Is A Fact Table Normalized Or De-normalized?

A fact table is always DENORMALISED table. It consists of data from dimension table (Primary Key's) and Fact table has foreign keys and measures.

What are the types of Fact Tables?

The types of Fact Tables are:

1. **Cumulative Fact Table:** This type of fact table generally describes what was happened over the period. They contain additive facts.
2. **Snapshot Fact Table:** This type of fact table deals with a particular period. They contain non-additive and semi-additive facts.

How Many Types Of Facts Are There And What Are They?

There are three types of facts

1. **Additive fact:** a fact which can be summarized by any one of dimension or all dimensions
EX: QTY, REVENUE
2. **Semi additive fact:** a fact which can be summarized for few dimensions not for all dimensions. ex: current balance
3. **Non additive fact:** a fact which cannot be summarized by any of dimensions. ex: percentage of profit

What Is The Procedure To Load The Fact Table. Give In Detail?

Based on the requirement to your fact table, choose the sources and data and transform it based on your business needs. For the fact table, you need a primary key so use a sequence generator transformation to generate a unique key and pipe it to the target (fact) table with the foreign keys from the source tables.

What Are The Measure Objects?

Aggregate calculation like sum, avg, max, min these are the measure objects.

What is the Factless Fact table?

The Fact Table which does not contain facts is called Fact Table. Generally when we need to combine two data marts, then one data mart will have a factless fact table and another one with the common fact table.

Explain what factless fact schema is and what is Measures?

A fact table without measures is known as a Factless fact table. It can view the number of occurring events. For example, it is used to record an event such as employee count in a company.

The numeric data based on columns in a fact table is known as Measures.

What is Denormalization?

Denormalization means a table with a multi duplicate key. The dimension table follows the Denormalization method with the technique of surrogate key.

Name the other tools used for scheduling purpose other than Workflow Manager and pmcmd?

'CONTROL M' is a third-party tool used for scheduling purpose.

What are snapshots? What are materialized views & where do we use them? What is a materialized view?

A materialized view is a view in which data is also stored in some temp tables. i.e if we will go with the View concept in DB in that we only store query and once we call View it extract data from DB. But In materialized View data is stored in some temp tables.

What is the difference between View and Materialized View?

Based upon on our understanding of View and Materialized View, Let's see, some short difference between them :

- 1) First difference between View and materialized view is that, In Views query result is not stored in the disk or database but Materialized view allow to store query result in disk or table.
- 2) Another difference between View vs materialized view is that, when we create view using any table, RowId of view is same as original table but in case of Materialized view RowId is different.
- 3) One more difference between View and materialized view in database is that, In case of View we always get latest data but in case of Materialized view we need to refresh the view for getting latest data.
- 4) Performance of View is less than Materialized view.

How You Will Create Header And Footer In Target Using Informatica?

If you are focus is about the flat files then one can set it in file properties while creating a mapping or at the session level in session properties.

How many fact tables we can use in one project?

we can use 'N' number of fact tables in a Project. but, it depends as per our requirements or client specifications so, be sure that what r the client specifications its all as per my knowledge.

n number of fact tables we can use,

What is meant by centralised data warehousing?

The data warehousing where all the operations activities maintained, so that you can access the analysed data from a different place.

How can we schedule our jobs in third party tool - AutoSys?

We need to write shell script to run the job and write JIL on AutoSys to schedule the job.

What is load frequency and load window in etl?

ETL Load frequency is how often the ETL processes (that load data into the data warehouse) are run.

Load Window is the time allocated for ETL load to execute.

e.g. If the load is run daily, and the time allowed for execution is 8 hours then:

ETL load frequency = daily

Load window = 8 hours

Flat file is having 10 records as input and I want to push 5 records to the target?

one variable with the name \$\$count at mapping variable with a Count aggregation type and initial value for that \$\$count=0. Create one expression and use

SETCOUNTVARIABLEv_count=SetCountVariable (\$\$Count) Create one filter after expression and put condition v_count <=5

What are the different ways to filter rows using Informatica transformations?

- Source Qualifier
- Joiner
- Filter
- Router

What is update strategy transformation?

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

Describe two levels in which update strategy transformation sets?

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.

What is the use of control break statements?

They execute a set of codes within the loop and end loop.

Which tool you use to create and manage sessions and batches and to monitor and stop the Informatica server?

Informatica server manager.

Why we use partitioning the session in Informatica?

Partitioning achieves the session performance by reducing the time period of reading the source and loading the data into target.

To achieve the session partition what are the necessary tasks you have to do?

Configure the session to partition source data.

Install the Informatica server on a machine with multiple CPU's.

What is Data driven?

The Informatica server follows instructions coded into update strategy transformations within the session mapping which determine how to flag records for insert, update, delete or reject. If we do not choose data driven option setting, the Informatica server ignores all update strategy transformations in the mapping.

What is polling

It displays the updated information about the session in the monitor window. The monitor window displays the status of each session when you poll the Informatica server.

What are the types of mapping wizards that are provided in Informatica?

The designer provide two mapping wizard.

Getting Started Wizard creates mapping to load static facts and dimension tables as well as slowly growing dimension tables.

Slowly Changing Dimensions Wizard, creates mappings to load slowly changing dimension tables based on the amount of historical dimension data we want to keep and the method we choose to handle historical dimension data.

Can you generate reports in Informatica?

Yes. By using Metadata reporter we can generate reports in Informatica.

What is metadata reporter?

It is a web based application that enables you to run reports against repository metadata.

With a Meta data reporter you can access information about your repository without having knowledge of sql, transformation language or underlying tables in the repository.

What does a Mapping document contain?

The Mapping document contains the following information:

Source Definition – from where the database has to be loaded

Target Definition – to where the database has to be loaded

Business Logic – what logic has to be implemented in the staging area

I have two different source structure tables, but I want to load into single target table?

How do I go about it? Explain in detail through mapping flow?

- We can use joiner, if we want to join the data sources. Use a joiner and use the matching column to join the tables.
- We can also use a Union transformation, if the tables have some common columns and we need to join the data vertically. Create one union transformation add the matching ports from the two sources, to two different input groups and send the output group to the target.

The basic idea here is to use, either Joiner or Union transformation, to move the data from two sources to a single target. Based on the requirement, we may decide, which one should be used.

What are the types of OLAP?

DOLAP: The OLAP tool in which words with desktop databases are called DOLAP. Example: Cognos EP 7 Series and Business Objects, Micro strategy.

ROLAP: The OLAP which works with Relational databases are called ROLAP. Example: Business Object, Micro strategy, Cognos ReportNet, and BRIO.

MOLAP: The OLAP which is responsible for creating multi-dimensional structures called cubes are called MOLAP. Example: Cognos ReportNet.

HOLAP: The OLAP which uses the combined features of ROLAP and MOLAP are called HOLAP. Example Cognos ReportNet.

What are the differences between OLTP and OLAP?

OLTP	OLAP
Application Oriented	Subject Oriented
Used to run business	Used to Analyze business
Detailed data	Summarized and Refined
Current up-to-date	Snapshot data
Isolated data	Integrated Data
Repetitive access	Ad-hoc access
Clerical User	Knowledge User (Manager)
Performance Sensitive	Performance Relaxed
Few records accessed at a time (Tens)	Large volumes accessed at a time (Millions)
Read / Update Access	Mostly Read (Batch Update)
No data redundancy	Redundancy present
DB size : 100 MB – 100 GB	DB Size : 100 GB – Few TBs
Thousands of Users	Hundreds of Users

What is SUBSTR in INFORMATICA?

SUBSTR is a function that extracts or removes a set of characters from a larger character set.

Syntax: SUBSTR(string, start [,length])

Where,

string defines the character that we want to search.

start is an integer that is used to set the position where the counting should get started.

Length is an optional parameter that is used to count the length of a string to return from its starting position.

For Example, SUBSTR(Contact, 5, 8), where we start at the 5th character of our contact and returns to the next 8 characters.

What are the different transformations where you can use a SQL override?

- Source Qualifier
- Lookup
- Target

From where can we find the throughput option in Informatica?

In Workflow Monitor, we can find the throughput option. By right-clicking on session, then pressing on get run properties, and, under source/target statistics, we can find this option.

Can two flat files be joined with Joiner Transformation?

Yes, joiner transformation can be used to join data from two flat file sources.

Can a lookup be done on Flat Files?

Yes.

Which transformation should you need while using the cobol sources as source definitions?

Normalizer transformation which is used to normalize the data. Since Cobol sources are often consists of De normalized data.

What are the types of data that passes between informatica server and stored procedure?

3 types of data

- Input/Out put parameters
- Return Values
- Status code.

What is the status code?

Status code provides error handling for the informatica server during the session. The stored procedure issues a status code that notifies whether or not stored procedure completed successfully. This value can not seen by the user. It only used by the informatica server to determine whether to continue running the session or stop.

What is Target Update override?

It overrides the default update statement in the target properties.

How are Informatica object files stored?

They are stored as XML in the Informatica Repository.

If there is no PK or FK in the target table, how do we update or insert value into the table?

We take a dynamic lookup on the target and do a comparison with source in an expression and flag it.

How many ways you can update a relational source definition and what are they?

Two ways 1. Edit the definition 2. Reimport the definition

What are partition points?

Partition points mark the thread boundaries in a source pipeline and divide the pipeline into stages

How can you recognize whether or not the newly added rows in the source are get inserted in the target?

In the Type2 mapping we have three options to recognize the newly added rows 1. Version

- number
- 2. Flag value
- 3. Effective date Range

What are two types of processes that Informatica runs the session?

Load manager Process: Starts the session, creates the DTM process, and sends post-session email when the session completes. The DTM process. Creates threads to initialize the session, read, write, and transform data, and handle pre and post-session operations.

What is parameter file?

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

You can define the following values in parameter file:-

- Mapping parameters
- Mapping variables
- session parameters.

Differences between Normalizer and Normalizer transformation.

Normalizer: It is a transformation mainly using for Cobol sources, it's change the rows into columns and columns into rows

Normalization: To remove the redundancy and inconsistency

What is update strategy transformation?

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

How can You create or import flat file definition in to the warehouse designer?

You can not create or import flat file definition in to warehouse designer directly. Instead You must analyse the file in source analyser, then drag it into the warehouse designer.

What are the different options used to configure the sequential batches?

Two options

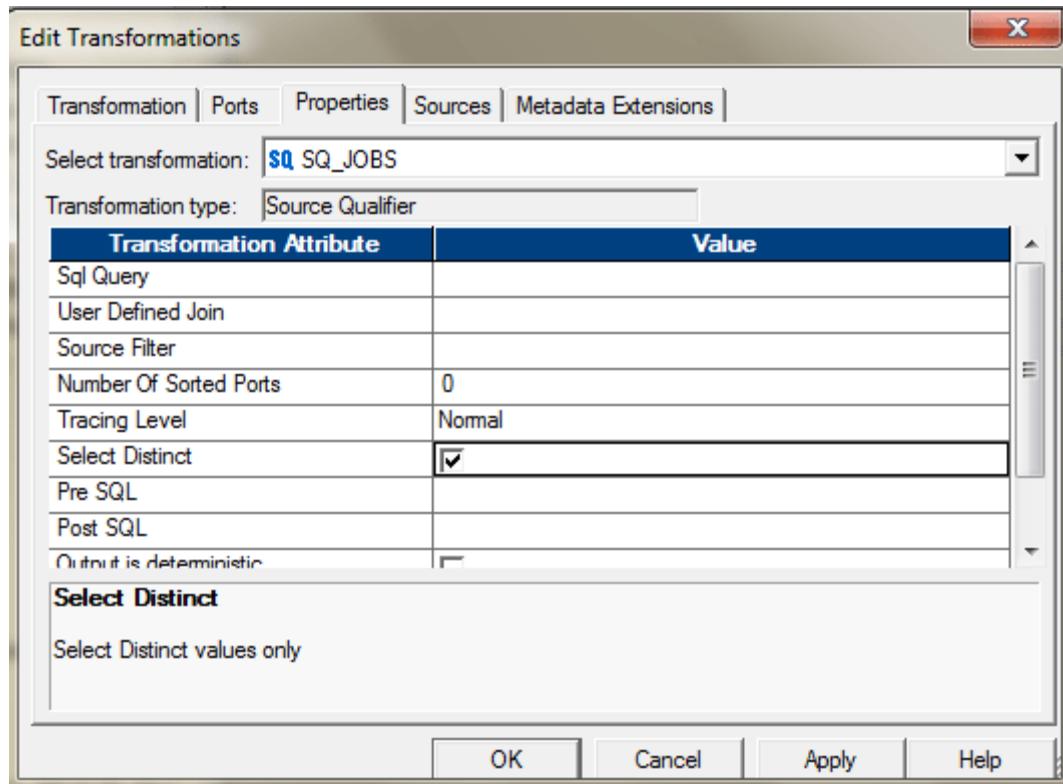
Run the session only if previous session completes successfully.

Always runs the session.

How do you remove Duplicate records in Informatica? And how many ways are there to do it?

There are several ways to remove duplicates.

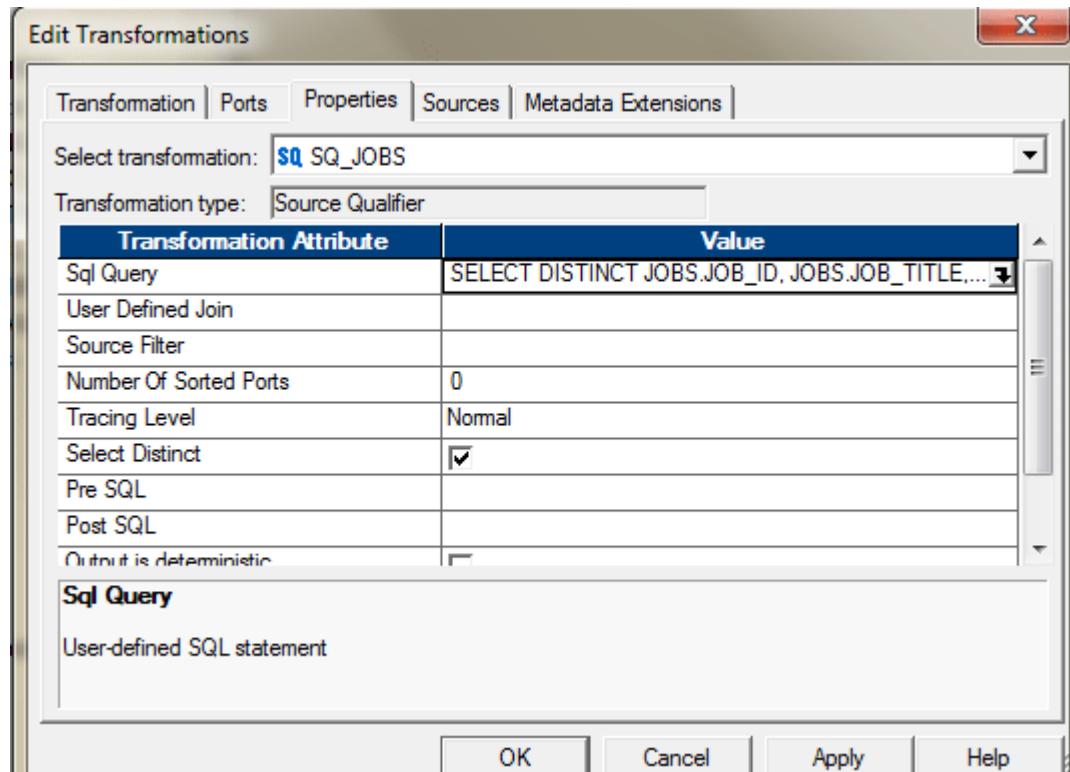
- i. If the source is DBMS, you can use the property in Source Qualifier to select the distinct



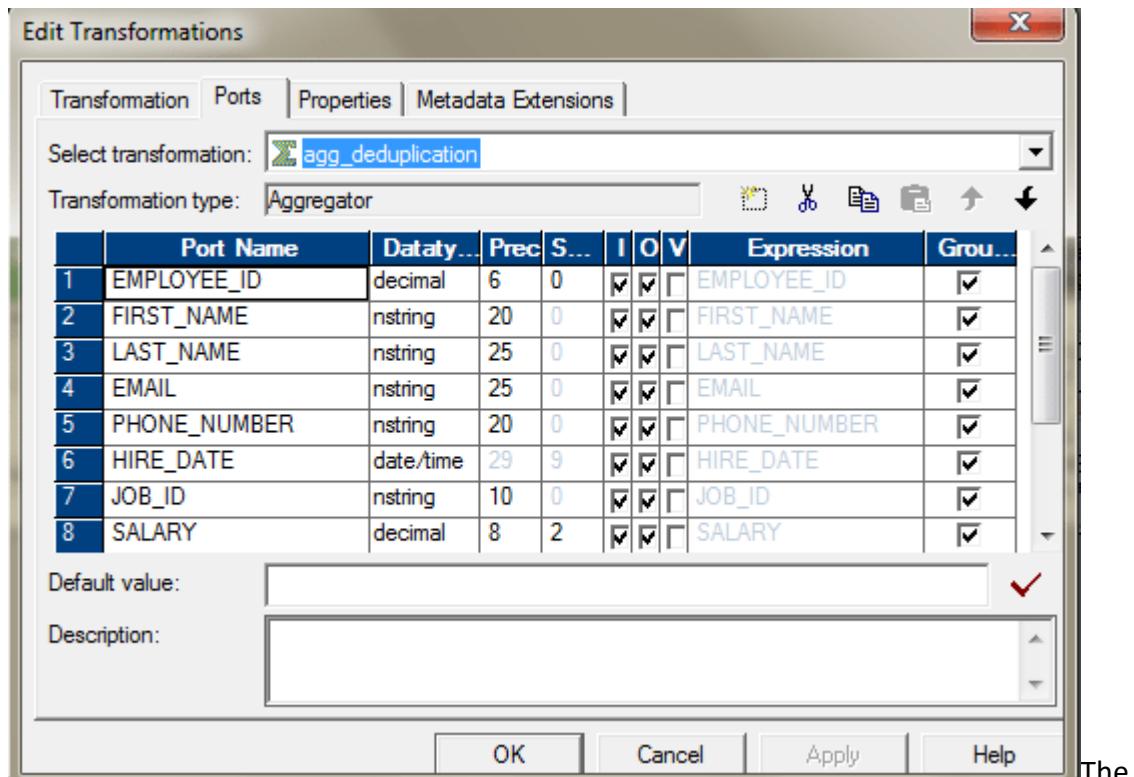
records.

Or

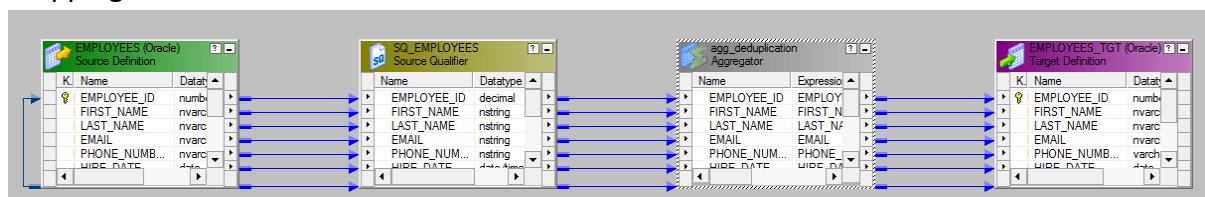
you can also use the SQL Override to perform the same.



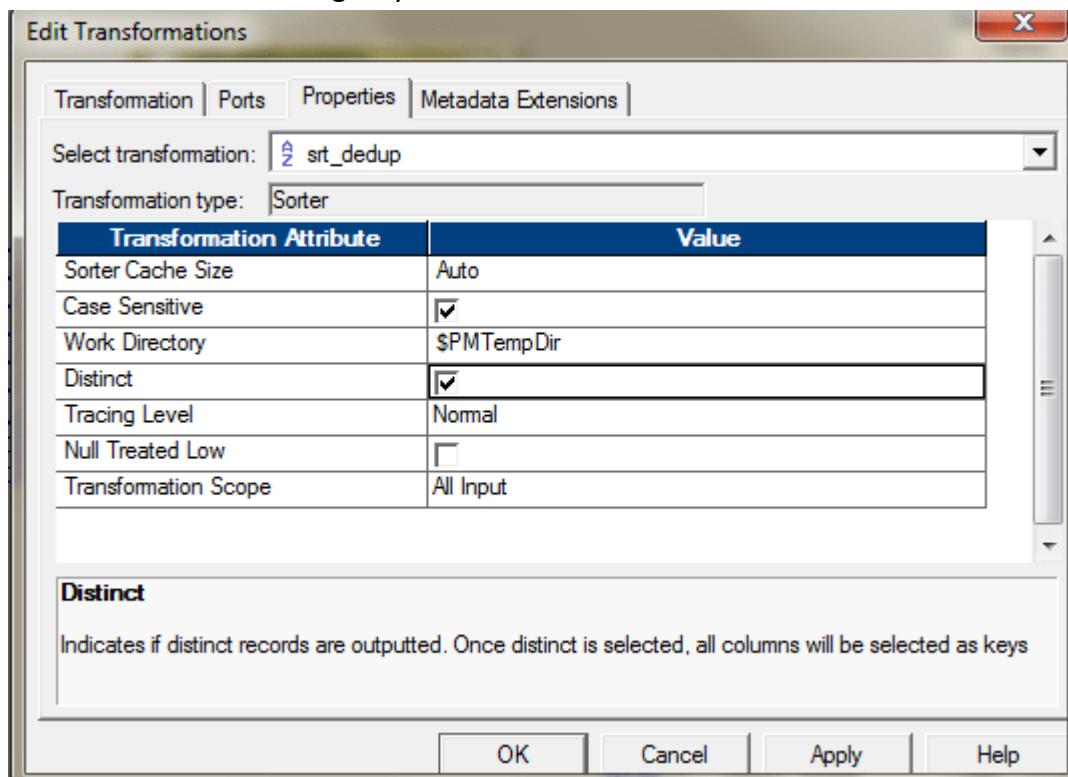
- ii. You can use, Aggregator and select all the ports as key to get the distinct values. After you pass all the required ports to the Aggregator, select all those ports , those you need to select for de-duplication. If you want to find the duplicates based on the entire columns, select all the ports as group by key.



Mapping will look like this.

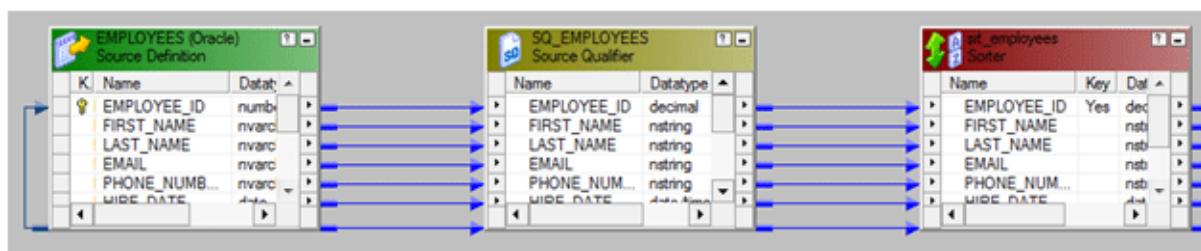


- iii. You can use Sorter and use the Sort Distinct Property to get the distinct values. Configure the sorter in the following way to enable this.



- iv. You can use, Expression and Filter transformation, to identify and remove duplicate if your data is sorted. If your data is not sorted, then, you may first use a sorter to sort the data and then apply this logic:

- Bring the source into the Mapping designer.
- Let's assume the data is not sorted. We are using a sorter to sort the data. The Key for sorting would be Employee_ID.



Configure the Sorter as mentioned below.

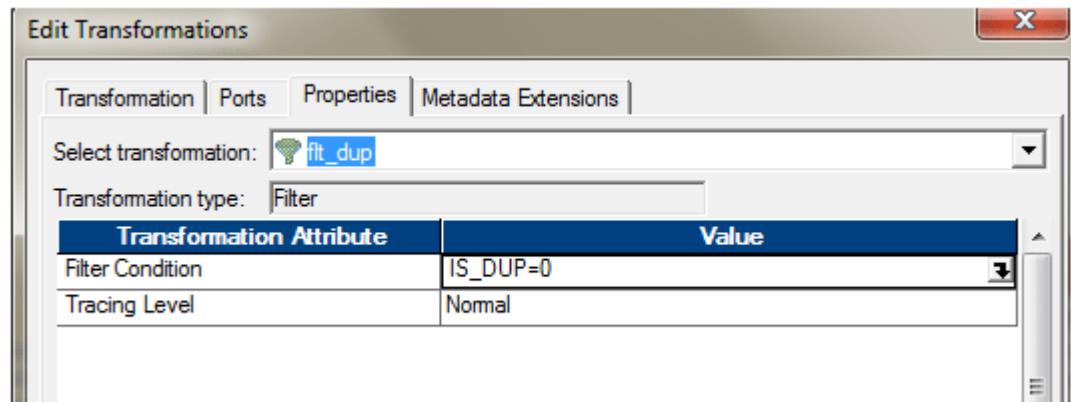
	Port Name	Datatype	Prec	Scale	I	O	K	Direction
1	EMPLOYEE_ID	decimal	6	0	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	Ascending
2	FIRST_NAME	nstring	20	0	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	- N/A -
3	LAST_NAME	nstring	25	0	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	- N/A -
4	EMAIL	nstring	25	0	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	- N/A -
5	PHONE_NUMBER	nstring	20	0	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	- N/A -
6	HIRE_DATE	date/time	29	9	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	- N/A -
7	JOB_ID	nstring	10	0	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	- N/A -
8	SALARY	decimal	8	2	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	- N/A -

- Use one expression transformation to flag the duplicates. We will use the variable ports to identify the duplicate entries, based on Employee_ID.

	Port Name	Datat...	Prec	Scale	I	O	V	Expression
1	EMPLOYEE_ID	decimal	6	0	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	EMPLOYEE_ID
2	FIRST_NAME	nstring	20	0	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	FIRST_NAME
3	LAST_NAME	nstring	25	0	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	LAST_NAME
4	EMAIL	nstring	25	0	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	EMAIL
5	PHONE_NUMBER	nstring	20	0	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	PHONE_NUMBER
6	HIRE_DATE	date/time	29	9	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	HIRE_DATE
7	JOB_ID	nstring	10	0	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	JOB_ID
8	SALARY	decimal	8	2	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	SALARY
9	v_IS_DUP	integer	10	0	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	IIF(v_PREV_EMPLOYEE_ID=EMPLOYEE_ID,1,0)
10	IS_DUP	integer	10	0	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	v_IS_DUP
11	v_PREV_EMPLOYEE_ID	integer	10	0	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	EMPLOYEE_ID

- Use a filter transformation, only to pass IS_DUP = 0. As from the previous expression transformation, we will have IS_DUP = 0 attached to only records, which are unique. If

`IS_DUP > 0`, that means, those are duplicate entries.



- Add the ports to the target. The entire mapping should look like this.



v. When you change the property of the Lookup transformation to use the Dynamic Cache, a new port is added to the transformation. `NewLookupRow`.

The Dynamic Cache can update the cache, as and when it is reading the data.

If the source has duplicate records, you can also use Dynamic Lookup cache and then router to select only the distinct one.

How To Get The First 100 Rows From The Flat File Into The Target?

create one procedure and declare the sequence inside the procedure, finally call the procedure in informatica with the help of stored procedure transformation.

What is Complex Mapping?

Complex Mapping is a mapping with huge requirements based on many dependencies. It doesn't necessarily need to have hundreds of 100 transformations, it can be a complex map even with 5 odd transformations. In case the requirement has many business restrictions and constraints, it is complex mapping.

How can you stop a batch?

By using server manager or pmcmd.

Can you use the mapping parameters or variables created in one mapping into any other reusable transformation?

Yes. Because reusable transformation is not contained with any mapplet or mapping .

How can you recover the session in sequential batches?

If you configure a session in a sequential batch to stop on failure, you can run recovery starting with the failed session.

How to recover sessions in concurrent batches?

If multiple sessions in a concurrent batch fail, you might want to truncate all targets and run the batch again. However, if a session in a concurrent batch fails and the rest of the sessions complete

Can you start a session inside a batch individually?

We can start our required session only in case of sequential batch.

When the Informatica server marks that a batch is failed?

If one of session is configured to "run if previous completes" and that previous session fails.

To achieve the session partition what are the necessary tasks you have to do?

Configure the session to partition source data. Install the Informatica server on a machine with multiple CPU's.

How To Load Time Dimension?

We can use SCD Type 1/2/3 to load any Dimensions based on the requirement. We can also use procedure to populate Time Dimension

What is the difference between Static and Dynamic Lookup Cache?

We can configure a Lookup transformation to cache the corresponding lookup table. In case of static or read-only lookup cache the Integration Service caches the lookup table at the beginning of the session and does not update the lookup cache while it processes the Lookup transformation.

In case of dynamic lookup cache the Integration Service dynamically inserts or updates data in the lookup cache and passes the data to the target. The dynamic cache is synchronized with the target.

What is Persistent Lookup Cache?

Lookups are cached by default in Informatica. Lookup cache can be either non-persistent or persistent. The Integration Service saves or deletes lookup cache files after a successful session run based on whether the Lookup cache is checked as persistent or not.

What are the features of Complex Mapping?

The features of Complex Mapping are –

- Complicated and huge requirements
- Complex business logic
- Multiple transformations

What are the transformations that are not supported in Maplet?

Normalizer, Cobol sources, XML sources, XML Source Qualifier transformations, Target definitions, Pre- and post- session Stored Procedures, Other Mapplets.

What are the ERROR tables present in Informatica?

- **PMERR_DATA**- Stores data and metadata about a transformation row error and its corresponding source row.
- **PMERR_MSG**- Stores metadata about an error and the error message.
- **PMERR_SESS**- Stores metadata about the session.
- **PMERR_TRANS**- Stores metadata about the source and transformation ports, such as name and datatype, when a transformation error occurs.

What is the difference between STOP and ABORT?

When we issue the STOP command on the executing session task, the Integration Service stops reading data from source. It continues processing, writing and committing the data to targets. If the Integration Service cannot finish processing and committing data, we can issue the abort command.

In contrast ABORT command has a timeout period of 60 seconds. If the Integration Service cannot finish processing and committing data within the timeout period, it kills the DTM process and terminates the session.

Can we copy a session to new folder or new repository?

Yes we can copy session to new folder or repository provided the corresponding Mapping is already in there.

What type of join does Lookup support?

Lookup is just similar like SQL LEFT OUTER JOIN.

What are the benefits of data warehousing?

- ❑ Historical information for comparative and competitive analysis.

- ❑ Enhanced data quality and completeness.
- ❑ Supplementing disaster recovery plans with another data back up source.

What are the types of data warehouse?

There are mainly three type of Data Warehouse are :

- ❑ Enterprise Data Warehouse
- ❑ Operational data store
- ❑ Data Mart

Suppose we have 100 records coming from the source. Now for a target column population we used a Sequence generator.

Suppose the Current Value is 0 and End Value of Sequence generator is set to 80. What will happen?

End Value is the maximum value the Sequence Generator will generate. After it reaches the End value the session fails with the following error message:

TT_11009 Sequence Generator Transformation: Overflow error.

Failing of session can be handled if the Sequence Generator is configured to **Cycle** through the sequence, i.e. whenever the Integration Service reaches the configured end value for the sequence, it wraps around and starts the cycle again, beginning with the configured Start Value.

What are the changes we observe when we promote a non reusable Sequence Generator to a reusable one? And what happens if we set the Number of Cached Values to 0 for a reusable transformation?

When we convert a non reusable sequence generator to reusable one we observe that the **Number of Cached Values** is set to 1000 by default; And the **Reset** property is disabled.

When we try to set the **Number of Cached Values** property of a Reusable Sequence Generator to 0 in the Transformation Developer we encounter the following error message:
The number of cached values must be greater than zero for reusable sequence transformation.

What Is Hash Table Informatica?

In hash partitioning, the Informatica Server uses a hash function to group rows of data among partitions. The Informatica Server groups the data based on a partition key. Use hash partitioning when you want the Informatica Server to distribute rows to the partitions by

group. For example, you need to sort items by item ID, but you do not know how many items have a particular ID number.

What are the applications of data warehouse?

- ❑ Datawarehouse are used extensively in banking and financial services, consumer goods.
- ❑ Datawarehouse is mainly used for generating reports and answering predefined queries.
- ❑ Datawarehouse is used for strategic purposes, performing multidimensional analysis.
- ❑ Datawarehouse is used for knowledge discovery and strategic decision making using data mining tools.

What is meant by grain?

Granularity of data is a very important factor in the design of fact table. In fact table, we can represent data at different atomic levels, called grains.

During the running session, what all output files are created by the Informatica server?

Mentioned below are the few output files:

- **Cache files:** These files are created at the time of memory cache creation. For circumstances like Lookup transformation, Aggregator transformation, etc index and data cache files are created by the Informatica server.
- **Session detail file:** As the name defines, this file contains load statistics like table name, rows rejected or written for each target in mapping and can be viewed in the monitor window.
- **Performance detail file:** This file is a part of the session property sheet and contains session performance information in order to determine improvement areas.
- **INFORMATICA server log:** The server creates a log for all status and error messages and can be seen in the home directory.
- **Session log file:** For each session, the server creates a session log file depending on the set tracing level. The information that can be seen in log files about sessions can be:
 - Session initialization process,
 - SQL commands creation for reader and writer threads,
 - List of errors encountered and
 - Load summary
- **Post-session email:** This helps in communicating the information about the session (session completed/session failed) to the desired recipients automatically.

- **Reject file:** This file contains information about the data that has not been used/written to targets.
- **Control file:** In case, when the session uses the external loader, the control file consists of loading instructions and data format about the target file.
- **Indicator file:** This file basically contains a number that highlights the rows marked for INSERT/UPDATE/DELETE or REJECT.
- **Output file:** The output file is created based on the file properties.

Can you copy the session to a different folder or repository?

Yes. By using copy session wizard, you can copy a session in a different folder or repository.

What is a command that used to run a batch?

pmcmd is used to start a batch.

What are the types of mapping in Getting Started Wizard?

Simple Pass through mappings

Slowly growing target mapping

What are the different types of Type2 slowly changing dimensions?

There are three types of slowly changing dimensions

SCD with versioning

SCD with flags

SCD with Date

What are the active and passive transformations?

An active transformation can change the number of rows that pass through it.

A passive transformation does not change the number of rows that pass through it.

What are the scheduling options to run a session?

Different options of scheduling are

Run only on demand: Informatica server runs the session only when user starts session explicitly

Run once: Informatica server runs the session only once at a specified date and time.

Run every: Informatica server runs the session at regular intervals as you configured.

Customized repeat: Informatica server runs the session at the date and time specified in the repeat dialog box.

How can we store previous session logs?

Just run the session in time stamp mode then automatically session log will not overwrite current session log.

What is the difference between Mapping and Mapplet?

Mapping is the collection of Source Definition, Target Definition, Transformation(s) and/or Mapplet.

Mapplet is a set of reusable transformations for applying same business logic.

What is the use of Repository Manager?

Repository Manager is used to manage folders, view Source/Target dependencies, Mapping Dependencies, adding and removing repositories, Versioning etc.

What are the various Tasks in Informatica Workflow Manager?

The various Tasks in Informatica are

Assignment Task

Command Task

Control Task

Decision Task

E-mail Task

Event Raise Task

Event Wait Task

Session Task

Timer Task

Link Task

Can we use Reusable Sequence Generator transformation in Mapplet?

No

Name the scenario in which Informatica server reject files?

When the server faces a rejection of the update strategy transformation, it rejects files. The database consisting of the information and data also gets disrupted. This is a rare case scenario.

What Is A Node In Informatica?

A node in Informatica is a logical representation of a machine.

What are the configurable commit types?

There are three configurable commit types: target-based, source-based and user-defined.

How To Go To The Older Version For A Mapping?

There is no direct approach to revert back a mapping to the previous version however it can be achieved by the procedure defined below.

Right-click on the mapping->version-View history

All the versions will be listed out. Select the version you want to revert back to and export or save its definition in local drive. Rename the last version with a new name and then import the file.

What Is The Difference Between Normal Load And Bulk Load?

Normal Load : Normal load will write information to the database log file so that if any recovery is needed it is will be helpful. when the source file is a text file and loading data to a table, in such cases we should you normal load only, else the session will be failed.

Bulk Mode : Bulk load will not write information to the database log file so that if any recovery is needed we can't do any thing in such cases. comparatively Bulk load is pretty faster than normal load.

Discuss Which Is Better Among Incremental Load, Normal Load And Bulk Load?

It depends on the requirement. Otherwise Incremental load which can be better as it takes one that data which is not available previously on the target.

According to performance bulk is better than normal. But both having some conditions in source data.

Conditions are like:

1. Does not contain any constraint in data.
2. Don't use the double datatype if necessary to use then use it as last row of the table.
3. It does not support the CHECK CONSTRAINT.

Can We Create Multiple Integration Service On Single Repository?

Yes it is possible to create multiple integration service on a single repository. This is used when you have to choose different code page for different integration service in the same repository

Eg: Repository: ASIA_PACIFIC_Repo is used to load the data for countries specific to ASIA PACIFIC to target warehouse. Different Integration service created for non-English countries to support different code page (Malaysia, China, Japan, Thailand)

What is a Sequence Generator Transformation?

A **Sequence Generator** transformation is a **Passive** and **Connected** transformation that generates numeric values. It is used to create unique primary key values, replace missing primary keys, or cycle through a sequential range of numbers. This transformation

by default contains **ONLY Two OUTPUT** ports namely **CURRVAL** and **NEXTVAL**. We cannot edit or delete these ports neither we cannot add ports to this unique transformation. We can create approximately two billion unique numeric values with the widest range from 1 to 2147483647.

Define the Properties available in Sequence Generator transformation in brief?

Sequence Generator Properties	Description
Start Value	Start value of the generated sequence that we want the Integration Service to use if we use the Cycle option. If we select Cycle, the Integration Service cycles back to this value when it reaches the end value. Default is 0.
Increment By	Difference between two consecutive values from the NEXTVAL port. Default is 1.
End Value	Maximum value generated by SeqGen. After reaching this value the session will fail if the sequence generator is not configured to cycle. Default is 2147483647.
Current Value	Current value of the sequence. Enter the value we want the Integration Service to use as the first value in the sequence. Default is 1.
Cycle	If selected, when the Integration Service reaches the configured end value for the sequence, it wraps around and starts the cycle again, beginning with the configured Start Value.

Number of Cached Values	Number of sequential values the Integration Service caches at a time. Default value for a standard Sequence Generator is 0. Default value for a reusable Sequence Generator is 1,000.
Reset	Restarts the sequence at the current value each time a session runs. This option is disabled for reusable Sequence Generator transformations.

What Is The Difference Between Informatica Powercenter Server, Repository server And Repository?

Repository is a database in which all informatica components are stored in the form of tables. The repository server controls the repository and maintains the data integrity and Consistency across the repository when multiple users use Informatica. Powercenter Server/Infa Server is responsible for execution of the components (sessions) stored in the repository.

What is the use of Shared Object?

It is easier to explain this using an example.

Suppose there is an object which is used by many users in the company.

If the object undergoes a change, this change has to be updated to each and every user. Instead, if the object is made as shared, then the update has to be done to the object and all other users get the update.

What is Dual table?

Dual table is a table that is created by oracle along with data dictionary. It gives output of exactly one column name dummy and one record 'x'.

Select * from Dual;

DUMMY

X

Can we install Power Center Informatica 8.X version on Windows 7?

No we can't install Informatica on Power Center. We can Install Informatica 9.x on Windows 7.

What are the types of variables in Informatica?

There are three types of variables in Informatica

Predefined variable represented by \$

User defined variable represented with \$\$

System variable denoted by \$\$\$

Difference between Informatica normal join and Oracle Equi join?

Equi join in Oracle is performed on oracle sources (relational sources) while Informatica Equi joins can be performed on non-relational sources too (oracle and flat file etc).

What is degenerate dimension?

Dimension which has no dimension table of its own and is derived from the fact table.

What is requirements gathering?

It is carried out by Business Analyst. It is nothing but interacting with end users and getting to know what his requirements are. Based on his requirements, the rest of the phases Analysis, Design, Implementation and Testing and finally Maintenance are carried on.

What is Junk dimension?

The dimension that is formed by lumping of smaller dimensions is called Junk dimension.

What is Staging Area?

Staging Area is indeed a database where data from different source systems are brought together and this database acts as an input to Data Cleansing.

What are the types of joins in Informatica and in Oracle?

There are four types of joins in oracle

equi join

non equi join

self join

outer join

Joins in informatica

master join (right outer join)

detailed join (left outer join)

normal join

What is the file extension or format of files for the Informatica Objects like sessions, mappings etc. in Repository?

The format of files for Informatica Objects in Repository is XML

Where can we find Versioning in Informatica? What happens if Versioning is turned off?

In Informatica, we can find Versioning in Repository Manager. If Versioning is turned off, we will not be able to track the changes for the respective Sessions/Mappings/Workflows.

What is tracing level? What are the types of tracing level?

Tracing level is the amount of information that Informatica server writes into a log file.

Types of tracing level

Normal

Terse

Verbose init

Verbose data

In joiner transformation, we take the table with lesser number of rows as master while the more number of rows as detailed. Why?

In joiner, each and every row of the master is compared with every row of the detailed and so, the less number of rows in master, the less is the number of iterations and so better is the performance of the system.

What are the Limitations of Pushdown Optimization?

1. Rank T/R cannot be pushed
2. Transaction control T/R
3. Sorted aggregation.

Procedure:

1. Design a mapping with filter, rank and expression T/R.
2. Create a session --> Double click the session select properties tab.

Attribute	Value
Pushdown optimization	Full

3. Select the mapping tab --> set reader, writer connection with target load type normal.
4. Click apply --> click ok --> save the session.
5. Create & start workflow.

Pushdown Optimization Viewer:-

Double click the session --> Select the mapping tab from left window --> select pushdown optimization.

What is an overview window?

It is a window where you can see all the transformations in a mapping.

In how many ways we can update a source definition?

Two ways

We can reimport the source definition

We can edit the source definition

What is mapping?

Mapping is nothing but data flow between source and target.

What is a session?

Session is a set of instructions that tells the Informatica server when to and how to move the data from source to targets.

What are the different things we can do using PMCMD command?

We can START, STOP and ABORT session using PMCMD command.

What is the use of bitmap indexes?

Bitmap indexes are used to join large fact tables to smaller dimension tables.

How can we delete duplicate rows from flat files?

We can make use of sorter transformation and select distinct option.

If a session fails after loading 10000 records into the target how can we start loading into the target from the 10001th record?

We can run the session with the recovery strategy mode.

What is the limit of joiner transformations?

We cannot use sequence generator or update strategy transformations before or after joiner transformations.

How does server recognize the source and target databases?

By using ODBC if they are relational, FTP if they are flat files.

What is rank index in a group?

Power Center Designer automatically creates a RANK INDEX port while using Rank transformation. The purpose of this RANK INDEX port is to store the ranking for the

column(s) we are interested in.

What are the constants or flags for each database operation and their numeric equivalent in Update Strategy?

Insert DD_INSERT 0

Update DD_UPDATE 1

Delete DD_DELETE 2

Reject DD_REJECT 3

Can you generate reports using Informatica?

No, we cannot generate reports using Informatica, it is just an ETL tool but we can generate metadata reports.

Yes, Using Informatica Data Analyzer Tool we can generate reports.

Can you start batches within a batch?

No, you can't

What are the types of groups in Router Transformation?

There are three types of groups in Router Transformation namely

Input group

Output group

Default group

What are batches? What are the types of batches?

Batches provide a way to group sessions for either sequential or parallel execution by Informatica server.

Concurrent batches – which run at the same time

Sequential batches – which run one after the other

What is Power Center Repository?

Power Center Repository allows you to share metadata among different repositories and to create a DataMart domain.

What are the types of metadata that stores in repository?

Source definition

Target definition

Mappings

Maplet

Transformations

Differences between dynamic cache and static cache

In case of dynamic cache, if we want to insert a new row, then it will first look in the lookup cache and if the row is not present in the cache, it inserts the row into the cache as well as target but in case of static cache, it stores only into target table and not in cache.

What is the use of source qualifier?

Source qualifier is used to convert different data types to Informatica compatible data types.

What is page code compatibility?

It is nothing but compatibility of code for maintaining data accuracy. It comes into picture when data is in different languages.

What are Synonyms?

Synonyms are alternative names for database objects such as tables, views, stored procedures etc.

Syntax

Create [replace] synonym for [schema.]object name

Types of Lookup Cache?

Static cache

Dynamic cache

Persistent cache

Recache from database

Shared cache

What are various stages of SDLC?

Requirements Gathering, Analysis, Design, Implementation and Testing and Maintenance.

What is Star Schema?

Star Schema is a simplest form of schema which has one fact table and at least one dimension table. The dimensions here are denormalized.

What is Fact table?

It is a centralized table in Star Schema. Fact table has two types of columns. First type is measures and second type is the foreign keys for the dimension tables.

What is Snowflake Schema?

In Snowflake Schema, the dimensions are further divided into sub dimensions. The dimensions here are normalized.

What is Dimension table?

A dimension table is one that describes the business entities of an enterprise.

What is operational data store (ODS)?

Operational data store is defined to be structure that is

Subject-oriented

Integrated

Volatile

and current data that is a day or perhaps a month old.

What is Logical Data Modeling?

Logical Data Modeling is a type of data modeling which represents business requirements of an organization.

What is Physical Data Modeling?

Physical Data Modeling is a type of data modeling which includes all required tables, columns, and relationships for the physical implementation of a database.

What are the types of files created by Informatica server during the session running?

Types of files created are

Cache file

Session log file

Informatica server log file

Output file

Reject file

What is throughput in Informatica?

Throughput is nothing but the rate at which Informatica server reads the data from sources and writes them successfully to the target.

Where can we find the throughput option in Informatica?

We can view this in workflow monitor

In workflow monitor, right click on session, then click on Get Run Properties and under Source/Target statistics we can find throughput option

What are types of loading in Informatica?

The two types of loading available in Informatica are

Bulk Loading

Normal Loading

What is the difference between local index and global index?

Global index is nothing but a single index covering all partitions whereas local index has

separate index for each partition.

What is Complex Mapping?

Complex Mapping will have the following features

Difficult requirements

Many number of transformations

Complex business logic

How many number of sessions can you group in batches?

Any number of sessions but the lesser the number of sessions in a batch, the easier the migration.

What is the difference between Aggregator transformation and Expression Transformation?

Aggregator transformation use aggregator functions and performs calculation on entire group whereas in Expression transformation performs calculation on row by row basis.

What is the difference between Filter transformation and Router transformation?

Filter transformation drops the data that do not meet the condition whereas Router transformation captures the data even though the condition is not met and saves it in Default output group.

Filter transformation works on single condition only while Router transformation works on multiple conditions as well.

Filter transformation gives only one output. Router transformation can give more than one output.

What is the similarity between Router and Filter transformations?

Router transformation and Filter transformations are used to filter the data based on condition.

Both Filter and Router transformation are Active transformations.

Both Filter and Router transformation are connected.

What is the difference between Source Qualifier transformation and Joiner transformation?

Source Qualifier transformation is used to join the data from homogeneous sources while Joiner transformation is used to join data from heterogeneous sources as well as homogenous sources from different schemas.

We need matching keys to join two relational sources in Source Qualifier transformation and is not the case with joiner transformation.

What is PARAM file?

Param file is an ordinary text file where we can define the value for the parameter which is defined in session.

These parameter files are defined in session properties.

What are the various transformations which we cannot use in Maplet?

Transformations which we cannot use in Maplet are

Normalizer Transformation

XML Source Qualifier Transformation

Target Definition

Cobol Sources

Pre and Post Session Stored Procedures

What is the difference between Joiner transformation and Lookup transformation?

Joiner is active transformation while Lookup is passive transformation.

Joiner works on source data only while Lookup works on source as well as target data.

Joiner transformation supports equi joins only while Lookup supports equi join as well as non equi joins.

Joiner transformation is connected while Lookup transformation can be either connected or unconnected.

What is the use of Lookup transformation?

Lookup transformation is used to check whether matching record exists in the target table and if the matching record doesn't exist, it inserts accordingly.

How can we improve the session performance in Aggregator transformation?

We can increase the session performance by sending the sorted input to the aggregator transformation.

What is aggregated cache?

Aggregator cache is a temporary location which stores the input data values while the aggregation calculation is being carried out.

Difference between connected Lookup and Unconnected Lookup?

Connected lookup receives input directly from mapping pipeline whereas unconnected lookup receives input from :LKP expression of another transformation.

Connected lookup returns more than one column in a row whereas unconnected lookup returns only one column in each row.

Connected lookup Supports user-defined values while unconnected doesn't.

Connected lookup is not reusable whereas unconnected is.

Performance of connected lookup is lower compared to unconnected lookup.

Can Lookup be done on flat files?

Yes

What is the transformation used in loading 4 flat files of similar structure to a single target?

We can make use of Union transformation

Difference between direct and indirect loading options in sessions?

Direct loading can be used on single transformations while indirect loading can be used on multiple transformations

In direct loading, we can perform recovery process while in indirect loading, we cannot perform recovery process.

What are the various techniques for implementing a dimensional model?

Star schema

Snowflake schema

What are the types of dimensions?

There are three types of dimensions

Slowly changing dimensions

Confirmed dimensions

Casual dimensions

What is SQL override?

It is nothing but overriding SQL in source qualifier or lookup for additional logic.

What are the default values for variables?

The default variables for

Number=0

Variable=NULL

Date=1/1/1753

What are the types of data movement in Informatica server?

There are two types of modes in Informatica

Ascii mode

Unicode mode

What happens to the discarded records in Filter transformation?

Discarded rows do not appear in session logs and reject files.

How to increase the performance of session in using Filter transformation?

To increase the performance of session, we need to use Filter transformation as close as possible to the sources in the mapping.

In case of flat files (which comes through FTP) haven't arrived what happens?

The session is going to fail because of fatal error

Why do we need to reinitialize aggregate cache?

We need to reinitialize the aggregate cache only to remove the previous data present in the aggregator cache and to be used by the new source.

What are the various mapping objects available in Informatica?

Mapping objects that are available in Informatica are

Source Definition

Target Definition

Link

Transformation

Maplet

What is the default source option for Update Strategy?

Data Driven

What is the use of Update Strategy transformation?

Update Strategy is used to perform DML operations like Insert, Update, Delete and Reject on already populated targets.

What is the default join that the Joiner transformation provides?

Normal join

What are the differences between unique key and primary key?

Primary key cannot contain null value whereas unique key can contain one and only one null value

A unique key is similar to primary key but we can have more than one unique key per table

What is the difference between RowId and RowNum?

RowId is the physical address of a row. If we know the RowId, we can read entire row.

RowNum is the temporary number allocated during query execution

What are pseudo columns? What are the various types of pseudo columns?

Pseudo columns are columns which are not in the table but they can be used in sequel queries as if they are part of the table

RowNum

RowId

Sysdate

User

Curval

Nextval

What are the various types of Statements in Oracle?

The various Statements that are available in Oracle are

Data Manipulation Language (DML) Statements

Data Definition Language (DDL) Statements

Transaction Control (TC) Statements

Session Control (SC) Statements

System Control (SC) Statements

What are the various types of DML Statements?

Select

Update

Delete

Insert

Merge

What are various AGGREGATOR FUNCTIONS?

SUM

AVG

MIN

MAX

COUNT

STDDEV

VARIANCE

FIRST

LAST

Difference between char and varchar?

Char is a fixed length data type whereas varchar is a variable length data type. So in case of char, though we entered less size for a column, it is going to allocate max size and in case of varchar, it is not going to allocate the maximum size.

Difference between varchar and varchar2

Varchar and varchar2 are variable length datatypes. Varchar datatype has a size of 2000

bytes whereas varchar2 has a size of 4000 bytes.

Varchar is in ascii whereas varchar2 is in Unicode.

What is Normalization? Define 1NF, 2NF and 3NF.

It is the process of reducing the complex data structure into a simpler one by removing the redundancy.

First normal form: First normal form state that each field is atomic.

Second normal form: Second normal form states that data redundancy can be reduced if all the non key attributes which are dependent on one of the primary keys of a composite primary key are put to a separate table along with depended primary key . This should also satisfy the 1nf. For example, if a table has attributes like partid, city, state and country and partid, country for a composite primary key and city, state depend on country then the table is separated as two different tables as attributes having partid, country and country, state, city.

Third normal form: Third normal form states that if a dependency exist between non key attributes, then these attributes are isolated to a different table. This should also satisfy the 1nf + 2nf.

What Is Difference Between Maplet And Reusable Transformation?

Maplet :

- one or more transformations.
- set of transformations that are reusable.

Reusable transformation :

- only one transformation
- Single transformation which is reusable.

What is an Aggregator Transformation?

An aggregator is an Active, Connected transformation which performs aggregate calculations

like **AVG, COUNT, FIRST, LAST, MAX, MEDIAN, MIN, PERCENTILE, STDDEV, SUM** and **VARIANC**E.

How an Expression Transformation differs from Aggregator Transformation?

An Expression Transformation performs calculation on a **row-by-row** basis. An Aggregator Transformation performs calculations **on groups**.

Does an Informatica Transformation support only Aggregate expressions?

Apart from aggregate expressions Informatica Aggregator also supports non-aggregate expressions and conditional clauses.

How does Aggregator Transformation handle NULL values?

By default, the aggregator transformation treats null values as NULL in aggregate functions. But we can specify to treat null values in aggregate functions as NULL or zero.

What is Incremental Aggregation?

We can enable the session option, Incremental Aggregation for a session that includes an Aggregator Transformation. When the Integration Service performs incremental aggregation, it actually passes changed source data through the mapping and uses the historical cache data to perform aggregate calculations incrementally.

What are the performance considerations when working with Aggregator Transformation?

- Filter the unnecessary data before aggregating it. Place a Filter transformation in the mapping before the Aggregator transformation to reduce unnecessary aggregation.
- Improve performance by connecting only the necessary input/output ports to subsequent transformations, thereby reducing the size of the data cache.
- Use Sorted input which reduces the amount of data cached and improves session performance.

What differs when we choose Sorted Input for Aggregator Transformation?

Integration Service creates the index and data caches files in memory to process the Aggregator transformation. If the Integration Service requires more space as allocated for the index and data cache sizes in the transformation properties, it stores overflow values in cache files i.e. paging to disk. One way to increase session performance is to increase the index and data cache sizes in the transformation properties. But when we check Sorted Input the Integration Service uses memory to process an Aggregator transformation it does not use cache files.

Under what conditions selecting Sorted Input in aggregator will still not boost session performance?

- Incremental Aggregation, session option is enabled.
- The aggregate expression contains nested aggregate functions.
- Source data is data driven.

Under what condition selecting Sorted Input in aggregator may fail the session?

- If the input data is not sorted correctly, the session will fail.
- Also if the input data is properly sorted, the session may fail if the sort order by ports and the group by ports of the aggregator are not in the same order.

Suppose we do not group by on any ports of the aggregator what will be the output?

If we do not group values, the Integration Service will return **only the last row** for the input rows.

What is the expected value if the column in an aggregator transform is neither a group by nor an aggregate expression?

Integration Service produces one row for each group based on the group by ports. The columns which are neither part of the key nor aggregate expression will return the corresponding value of last record of the group received. However, if we specify particularly the FIRST function, the Integration Service then returns the value of the specified first row of the group. So default is the **LAST** function.

Give one example for each of Conditional Aggregation, Non-Aggregate expression and Nested Aggregation?

Use conditional clauses in the aggregate expression to reduce the number of rows used in the aggregation. The conditional clause can be any clause that evaluates to TRUE or FALSE.
SUM(SALARY, JOB = CLERK)

Use non-aggregate expressions in group by ports to modify or replace groups.

IIF(PRODUCT = Brown Bread, Bread, PRODUCT)

The expression can also include one aggregate function within another aggregate function, such as:

MAX(COUNT(PRODUCT))

How can we create Index after completion of Load Process?

Command Tasks are used to create the Index. Command Task scripts can be used in a session of the workflow to create an index.

What Is The Difference Between Stop And Abort?

Stop : If the session u want to stop is a part of batch you must stop the batch, if the batch is part of nested batch, Stop the outer most batch.

Abort : You can issue the abort command, it is similar to stop command except it has 60 second time out.

What is a predefined event?

Predefined event is the file-watch event. It waits for a specific file to arrive at a specific location.

What are Sessions?

Session is a Set of Instructions that are used while moving data from the source to destination. We can partition the session to implement several sequences of sessions to improve server performance.

After creating session we can use server manager or command-line program pmcmd to stop or start the session.

How To Recover Sessions In Concurrent Batches?

If multiple sessions in a concurrent batch fail, you might want to truncate all targets and run the batch again. However, if a session in a concurrent batch fails and the rest of the sessions complete successfully, you can recover the session as a standalone session.

To recover a session in a concurrent batch:

1. Copy the failed session using Operations-Copy Session.
2. Drag the copied session outside the batch to be a standalone session.
3. Follow the steps to recover a standalone session.
4. Delete the standalone copy.

How you will recover sessions?

If we stop a session or if an error causes a session to stop, refer to the session and error logs to determine the cause of failure. Correct the errors, and then complete the session. The method we use to complete the session depends on the properties of the mapping, session, and Informatica Server configuration.

Use one of the following methods to complete the session:

1. Run the session again if the Informatica Server has not issued a commit.
2. Truncate the target tables and run the session again if the session is not recoverable.
3. Consider performing recovery if the Informatica Server has issued at least one commit.

When to use Abort, Decode functions?

Abort can be used to Abort / stop the session on an error condition. If the primary key column contains NULL, and you need to stop the session from continuing then you may use ABORT function in the default value for the port. It can be used with IIF and DECODE function to Abort the session.

What is constraint based loading ?

Constraint based loading. the data was loaded into the target table based on the Constraints .i.e if we want to load the EMP&DEPT data, first it loads the data of DEPT then EMP because

DEPT is PARENT table EMP is CHILD table. In simple terms, it loads PARENT table first then CHILD table.

How Can You Complete Unrecoverable Sessions?

Under certain circumstances, when a session does not complete, you need to truncate the target tables and run the session from the beginning. Run the session from the beginning when the Informatica Server cannot run recovery or when running recovery might result in inconsistent data.

If there is no recovery mode on in session and workflow failed in mid of execution then

1. Don't truncate table immediately.
2. If there is large volume of data is performing by the load and more than 25% data has loaded then-if same workflow has multiple session then check particular session which caused to be failed and fire the delete command only to delete particular session data which has loaded and copy the session into new workflow and run only that session or dependent others.

Explain these terms Session, Worklet, Mapplet, and Workflow?

- Mapplet: It arranges or creates sets of transformation
- Worklet: It represents a specific set of tasks given
- Workflow: It's a set of instructions that tell the server how to execute tasks
- Session: It is a set of parameters that tells the server how to move data from sources to target

Can We Run A Group Of Sessions Without Using Workflow Manager?

It is possible two run two session only (by precession, post session) using pmcmd without using workflow. Not more than two.

How can we create indexes after completing the load process?

With the help of command task at the session level, we can create indexes after the load procedure.

What is a Rank Transform?

Rank is an Active Connected Informatica transformation used to select a set of top or bottom values of data.

How does a Rank Transform differ from Aggregator Transform functions MAX and MIN?

Like the Aggregator transformation, the Rank transformation lets us group information. The Rank Transform allows us to select a group of top or bottom **values**, not just **one value** as in case of Aggregator MAX, MIN functions.

What is a RANK port and RANKINDEX?

Rank port is an input/output port used to specify the column for which we want to rank the source values. By default Informatica creates an output port RANKINDEX for each Rank transformation. It stores the ranking position for each row in a group.

How can you get ranks based on different groups?

Rank transformation lets us group information. We can configure one of its input/output ports as a group by port. For each unique value in the group port, the transformation creates a group of rows falling within the rank definition (top or bottom, and a particular number in each rank).

What happens if two rank values match?

If two rank values match, they receive the same value in the rank index and the transformation skips the next value.

What are the restrictions of Rank Transformation?

- We can connect ports from only one transformation to the Rank transformation.
- We can select the top or bottom rank.
- We need to select the Number of records in each rank.
- We can designate only one Rank port in a Rank transformation.

How does a Rank Cache works?

During a session, the Integration Service compares an input row with rows in the data cache. If the input row out-ranks a cached row, the Integration Service replaces the cached row with the input row. If we configure the Rank transformation to rank based on different groups, the Integration Service ranks incrementally for each group it finds. The Integration Service creates an index cache to store the group information and data cache for the row data.

How does Rank transformation handle string values?

Rank transformation can return the strings at the top or the bottom of a session sort order. When the Integration Service runs in Unicode mode, it sorts character data in the session using the selected sort order associated with the Code Page of IS which may be French, German, etc. When the Integration Service runs in ASCII mode, it ignores this setting and uses a binary sort order to sort character data.

If You Want To Create Indexes After The Load Process Which Transformation You Choose?

It's usually not done in the mapping (transformation) level. It's done in session level. Create a command task which will execute a shell script (if Unix) or any other scripts which contains

the create index command. Use this command task in the workflow after the session or else, You can create it with a post session command.

Why is Sorter an Active Transformation?

When the Sorter transformation is configured to treat output rows as distinct, it assigns all ports as part of the sort key. The Integration Service discards duplicate rows compared during the sort operation. The number of Input Rows will vary as compared with the Output rows and hence it is an Active transformation.

How does Sorter handle Case Sensitive sorting?

The Case Sensitive property determines whether the Integration Service considers case when sorting data. When we enable the Case Sensitive property, the Integration Service sorts uppercase characters higher than lowercase characters.

How does Sorter handle NULL values?

We can configure the way the Sorter transformation treats null values. Enable the property Null Treated Low if we want to treat null values as lower than any other value when it performs the sort operation. Disable this option if we want the Integration Service to treat null values as higher than any other value.

How does a Sorter Cache works?

The Integration Service passes all incoming data into the Sorter Cache before Sorter transformation performs the sort operation.

The Integration Service uses the Sorter Cache Size property to determine the maximum amount of memory it can allocate to perform the sort operation. If it cannot allocate enough memory, the Integration Service fails the session. For best performance, configure Sorter cache size with a value less than or equal to the amount of available physical RAM on the Integration Service machine.

If the amount of incoming data is greater than the amount of Sorter cache size, the Integration Service temporarily stores data in the Sorter transformation work directory. The Integration Service requires disk space of at least twice the amount of incoming data when storing data in the work directory.

What is the worklet?

The worklet is a group of sessions. To execute the worklet we have to create the workflow.

How many number of sessions can we have in one group?

We can have any number of sessions, but it is advisable to have lesser number of sessions in a batch because it will become easier for migration.

What is difference between Rename and Alias?

Rename is a permanent name given to a table or column whereas Alias is a temporary name given to a table or column which do not exist once the SQL statement is executed.

What is a view?

A view is stored procedure based on one or more tables, its a virtual table.

Can a primary key contain more than one columns?

Yes

How you will avoid duplicating records in a query?

By using DISTINCT

What is INFORMATICA Worklet, explain in detail?

Worklet works as a Mapplet with the feature of Reusability, the only difference is that we can apply worklet to any number of workflows in INFORMATICA, unlike mapplet. Worklet saves the logic and tasks at a single place to reuse.

Worklet is much similar to the Mapplet and is defined as the group of tasks that can be either reusable or non-reusable at the workflow level. It can be added to as much number of workflows as required. With its reusability feature, much time is saved as reusable logic can be developed once and can be placed from where it can be reused.

In the INFORMATICA Power center environment, Mapplets are considered as the most advantageous feature. Mapplets are created in Mapplet designer and are a part of the Designer tool.

It basically contains a set of transformations that are designed to be reused in multiple mapping.

Mapplets are said to be reusable objects which simplify mapping by:

Including multiple transformations and source definitions.

Not required to connect to all input and output ports.

Accept data from sources and pass to multiple transformations

Well, overall when it is required to reuse the mapping logic then the logic should be placed in Mapplet.

What is Worklet and types of worklets?

1. A worklet is defined as group of related tasks.
2. There are 2 types of the worklet:
 - Reusable worklet
 - Non-Reusable worklet

3. Worklet expands and executes the tasks inside the workflow.
4. A workflow which contains the worklet is known as Parent Workflow.

(a) Reusable Worklet:-

Created using worklet designer tool.

Can be assigned to Multiple workflows.

(b) Non-Reusable Worklet:-

Created using workflow designer tool.

Created Specific to workflow.

What Is Change Data Capture?

Change data capture (CDC) is a set of software design patterns used to determine the data that has changed in a database so that action can be taken using the changed data.

Which datatype is used for storing graphics and images?

LONG RAW data type is used for storing BLOB's (binary large objects).

How will you delete duplicating rows from a base table?

```
DELETE FROM table_name A WHERE rowid > (SELECT min(rowid) from table_name B where  
B.table_no=A.table_no);
```

```
CREATE TABLE new_table AS SELECT DISTINCT * FROM old_table;
```

What is Code page Compatibility?

Transfer of data takes place from one code page to another keeping that both code pages have the same character sets then data failure cannot occur.

What Are Partition Points?

Partition points mark the thread boundaries in a source pipeline and divide the pipeline into stages.

How can we identify whether a mapping is correct or not without a connecting session?

With the help of the debugging option, we can identify whether a mapping is correct or not without connecting sessions.

What is difference between SUBSTR and INSTR?

SUBSTR returns a specified portion of a string eg SUBSTR('BCDEF',4) output BCDE INSTR provides character position in which a pattern is found in a string.

eg INSTR('ABC-DC-F','-',2) output 7 (2nd occurrence of '-')

There is a string '120000 12 0 .125', how you will find the position of the decimal place?

INSTR('120000 12 0 .125','.',1)

output 13

There is a '%' sign in one field of a column. What will be the query to find it?

IN LIKE Operator " Should be used before '%'.

When you use WHERE clause and when you use HAVING clause?

HAVING clause is used when you want to specify a condition for a group function and it is written after GROUP BY clause The WHERE clause is used when you want to specify a condition for columns, single row functions except group functions and it is written before GROUP BY clause if it is used.

Which is more faster – IN or EXISTS?

EXISTS is more faster than IN because EXISTS returns a Boolean value whereas IN returns a value.

Result of the subquery is small Then “IN” is typically more appropriate. and Result of the subquery is big/large/long Then “EXIST” is more appropriate.

What is a OUTER JOIN?

Outer Join–Its a join condition used where you can query all the rows of one of the tables in the join condition even though they don't satisfy the join condition.

How you will avoid your query from using indexes?

SELECT * FROM emp Where emp_no+' '=12345;

What is a pseudo column. Give some examples

It is a column that is not an actual column in the table.

eg USER, UID, SYSDATE, ROWNUM, ROWID, NULL, AND LEVEL.

Suppose customer table is there having different columns like customer no, payments.

What will be the query to select top three max payments.

What is the benefit of Session Partitioning?

While Integration Service is running in the environment the workflow is partitioned for better performance. These partitions are then used to perform Extraction, Transformation, and Loading.

Why should we partition a Session?

Partition not only helps in optimizing a Session but also helps in loading colossal volume of data and improve the server's operation and efficiency.

Why We Use Partitioning The Session In Informatica?

Performance can be improved by processing data in parallel in a single session by creating multiple partitions of the pipeline.

Informatica server can achieve high performance by partitioning the pipeline and performing the extract , transformation, and load for each partition in parallel.

Can we use mapping parameters or variables, developed in one mapping, into any other reusable transformation?

Yes, we can use mapping parameters or variables into any other reusable transformation because they doesn't have any mapplet.

Write the prerequisite tasks to achieve the session partition?

In order to perform session partition, one needs to configure the session to partition source data and then installing the Informatica server machine in multifold CPU's.

What is the use of the aggregator cache file?

If extra memory is needed, aggregator provides extra cache files for keeping the transformation values. It also keeps the transitional value that are there in the local buffer memory.

What Is Difference Between Partitioning Of Relational Target And Partitioning Of File Targets?

Partition's can be done on both relational and flat files.

Informatica supports following partitions

1. Database partitioning
2. RoundRobin
3. Pass-through
4. Hash-Key partitioning
5. Key Range partitioning

All these are applicable for relational targets. For flat file only database partitioning is not applicable.

Informatica supports Navy partitioning. you can just specify the name of the target file and create the partitions, rest will be taken care by informatica session.

What is Data warehouse?

Data warehouse is relational database used for query analysis and reporting. By definition data warehouse is Subject-oriented, Integrated, Non-volatile, Time variant.

Subject oriented : Data warehouse is maintained particular subject.

Integrated : Data collected from multiple sources integrated into a user readable unique format.

Non volatile : Maintain Historical date.

Time variant : data display the weekly, monthly, yearly.

What is Data mart?

A subset of data warehouse is called Data mart.

Difference between Data warehouse and Data mart?

Data warehouse is maintaining the total organization of data. Multiple data marts used in data warehouse. where as data mart is maintained only particular subject.

Difference between OLTP and OLAP?

OLTP is Online Transaction Processing. This is maintained current transactional data. That means insert, update and delete must be fast.

Explain ODS?

Operational data store is a part of data warehouse. This is maintained only current transactional data. ODS is subject oriented, integrated, volatile, current data.

Difference between Power Center and Power Mart?

Power center receive all product functionality including ability to multiple register servers and metadata across the repository and partition data.

One repository multiple informatica servers. Power mart received all features except multiple register servers and partition data.

What is a staging area?

Staging area is a temporary storage area used for transaction, integrated and rather than transaction processing.

When ever your data put in data warehouse you need to clean and process your data.

Explain Additive, Semi-additive, Non-additive facts?

Additive fact: Additive Fact can be aggregated by simple arithmetical additions.

Semi-Additive fact: semi additive fact can be aggregated simple arithmetical additions along with some other dimensions.

Non-additive fact: Non-additive fact can't be added at all.

What is a Fact less Fact and example?

Fact table which has no measures.

Explain Surrogate Key?

Surrogate Key is a series of sequential numbers assigned to be a primary key for the table.

How many types of approaches in DHW?

Two approaches: Top-down(Immol approach), Bottom-up(Ralph Kimball)

Explain Star Schema?

Star Schema consists of one or more fact table and one or more dimension tables that are to foreignkeys. Dimension tables are De-normalized, Fact table-normalized

Advantages: Less database space & Simplify queries.

Explain Snowflake schema?

Snowflake schema is a normalize dimensions to eliminate the redundancy. The dimension data has been grouped into one large table. Both dimension and fact tables normalized.

What is confirm dimension?

If both data marts use same type of dimension that is called confirm dimension. If you have same type of dimension can be used in multiple fact that is called confirm dimension.

When you use for dynamic cache.

Your target table is also look up table then you go for dynamic cache .In dynamic cache multiple matches return an error. Use only = operator.

What is lookup override?

Override the default SQL statement. You can join multiple sources use lookup override. By default informatica server add the order by clause.

We can pass the null value in lookup transformation?

Lookup transformation returns the null value or equal to null value.

What is the target load order?

You specify the target load order based on source qualifiers in a mapping. If u have the multiple source qualifiers connected to the multiple targets you can designate the order in which informatica server loads data into the targets.

What is default join that source qualifier provides?

Inner equi join.

What is data driven?

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

if u do not choose data driven option setting , the informatica server ignores all update strategy transformations in the mapping.

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

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

Difference between the source filter and filter?

Source filter is filtering the data only relational sources. Where as filter transformation filter the data any type of source.

What is a tracing level?

Amount of information sent to log file.

What are the types of tracing levels?

Normal, Terse, verbose data, verbose initialization.

Can you connect multiple ports from one group to multiple transformations?

Yes

Can you connect more than one group to the same target or transformation?

NO

what are the methods for creating reusable transformation?

Two methods

- 1) Design it in the transformation developer.
- 2) Promote a standard transformation from the mapping designer. After you add a transformation to the mapping, you can promote it to status of reusable transformation. Once you promote a standard transformation to reusable status, you can demote it to a standard transformation at any time.

If u change the properties of a reusable transformation in mapping , you can revert it to the original reusable transformation properties by clicking the revert.

How does a pipeline partition improve performance?

A pipeline partition lets you divide a pipeline into different reader/transformation/writer threads. The integration service can run the different partitions within the mapping at the same time, increasing efficiency

Can you use the mapping parameters or variables created in one mapping into another mapping?

NO, we can use mapping parameters or variables in any transformation of the same mapping or mapplet in which have created mapping parameters or variables.

Can you use the mapping parameters or variables created in one mapping into any other result transformation.

Yes because the reusable transformation is not contained with any mapplet or mapping.

What are the different ways to implement partitioning/parallel processing in Informatica?

We can implement parallel processing using various types of partition algorithms:

Database partitioning: The Integration Service queries the database system for table partition information. It reads partitioned data from the corresponding nodes in the database.

Round-Robin Partitioning: Using this partitioning algorithm, the Integration service distributes data evenly among all partitions. It makes sense to use round-robin partitioning when you need to distribute rows evenly and do not need to group data among partitions.

Hash Auto-Keys Partitioning: The Powercenter Server uses a hash function to group rows of data among partitions. When the hash auto-key partition is used, the Integration Service uses all grouped or sorted ports as a compound partition key. You can use hash auto-keys partitioning at or before Rank, Sorter, and unsorted Aggregator transformations to ensure that rows are grouped properly before they enter these transformations.

Hash User-Keys Partitioning: Here, the Integration Service uses a hash function to group rows of data among partitions based on a user-defined partition key. You can individually choose the ports that define the partition key.

Key Range Partitioning: With this type of partitioning, you can specify one or more ports to form a compound partition key for a source or target. The Integration Service then passes data to each partition depending on the ranges you specify for each port.

Pass-through Partitioning: In this type of partitioning, the Integration Service passes all rows from one partition point to the next partition point without redistributing them.

Explain the WORKING of EVENT TASKS?

We can define events in the workflow to specify the sequence of task execution.

Types of Events:

Pre-defined event: A pre-defined event is a file-watch event. This event Waits for a specified file to arrive at a given location.

User-defined event: A user-defined event is a sequence of tasks in the Workflow. We create events and then raise them as per need.

Types of Events Tasks:

EVENT RAISE: Event-Raise task represents a user-defined event. We use this task to raise a user defined event.

EVENT WAIT: Event-Wait task waits for a file watcher event or user defined event to occur before executing the next session in the workflow.

What is TIMER TASK?

The Timer task allows us to specify the period of time to wait before the Power Center Server runs the next task in the workflow. The Timer task has two types of settings:

Absolute time: We specify the exact date and time or we can choose a user-defined workflow variable to specify the exact time. The next task in workflow will run as per the date and time specified.

Relative time: We instruct the Power Center Server to wait for a specified period of time after the Timer task, the parent workflow, or the top-level workflow starts.

What is DECISION TASK?

The Decision task allows us to enter a condition that determines the execution of the workflow,

What is CONTROL TASK?

We can use the Control task to stop, abort, or fail the top-level workflow or the parent workflow based on an input link condition.

What is ASSIGNMENT TASK?

The Assignment task allows us to assign a value to a user-defined workflow variable.

What is Scheduler?

We can schedule a workflow to run continuously, repeat at a given time or interval, or we can manually start a workflow.

Odd number of records?

sql> select * from emp where (rowid,1) in (select rowid, mod(rownum,2) from emp);

Even number of records?

sql> select * from emp where (rowid,0) in (select rowid, mod(rownum,2) from emp);

What is the rank index in rank transformation?

The designer automatically creates a RANKINDEX port for each Rank transformation. The informatica server uses the Rank Index port to store the ranking position for each record in a group. For example, if you create a Rank transformation that ranks the top 5 sales persons for each quarter, the rank index number the salespeople from 1 to 5.

What is a parameter a file?

Parameter file defines the values for parameter and variables.

Explain a work flow process?

The power center server uses both process memory and system shared memory to perform these tasks.

Load manager process: stores and locks the workflow tasks and start the DTM run the sessions.

Data Transformation Process DTM: Perform session validations, create threads to initialize the session, read, write and transform data, and handle pre and post session operations.

The default memory allocation is 12,000,000 bytes.

How many sessions can be grouped in a batch?

There is no limit to the number of sessions that can comprise a batch. But the fewer the sessions, the easier the migration.

Explain work flow manager tools?

- 1) Task developer.
- 2) Work flow designer.
- 3) Worklet designer.

Explain work flow schedule.

You can schedule a work flow to run continuously, repeat at given time or interval or you manually start a work flow. By default the workflow runs on demand.

Can You Start A Batches With In A Batch?

You cannot. If you want to start batch that resides in a batch, create a new independent batch and copy the necessary sessions into the new batch.

Explain bulk loading?

You can use bulk loading to improve performance of a session that inserts a large amount of data to a db2,sysbase,oracle or MS SQL server database.

When bulk loading the power center server bypasses the database log, which speeds performance.

Without writing to the database log, however the target database can't perform rollback. As a result you may not be able to perform recovery.

How can we use Batches?

Batches are the collection of sessions which is used to migrate the data from the source to target on a server. Batches can have the largest number of sessions in it but they cause more network traffic whereas less number of sessions in a batch can be moved rapidly.

What is transformation?

Transformation is repository object that generates, modifies or passes data.

What are the types of groups in router transformation?

Router transformation has 2 groups 1. Input group 2. Output groups.

Output groups are of 2 types. 1. user defined group 2. default group.

How can you improve the session performance in aggregate transformation?

Use stored input.

What is aggregate cache in aggregate transformation?

The aggregate stores data in the aggregate cache until it completes aggregate calculations. When you run a session that uses an aggregate transformation, the Informatica server creates index and data caches in memory to process the transformation. If the Informatica server requires more space it stores overview values in cache files.

What is the difference between partitioning of relational target and partitioning of file targets ?

If you partition a session with a relational target Informatica server creates multiple connections to the target database to write target data concurrently. If you partition a session with a file target the Informatica server creates one target file for each partition. You can configure session properties to merge these target files.

What are the settings that you use to configure the joiner transformation?

Master and detail source.

Type of join

Condition of the join

Difference between primary key and unique key?

Primary key is Not null unique

Unique accept the null values.

Write the query odd and even numbers?

Select * from emp where (rowid,1) in (select rowid, mod(rownum,2) from emp)

What are the transformation that restricts the partitioning of sessions?

Advanced External procedure transformation and External procedure

transformation: This transformation contains a check box on the properties tab to allow partitioning.

Aggregator Transformation : If you use sorted ports you can not partition the associated source

Joiner Transformation: you can not partition the master source for a joiner transformation .Normalizer Transformation XML targets.

What are the types of Approach in DWH?

Bottom up approach: first we need to develop data mart then we integrate these data mart into EDW.

Bottom up

OLTP ETL Data mart DWH OLAP

Bottom up

Planning & Designing the Data Marts without waiting for the Global warehouse design

- Immediate results from the data marts
- Tends to take less time to implement
- Errors in critical modules are detected earlier.
- Benefits are realized in the early phases.
- It is a Best Approach

Top down approach: first we need to develop EDW then form that EDW we develop data mart

Top down

OLTP ETL DWH Data mart OLAP

Top down

- Cost of initial planning & design is high
- Takes longer duration of more than an year

What are different types of Data Modelling?

Conceptual Data Modelling
Logical Data Modelling
Physical Data Modelling
Dimensional Data Modelling

What are the Important aspects of Star Schema & Snow Flake Schema?

In a star schema every dimension will have a primary key.
In a star schema, a dimension table will not have any parent table.
Whereas in a snow flake schema, a dimension table will have one or more parent tables.
Hierarchies for the dimensions are stored in the dimensional table itself in star schema.
Whereas hierarchies are broken into separate tables in snow flake schema. These hierarchies help to drill down the data from topmost hierarchies to the lowermost hierarchies.

What is Incremental Aggregation?

Whenever a session is created for a mapping Aggregate Transformation, the session option for Incremental Aggregation can be enabled. When PowerCenter performs incremental aggregation, **it passes new source data through the mapping and uses historical cache data to perform new aggregation calculations incrementally.**

Can two flat files be joined with Joiner Transformation?

Yes, joiner transformation can be used to join data from two flat file sources.

Can a lookup be done on Flat Files?

Yes.

What are the types of loading in Informatica?

There are two types of loading, **1. Normal loading and 2. Bulk loading.**
In normal loading, it loads record by record and writes log for that. It takes comparatively a longer time to load data to the target.
In bulk loading, it loads number of records at a time to target database. It takes less time to load data to target.

What is aggregate cache in aggregator transformation?

The aggregator stores data in the aggregate cache until it completes aggregate calculations. When you run a session that uses an aggregator transformation, the informatica server creates index and data caches in memory to process the transformation. If the informatica server requires more space, it stores overflow values in cache files.

What type of repositories can be created using Informatica Repository Manager?

Informatica PowerCenter includes following type of repositories:

- **Standalone Repository:** A repository that functions individually and this is unrelated to any other repositories.
- **Global Repository:** This is a centralized repository in a domain. This repository can contain shared objects across the repositories in a domain. The objects are shared through global shortcuts.
- **Local Repository:** Local repository is within a domain and its not a global repository. Local repository can connect to a global repository using global shortcuts and can use objects in its shared folders.
- **Versioned Repository:** This can either be local or global repository but it allows version control for the repository. A versioned repository can store multiple copies, or versions of an object. This feature allows efficiently developing, testing and deploying metadata in the production environment.

What is a code page?

A code page contains encoding to specify characters in a set of one or more languages. The code page is selected based on source of the data. For example if source contains Japanese text then the code page should be selected to support Japanese text.

When a code page is chosen, the program or application for which the code page is set, refers to a specific set of data that describes the characters the application recognizes. This influences the way that application stores, receives, and sends character data.

How to execute PL/SQL script from Informatica mapping?

A. Stored Procedure (SP) transformation can be used to execute PL/SQL Scripts. In SP Transformation PL/SQL procedure name can be specified. Whenever the session is executed, the session will call the pl/sql procedure.

What is Data Driven?

The informatica server follows instructions coded into update strategy transformations within the session mapping which determine how to flag records for insert, update, delete or reject. If we do not choose data driven option setting, the informatica server ignores all update strategy transformations in the mapping.

What are the types of mapping wizards that are provided in Informatica?

The designer provide two mapping wizard.

1. **Getting Started Wizard** - Creates mapping to load static facts and dimension tables as well as slowly growing dimension tables.
2. **Slowly Changing Dimensions Wizard** - Creates mappings to load slowly changing dimension tables based on the amount of historical dimension data we want to keep and the method we choose to handle historical dimension data.

What is Session and Batches?

Session - A Session Is A set of instructions that tells the Informatica Server How And When To Move Data From Sources To Targets. After creating the session, we can use either the server manager or the command line program pmcmd to start or stop the session.

Batches - It Provides A Way to Group Sessions For Either Serial Or Parallel Execution By The Informatica Server. There Are Two Types Of Batches:

1. **Sequential** - Run Session One after the Other.
2. **Concurrent** - Run Session At The Same Time.

How many ways you can update a relational source definition and what are they?

- A. Two ways
1. Edit the definition
 2. Reimport the definition

What is a transformation?

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

What are the designer tools for creating transformations?

- A. Mapping designer
Transformation developer
Maplet designer

In how many ways can you create ports?

- A. Two ways
1. Drag the port from another transformation
 2. Click the add button on the ports tab.

What are reusable transformations?

A. A transformation that can be reused is called a reusable transformation
They can be created using two methods:
1. Using transformation developer
2. Create normal one and promote it to reusable

What r the settings that u use to configure the joiner transformation?

- **Master and detail source**
- **Type of join**
- **Condition of the join**

What are the join types in joiner transformation?

Normal (Default) -- only matching rows from both master and detail

Master outer -- all detail rows and only matching rows from master

Detail outer -- all master rows and only matching rows from detail

Full outer -- all rows from both master and detail (matching or non matching)

What are the types of lookup caches?

- **Static cache:** You can configure a static or read-only cache for only lookup table. By default Informatica server creates a static cache. It caches the lookup table and lookup values in the cache for each row that comes into the transformation. When the lookup condition is true, the Informatica server does not update the cache while it processes the lookup transformation.
- **Dynamic cache:** If you want to cache the target table and insert new rows into cache and the target, you can create a look up transformation to use dynamic cache. The Informatica server dynamically inserts data to the target table.
- **Persistent cache:** You can save the lookup cache files and reuse them the next time the Informatica server processes a lookup transformation configured to use the cache.
- **Recache from database:** If the persistent cache is not synchronized with the lookup table, you can configure the lookup transformation to rebuild the lookup cache.
- **Shared cache:** You can share the lookup cache between multiple transactions. You can share unnamed cache between transformations in the same mapping.

What kinds of sources and of targets can be used in Informatica?

- Sources may be Flat file, relational db or XML.
- Target may be relational tables, XML or flat files.

What is the different type of tracing levels?

Tracing level represents the **amount of information that Informatica Server writes in a log file**. Tracing levels store information about mapping and transformations. There are 4 types of tracing levels supported

1. **Normal:** It specifies the initialization and status information and summarization of the success rows and target rows and the information about the skipped rows due to transformation errors.
2. **Terse:** Specifies Normal + Notification of data
3. **Verbose Initialization:** In addition to the Normal tracing, specifies the location of the data cache files and index cache files that are treated and detailed transformation statistics for each and every transformation within the mapping.
4. **Verbose Data:** Along with verbose initialization records each and every record processed by the informatica server.

How do you identify existing rows of data in the target table using lookup transformation?

There are two ways to lookup the target table to verify a row exists or not :

1. Use connect dynamic cache lookup and then check the values of NewLookuprow Output port to decide whether the incoming record already exists in the table / cache or not.
2. Use Unconnected lookup and call it from an expression transformation and check the Lookup condition port value (Null/ Not Null) to decide whether the incoming record already exists in the table or not.

What is a level of Granularity of a fact table?

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

What is worklet?

Worklet are objects that represent a set of workflow tasks that allow to reuse a set of workflow logic in several window.

Use of Worklet: You can bind many of the tasks in one place so that they can easily get identified and also they can be of a specific purpose.

Why cannot we use sorted input option for incremental aggregation?

In incremental aggregation, the aggregate calculations are stored in historical cache on the server. In this historical cache the data need not be in sorted order. If you give sorted input, the records come as pre-sorted for that particular run but in the historical cache the data may not be in the sorted order. That is why this option is not allowed.

How can we store previous session logs?

Go to Session Properties → Config Object → Log Options

Select the properties as follows....

Save session log by → Session Runs

Save session log for these runs → Change the number that you want to save the number of log files (Default is 0)

If you want to save all of the logfiles created by every run, and then select the option

Save session log for these runs → Session Timestamp

You can find these properties in the session/workflow Properties.

What is an indicator file? and how it can be used?

Indicator file is used for Event Based Scheduling when you don't know when the Source Data is available. A shell command, script or a batch file creates and send this indicator file to the directory local to the Informatica Server. Server waits for the indicator file to appear before running the session.

If session fails after loading 10000 records in the target, how can we load 10001th record when we run the session in the next time?

Select the **Recovery Strategy** in session properties as “**Resume from the last check point**”.

Note – Set this property before running the session

What is “Insert Else Update” and “Update Else Insert”?

These options are used when dynamic cache is enabled.

- **Insert Else Update** option applies to rows entering the lookup transformation with the row type of insert. When this option is enabled the integration service inserts new rows in the cache and updates existing rows. When disabled, the Integration Service does not update existing rows.
- **Update Else Insert** option applies to rows entering the lookup transformation with the row type of update. When this option is enabled, the Integration Service updates existing rows, and inserts a new row if it is new. When disabled, the Integration Service does not insert new rows.

What are the Different methods of loading Dimension tables?

Conventional Load - Before loading the data, all the Table constraints will be checked against the data.

Direct load (Faster Loading) - All the Constraints will be disabled. Data will be loaded directly. Later the data will be checked against the table constraints and the bad data won't be indexed.

What are the different types of Commit intervals?

The different commit intervals are:

- **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.
- **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.

How to add source flat file header into target file?

Edit Task-->Mapping-->Target-->Header Options--> Output field names

How to load name of the file into relation target?

Source Definition-->Properties-->Add currently processed file name port

How to return multiple columns through un-connect lookup?

Suppose your look table has f_name,m_name,l_name and you are using unconnected lookup. In **override SQL of lookup** use f_name||~||m_name||~||l_name you can easily get

this value using unconnected lookup in expression. Use substring function in expression transformation to separate these three columns and make them individual port for downstream transformation /Target.

What is Factless fact table? In which purpose we are using this in our DWH projects? Plz give me the proper answer?

It is a fact table which does not contain any measurable data.

EX: Student attendance fact (it contains only Boolean values, whether student attended class or not ? Yes or No.)

A Factless fact table contains only the keys but there is no measures or in other way we can say that it contains no facts. Generally it is used to integrate the fact tables

Factless fact table contains only foreign keys. We can have two kinds of aggregate functions from the factless fact one is count and other is distinct count.

2 purposes of factless fact

1. Coverage: to indicate what did NOT happen. Like to

Like: which product did not sell well in a particular region?

2. Event tracking: To know if the event took place or not.

Like: Fact for tracking student's attendance will not contain any measures.

How can we store previous session logs?

If you run the session in the time stamp mode then automatically session log out will not overwrite the current session log.

Go to Session Properties → Config Object → Log Options

Select the properties as follows:

Save session log by → SessionRuns

Save session log for these runs → Change the number that you want to save the number of log files (Default is 0)

If you want to save all of the log files created by every run, and then select the option Save session log for these runs → SessionTimeStamp

You can find these properties in the session/workflow Properties.

What is use of batch file in informatica? How many types of batch file in informatica?

With the batch file, we can run sessions either in sequential or in concurrently.

Grouping of Sessions is known as Batch.

Two types of batches:

1) Sequential: Runs Sessions one after another.

2) Concurrent: Run the Sessions at the same time.

If u have sessions with source-target dependencies u have to go for sequential batch to start the sessions one after another. If u have several independent sessions u can use concurrent batches Which run all the sessions at the same time

What is joiner cache?

When we use the joiner transformation an integration service maintains the cache, all the records are stored in joiner cache. Joiner caches have 2 types of cache **1.Index cache 2.**

Joiner cache.

Index cache stores all the port values which are participated in the join condition and data cache have stored all ports which are not participated in the join condition.

In informatics server Which files are created during the session runs?

Errors log, Bad file, Workflow log and session log namely files are created during the session runs. Company.

How to delete the data in the target table after loaded.

SQ--> Properties tab-->Post SQL
delete from target_tablename

SQL statements executed using the source database connection, after a pipeline is run write post sql in target table as truncate table name. we have the property in session truncate option.

What is polling in informatica?

It displays the updated information about the session in the monitor window. The monitor window displays the status of each session when you poll the Informatica server.

How i will stop my workflow after 10 errors

Session level property **error handling** mention condition stop on errors: 10

--->Config object -> Error Handling -> Stop on errors

"Stop On Error" property means that informatica will fail the session after n number of errors.

So if you set it to 10 than Informatica will skip first 10 errors and than it will fail session.

So if set it to 0 so it wont stop on non-fatal errors but it will stop as soon as it encounters a Fatal Error. This property is not applicable for Fatal Errors.

How can we calculate fact table size?

A fact table is multiple of combination of dimension tables

ie if we want to find the fact table size of 3 years of historical data with 200 products and 200 stores

$3 \times 365 \times 200 \times 200 = \text{fact table size}$

Without using emaitask how will send a mail from informatica?

by using 'mailx' command in Unix or shell scripting

How will compare two mappings in two different repositories?

in the designer client , goto mapping tab there is one option that is 'compare', here we will compare two mappings in two different repository in informatica designer go to mapping tab--->compare..
we can compare 2 folders within the same repository ..
we can compare 2 folders within different repository ..

What is constraint based load order

Constraint based load order defines the order in which data loads into the multiple targets based on primary key and foreign key relationship.

What is target load plan

Suppose i have 3 pipelines in a single mapping designer

emp source--->sq--->tar1

dept source--->sq--->tar2

bonus source--->sq--->tar3

my requirement is to load first in tar2 then tar1 and then finally tar3

for this type of loading to control the extraction of data from source by source qualifier we use target load plan.

What is meant by data driven.. in which scenario we use that..?

Data driven is available at session level. it says that when we are using update strategy t/r ,how the integration service fetches the data and how to update/insert row in the database log.

Data driven is nothing but instruct the source rows that should take action on target i.e(update, delete, reject, insert). If we use the update strategy transformation in a mapping then will select the data driven option in session.

How to run workflow in Unix?

Syntax: pmcmd startworkflow -sv <service name> -d <domain name> -u <user name> -p <password> -f <folder name> <workflow name>

Example

Pmcmd start workflow –service

`${INFA_SERVICE} -domain`

```
 ${INFA_DOMAIN} -uv xxx_PMCMD_ID -pv PSWD -folder  
 ${ETLFolder} -wait ${ETLWorkflow} \
```

What is the main difference between a Joiner Transformation and Union Transformation?

Joiner Transformation merge horizontally

Union Transformation merge vertically

A joiner Transformation is used to join data from heterogeneous database ie (Sql database and flat file) where has Union transformation is used to join data from the same relational sources (oracle table and another Oracle table)

Join Transformation combines data record horizontally based on join condition.

And combine data from two different sources having different metadata.

Join transformation supports heterogeneous, homogeneous data source.

Union Transformation combines data record vertically from multiple sources, having same metadata.

Union transformation also support heterogeneous data source.

Union transformation functions as UNION ALL set operator.

Difference between top down(W.H Inmon)and bottom up(Ralph Kimball)approach?

Top Down approach:-

As per W.H.INWON, first we need to build the Data warehouse after that we need to build up the DataMart but this is so what difficult to maintain the DWH.

Bottom up approach;-

As per Ralph Kimball, first we need to build up the Data Marts then we need to build up the Datawarehouse..

this approach is most useful in real time while creating the Data warehouse.

What are the different caches used in informatica?

- Static cache
- Dynamic cache
- Shared cache
- Persistent cache

What is the command to get the list of files in a directory in Unix?

\$ls -lrt

How to import multiple flat files in to single target where there is no common column in the flat files

in workflow session properties in Mapping tab in properties choose **Source filetype - Indirect**

Give the Source filename : <file_path>

This <file_path> file should contain all the multiple files which you want to Load

How to connect two or more table with single source qualifier?

Create a Oracle source with how much ever column you want and write the join query in SQL query override. But the column order and data type should be same as in the SQL query.

How to call unconnected lookup in expression transformation?

:LKP.LKP_NAME(PORTS)

What is diff between connected and unconnected lookup?

Connected lookup:

It is used to join the two tables

it returns multiple rows

it must be in mapping pipeline

u can implement lookup condition

using connect lookup u can generate sequence numbers by enabling dynamic lookup cache.

Unconnected lookup:

it returns single output through return port

it acts as a lookup function(:lkp)

it is called by another t/r.

not connected either source r target.

CONNECTED LOOKUP:

>> It will participated in data pipeline

>> It contains multiple inputs and multiple outputs.

>> It supported static and dynamic cache.

UNCONNECTED LOOKUP:

>> It will not participated in data pipeline

>> It contains multiple inputs and single output.

>> It supported static cache only.

Types of partitioning in Informatica?

Partition 5 types

1. Simple pass through

2. Key range

- 3. Hash
- 4. Round robin
- 5. Database

Which transformation uses cache?

- 1. Lookup transformation
- 2. Aggregator transformation
- 3. Rank transformation
- 4. Sorter transformation
- 5. Joiner transformation

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

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

How to identify this row for insert and this row for update in dynamic lookup cache?

Based on NEW LOOKUP ROW Informatica server indicates which one is insert and which one is update.

Newlookuprow- 0...no change

Newlookuprow- 1...Insert

Newlookuprow- 2...update

How many ways can we implement SCD2?

- 1) Date range
- 2) Flag
- 3) Versioning

How will you check the bottle necks in informatica? From where do you start checking?

You start as per this order

- 1. Target
- 2. Source
- 3. Mapping
- 4. Session
- 5. System

What is incremental aggregation?

When the aggregator transformation executes all the output data will get stored in the temporary location called aggregator cache. When the next time the mapping runs the aggregator transformation runs for the new records loaded after the first run. These output values will get incremented with the values in the aggregator cache. This is called incremental aggregation. By this way we can improve performance.

Incremental aggregation means applying only the captured changes in the source to aggregate calculations in a session.

When the source changes only incrementally and if we can capture those changes, then we can configure the session to process only those changes. This allows informatica server to update target table incrementally, rather than forcing it to process the entire source and recalculate the same calculations each time you run the session. By doing this obviously the session performance increases.

How can i explain my project architecture in interview..? Tell me your project flow from source to target..?

Project architecture is like

1. **Source Systems:** Like Mainframe, Oracle, People soft, DB2.
2. **Landing tables:** These are tables act like source. Used for easy to access, for backup purpose, as reusable for other mappings.
3. **Staging tables:** From landing tables we extract the data into staging tables after all validations done on the data.
4. **Dimension/Facts:** These are the tables those are used for analysis and make decisions by analysing the data.
5. **Aggregation tables:** These tables have summarized data useful for managers who wants to view monthly wise sales, year wise sales etc.
6. **Reporting layer:** 4 and 5 phases are useful for reporting developers to generate reports. I hope this answer helps you.

What type of transformation is not supported by mapplets?

- Normalizer transformation
- COBOL sources, joiner
- XML source qualifier transformation
- XML sources
- Target definitions
- Pre & Post Session stored procedures
- Other mapplets

How informatica recognizes mapping?

All are organized by Integration service.

Power center talks to Integration Service and Integration service talk to session. Session has mapping Structure. These are flow of Execution.

How to create reusable transformation?

Except source qualifier transformation, all transformations support reusable property.

Reusable transformation developed in two ways.

1. In mapping which transformation do you want to reuse, select the transformation and double click on it, there you got option like make it as reusable transformation option. There you need to check the option for converting non reusable to reusable transformation. but except for source qualifier tr

2. By using transformation developer

What is Pre Sql and Post Sql?

Pre SQL means that the integration service runs SQL commands against the source database before it reads the data from source.

Post SQL means integration service runs SQL commands against target database after it writes to the target.

Insert else update option in which situation we will use?

if the source table contain multiple records .if the record specified in the associated port to insert into lookup cache. it does not find a record in the lookup cache when it is used find the particular record & change the data in the associated port.

We set this property when the lookup TRFM uses dynamic cache and the session property TREAT SOURCE ROWS AS "Insert" has been set.

This option we use when we want to maintain the history.

If records are not available in target table then it inserts the records in to target and records are available in target table then it updates the records.

Select max(sal), deptno from emp where max(sal)>2000 group by deptno this Query is right or wrong?

select max(sal), deptno from emp group by deptno having max(sal)>2000 it is the right one.

What is the diff b/w where and having clause?

where can restrict each row or record

having clause restrict group of records.

Can we use transaction control transformation on a flat file target?

we can't use, Transaction Control transformations connected to any target other than relational, XML, or dynamic MQSeries targets are ineffective for those targets.

Which one given better performance static or dynamic lkp?

Performance wise static lookup is better than dynamic lkp, dynamic Lookup is that they are slow as the caches are updated frequently based on the transaction posted in the database.

When we use dynamic lookup?

if the table contains more no.of duplicate records then we use dynamic lkp to eliminate duplicates.

I have 10 records in my source table i can take filter transformation i can give the condition true then how many records loaded into the target tale ?

10 records, every active transformation by default act as passive transformation.

How to implement Security Measures using Repository manager?

There are 3 ways to implement security measures.

They are:

Folder Permission within owner, groups, and users.

Locking (Read, Write, Retrieve, Save and Execute).

Repository Privileges viz.

Browse Repository.

Use Workflow Manager(To create session and batches and set its properties).

Workflow Operator(To execute Session and batches).

Use Designer, Admin Repository(Allows any user to create and manage repository).

Admin User(Allows the user to create a repository server and set its properties).

Superuser(All the privileges are granted to the user).

What is rank and dense rank in informatica with any examples and give sql query for this both ranks?

for eg: the file contains the records with column

100

200(repeated rows)

200

300

400

500

the rank function gives output as

1

2

2

4

5

6

and dense rank gives

1

2

2

3

4

5

for eg: the file contains the records with column

EmpNo sal

100 1000

200(repeated rows) 2000

200 3000

300 4000

400 5000

500 6000

Rank :

select rank() over (partition by EmpNo order by sal) from emp

1

2

2

4

5

6

Dense Rank

select dense_rank() over (partition by empno order by sal) from emp

and dense rank gives

1

2

2

3

4

5

What is session parameter?

Parameter file is a text file where we can define the values to the parameters .session parameters are used for assign the database connection values

What are different type of repositories that can be created using Informatica Repository Manager?

1. Standalone Repository : A repository which functions individually and is unrelated to any other repositories.
2. Global Repository : This is a centralized repository in a domain. This repository can contain shared objects across the repositories in a domain. The objects are shared through global shortcuts.
3. Local Repository : Local repository is within a domain . Local repository can connect to a global repository using global shortcuts and can use objects in it's shared folders.

What is parameter file?

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

- Mapping parameters
- Mapping variables
- Session parameters

What is session override?

Session override is an option in informatica at session level. Here we can manually give a sql query which is issued to the database when the session runs. It is nothing but over riding the default sql which is generated by a particular transformation at mapping level.

What is the difference between the mapping parameter and variable?

A Mapping Parameter is a static value that you define before running the session and its value remains until the end of the session. When we run the session PowerCenter evaluates the value from the parameter and retains the same value throughout the session. When the session run again it reads from the file for its value.

A Mapping Variable is dynamic or changes anytime during the session. PowerCenter reads the initial value of the variable before the start of the session and changes its value by using variable functions and before ending the session its saves the current value (last value held by the variable). Next time when the session runs the variable value is the last saved value in the previous session.

Variable Port	Mapping Variable
---------------	------------------

1. Local to the T/R	1. Local to the Mapping
2. Values are non-persistent	2. Values are persistent
3. Can't be used with SQL override	3. Can be used with SQL override

- Mapping variables is used for incremental extraction.
- In mapping variables no need to change the data. It automatically changed.
- In mapping parameter you have to change the data and time.

What is threshold error in informatica?

When the target is used by the update strategy DD_REJECT, DD_UPDATE and some limited count, then if the number of rejected records exceed the count then the session ends with failed status. This error is called Threshold Error.

How to delete duplicate record in Informatica?

Following are ways to remove duplicate records

1. In source qualifier use select distinct
2. Use Aggregator and group by all fields
3. Override SQL query in Source qualifier

How To Read Rejected Data Or Bad Data From Bad File And Reload It To Target?

Correction the rejected data and send to target relational tables using load order utility.

Find out the rejected data by using column indicator and row indicator.

What is meant by inline view?

inline view is a SELECT statement in the FROM-clause of another SELECT statement.

In-line views are commonly used to simplify complex queries by removing join operations and condensing several separate queries into a single query.

What is conformed dimension?

A dimension table can be shared by more than one fact table and is known as conformed dimension.

ex: date dimension.

What is fact less fact table?

factless fact table captures many-to-many relationships between dimensions, and it doesn't contain any numeric values.

ex: students attendance tracking.
store that did not sell a product over a period of time.

I have 3 flat files with same structure how to join without using joiner transformation?
by using file list concept with file type as indirect .

How to delete the duplicate records in unix?

sort filename.txt | uniq -u

What is surrogate key?

surrogate key is a system generated sequence number to be used as a primary key.

What is diff b/w surrogate key and primary key?

surrogate key	primary key
1.artificialkey	1.naturalkey
2.generated by system sequence number	2.generated by database

What is subquery & corelated subquery?

subquery:-A query nested inside a SELECT statement is known as a subquery and is an alternative to complex join statements.

Corelated Subquery:-It is a sub-query (a query nested inside another query) that uses values from the outer query in its WHERE clause. The sub-query is evaluated once for each row processed by the outer query.

How to eliminate duplicate records in informatica and in sql?

in informatica by using distinct property in source qualifier or by using sorter t/r.

in sql delete from emp where rowid not in(select max(rowid) from emp group by empno);

What is a candidate key?

A candidate key is a combination of attributes that can be uniquely used to identify a database record without any extraneous data (unique). Each table may have one or more candidate keys. One of these candidate keys is selected as the table primary key else are called Alternate Key.

What is the difference between Bitmap and B tree index?

Bitmap index is used for repeating values.

ex: Gender: male/female

Account status: Active/Inactive

Btree index is used for unique values.

ex: empid.

What is Throughput in Informatica?

Throughput is the rate at which power centre server read the rows in bytes from source or writes the rows in bytes into the target per second.

You can find this option in workflow monitor. Right click on session choose properties and **Source/Target Statistics** tab you can find throughput details for each instance of source and target.

What are set operators in Oracle

UNION

UNION ALL

MINUS

INTERSECT

What is a candidate key?

A candidate key is a combination of attributes that can be uniquely used to identify a database record without any extraneous data (unique). Each table may have one or more candidate keys. One of these candidate keys is selected as the table primary key else are called Alternate Key.

Can anyone tell me the difference between persistence and dynamic caches? On which conditions we are using these caches?

Dynamic

- 1)When you use a dynamic cache, the Informatica Server updates the lookup cache as it passes rows to the target.
- 2)In Dynamic, we can update catch will New data also.
- 3) Dynamic cache, Not Reusable
(when we need Updated cache data, That only we need Dynamic Cache)

Persistent

- 1)a Lookup transformation to use a non-persistent or persistent cache. The PowerCenter Server saves or deletes lookup cache files after a successful session based on the Lookup Cache Persistent property.
- 2) Persistent, we are not able to update the catch with New data.
- 3) Persistent catch is Reusable.
(When we need Previous Cache data, That only we need Persistent Cache)

1. Dynamic lookup allows modifying cache where as Persistent lookup does not allow us to modify cache.

2. Dynamic lookup use 'newlookup row', a default port in the cache but persistent does use any default ports in cache.
3. As session completes dynamic cache removed but the persistent cache saved in informatica power centre server.

How to obtain performance data for individual transformations?

There is a property at session level "**Collect Performance Data**", you can select that property. It gives you performance details for all the transformations.

What will happen if we connect only current value port from seq generator to next transformation (without connecting nextval)?

Each target will get the value 1.

What Is Data Movement Mode In Informatica And Difference Between Them?

The data movement mode depends on whether Informatica server will process a single byte or multi-byte character data set.

There are two data movement modes in Informatica:

1. **Unicode** - IS allows 2 bytes for each character and uses additional byte for each non-ascii character (such as Japanese characters)
2. **ASCII** - IS holds all data in a single byte. The IS data movement mode can be changed in the Informatica Server configuration parameters. This comes into effect once you restart the Informatica Server.

Eliminating of duplicate records without using dynamic lookups?

Hi U can eliminate duplicate records by an simple one line SQL Query.

Select id, count (*) from seq1 group by id having count (*)>1;

Below are the ways to eliminate the duplicate records:

1. **By enabling the option in Source Qualifier transformation as select distinct.**
2. **By enabling the option in sorter transformation as select distinct.**
3. **By enabling all the values as group by in Aggregator transformation.**

Can anyone give idea on how do we perform test load in informatica? What do we test as part of test load in informatica?

With a test load, the Informatica Server reads and transforms data without writing to targets. The Informatica Server does everything, as if running the full session. The Informatica Server writes data to relational targets, but rolls back the data when the session completes. So, you can enable collect performance details property and analyse the how efficient your mapping is. If the session is running for a long time, you may like to find out the bottlenecks that are existing. It may be bottleneck of type target, source, mapping etc.

The basic idea behind test load is to see the behaviour of Informatica Server with your session.

What are the transformations that cannot be placed between the sort origin and the Joiner transformation so that we do not lose the input sort order.

The best option is to place the Joiner transformation directly after the sort origin to maintain sorted data. However do not place any of the following transformations between the sort origin and the Joiner transformation:

- Custom
- **Unsorted**Aggregator
- Normalizer
- Rank
- Union transformation
- sXML Parser transformation
- XML Generator transformation
- Mapplet [if it contains any one of the above mentioned transformations]

What Is A Shortcut And Copy In Informatica And How Two Are Different With Each Other?

Copy	Shortcut
1. Copy an object to another folder	1. Dynamic link to an object in the folder
2. Changes to original object doesn't reflect	2. Dynamically reflects the changes to an original object
3. Duplicate's the space	3. Preserves the space
4. Created from unshared folders	4. Created from shared folders

What are business components in Informatica?

- Domains
- Nodes
- Services

WHAT IS VERSIONING?

It's used to keep history of changes done on the mappings and workflows

1. **Check in:** You check in when you are done with your changes so that everyone can see those changes.
2. **Check out:** You check out from the main stream when you want to make any change to the mapping/workflow.
3. **Version history:** It will show you all the changes made and who made it.

Difference between \$\$\$sessstarttime and sessstarttime?

\$\$\$\$SessStartTime - Returns session start time as a string value (String datatype)
SESSSTARTTIME - Returns the date along with date timestamp (Date datatype)

Difference between \$,\$,\$,\$\$ in Informatica?

1. **\$ Refers**

These are the system variables/Session Parameters like \$Bad file, \$input file, \$output file, \$DB connection, \$source, \$target etc..

2. **\$\$ Refers**

User defined variables/Mapping Parameters like \$\$State,\$\$Time, \$\$Entity, \$\$Business_Date, \$\$SRC,etc.

3. **\$\$\$ Refers**

System Parameters like \$\$\$SessStartTime

Finding Duplicate Rows based on Multiple Columns?

```
SELECT firstname, COUNT(firstname), surname, COUNT(surname), email, COUNT(email)
FROM employee
GROUP BY firstname, surname, email
HAVING (COUNT(firstname) > 1) AND (COUNT(surname) > 1) AND (COUNT(email) > 1);
```

Finding Nth Highest Salary in Oracle?

Pick out the Nth highest salary, say the 4th highest salary.

Select * from

```
(select ename, sal, dense_rank() over (order by sal desc) emp_rank from emp)
where emp_rank=4;
```

Find out the third highest salary?

```
SELECT MIN(sal) FROM emp WHERE
sal IN (SELECT distinct TOP 3 sal FROM emp ORDER BY sal DESC);
```

How do you handle error logic in Informatica? What are the transformations that you used while handling errors? How did you reload those error records in target?

Row indicator: It generally happens when working with update strategy transformation. The writer/target rejects the rows going to the target

Column indicator:

D -Valid

o - Overflow

n - Null

t - Truncate

When the data is with nulls, or overflow it will be rejected to write the data to the target

The reject data is stored on reject files. You can check the data and reload the data in to the target using reject reload utility.

Difference between STOP and ABORT?

Stop - If the Integration Service is executing a Session task when you issue the stop command, the Integration Service stops reading data. It continues processing and writing data and committing data to targets. If the Integration Service cannot finish processing and committing data, you can issue the abort command.

Abort - The Integration Service handles the abort command for the Session task like the stop command, except it has a timeout period of 60 seconds. If the Integration Service cannot finish processing and committing data within the timeout period, it kills the DTM process and terminates the session.

WHAT IS INLINE VIEW?

An inline view is term given to sub query in FROM clause of query which can be used as table. Inline view effectively is a named sub query

Ex : Select Tab1.col1,Tab1.col.2,Invview.col1,Invview.Col2

From Tab1, (Select statement) In view

Where Tab1.col1=Invview.col1

SELECT DNAME, ENAME, SAL FROM EMP ,
(SELECT DNAME, DEPTNO FROM DEPT) D

WHERE A.DEPTNO = B.DEPTNO

In the above query (SELECT DNAME, DEPTNO FROM DEPT) D is the inline view.

Inline views are determined at runtime, and in contrast to normal view they are not stored in the data dictionary,

Disadvantage of using this is

1. Separate view need to be created which is an overhead
2. Extra time taken in parsing of view

This problem is solved by inline view by using select statement in sub query and using that as table.

Advantage of using inline views:

1. Better query performance
2. Better visibility of code

Practical use of Inline views:

1. Joining Grouped data with non-grouped data
2. Getting data to use in another query

WHAT IS GENERATED KEY AND GENERATED COLUMN ID IN NORMALIZER TRANSFORMATION?

- The integration service increments the generated key (**GK**) sequence number each time it processes a source row. When the source row contains a multiple-occurring column or a multiple-occurring group of columns, the normalizer transformation returns a row for each occurrence. Each row contains the same generated key value.
- The normalizer transformation has a generated column ID (**GCID**) port for each multiple-occurring column. The GCID is an index for the instance of the multiple-occurring data. For example, if a column occurs 3 times in a source record, the normalizer returns a value of 1, 2 or 3 in the generated column ID.

DIFFERENCE BETWEEN SURROGATE KEY AND PRIMARY KEY?

Surrogate key:

1. Query processing is fast.
2. It is only numeric
3. Developer develops the surrogate key using sequence generator transformation.
4. Eg: 12453

Primary key:

1. Query processing is slow
2. Can be alpha numeric
3. Source system gives the primary key.
4. Eg: C10999

How does the server recognize the source and target databases?

If it is relational - By using ODBC connection

FTP connection - By using flat file

What is the use of Shared Folder?

Shared folders allow users to create shortcuts to objects in the folder. If you have an object that you want to use in several mappings or across multiple folders, we can place the object in a shared folder. We can then access the object from other folders by creating a shortcut to the object. Shortcuts inherit changes to their shared object. Shared folders in global repositories can be used by any folder in the domain. Once you make a folder shared, you cannot reverse it.

WHAT ARE ALL THE TRANSFORMATION YOU USED IF SOURCE AS XML FILE?

- XML Source Qualifier
- XML Parser
- XML Generator

List the files in ascending order in UNIX?

ls -lt (sort by last date modified)

ls -ltr (reverse)

ls -IS (sort by size of the file)

What are the different ways to migrate from one environment to another in Informatica?

1. We can export repository and import into the new environment

2. We can use informatica deployment groups

3. We can Copy folders/objects

4. We can Export each mapping to xml and import in new environment

How would you join a node to the already existing domain?

Joining a node to an existing domain can be archived while installing Informatica.

Eg: Server is on machine1 and the new node is on machine 2. Then go to machine 2-Open Installer-> join node to the domain while configuring domain and provide the existing domain details.

How to display the first 10 lines of a file?

```
# head -10 logfile
```

How to display the last 10 lines of a file?

```
# tail -10 logfile
```

How did you schedule sessions in your project?

1. **Run once** – Set 2 parameter date and time when session should start.
2. **Run Every** – Informatica server run session at regular interval as we configured, parameter Days, hour, minutes, end on, end after, forever.
3. **Customized repeat** – Repeat every 2 days, daily frequency hr, min, every week, every month.

What is Load Order?

Design mapping applications that first loads the data into the dimension tables. And then load the data into the fact table.

- **Load Rule:-** If all dimension table loadings are success then load the data into fact table.
- **Load Frequency:-** Database gets refreshed on daily loads, weekly loads and monthly loads.

How to get sequence value using Expression?

```
v_temp = v_temp+1
```

```
o_seq = IIF(ISNULL(v_temp), 0, v_temp)
```

What Is Primary And Backup Node?

The primary node is a node where a current node or applications are running and the backup node is defined to run the domain or applications services while the primary node goes down

What is difference between grep and find?5ds

Grep is used for finding any string in the file.

Syntax - grep <String> <filename>

Example - grep 'compu' details.txt

Display the whole line, in which line compu string is found.

Find is used to find the file or directory in given path,

Syntax - find <filename>

Example - find compu*

Display all file names starting with computer

WHAT ARE THE DIFFERENCE BETWEEN DDL, DML AND DCL COMMANDS?

DDL is Data Definition Language statements

- CREATE – to create objects in the database
- ALTER – alters the structure of the database
- DROP – delete objects from the database
- TRUNCATE – remove all records from a table, including all spaces allocated for the records are removed
- COMMENT – add comments to the data dictionary
- GRANT – gives user's access privileges to database
- REVOKE – withdraw access privileges given with the GRANT command

DML is Data Manipulation Language statements

- SELECT – retrieve data from the a database
- INSERT – insert data into a table
- UPDATE – updates existing data within a table
- DELETE – deletes all records from a table, the space for the records remain
- CALL – call a PL/SQL or Java subprogram
- EXPLAIN PLAN – explain access path to data
- LOCK TABLE – control concurrency

DCL is Data Control Language statements

- COMMIT – save work done
- SAVEPOINT – identify a point in a transaction to which you can later roll back

- ROLLBACK – restore database to original since the last COMMIT
- SET TRANSACTION – Change transaction options like what rollback segment to use

What do you understand by SOA of Informatica?

SOA or service-oriented architecture allows the Informatica to have services on a distributed system. You can define a domain on a node say node1 and repository service on node2, repository service on node3. All nodes can be anywhere within a network. Once a request is made from the client it goes to the domain or the gateway node and then it directs the request to the appropriate nodes or services.

What is View?

A view is a tailored presentation of the data contained in one or more tables (or other views). 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 the view references. These tables are called base tables.

Views present a different representation of the data that resides within the base tables. Views are very powerful because they allow you to tailor the presentation of data to different types of users.

Views are often used to:

- Provide an additional level of table security by restricting access to a predetermined set of rows and/or columns of a table
- Hide data complexity
- Simplify commands for the user
- Present the data in a different perspective from that of the base table
- Isolate applications from changes in definitions of base tables
- Express a query that cannot be expressed without using a view

What is Trigger?

A trigger is a SQL procedure that initiates an action when an event (INSERT, DELETE or UPDATE) occurs.

Nested Trigger: A trigger that contains data modification logic within itself is called a nested trigger.

What is Stand alone Email task?

1. It can be used anywhere in the workflow, defined will Link conditions to notify the success or failure of prior tasks.
2. Visible in Flow Diagram.
3. Email Variables can be defined with stand alone email tasks.

Can we create two repositories on the same database instance in oracle?

It is not possible to create two repositories on the same instance of the database in oracle as it will be creating the metadata tables for the repository. However, it is possible to create two repositories in oracle on the same server if the database instance is different.

What is the difference between a HAVING CLAUSE and a WHERE CLAUSE?

1. Specifies a search condition for a group or an aggregate. HAVING can be used only with the SELECT statement.
2. HAVING is typically used in a GROUP BY clause. When GROUP BY is not used, HAVING behaves like a WHERE clause.
3. Having Clause is basically used only with the GROUP BY function in a query. WHERE Clause is applied to each row before they are part of the GROUP BY function in a query.

What is RANK CACHE?

Sample Rank Mapping

When the Power Center Server runs a session with a Rank transformation, it compares an input row with rows in the data cache. If the input row out-ranks a Stored row, the Power Center Server replaces the stored row with the input row.

Example: Power Center caches the first 5 rows if we are finding top 5 salaried Employees. When 6th row is read, it compares it with 5 rows in cache and places it in Cache if needed.

1) RANK INDEX CACHE:

The index cache holds group information from the **group by ports**. If we are Using Group By on DEPTNO, then this cache stores values 10, 20, 30 etc.

All Group By Columns are in RANK INDEX CACHE. Ex. DEPTNO

2) RANK DATA CACHE:

It holds row data until the Power Center Server completes the ranking and is generally larger than the index cache. To reduce the data cache size, connect only the necessary input/output ports to subsequent transformations.

All Variable ports if there, Rank Port, All ports going out from RANK Transformations are stored in RANK DATA CACHE.

Example: All ports except DEPTNO In our mapping example.

What is Aggregator Caches?

1. The Power Center Server stores data in the aggregate cache until it completes Aggregate calculations.
2. It stores group values in an index cache and row data in the data cache. If the Power Center Server requires more space, it stores overflow values in cache files.

Note: The Power Center Server uses memory to process an Aggregator transformation with sorted ports. It does not use cache memory. We do not need to configure cache memory for Aggregator transformations that use sorted ports.

1) Aggregator Index Cache:

The index cache holds group information from the **group by ports**. If we are using Group By on DEPTNO, then this cache stores values 10, 20, 30 etc.

- All Group By Columns are in AGGREGATOR INDEX CACHE. Ex. DEPTNO

2) Aggregator Data Cache:

DATA CACHE is generally larger than the AGGREGATOR INDEX CACHE.

Columns in Data Cache:

- Variable ports if any
- Non group by input/output ports.
- Non group by input ports used in non-aggregate output expression.
- Port containing aggregate function

What is JOINER CACHE?

Joiner always caches the MASTER table. We cannot disable caching. It builds Index cache and Data Cache based on MASTER table.

1) Joiner Index Cache:

All Columns of MASTER table used in Join condition are in JOINER INDEX CACHE.

Example: DEPTNO in our mapping.

2) Joiner Data Cache:

Master column not in join condition and used for output to other transformation or target table are in Data Cache.

Example: DNAME and LOC in our mapping example.

Lookup Cache Files

1. Lookup Index Cache:

Stores data for the columns used in the lookup condition.

2. Lookup Data Cache:

- For a connected Lookup transformation, stores data for the connected output ports, not including ports used in the lookup condition.
- For an unconnected Lookup transformation, stores data from the return port.

How Do You Take Back-Up Of Repository?

Login to the admin console. Select the repository-> Action->Backup Content

The backup content will be stored in an XML file and the location of the backup directory is defined in the domain configuration.

Also, you can create a script and schedule it to back up the content on a certain date and time.

In script first, connect to the repository using pmrep command and then back up the content with the below command.

What is Mapping Debugger?

- Debugger is a tool. By using this we can identify records are loaded or not and correct data is loaded or not from one T/R to other T/R.
- Session succeeded but records are not loaded. In this situation we have to use Debugger tool.

What Do You Mean By Parameter File? Why Do We Use It And What All Things We Can Define In A Parameter Files?

Parameter file is a file assigned to a session which has configuration details in it. The parameter file is used for reusability. Below things are defined in it.

Connections, mapping parameter, mapping variable.

What is an Integration service in Informatica?

Integrations service is an application service in Informatica which run the workflows and session. It tells the Informatica server how to flow the data from source to target.

Analogy: Integration service is like a carrier truck which carries the goods (data) from one place to another.

What is the difference between a gateway node and worker node?

A gateway node is a node on which domain is defined and it acts as a router or administrator to all services. All other nodes in a domain are worker nodes where application services like Informatica services, repository services will be running.

What Do You Mean By Load Balancer?

Load Balancer: The Integration Service uses the Load Balancer to dispatch tasks. The Load Balancer dispatches tasks to achieve optimal performance. It may dispatch tasks to a single node or across the nodes in a grid.

What is DTM

DTM (Data Transformation Manager) is the process associated with the session task. The Load Manager creates one DTM process for each session in the workflow. The DTM process performs the following tasks:

1. Reads session information from the repository.
2. Expands the server, session, and mapping variables and parameters.
3. Creates the session log file. Validates source and target code pages.
4. Verifies connection object permissions.

- 5.Runs pre-session shell commands, stored procedures and SQL.
- 6Creates and runs mapping, reader, writer, and transformation threads to extract, transform, and load data.
7. Runs post-session stored procedures, SQL, and shell commands. Sends post-session email.

What is resilience time?

Resilience time is the time defined in Informatica to wait before it loses the connection on-network failure.

Eg: If your resilience time is 180 sec and the network goes down so Informatica will wait for 180 sec and if the network is established in between, the connection will be established. If the network is not up within 180 sec so the unsaved data will be lost and the client will be disconnected from the domain or repository.

What Is Deployment Group?

A deployment group is a global object that consists of versioned objects from one or more folders. You use a deployment group to copy the versioned objects to another folder or repository. You can use a deployment group when you want to copy some, but not all, of the objects in a folder. You can also use a deployment group to copy objects from multiple folders.

Pmrep and pmcmd command?

pmrep and pmcmd commands are the command-line interfaces to perform Informatica related task. These commands are used mostly in the form of scripts. Pmrep has used for all repository related tasks. Pmcmd is used for Informatica monitor and server related tasks

Eg: pmrep connect -r Repository_name -d Domain_name -n Username -x password

The above command connects to the repository. Once connected you can create a folder in the repository by below command

Pmrep create a folder -f Foldername -o OwnerName company

Some of the pmcmd commands:

I. Start workflow

II. Stop workflow

III. Wait workflow

IV. Recover workflow

V. Schedule workflow

VI. Start task

VII. Stop task

What Could Be The Possible Reasons Of Locks By User?

Network problems occur.

A PowerCenter Client, Integration Service, Repository Service, or database machine shuts down improperly.

How to use PowerCenter Command Line in Informatica?

The transformation language provides two comment specifiers to let you insert comments in expression:

- Two Dashes (--)
- Two Slashes (//)

The Power center integration service ignores all text on a line preceded by these two comment specifiers.

Can We Create A Integration Service Without A Repository Service And Vice Versa?

It is not possible to create an Integration service without a repository service however a repository can be created without an integration service.

What will happen if we connect only current value port from seq generator to next transformation (without connecting nextval)

Ans Each target will get the value 1.

What is a Surrogate Key?

A surrogate key is a sequentially generated integer value which is used as another substitute or replacement for the primary key which is required as a unique identification of each row in a table.

The primary key can be changed frequently as per the need which makes the update process more difficult for a future requirement, Surrogate key is the only solution for this problem.

Difference between Informatica Vs Talend ?

Informatica	Talend
Provides only commercial data integration	Available open source and commercial editions
Founded way back in 1993	Founded in the year 2006
Charges applicable per customer	Open source is for free

RDBMS repository stores metadata generated	Implemented on any java supported platforms
Integrating code is not so effective	Code customization is effective
No prior knowledge is required	knowledge on java is preferred
Automated deployment is not up to the mark	Deployment made easy
Transformations are re-usable	Components are re-usable

What is a Session?

A session in Informatica is a set of instructions to be followed when data is being transferred from source to target using Session Command. A Session Command can be a pre-session command or a post-session command.

How do pre- and post-session shell commands function?

A command task can be called as a pre- or post-session shell command for a session task. Users can run it as a pre-session command, a post-session success command, or a post-session failure command. Based on use cases, the application of shell commands can be changed or altered.

How to update Source Definition?

There are two ways to update the source definition.

They are:

- You can edit the existing source definition.
- You can import a new source from the database.

What is a User-Defined Event?

A User-Defined Event is a flow of tasks in a workflow. It allows users to create and name an event.

These are some of the popular questions asked during an Informatica interviews. Always be prepared to answer all types of questions — technical skills, interpersonal, leadership or methodology. If you have recently started your career in a data profile, you can always get certified to understand the industry-related terminology, skills and methodologies.

What is the Event and what are the tasks related to it?

The event can be any action or function that occurs in the workflow.

There are two tasks related to it, which includes:

- **Event Wait Task:** This task waits until an event occurs, once the event is triggered this task gets accomplished and assigns the next task.
- **Events Raise Task:** Event Raise task triggers the specific event in the workflow.

What is a pre-defined event and User-defined event?

Predefined events are system-defined events that wait until the arrival of a specific file at a specific location. It is also called as File-Watcher event.

User-Defined events are created by the user to raise anytime in the workflow once created.

Explain different Components under Workflow Manager Tool?

Components under Workflow Manager Tool

1. Workflow: This is top level object and the entire task (process) for data loading has to be defined under the workflow. It is like a Mapping that integrates different kind of tasks as a Unit.



2. Task: A task is an individual process to perform a very specific activity during data loading. There are 10 different kinds of tasks that can be grouped under a Workflow:

2.1 Session:

This is a compulsory task for data loading.

A session is an instance of Mapping Program or in other words a running instance of a mapping is referred as Session. For one Mapping Program we can create one or more Sessions. Generally we require one session for one mapping but for Parallel data loading we may create multiple sessions.

2.2 Command:

To execute operating System Commands or Programs. For example: If we need to inform all the users about data loading process, we can write Shell Script at Operating System and execute them via 'Command Task' just before the session execution.

2.3 Email:

To send emails to users using Mail Server (if configured). This job can be done via Command Task also but Email Task is integrated part of Workflow Manager and is much simple compared to Command Task.

2.4 Decision:

This task is used to evaluate condition based on other tasks' values to decide next course of actions. It is like IF statement.

2.5 Control:

This task is to control the flow of tasks within the Workflow. For example: If we need that control should not reach to specific task (like a Command Task) when a condition fails then we can use Control Task.

2.6 Event Wait:

This task is used to define an event and when the particular event fires (activates) then process continues.

2.7 Event Raise:

This task is used to fire (activate) an event forcefully.

2.8 Assignment:

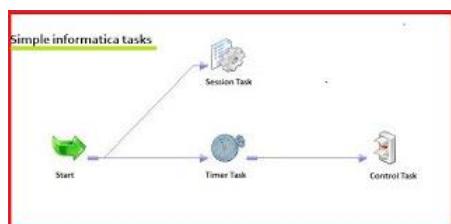
This task is used to assign values to Parameters and Variables used within workflow.

2.9 Timer:

This task is used to specify time of execution (delay) for a task.

2.10 Worklet:

This task is to define reusable Workflow. If we need to execute set of tasks again and again under different Workflows then it is better to define them as Worklet and use under different Workflows.



Note:

Three types of tasks (Session, Command and Email) can be defined as Reusable Tasks. Reusable task means a task is created as an Independent Task and it is used within WorkFlow or Worklet. So if a task is created within workflow or worklet directly then it is non-reusable task.

So if we need a task to be executed once within a Workflow or Worklet then create it as non-reusable task otherwise create them as reusable (independent) task.

Worklet Task also can be defined as Reusable Task via separate Menu Interface.

What is ETL?

ETL stands for extract, transform, and load. It can consolidate the scattered data for any organization while working with different departments. It can very well handle the data coming from different departments.

What is Datamart?

A Data Mart is a subset of a data warehouse that can provide data for reporting and analysis on a section, unit or department like Sales Dept, HR Dept, etc. The Data Mart is sometimes also called as HPQS (Higher Performance Query Structure).

What is a Power Center Repository?

The Power Center Repository allows you to share metadata across repositories to create a data mart domain. In a data mart domain, you can create a single global repository to store metadata used across an enterprise and several local repositories to share the global metadata as needed.

How will you define Tracing Level?

Tracing Level refers to the amount of information that the server writes in the session log. Tracing Level is created and configured either at –

- The transformation level
- The session-level
- Else at both the levels

Different types of Tracing Level are –

- None
- Terse
- Verbose Initialization

- Verbose Data

What is the difference between Data Warehousing and Data Mining?

Data mining, the operational information is investigated utilizing analytical procedures and clustering methods to detect unknown designs and trends. So, the data works do an amazing variety of summarization of the data and can be applied by data repositories for active analytical processing for marketing intellect.

The data repository may produce the utility of a data mine for penetrating the processing of the data in a faster process.

Enlist the differences between Database and Data Warehouse?

Database	Data Warehouse
It stores/records current and up to date which is used in daily operations	It stores/analyze historical data which is used for information support on a long-term basis.
Its orientation is on Online Transactional processing which includes simple and short transactions.	Its orientation is on Online Analytical Processing which includes complex queries.
It consists of detailed and primitive data where its view is flat relational.	It consists of summarized a consolidated data where its view is multidimensional.
Low performance is observed for Analytical queries.	Analytical queries are judged here as high performance.
Efficiency is determined by measuring transaction throughput.	Efficiency is determined by measuring query throughput and response time.

What happens to a mapping if we alter the datatypes between Source and its corresponding Source Qualifier?

The Source Qualifier transformation displays the transformation datatypes. The transformation datatypes determine how the source database binds data when the Integration Service reads it.

Now if we alter the datatypes in the Source Qualifier transformation or the **datatypes in the source definition and Source Qualifier transformation do not match**, the Designer marks the **mapping as invalid** when we save it.

Suppose we have used the Select Distinct and the Number Of Sorted Ports property in the SQ and then we add Custom SQL Query. Explain what will happen?

Whenever we add Custom SQL or SQL override query it **overrides** the User-Defined Join, Source Filter, Number of Sorted Ports, and Select Distinct settings in the Source Qualifier transformation. Hence only the user defined SQL Query will be fired in the database and all the **other options will be ignored**.

Describe the situations where we will use the Source Filter, Select Distinct and Number Of Sorted Ports properties of Source Qualifier transformation?

Source Filter option is used basically to reduce the number of rows the Integration Service queries so as to improve performance.

Select Distinct option is used when we want the Integration Service to select unique values from a source, filtering out unnecessary data earlier in the data flow, which might improve performance.

Number Of Sorted Ports option is used when we want the source data to be in a sorted fashion so as to use the same in some following transformations like Aggregator or Joiner, those when configured for sorted input will improve the performance.

What will happen if the SELECT list COLUMNS in the Custom override SQL Query and the OUTPUT PORTS order in SQ transformation do not match?

Mismatch or Changing the order of the list of selected columns to that of the connected transformation output ports may result in **session failure**.

What happens if in the Source Filter property of SQ transformation we include keyword WHERE say, WHERE CUSTOMERS.CUSTOMER_ID > 1000?

We use source filter to reduce the number of source records. If we include the string **WHERE** in the source filter, the Integration Service **fails the session**.

Describe the scenarios where we go for Joiner transformation instead of Source Qualifier transformation?

While joining Source Data of **heterogeneous sources** as well as to join **flat files** we will use the Joiner transformation. Use the Joiner transformation when we need to join the following types of sources:

- Join data from different Relational Databases.
- Join data from different Flat Files.
- Join relational sources and flat files.

What is the maximum number we can use in Number Of Sorted Ports for Sybase source system?

Sybase supports a maximum of **16** columns in an ORDER BY clause. So if the source is Sybase, do not sort more than 16 columns.

Define Workflow?

Workflow is a set of multiple tasks that enable a server to communicate and get the tasks implemented. These tasks are connected with start task link and trigger the required sequence to start a process.

What is difference between direct and indirect loading options in sessions?

Use file type direct when we are loading single file into Target. Use Indirect when we want to load multiple files through single session in the mapping.

When we create a target as flat file and source as oracle. How can we have first rows as column names in flat files?

We can add a union all clause in the Source Qualifier of the Relational source to add the header values in SELECT clause, so as to populate them as header.

How to capture the user information (Username) of the person, who is executing a workflow or session in workflow manager?

```
SELECT USER_NAME, WORKFLOW_NAME FROM REP_WFLOW_RUN;
```

How to join two tables, which reside in two different databases in the Source qualifier?

By using DB links in SQ ,but throughput will be very slow. Joiner will be better approach in this case.

What is Workflow? What are the components of the Workflow Manager?

Workflow is the way of a manner in which the task should be implemented. It is a collection of instructions that inform the server about how to implement the task.

Given below are the three major components of the Workflow Manager:

- Task Designer
- Task Developer
- Workflow Designer

What Is The Target Load Order?

A target load order group is the collection of source qualifiers, transformations, and targets linked together in a mapping.

You specify the target load order based on source qualifiers in a mapping. If you have the multiple source qualifiers connected to the multiple targets, you can designate the order in which Informatica server loads data into the targets.

Is there any way to handle sequence generation over 2 billion records, As sequence generator can generate max 2 billion records?

Method1

it will be done by unconnected lookup create lookup transformation on target just keep one field(id_field) and remove all remaining fields. Create i/p field with same data type as out field(id_field) select return port for id_field in look up override write a query as select nvl(max(id_field),0) from table and ' lookup condition is id_field>=in_field in exp transformation v_dummy(v)-->it always to be zero v_lkp_return iif(:lkp_target(v_dummy) --here u can pass v_dummy value or just zero o_id_field(0).v_lkp_return+1 _____._____.

Method 2

One way would be to use expression trans after sequence generator. Another way could be to use expression instead of sequence generator. Here you will use variable of agg type of count to create sequences.

What is difference between \$ and \$\$ parameters/variables

\$ are System defined and \$\$ are User defined variables

How to use data from SAP/Oracle Apps /Mainframe

Through power exchange.

What is an indicator file and how it can be used?

Indicator file is used for Event Based Scheduling when you don't know when the Source Data is available., A shell command ,script or a batch file creates and send this indicator file to the directory local to the Informatica Server. Server waits for the indicator file to appear before running the session.

What is the Session task?

Session task is a set of instructions that are to be applied while transferring data from source to target using session command. Session command can be either pre-session command or post-session command.

State the limitations where we cannot use Joiner in the mapping pipeline?

The Joiner transformation accepts input from most transformations. However, following are the limitations:

- Joiner transformation cannot be used when either of the input pipeline contains an **Update Strategy** transformation.
- Joiner transformation cannot be used if we connect a **Sequence Generator** transformation directly before the Joiner transformation.

What is a Command Task?

A Command Task is used to run the shell/UNIX commands in Windows during the workflow. It allows a user to specify UNIX commands in the command task to remove rejected files, create files, copy files, rename files and archive files, among others.

How does Joiner transformation treat NULL value matching?

The Joiner transformation **does not match null values**.

For example, if both EMP_ID1 and EMP_ID2 contain a row with a null value, the Integration Service does not consider them a match and does not join the two rows.

To join rows with null values, replace null input with **default values** in the Ports tab of the joiner, and then join on the default values.

Note: If a result set includes fields that do not contain data in either of the sources, the Joiner transformation populates the empty fields with null values. If we know that a field will return a NULL and we do not want to insert NULLs in the target, set a default value on the Ports tab for the corresponding port.

What is the Standalone command task?

The standalone command task can be used to run Shell command anywhere and anytime in the workflow.

Scenarios:

How to generate sequence / incremental numbers in Informatica?

Solution 1: In the expression transformation, create a variable port and increment it by 1.

Then assign the variable port to an output port.

In the expression transformation, the ports are:

V_COUNT=V_COUNT+1

O_COUNT=V_COUNT

Solution 2: Insert a sequence generator transformation drag NEXTVAL port from sequence generator to expression. This port will give you the incremental numbers.

Use **Start Value** property to 1 and **Increment By** property to 1 for a series like

1,2,3,4,5,6,7,8.....

Design a mapping to load the first 3 rows from a flat file into a target?

Solution:

You have to assign row numbers to each record. Generate the row numbers either using the expression transformation as mentioned above or use sequence generator transformation. Then pass the output to filter transformation and specify the filter condition as O_count <=3

What is the syntax of the INITCAP function?

This function is used to capitalize the first character of each word in the string and makes all other characters in lowercase.

Below is the Syntax:

INITCAP(string_name)

Design a mapping to load the last 3 rows from a flat file into the target?

Suppose the flat file in consideration has below data:

Column A

Aanchal

Priya

Karishma

Snehal

Nupura

Step1: Assign row numbers to each record. Generate row numbers using expression transformation by creating a variable port and incrementing it by 1. After this assign this variable port to output port. After expression transformation, the ports will be as –

Variable_count= Variable_count+1

O_count=Variable_count

Create a dummy output port for same expression transformation and assign 1 to that port.

This dummy port will always return 1 for each row.

Finally, the transformation expression will be as follows:

Variable_count= Variable_count+1

O_count=Variable_count

Dummy_output=1

The output of this transformation will be :

Column A O_count Dummy_output

Aanchal 1 1

Priya 2 1
Karishma 3 1
Snehal 4 1
Nupura 5 1

Step 2: Pass the above output to an aggregator and do not specify any group by condition. A new output port should be created as O_total_records in the aggregator and assign O_count port to it. The aggregator will return the last row. The final output of this step will have dummy port with value as 1 and O_total_records will have a total number of records in the source. The aggregator output will be: O_total_records, Dummy_output

5 1

Step 3: Pass this output to joiner transformation and apply a join on dummy port. The property sorted input should be checked in joiner transformation. Only then the user can connect both expression and aggregator transformation to joiner transformation. Joiner transformation condition will be as follows:

Dummy_output (port from aggregator transformation) = Dummy_output (port from expression transformation)

The output of joiner transformation will be

Column A O_count O_total_records

Aanchal 1 5

Priya 2 5

Karishma 3 5

Snehal 4 5

Nupura 5 5

Step 4: After the joiner transformation we can send this output to filter transformation and specify filter condition as O_total_records (port from aggregator)-O_count(port from expression) <=2

The filter condition, as a result, will be

O_total_records – O_count <=2

The final output of filter transformation will be :

Column A O_count O_total_records

Karishma 3 5

Snehal 4 5

Nupura 5 5

What is Workflow Monitor?

Workflow Monitor is used to monitor the execution of workflows or the tasks available in the workflow. It is mainly used to monitor the progress of activities such as Event log information, a list of executed workflows, and their execution time.

Workflow Monitor can be used to perform the following activities:

- You can see the details of execution
- You can see the history of workflow execution
- You can stop, abort, or restart the workflows.
- It displays the workflows that have been executed at least once.

It consists of the following windows:

- **Navigator window:** It displays the repositories, servers, and repositories objects that have been monitored.
- **Output window:** It displays messages coming from the Integration service and Repository service.
- **Time window:** It displays the progress of workflow execution.
- **Gantt Chart view:** It displays the progress of the workflow execution in a tabulated form.
- **Task view:** It displays the details about the workflow execution in a report format.

How To Delete Duplicate Row Using Informatica?

Assuming that the source system is a Relational Database, to eliminate duplicate records, we can check the Distinct option of the Source Qualifier of the source table and load the target accordingly.

What is Decode in INFORMATICA?

To understand Decode in an easy way, let's consider it as similar to the CASE statement in SQL. It is basically the function that is used by an expression transformation in order to search a specific value in a record.

There can be unlimited searches within the Decode function where a port is specified for returning result values. This function is usually used in cases where it is required to replace nested IF statements or to replace lookup values by searching in small tables with constant values.

Decode is a function that is used within Expression Transformation. It is used just like CASE Statement in SQL to search a specific record.

Below is a simple example of a CASE in SQL:

Syntax:

```
SELECT EMPLOYEE_ID,  
CASE  
WHEN EMPLOYEE_AGE <= 20 THEN 'Young' WHEN EMPLOYEE_AGE > 30 AND AGE <= 40  
THEN 'Knowledgeable' WHEN EMPLOYEE_AGE > 40 AND AGE = 60 THEN 'Wise'  
ELSE 'Very Wise'  
END AS EMPLOYEE_WISDOM  
FROM EMPLOYEE
```

What Is The Difference Between Router And Filter?

Router	Filter
Router transformation divides the incoming records into multiple groups based on some condition. Such groups can be mutually inclusive (Different groups may contain same record)	Filter transformation restricts or blocks the incoming record set based on one given condition.
Router transformation itself does not block any record. If a certain record does not match any of the routing conditions, the record is routed to default group	Filter transformation does not have a default group. If one record does not match filter condition, the record is blocked
Router acts like CASE.. WHEN statement in SQL (Or Switch().. Case statement in C)	Filter acts like WHERE condition in SQL.

What is Workflow Manager?

Workflow Manager is used to create Workflow and Worklet.

- **Workflow**
- Workflow is a set of instructions used to execute the mappings.
- The workflow contains various tasks such as session task, command task, event wait task, email task, etc. which are used to execute the sessions.
- It is also used to schedule the mappings.
- All the tasks are connected to each other through links inside a workflow.
- After creating the workflow, we can execute the workflow in the workflow manager and monitor its progress through the workflow monitor.
- **Worklet**
- Worklet is an object that groups a set of tasks which can be reused in multiple workflows.
- A worklet is similar to a workflow, but it does not have any scheduling information.
- In worklet, you can group the tasks in a single place so that it can be easily identified.

How to Schedule a Workflow?

1. A schedule is an automation of running the workflow at a given date and time.
2. There are 2 types of schedulers:

(i) Reusable scheduler: -

A reusable scheduler can be assigned to multiple workflows.

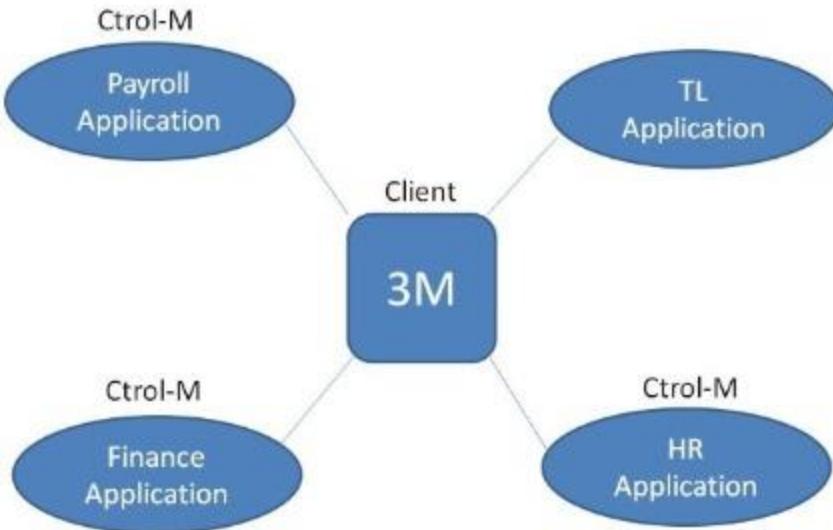
(ii) Non Reusable scheduler: -

- A non reusable scheduler is created specific to the workflow.
- A non reusable scheduler can be converted into a reusable scheduler.

The following are the 3rd party schedulers:

1. Cron (Unix based scheduling process)
 2. Tivoli
 3. Control M
 4. Autosys
 5. Tidal
 6. WLM (work load manager)
- 99% production people will do scheduling.

- Before we run the workflow manually. Through scheduling we run workflow this is called



Auto Running.

If informatica have own scheduler why using third party scheduler?

The client uses various applications (mainframes, oracle apps use Tivoli scheduling tool) and integrate different applications & scheduling that applications it is very easy by using third party schedulers.

What Type Of Repositories Can Be Created Using Informatica Repository Manager?

Informatica PowerCenter includes following type of repositories :

Standalone Repository : A repository that functions individually and this is unrelated to any other repositories.

Global Repository : This is a centralized repository in a domain. This repository can contain shared objects across the repositories in a domain. The objects are shared through global shortcuts.

Local Repository : Local repository is within a domain and it's not a global repository. Local repository can connect to a global repository using global shortcuts and can use objects in its shared folders.

Versioned Repository : This can either be local or global repository but it allows version control for the repository. A versioned repository can store multiple copies, or versions of an object. This feature allows to efficiently develop, test and deploy metadata in the production environment.

What is session?

- A session is a property in Informatica that has a set of instructions to define when and how to move the data from the source table to the target table.

- A session is like a task that we create in workflow manager. Any session that you create must have a mapping associated with it.
- Session must have a single mapping at a time, and it cannot be changed.
- In order to execute the session, it must be added to the workflow.
- A session can either be a reusable or non-reusable object where reusable means that we can use the data for multiple rows.

Design a Informatica mapping to load the First Half rows from a Relational table into a target?

Solution: In case of relational table you can go to source qualifier go to property and write the SQL query like

```
SELECT * from EMP  
WHERE ROWNUM <= (SELECT Count(*)/2 FROM EMP)
```

No additional filter required connect all ports to other transformation and finally to target, then run the mapping.

Convert single row from source to three rows in target?

Scenario1:

We have a source table containing 3 columns : Col1, Col2 and Col3. There is only 1 row in the table as follows:

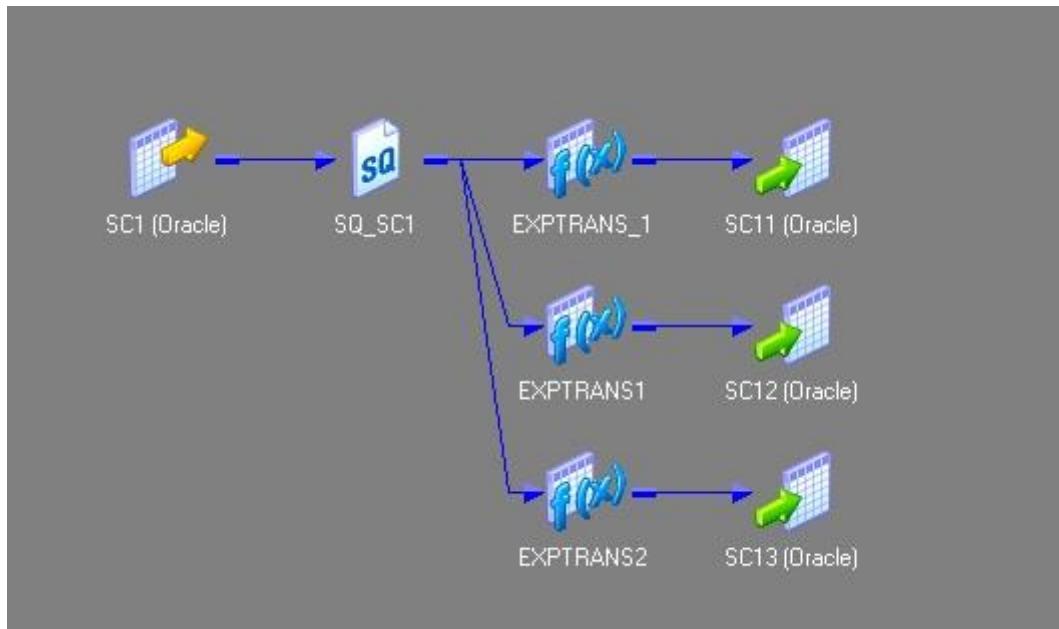
Col1	Col2	Col3
a	b	c

There is target table containing only 1 column Col. Design a mapping so that the target table contains 3 rows as follows:

Col
a
b
c

Without using normaliser transformation.

Solution:



Create 3 expression transformations exp_1,exp_2 and exp_3 with 1 port each. Connect col1 from Source Qualifier to port in exp_1. Connect col2 from Source Qualifier to port in exp_2. Connect col3 from source qualifier to port in exp_3. Make 3 instances of the target. Connect port from exp_1 to target_1. Connect port from exp_2 to target_2 and connect port from exp_3 to target_3.

Design a Informatica mapping to load the Second Half rows from a Relational table into a target?

Solution: In case of relational table you can go to source qualifier go to property and write the SQL query like

```
SELECT * from EMP  
MINUS  
SELECT * from EMP  
WHERE ROWNUM <= (SELECT Count(*)/2 FROM EMP)
```

No additional filter required connect all ports to other transformation and finally to target, then run the mapping.

Split the non-key columns to separate tables with key column in both / How to split the data of source table column-wise with respect to primary key. See the source and target tables below?

source table: ID is the key column, Name and Phone No are non-key columns.

ID	Name	Phone No
10	AAA	123
20	BBB	234
30	CCC	434
40	DDD	343
50	EEE	442

Target Table 1

ID	Name
10	AAA
20	BBB
30	CCC
40	DDD
50	EEE

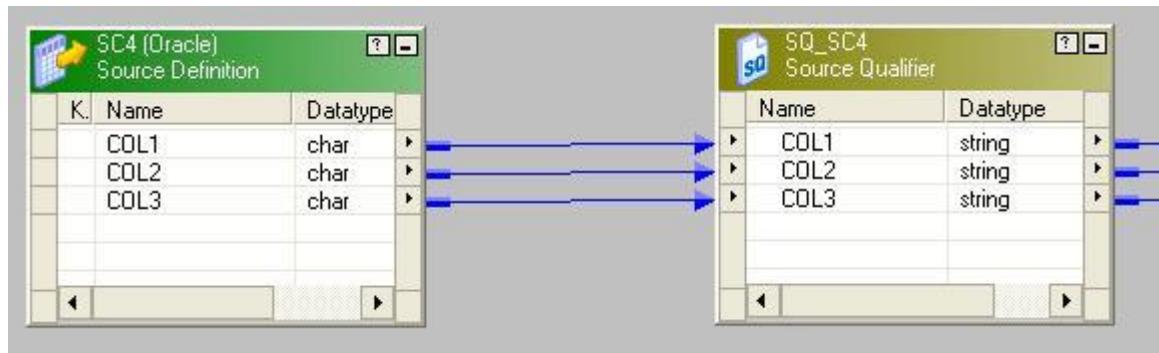
Target Table 2

ID	Phone No
10	123
20	234
30	434

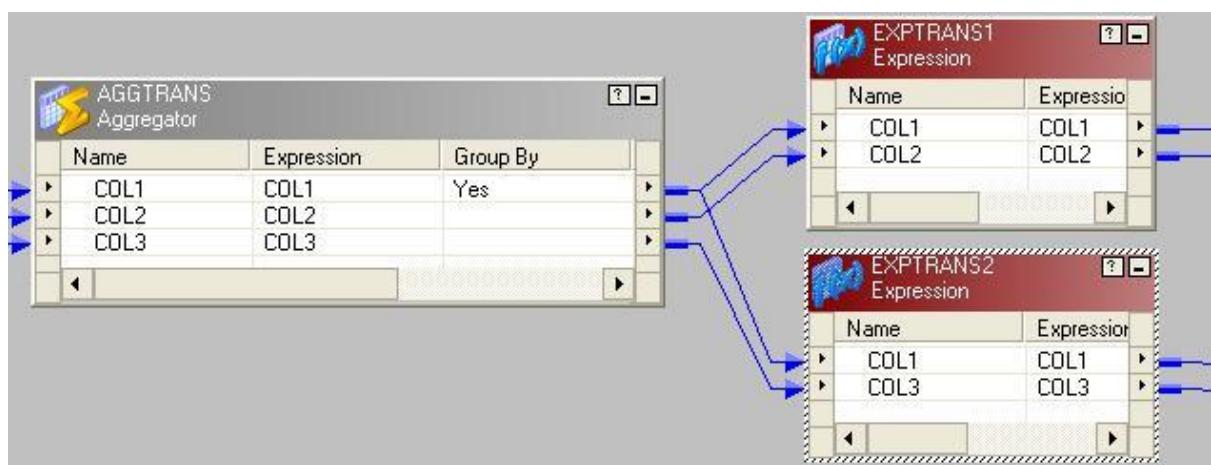
40	343
50	442

Solution:

Step 1: Source qualifier: get the source table to the mapping area. See image below.

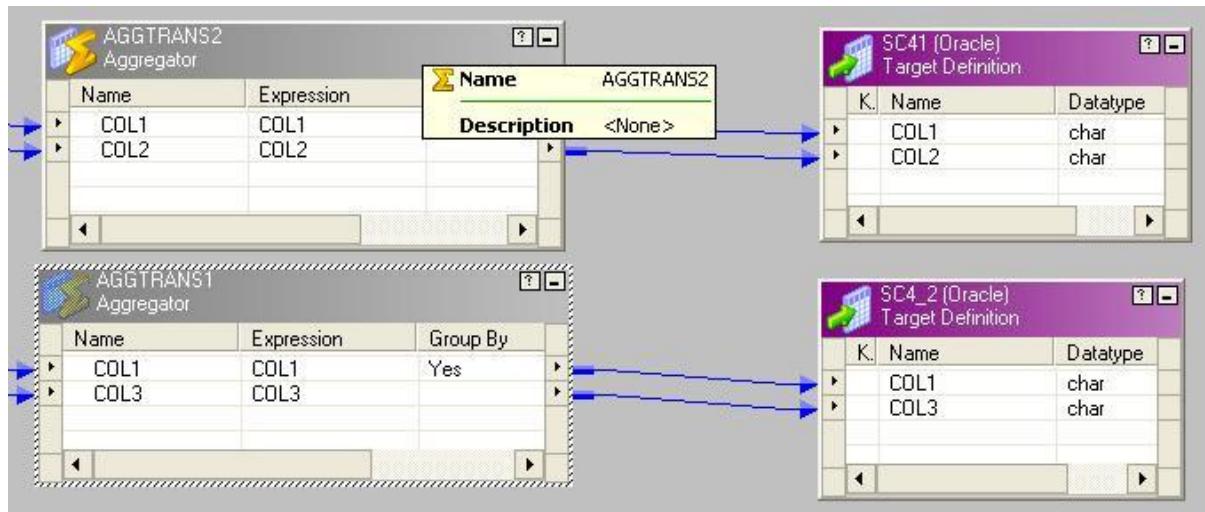


Step 2: Drag all the port from (from the previous step) to the Aggregator transformation and group by the key column. Since we have to split the columns to two different tables with the key column in each, so we are going to use two expression transformation, each will take the key column and one non-key column. Connect aggregator transformation with each of the expression transformation as follows.

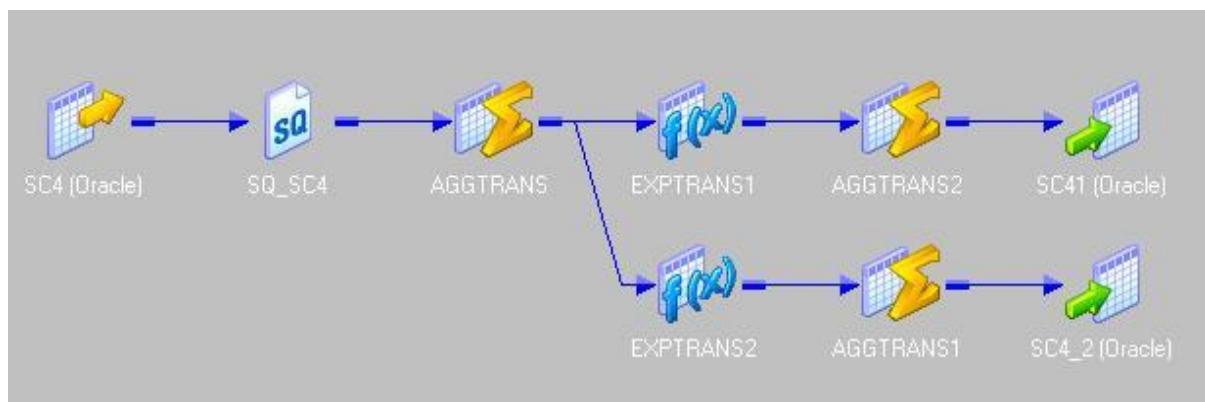


Step 3: We need another set of aggregator to be associated with each of the expression transformation from the previous step.

Step 4: In the final step connect the aggregators with the two target tables as follows.



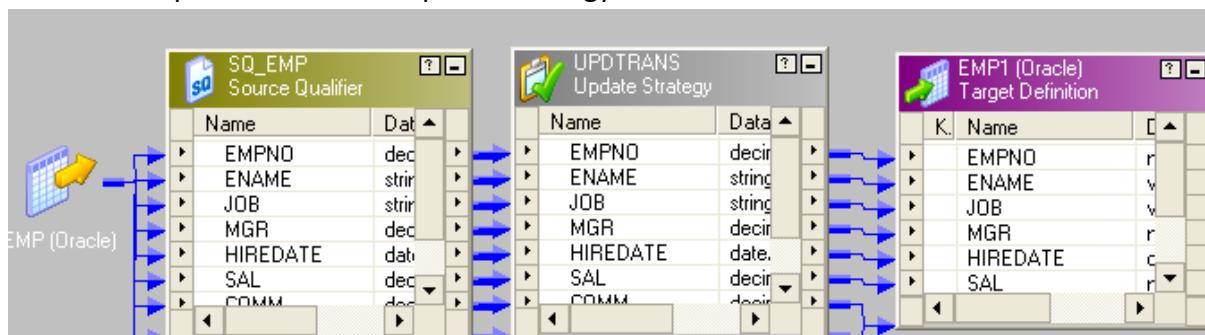
Here is the iconic view of the entire mapping.



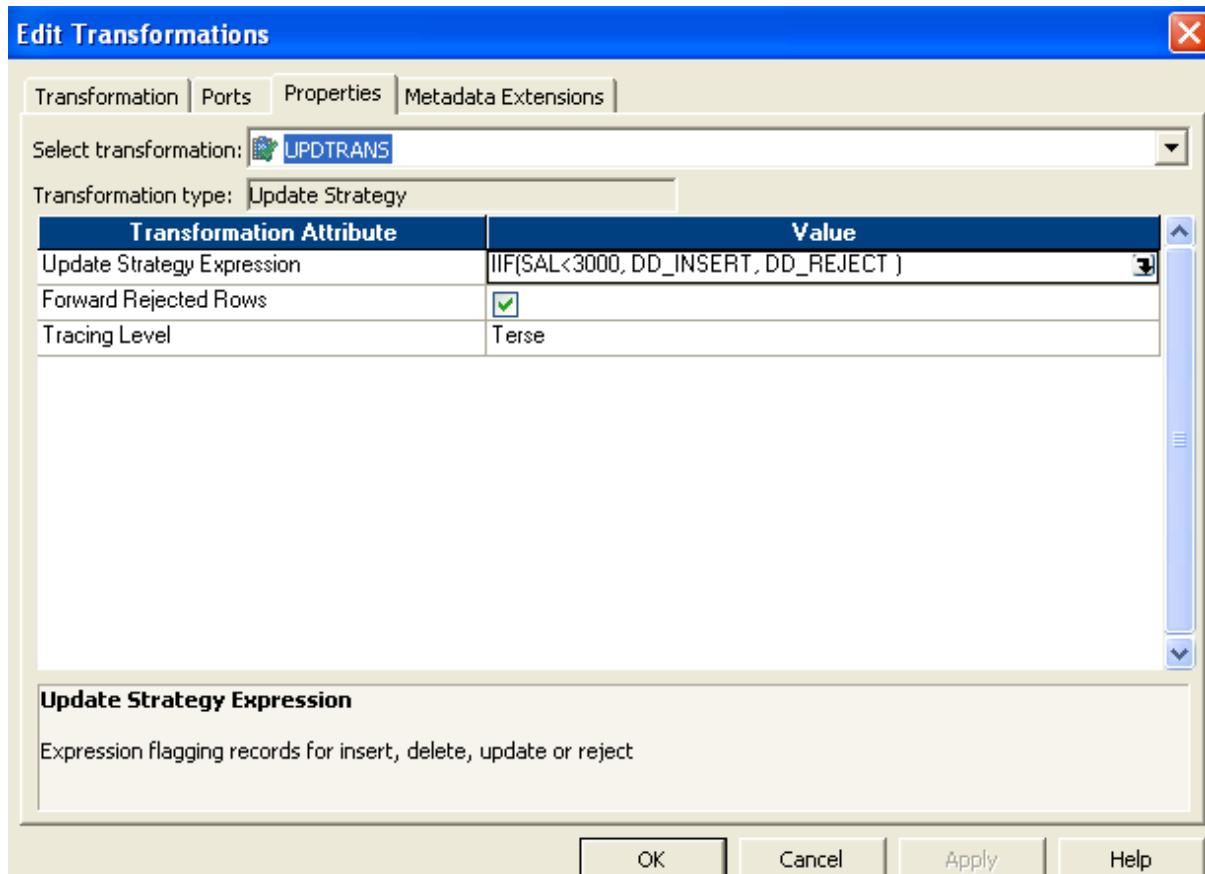
How to achieve this scenario? There is a emp table and from that table insert the data to target where sal<3000 and reject other rows?

Following are the steps for achieving it

1. connect out-puts from SQF to Update Strategy transformation.



2. In properties of Update Strategy write the condition like this



3. Connect the Update Strategy to target

Design a Informatica mapping to load the first 4 rows from a flat file into a target?

Solution: Since SQL Override will work for a flat file source then you have to assign row numbers to each record. Generate the row numbers either using the expression transformation as mentioned above or use sequence generator transformation. Then pass the output to filter transformation and specify the filter condition as ROW_NUMBER <=4

ID	COL1	COL2	ROW_NUMBER
1	A	10	1
2	B	20	2
3	C	30	3
4	D	10	4
5	E	30	5
6	F	10	6

How to segregate the duplicate and distinct rows from source table to separate target tables?

source table:

COL1	COL2	COL3
a	b	c
x	y	z
a	b	c
r	f	u
a	b	c
v	f	r
v	f	r

Target Table 1: Table containing all the unique rows

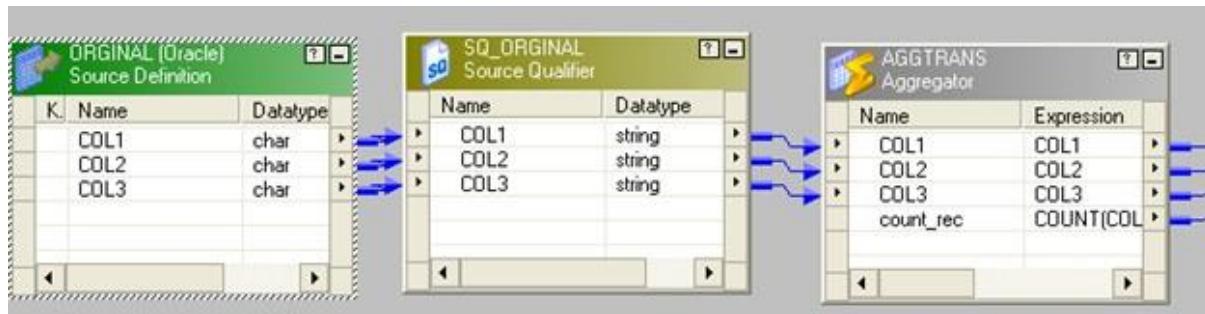
COL1	COL2	COL3
a	b	c
x	y	z
r	f	u
v	f	r

Target Table 2: Table containing all the duplicate rows

COL1	COL2	COL3
a	b	c
a	b	c
v	f	r

Solution:

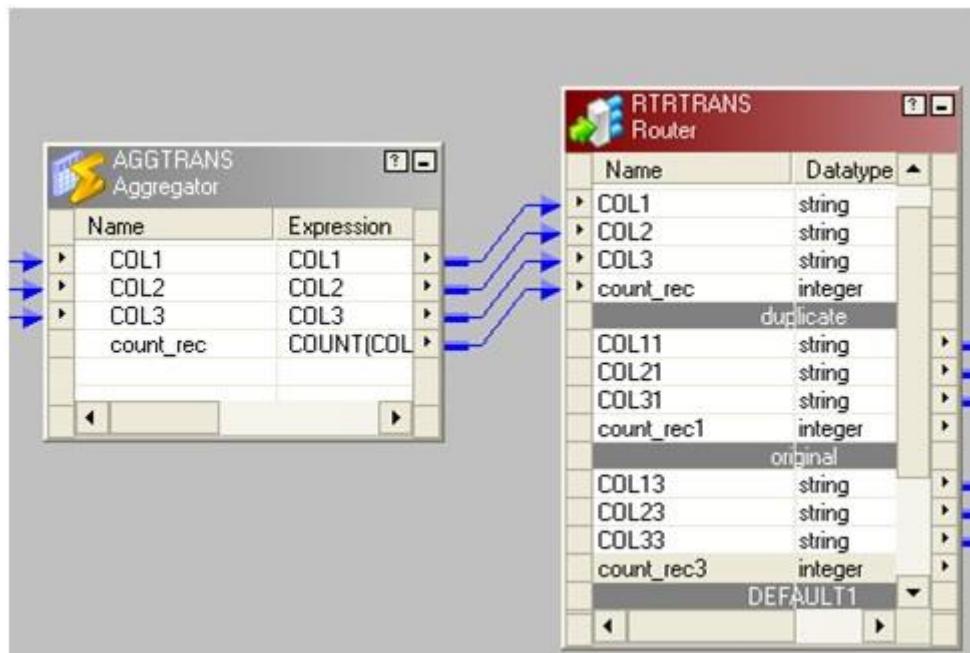
Step 1: Drag the source to mapping and connect it to an aggregator transformation.



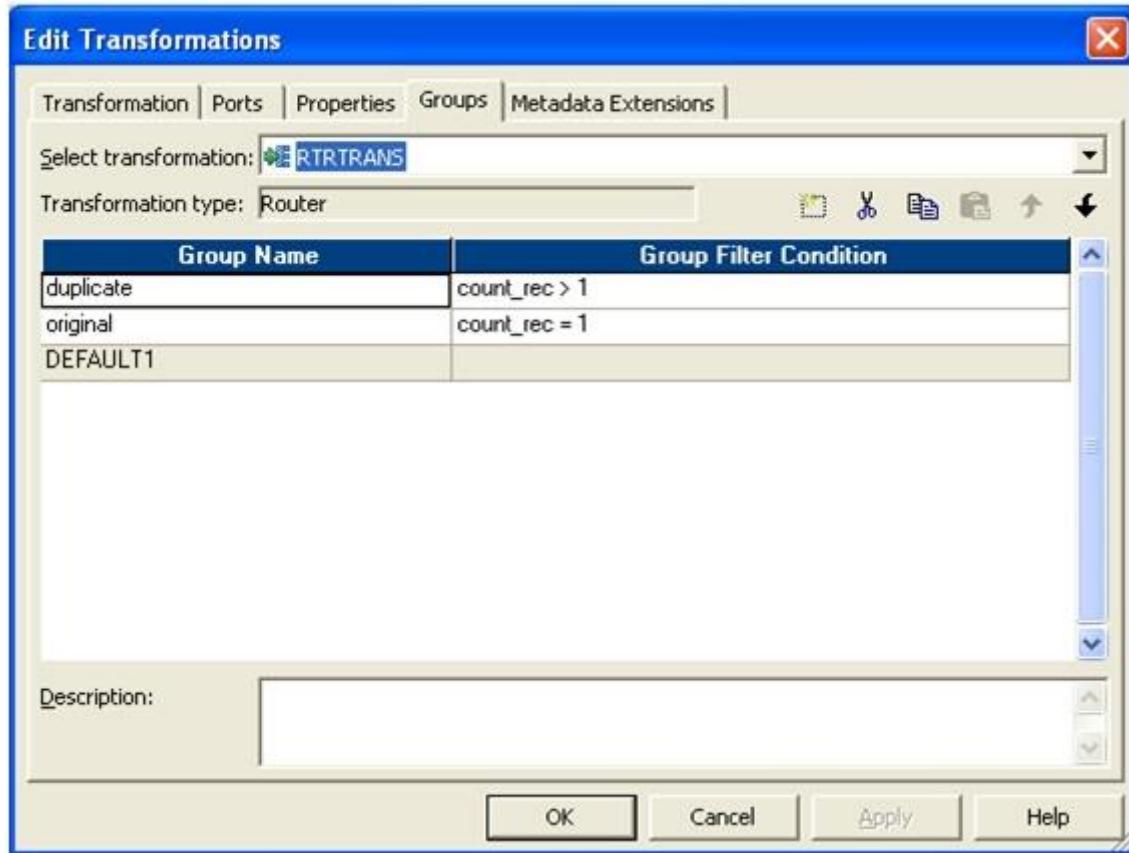
Step 2: In aggregator transformation, group by the key column and add a new port call it count_rec to count the key column.

Step 3: connect a router to the aggregator from the previous step. In router make two groups one named "original" and another as "duplicate"

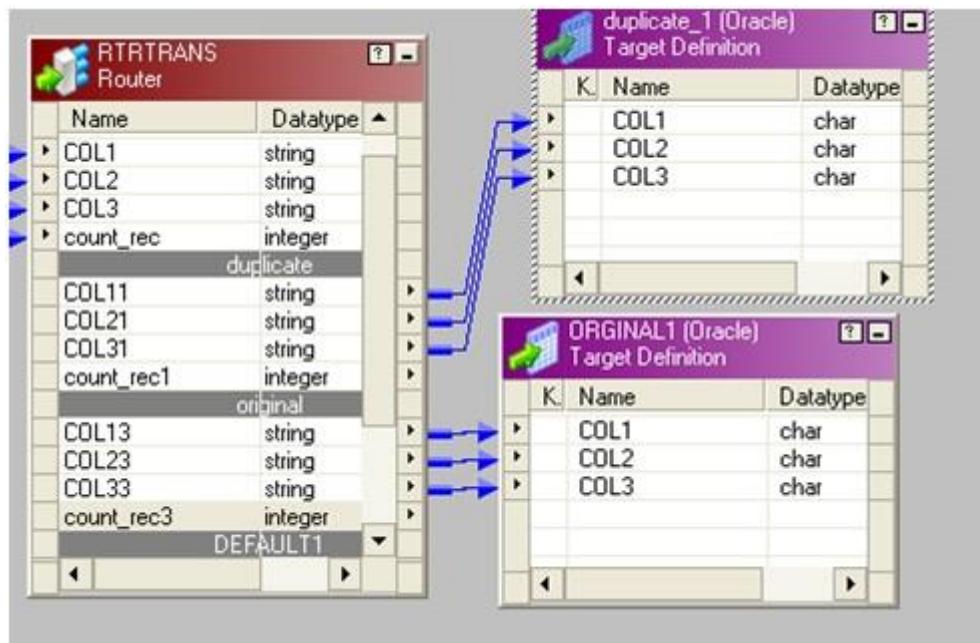
In original write count_rec=1 and in duplicate write count_rec>1.



The picture below depicting group name and the filter conditions



Step 4: Connect two group to corresponding target table.



How to achieve this scenario? Suppose you are importing a flat file emp.csv and hire date column is in numeric format, like 20101111 .Our objective is convert it to date,with a

format 'YYYYMMDD'?

source

EMPNO HIRE_DATE(numeric)

1 20101111

2 20090909

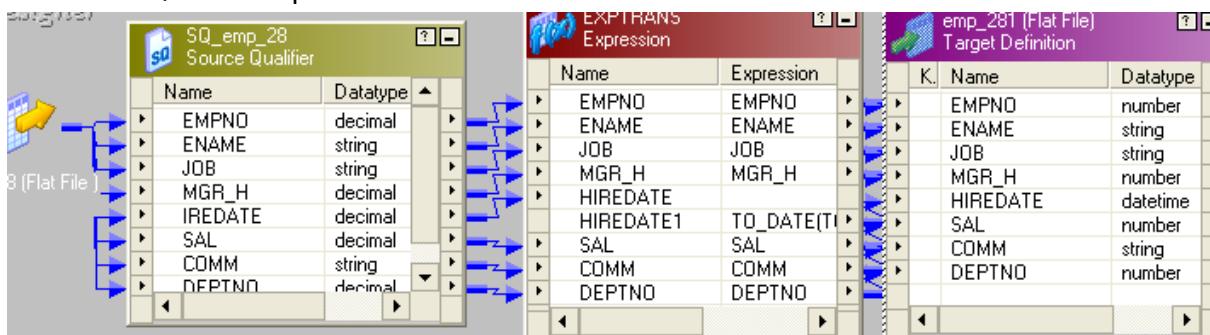
target

EMPNO HIRE_DATE (date)

1 11/11/2010

2 09/09/2009

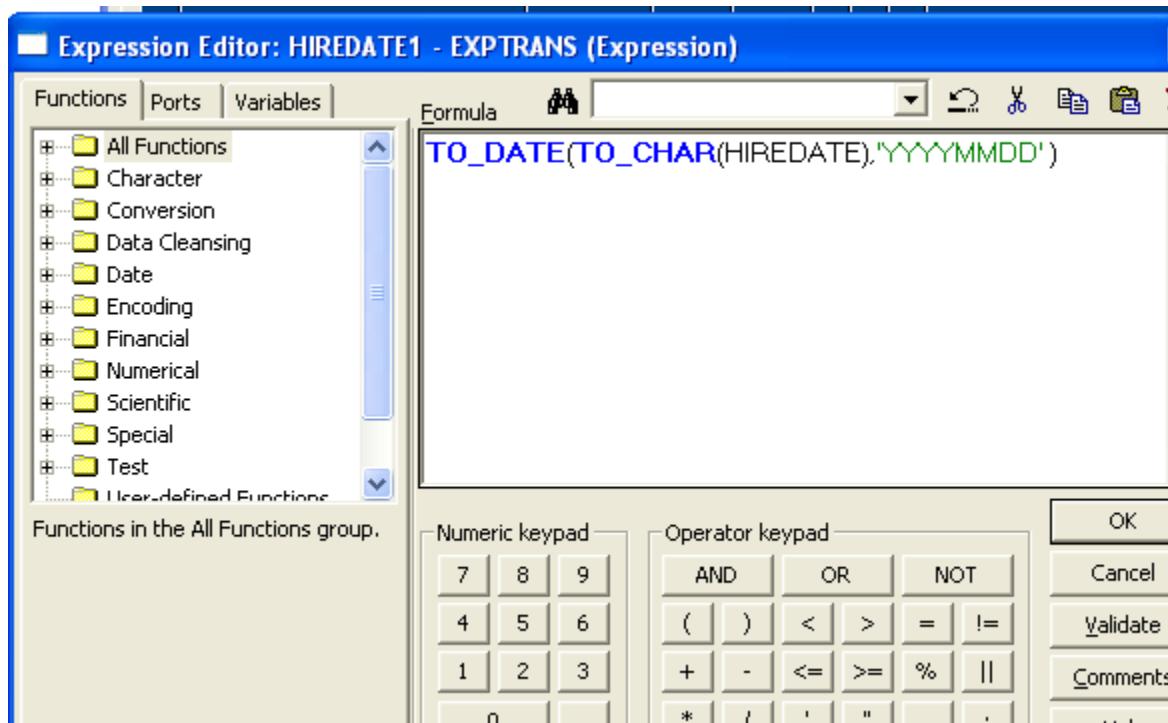
1. Connect SQF to an expression.



2. In expression make hire_date as input only and make another port hire_date1 as o/p port with date data type.

The screenshot shows the 'Transformation Editor' for the 'EXPRETRANS' transformation. The 'Ports' tab is selected, showing nine ports: EMPNO, ENAME, JOB, MGR_H, HIREDATE, HIREDATE1, SAL, COMM, and DEPTNO. The 'Expression' column for each port shows the mapping logic. For example, HIREDATE is mapped to HIREDATE1 using the expression TO_DATE(TO_CHAR(HIREDATE, 'YYYYMMDD')). The 'Ports' tab also includes columns for Port Name, Datatype, Prec, Scale, and various checkboxes for output types (I, O, V).

3. In o/p port of hire_date write condition like as below

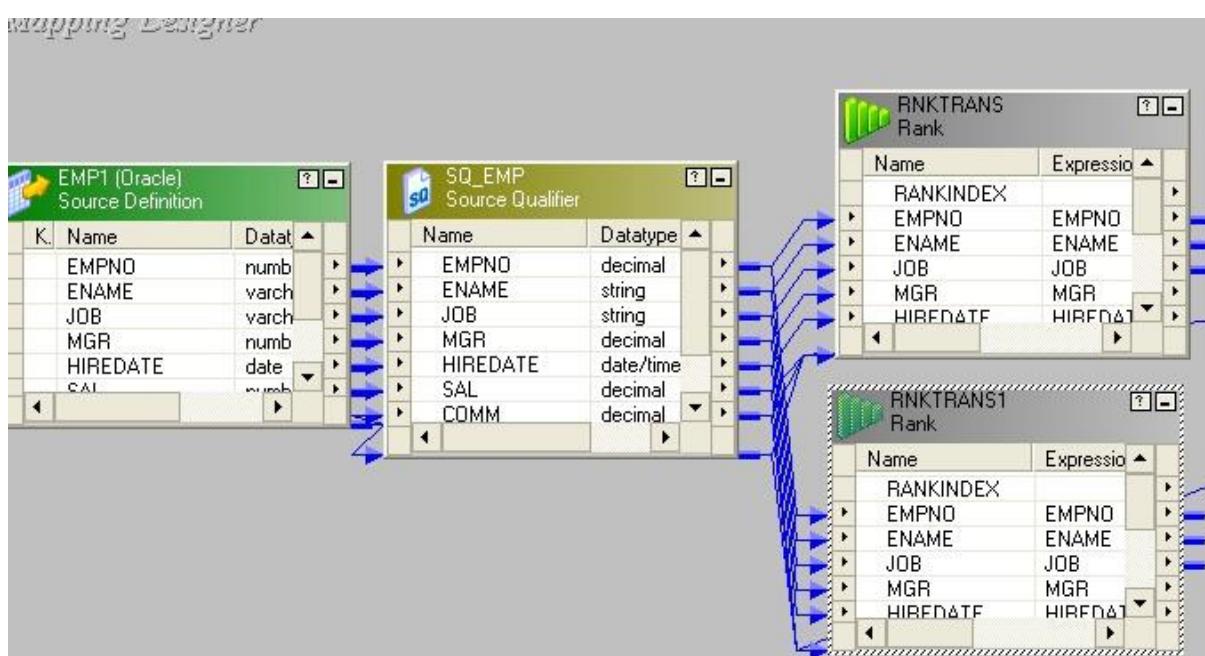


4. Finally send to target

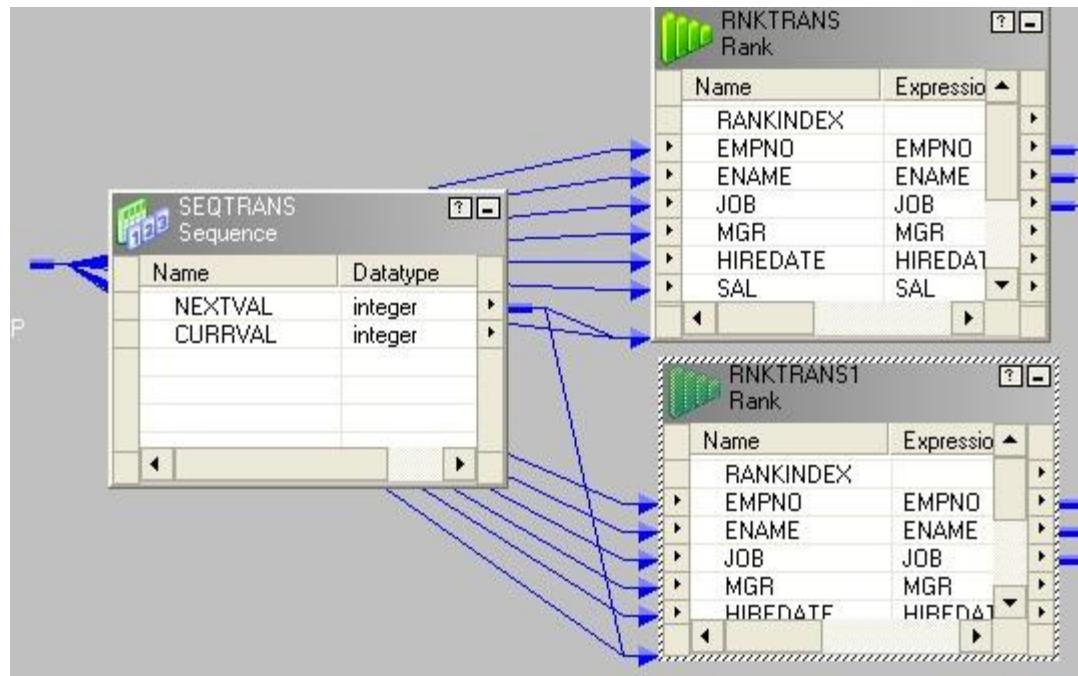
How to achieve this scenario? How to get first and last record from a table/file?

Solution:

Step 1: Drag and drop ports from source qualifier to two rank transformations.

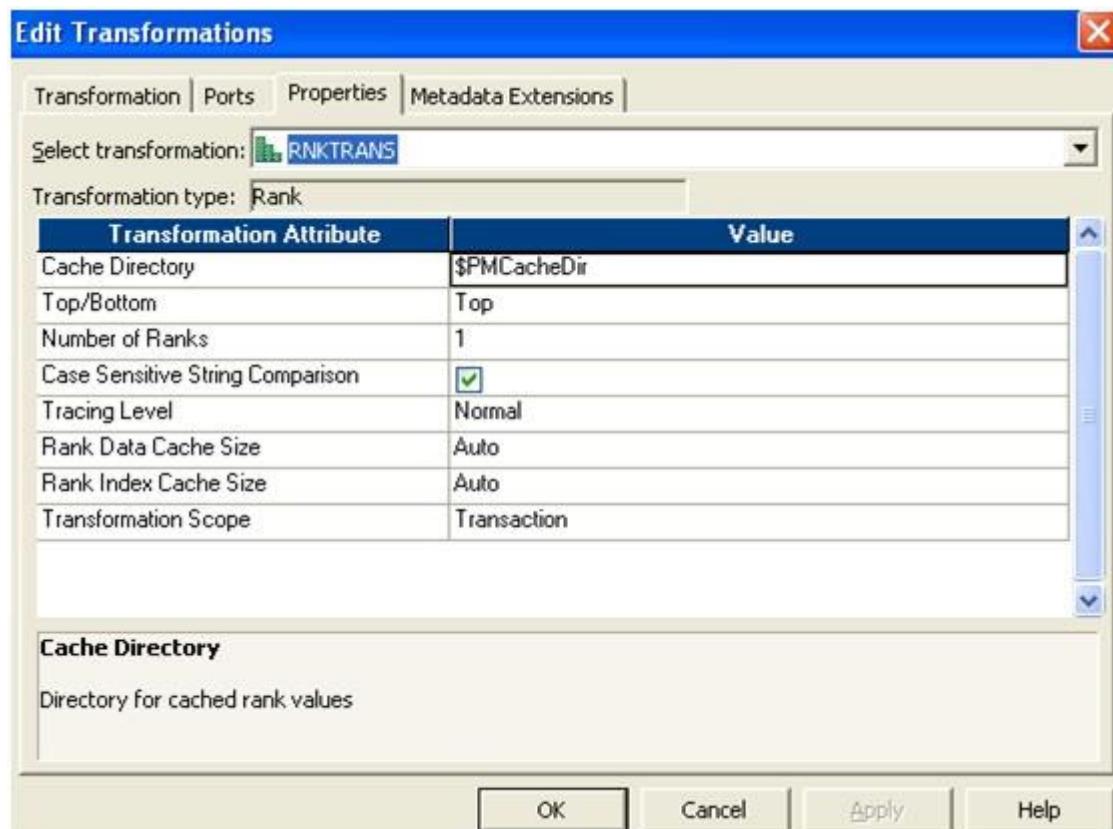


Step 2: Create a reusable sequence generator having start value 1 and connect the next value to both rank transformations.

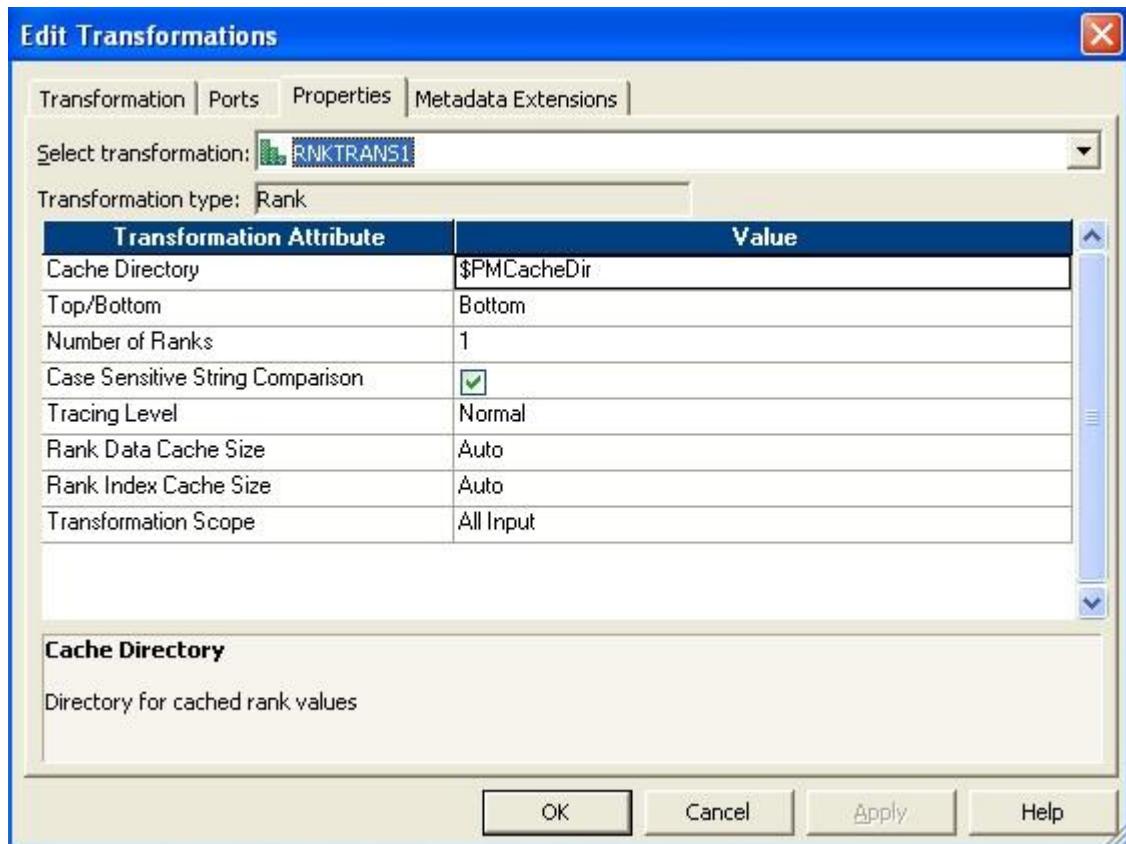


Step 3: Set rank properties as follows

In Rank1

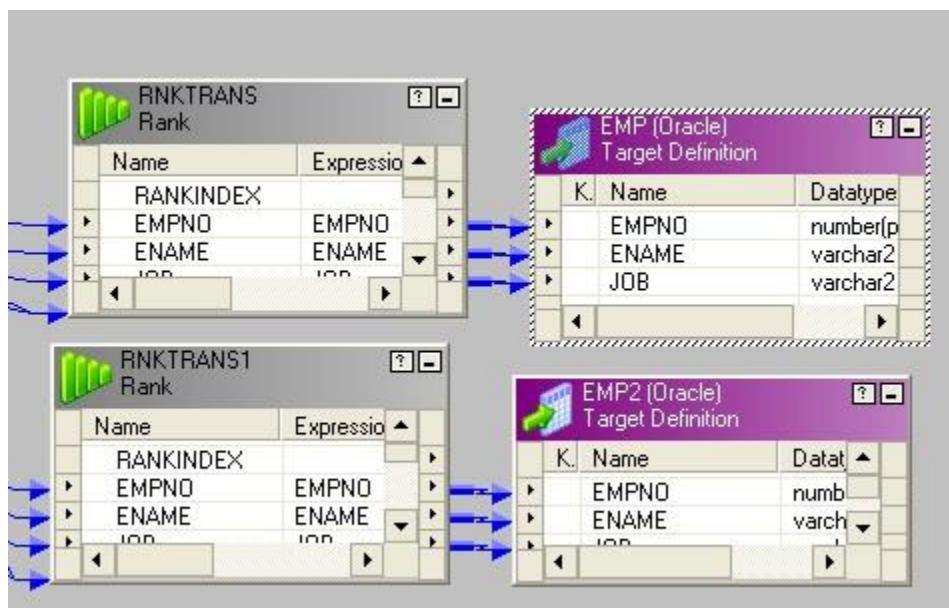


In Rank2



Step 4: Make two instances of the target.

Step 5: Connect the output port to target.



How to load only NULL records into the target? Explain using mapping flow?

Consider below data as a source

Emp_Id	Emp_Name	Salary	City	Pincode
619101	Aanchal Singh	20000	Pune	411051
619102	Nupura Pattihal	35000	Nagpur	411014
NULL	NULL	15000	Mumbai	451021

The target table also has table structure as a source. We will have two tables, one which will contain NULL values and other which would not contain NULL values.

The mapping can be as:

SQ → EXP → RTR → TGT_NULL/TGT_NOT_NULL

EXP – Expression transformation create an output port

O_FLAG= IIF ((ISNULL(emp_id) OR ISNULL(emp_name) OR ISNULL(salary) OR ISNULL(City)
OR ISNULL(Pincode)), 'NULL','NNULL')

RTR – Router transformation two groups

Group 1 connected to TGT_NULL (Expression O_FLAG='NULL')

Group 2 connected to TGT_NOT_NULL (Expression O_FLAG='NNULL')

How to achieve this scenario? How to remove footer from your file ?

For example the file content looks like as below:-

some Header here

col1 col2 col3 col4

data1 data2 data3 data4

data5 data6 data7 data8

data1 data2 data3 data4

data1 data2 data3 data4

footer

Just we have to remove footer from the file.

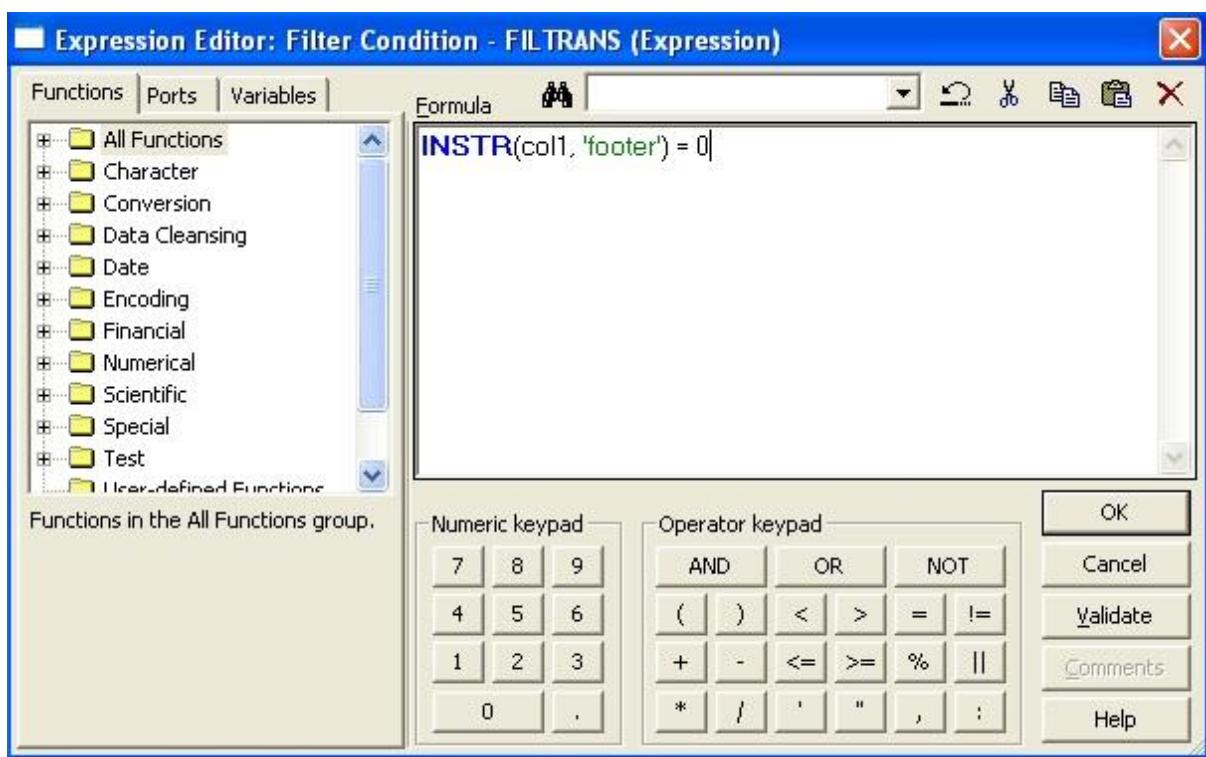
Solution:

Step1: Drag the source to mapping area.

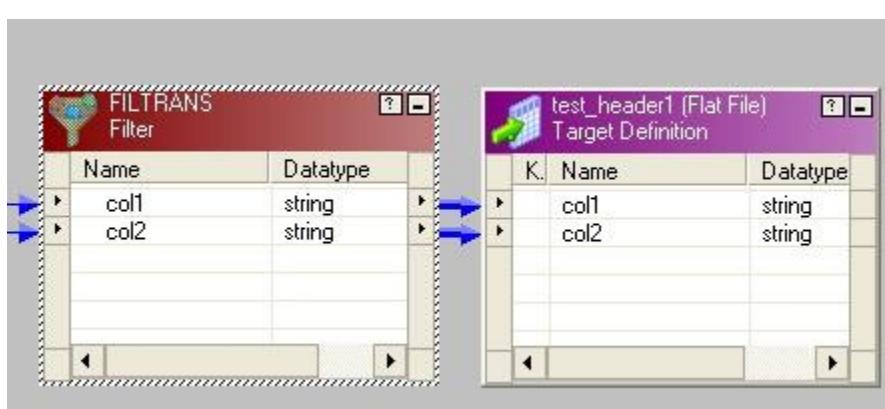


Step2: After that connect a filter or router transformation.

Step3: In filter write the condition like in the picture



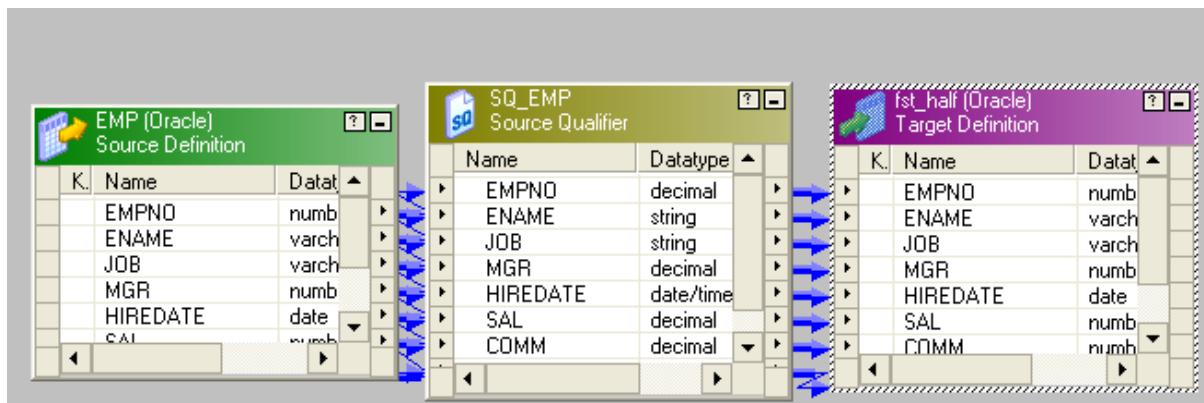
Step 4:Finally pass it over to target.



How to achieve this scenario? How to send first half record to target?

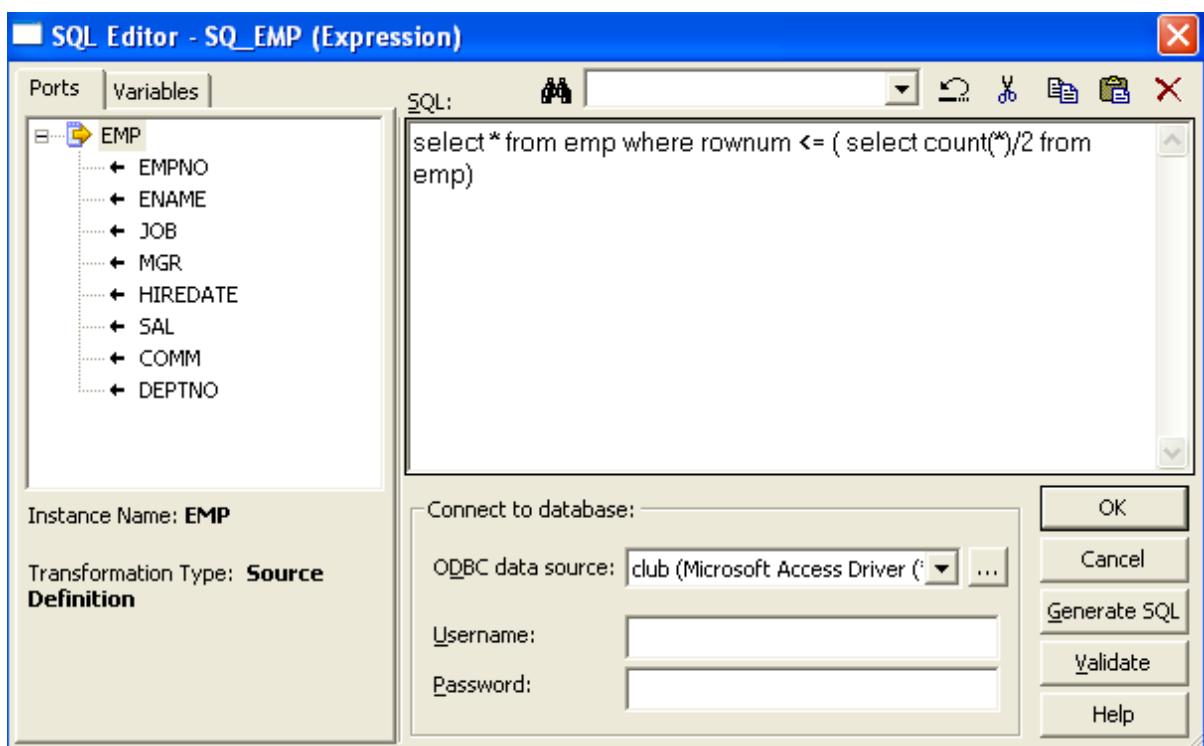
Solution:

1. Drag and drop the source to mapping.



2. Step:2 In source-Qualifier , go to property and write the SQL query like

1 select * from emp where rownum <= (select count(*)/2 from emp)



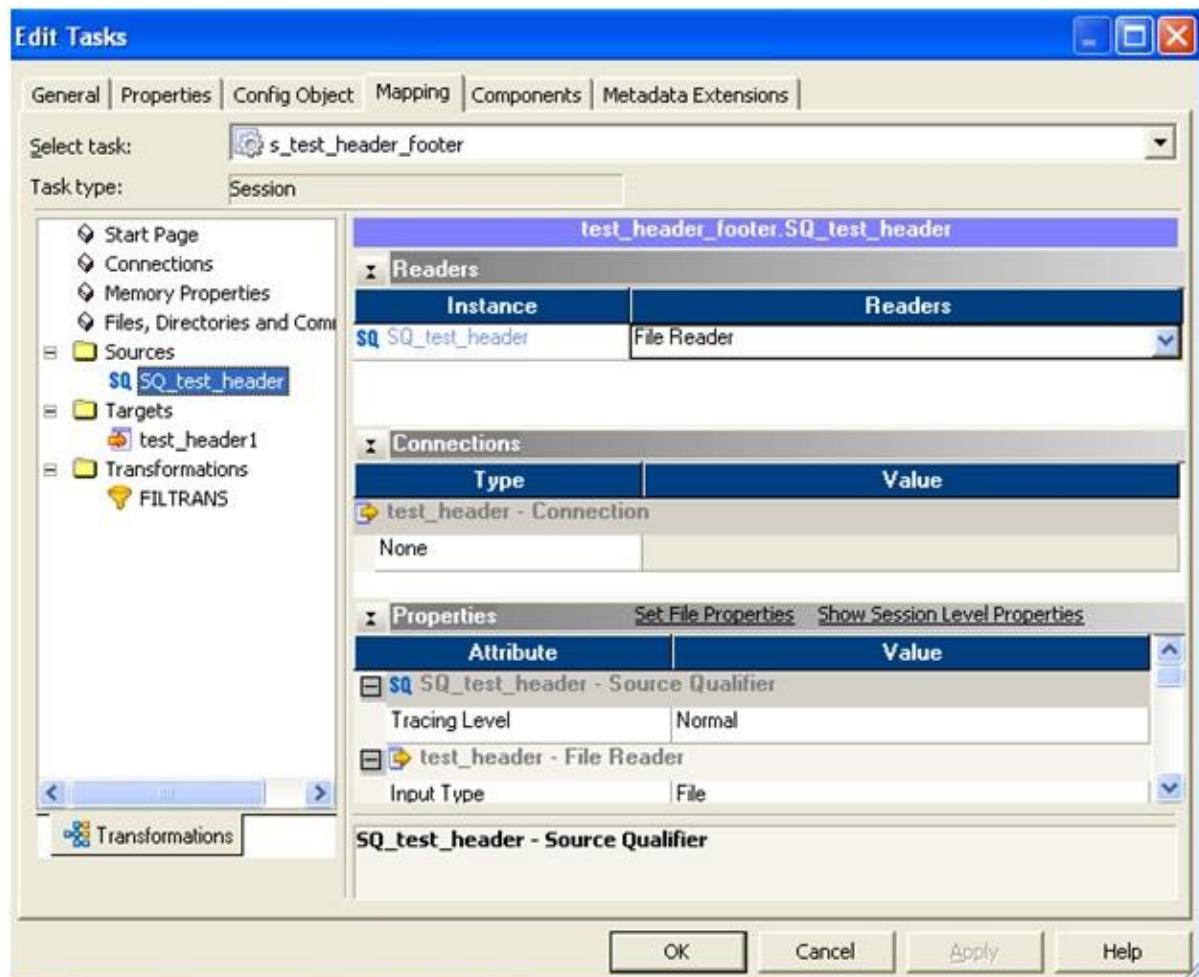
3. Then connect to target. Now you are ready to run the mapping to see it in action.

How to achieve this scenario? How to remove header from a file ?

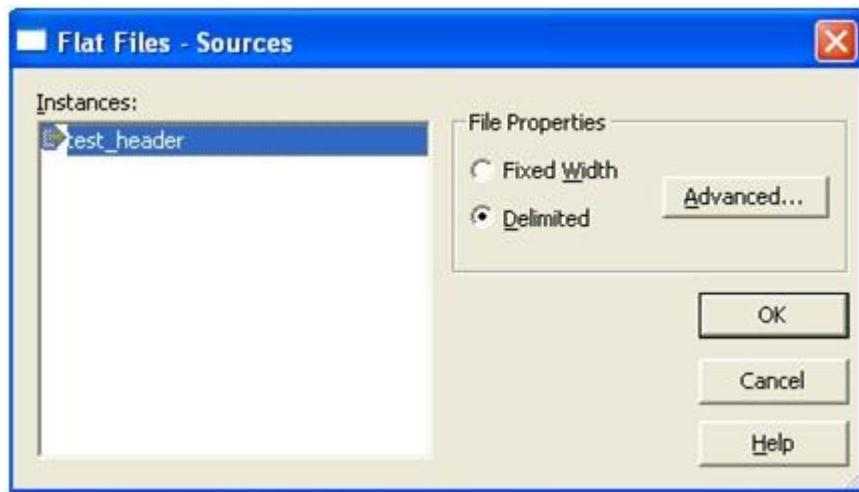
Solution

Step1: After mapping go to workflow and scheduled it.

Step2: Just double click on the session and go to mapping option.

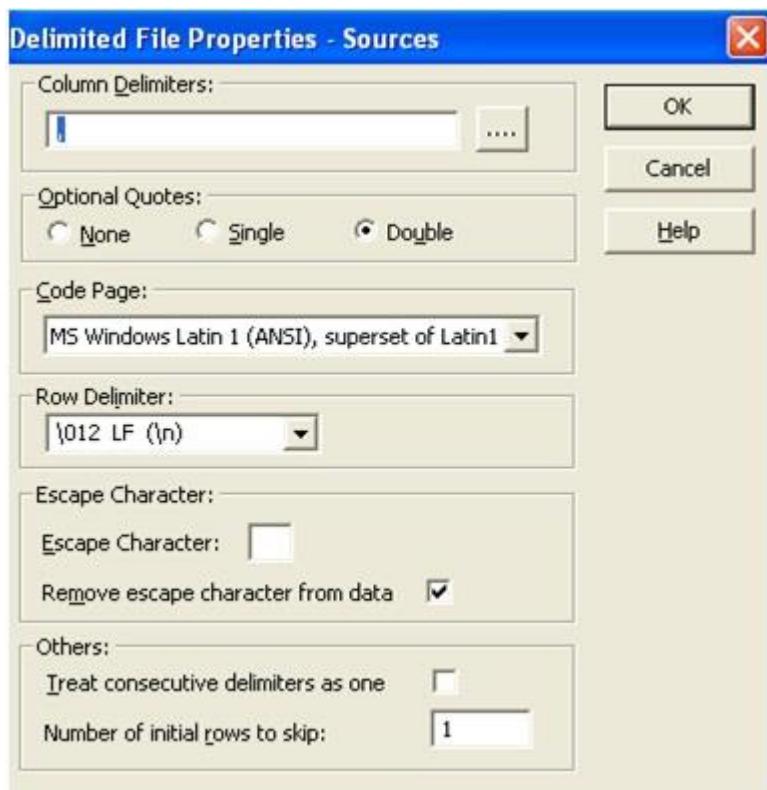


Step3: Select the source and go to the set file properties.



flat file properties

Step4: Choose the advance option. Set number of initial rows skip: 1 (it can be more as per requirement)



adv properties

It will skip the header.

Scenario 1

Consider the following source Data.

- A
- B
- C
- C
- B
- D
- B

1.1 Design a mapping to load all unique products in one table and the duplicate rows in another table.

The first table should contain the following output

- A
- D

The second target should contain the following output

- B
- B
- B
- C
- C

Solution: Use sorter transformation and sort the products data. Pass the output to an expression transformation and create a dummy port O_dummy and assign 1 to that port. So that, the DUMMY output port always return 1 for each row.

The output of expression transformation will be

Product O_dummy

- | | |
|---|---|
| A | 1 |
| B | 1 |
| B | 1 |
| B | 1 |
| C | 1 |
| C | 1 |
| D | 1 |

Pass the output of expression transformation to an aggregator transformation. Check the group by on product port. In the aggregator, create an output port O_count_of_each_product and write an expression count(product).

The output of aggregator will be

Product	O_count_of_each_product
A	1
B	3
C	2
D	1

Now pass the output of expression transformation, aggregator transformation to joiner transformation and join on the products port. In the joiner transformation check the property sorted input, then only you can connect both expression and aggregator to joiner transformation.

The output of joiner will be

Product	O_dummy	O_count_of_each_product
A	1	1
B	1	3
B	1	3
B	1	3
C	1	2
C	1	2
C	1	1

Now pass the output of joiner to a router transformation, create one group and specify the group condition as O_dummy=O_count_of_each_product. Then connect this group to one table. Connect the output of default group to another table.

1.2 Design a Informatica mapping to load original and duplicate records in two different tables / Separating duplicate and non-duplicate rows.

The first table should contain the following output

- A
- B
- C
- D

The second table should contain the following output

- B
- B
- C

Solution1: Use sorter transformation and sort the products data. Pass the output to an expression transformation and create a variable port, V_Current_product, and assign product port to it. Then create a V_Count port and in the expression editor write $V_Count=IIF(V_Current_product=V_Previous_product,V_Count+1,1)$. Create one more variable port V_Previous_product and assign product port to it. Now create an output port O_Count port and assign V_Count port to it.

In the expression transformation, the ports are

Product

V_Current_product=product
 $V_Count=IIF(V_Current_product=V_Previous_product,V_Count+1,1)$
V_Previous_product=product
O_Count=V_Count

The output of expression transformation will be

Product O_Count

A	1
B	1
B	2
B	3
C	1
C	2
D	1

Now Pass the output of expression transformation to a router transformation, create one group and specify the condition as O_count=1. Then connect this group to one table. Connect the output of default group to another table.

Solution2: Next approach to do this is use aggregator, First sort the data using sorter then connect it to aggregator transformation group by product with an additional port COUNT_RECORD in expression put COUNT(Product).

Now with a router transformation create two groups namely DUPLICATE & ORIGINAL and give the group condition COUNT_RECORD > 1 & COUNT_RECORD =1 respectively.

In the Aggregator transformation ports are

Product
COUNT_RECORD=COUNT(Product)

In the Router Transformation group Conditions are

DUPLICATE group (COUNT_RECORD>1)

ORIGINAL group (COUNT_RECORD=1)

Connect Table 1 to DUPLICATE group and Table 2 to Original Group. There you go you have duplicate and original data separated.

Scenario 2

Consider the following employees data as source

employee_id, salary

10, 1000

20, 2000

30, 3000

40, 5000

2.1 Design a mapping to load the cumulative sum of salaries of employees into target table?

The target table data should look like as

employee_id, salary, cumulative_sum

10, 1000, 1000

20, 2000, 3000

30, 3000, 6000

40, 5000, 11000

Solution:

Connect the source Qualifier to expression transformation. In the expression transformation, create a variable port V_cum_sal and in the expression editor write V_cum_sal+salary. Create an output port O_cum_sal and assign V_cum_sal to it.

2.2 Design a mapping to get the previous row salary for the current row. If there is no previous row exists for the current row, then the previous row salary should be displayed as null?

The output should look like as below.

employee_id, salary, pre_row_salary

10, 1000, Null

20, 2000, 1000

30, 3000, 2000

40, 5000, 3000

Solution:

Connect the source Qualifier to expression transformation. In the expression transformation, create a variable port V_count and increment it by one for each row entering the expression transformation. Also create V_salary variable port and assign the expression IIF(V_count=1,NULL,V_prev_salary) to it . Then create one more variable port V_prev_salary and assign Salary to it. Now create output port O_prev_salary and assign V_salary to it. Connect the expression transformation to the target ports.

In the expression transformation, the ports will be

```
employee_id  
salary  
V_count=V_count+1  
V_salary=IIF(V_count=1,NULL,V_prev_salary)  
V_prev_salary=salary  
O_prev_salary=V_salary
```

2.3 Design a mapping to get the next row salary for the current row. If there is no next row for the current row, then the next row salary should be displayed as null.

The output should look like as

```
employee_id, salary, next_row_salary  
10, 1000, 2000  
20, 2000, 3000  
30, 3000, 5000  
40, 5000, Null
```

Solution:

Step1: Connect the source qualifier to two expression transformation. In each expression transformation, create a variable port V_count and in the expression editor write V_count+1. Now create an output port O_count in each expression transformation. In the first expression transformation, assign V_count to O_count. In the second expression transformation assign V_count-1 to O_count.

In the first expression transformation, the ports will be

```
employee_id  
salary
```

V_count=V_count+1
O_count=V_count

In the second expression transformation, the ports will be

employee_id
salary
V_count=V_count+1
O_count=V_count-1

Step2: Connect both the expression transformations to joiner transformation and join them on the port O_count. Consider the first expression transformation as Master and second one as detail. In the joiner specify the join type as Detail Outer Join. In the joiner transformation check the property sorted input, then only you can connect both expression transformations to joiner transformation.

Step3: Pass the output of joiner transformation to a target table. From the joiner, connect the employee_id, salary which are obtained from the first expression transformation to the employee_id, salary ports in target table. Then from the joiner, connect the salary which is obtained from the second expression transformation to the next_row_salary port in the target table.

What Is The Method Of Loading 5 Flat Files Of Having Same Structure To A Single Target And Which Transformations I Can Use?

Two Methods.

1. write all files in one directory then use file repository concept (don't forget to type source file type as indirect in the session).
2. use union t/r to combine multiple input files into a single target.

Design a mapping to find the sum of salaries of all employees and this sum should repeat for all the rows?

The output should look like as.

employee_id, salary, salary_sum
10, 1000, 11000
20, 2000, 11000
30, 3000, 11000
40, 5000, 11000

Solution:

Step1: Connect the source qualifier to the expression transformation. In the expression transformation, create a dummy port and assign value 1 to it.

In the expression transformation, the ports will be

employee_id
salary
O_dummy=1

Step2: Pass the output of expression transformation to aggregator. Create a new port O_sum_salary and in the expression editor write SUM(salary). Do not specify group by on any port.

In the aggregator transformation, the ports will be

salary
O_dummy
O_sum_salary=SUM(salary)

Step3: Pass the output of expression transformation, aggregator transformation to joiner transformation and join on the DUMMY port. In the joiner transformation check the property sorted input, then only you can connect both expression and aggregator to joiner transformation.

Step4: Pass the output of joiner to the target table.

Suppose in flat_file some special symbols like @,%,\$,#,& has added in empno column along with the actual data. How to remove those special characters ?

empno in source

empno(in string format)

7@3%\$21

432#@1

324&*,\\$2

In target

empno

7321

4321

3242

Following are the steps for achieving this mapping

1. Connect 0/p columns of SQF to an expression transformation.
2. In expression make empno as input and create another port empno1 as output port with date datatype. And in empno1 write condition like this. and finally send it to target

The screenshot shows the 'Edit Transformations' window with the 'Transformation' tab selected. A table lists three ports: EMPNO, EMPNO1, and ENAME. The 'Expression' column for EMPNO1 contains the formula `REPLACESTR(1,EMPNO, '@', '#', '$', '%')`. Below this, the 'Expression Editor' window is open, showing the same formula in the 'Formula:' field and a tree view of available functions under 'All Functions'.

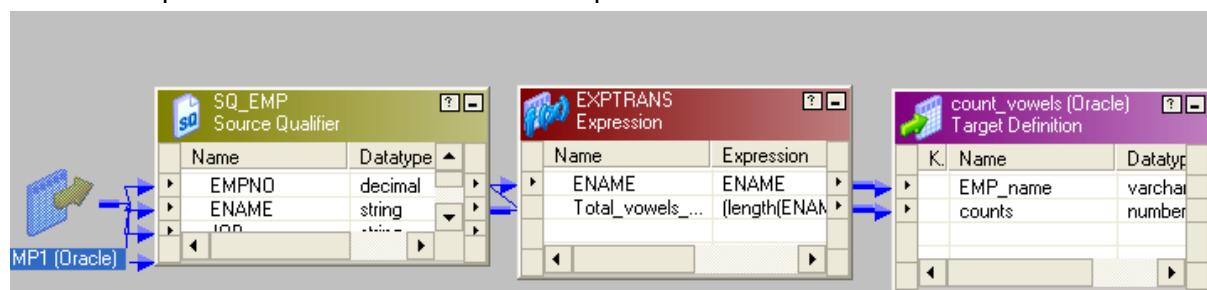
	Port Name	Datatype	Prec	Scale	I	O	V	Expression
1	EMPNO	string	6	0	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
2	EMPNO1	string	6	0	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	REPLACESTR(1,EMPNO, '@', '#', '\$', '%')
3	ENAME	string	8	0	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	ENAME

How to achieve? Count the no of vowels present in emp_name column of EMP table as shown below

emp_name	total_vowels_count
Allen	2
Scott	1
Ward	1

These are the steps to achieve it

1. Connect required columns from SQF to an expression transformation.



2. In Expression add 6 columns like in the picture as bellow. But You can make it two columns (One for all the vowels and one for the vowel counts). For better understanding I have added 6 columns, 5 for each of the vowels and one for the vowel count.

Edit Transformations

Select transformation: EXPTRANS

Transformation type: Expression

	Port Name	Datatype	Prec	Scale	I	O	V	Expression
1	ENAME	string	10	0	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	ENAME
2	A	string	10	0	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	REPLACECHR(0,ENAME,'a',N...)
3	E	string	10	0	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	REPLACECHR(0,ENAME,'e',N...)
4	I	string	10	0	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	REPLACECHR(0,ENAME,'i',N...)
5	O	string	10	0	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	REPLACECHR(0,ENAME,'o',N...)
6	U	string	10	0	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	REPLACECHR(0,ENAME,'u',N...)
7	Total_vowels_count	integer	10	0	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	(length(ENAME)-length(A))...

Default value:

Description:

OK Cancel Apply Help

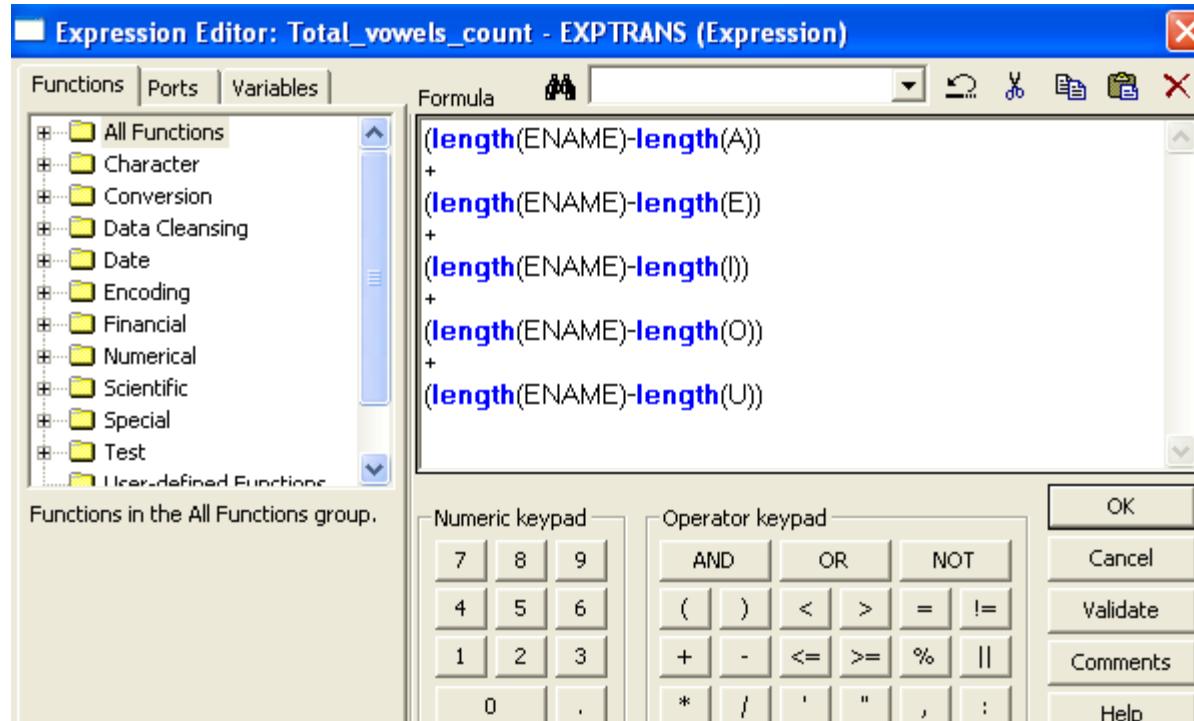
The way I achieved is for each of the vowels in ename , I replaced it with null and in port total vowel count , I subtract the vowel port from the ename length which gives me the individual count of vowels, after adding up for all vowels I found all the vowels present. Here are all the variable ports.

For A write REPLACECHR(0,ENAME,'a',NULL)
 For E write REPLACECHR(0,ENAME,'e',NULL)
 For I write REPLACECHR(0,ENAME,'i',NULL)
 For O write REPLACECHR(0,ENAME,'o',NULL)
 For U write REPLACECHR(0,ENAME,'u',NULL)

And for o/p column total_vowels_count write expression like this
 $(\text{length}(\text{ENAME})-\text{length}(A))$

+
 $(\text{length}(\text{ENAME})-\text{length}(E))$
 +
 $(\text{length}(\text{ENAME})-\text{length}(I))$
 +
 $(\text{length}(\text{ENAME})-\text{length}(O))$
 +

$(\text{length}(\text{ENAME}) - \text{length}(U))$



- Finally send to target.

Scenario 3:

Let us assume the following Sales Data

ID	PERSONAL_SALES
10	40000
20	80000
30	40000
40	60000
50	NULL
60	50000

3.1 Design an Informatica mapping to calculate the cumulative sum of PERSONAL_SALES and load that into target table?

The Expected Output in target should be

ID	PERSONAL_SALES	O_RETURN_VALUE
10	40000	40000
20	80000	120000
30	40000	160000
40	60000	220000
50	NULL	220000
60	50000	270000

Solution:

Connect the source Qualifier to expression transformation. In the expression transformation create a variable port named V_RETURN_VALUE and write the expression CUME(PERSONAL_SALES), again create a output port O_RETURN_VALUE and assign the expression V_RETURN_VALUE to it. Connect all the output ports to target and run the mapping.

In the expression transformation, the ports are

ID	// Input/Output Port
PERSONAL_SALES	// Input/Output Port
V_RETURN_VALUE = CUME(PERSONAL_SALES)	// Variable Port
O_RETURN_VALUE = V_RETURN_VALUE	// Output Port

Connect the output port from the expression transformation to the target and see the results.

3.2 Design a Informatica mapping to calculate the Moving sum of PERSONAL_SALES for two rows and load that into target table?

The Expected Output in target should be

ID	PERSONAL_SALES	O_RETURN_VALUE
10	40000	NULL
20	80000	120000
30	40000	120000
40	60000	100000
50	NULL	60000
60	50000	50000

Solution:

Connect the source Qualifier to expression transformation. In the expression transformation create a variable port named V_RETURN_VALUE and write the expression MOVINGSUM(PERSONAL_SALES,2) again create a output port O_RETURN_VALUE and assign the expression V_RETURN_VALUE to it. Connect all the output ports to target and run the mapping.

In the expression transformation, the ports are

ID	// Input/Output Port
PERSONAL_SALES	// Input/Output Port
V_RETURN_VALUE = MOVINGSUM(PERSONAL_SALES,2)	// Variable Port
O_RETURN_VALUE = V_RETURN_VALUE	// Output Port

Connect the output port to the target and see the results

3.3 Design a Informatica mapping to get the previous row value for the current row. If there is no previous row then load NULL as in case of first row there is no previous row

The Expected Output in target should be

ID	PERSONAL_SALES	O_RETURN_VALUE
10	40000	NULL
20	80000	40000
30	40000	80000
40	60000	40000
50	NULL	60000
60	50000	NULL

Solution:

Connect the source Qualifier to expression transformation. In the expression transformation, create a variable port V_COUNT and start incrementing it by one for each row entering the expression transformation. Also create V_SALES variable port and assign the expression IIF(V_COUNT=1,NULL,V_PREV_SALES) to it . Then Create one more variable port V_PREV_SALES and assign PERSONAL_SALES to it. Now create output port O_RETURN_VALUE and assign V_SALES to it. Connect the expression transformation to the target ports and run the mapping after creation of preceding session and workflow.

In the expression transformation, the ports will be

ID	// Input/Output Port
PERSONAL_SALES	// Input/Output Port
V_COUNT=V_COUNT+1	// Variable Port
V_SALES=IIF(V_COUNT=1,NULL,V_PREV_SALES)	// Variable Port
V_PREV_SALES=PERSONAL_SALES	// Variable Port
O_RETURN_VALUE=V_SALES	// Output Port

3.4 Design a mapping to get the next row Sales for the current row. If there is no next row for the current row, then the next row salary should be displayed as null.

The output should look like as

ID	PERSONAL_SALES	O_RETURN_VALUE
10	40000	80000
20	80000	40000
30	40000	60000
40	60000	NULL
50	NULL	50000

Solution:

1. Create a source qualifier transformation and connect it to two separate expression transformation. In each of the expression transformation create a variable port V_COUNT and increment it by 1 using expression V_COUNT+1. Then create an output port O_COUNT in each expression transformation. Assign the value V_COUNT in 1st expression and V_COUNT-1 in the 2nd expression as shown below.

In the 1st expression transformation, the ports will be

ID	// Input/Output Port
PERSONAL_SALES	// Input/Output Port
V_COUNT=V_COUNT+1	// Variable Port
O_COUNT=V_COUNT	// Output Port

1st Expression Data will look like

ID	PERSONAL_SALES	O_COUNT
10	40000	1
20	80000	2
30	40000	3
40	60000	4
50	NULL	5

In the 2nd expression transformation, the ports will be

ID	// Input/Output Port
PERSONAL_SALES	// Input/Output Port
V_COUNT=V_COUNT+1	// Variable Port
O_COUNT=V_COUNT-1	// Output Port

2nd Expression Data will look like

ID	PERSONAL_SALES	O_COUNT
10	40000	0
20	80000	1
30	40000	2
40	60000	3
50	NULL	4

2. Connect both the expression transformations ports to joiner transformation and join them on the port O_count. Consider the 1st expression transformation as Master and 2nd one as Detail Source. Specify the join type in Joiner transformation as Detail Outer Join. In the joiner transformation check the property sorted input this will allow you to connect both expression transformations to joiner transformation.
3. Pass the output of joiner transformation to a target table. In the Joiner transformation, connect the ID and PERSONAL_SALES from the 1st expression transformation to the ID and PERSONAL_SALES in target table. Connect the PERSONAL_SALES which is coming from the 2nd expression transformation to the next row sales port in the target table.

3.5 Design a Informatica Mapping to find the sum of all sales of an employee id and this sum should come in another column for all rows.

The output should look like as

ID	PERSONAL_SALES	O_RETURN_VALUE
10	40000	270000
20	80000	270000
30	40000	270000
40	60000	270000
50	NULL	270000
60	50000	270000

Solution:

Create a source qualifier transformation and connect it to the Aggregator transformation. In aggregator transformation create a Output port O_RETURN_VALUE and assign expression SUM(PERSONAL_SALES). Don not specify and group by for any filed and then connect all the other output ports to target and runt the mapping.

In the Aggregator transformation, the ports will be

ID	// Input/Output Port
PERSONAL_SALES	// Input/Output Port
O_RETURN_VALUE=SUM(PERSONAL_SALES)	// Output Port

How to achieve ? The Emp table contains the salary and commission in USD, in the target the com and sal will converted to a given currency prefix ex: Rs.

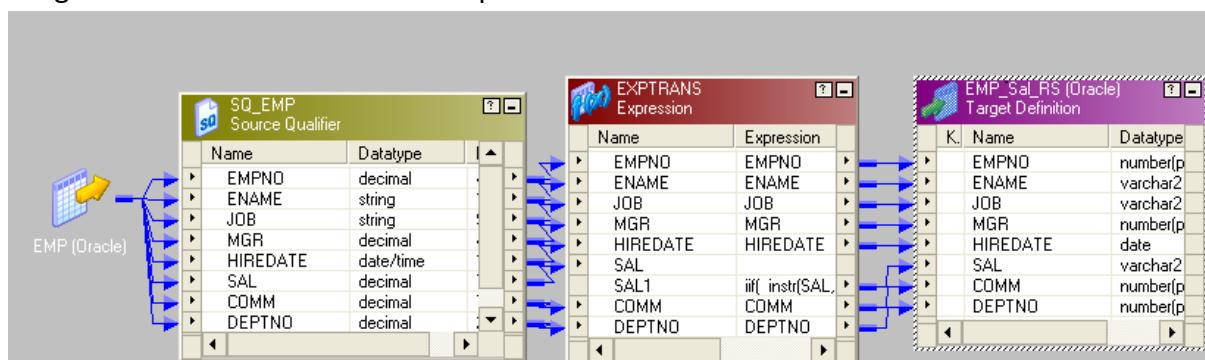
Source

EMPNO	ENAME	JOB	MGR	HIREDATE	SAL	DEPTNO
7369	SMITH	CLERK		7902 17-DEC-80	\$800	20
7499	ALLEN	SALESMAN		7698 20-FEB-81	\$1600	30

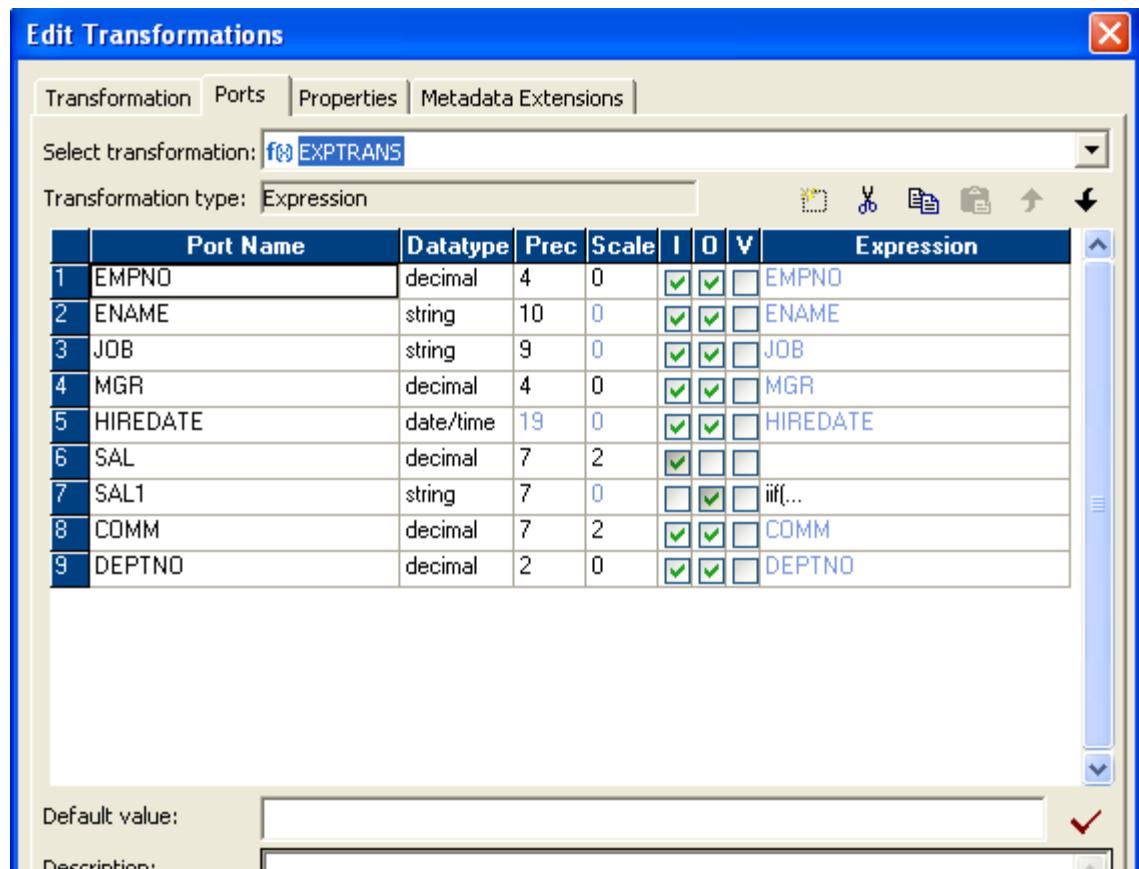
Target

EMPNO	ENAME	JOB	MGR	HIREDATE	SAL	DEPTNO
7369	SMITH	CLERK		7902 17-DEC-80	Rs.800	20
7499	ALLEN	SALESMAN		7698 20-FEB-81	RS.1600	30

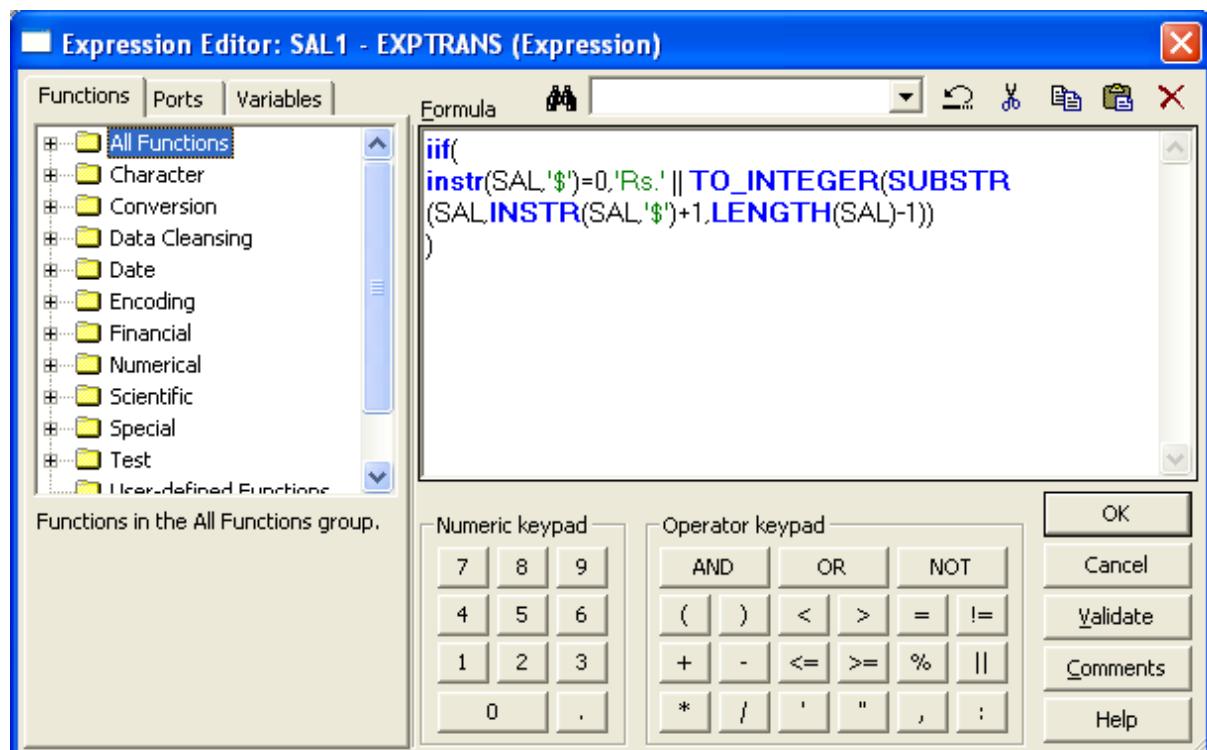
1. Drag the source and connect it to expression transformation



2. In expression make a output port sal1 and make sal as input port only.



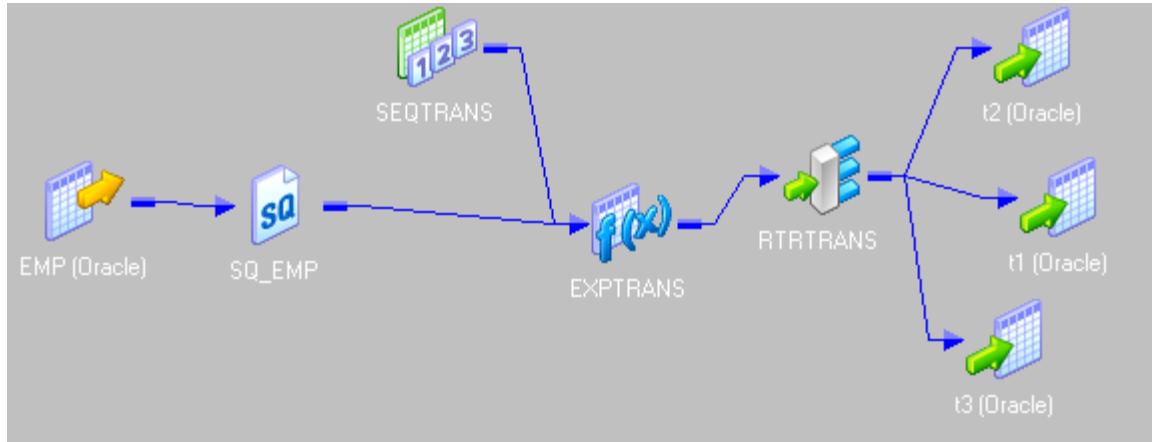
3. In sal1 write the condition as like bellow



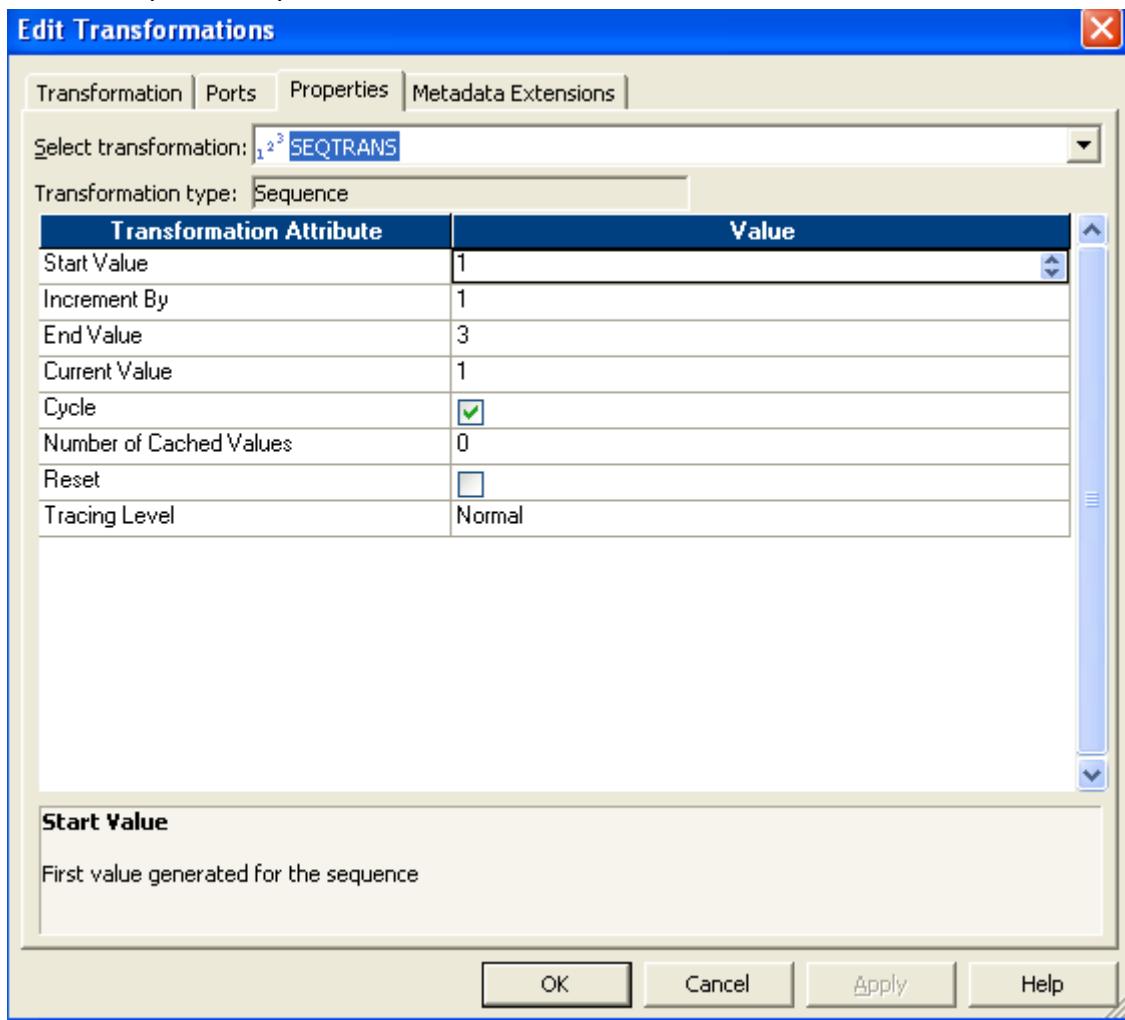
4. Then send it to target.

How to achieve ? In source there are some record. Suppose I want to send three targets. First record will go to first target, Second one will go to second target and third record will go to third target and then 4th to 1st,5th to 2nd , 6th to 3rd and so on.

- Put the source to mapping and connect it to an expression transformation.



- Drag an sequence generator transformation and set properties like this And connect the next value port to expression.



3. Drag all output port of expression to router. In router make three groups and give the conditions Like

Group Name	Group Filter Condition
NEWGROUP1	NEXTVAL = 1
NEWGROUP2	NEXTVAL = 2
NEWGROUP3	NEXTVAL = 3
DEFAULT1	

4. connect desire group to desire target .

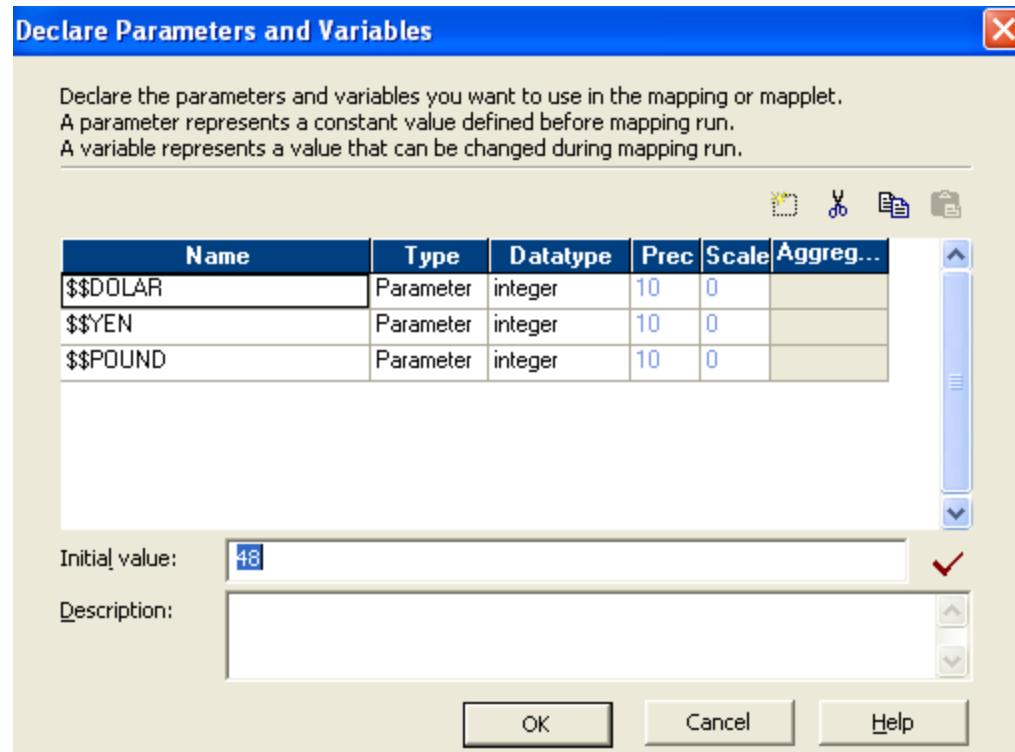
How to achieve ? Suppose that a source contains a column which holds the salary information prefixed with the currency code , for example

EMPNO	ENAME	JOB	MGR	HIREDATE	SAL	DEPTNO
7369	SMITH	CLERK	7902	17-DEC-80	\$300	20
7499	ALLEN	SALESMAN	7698	20-FEB-81	£1600	30
7521	WARD	SALESMAN	7698	22-FEB-81	¥8500	30

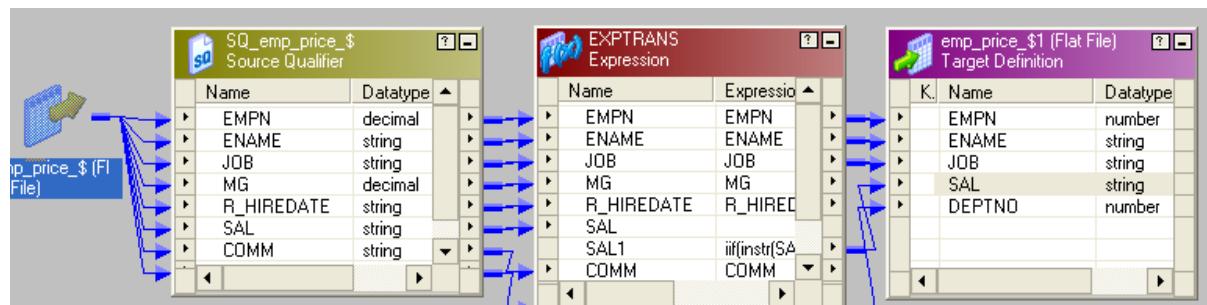
In the target different currency will evaluate to a single currency value, for example convert all to Rupees.

1. First thing we should consider that there are different types of currency like pound, dollar, yen etc. So it's a good idea to use mapping parameter or variable. Go to mapping=> mapping parameter and variables then create three parameters (for this example) and set its initial

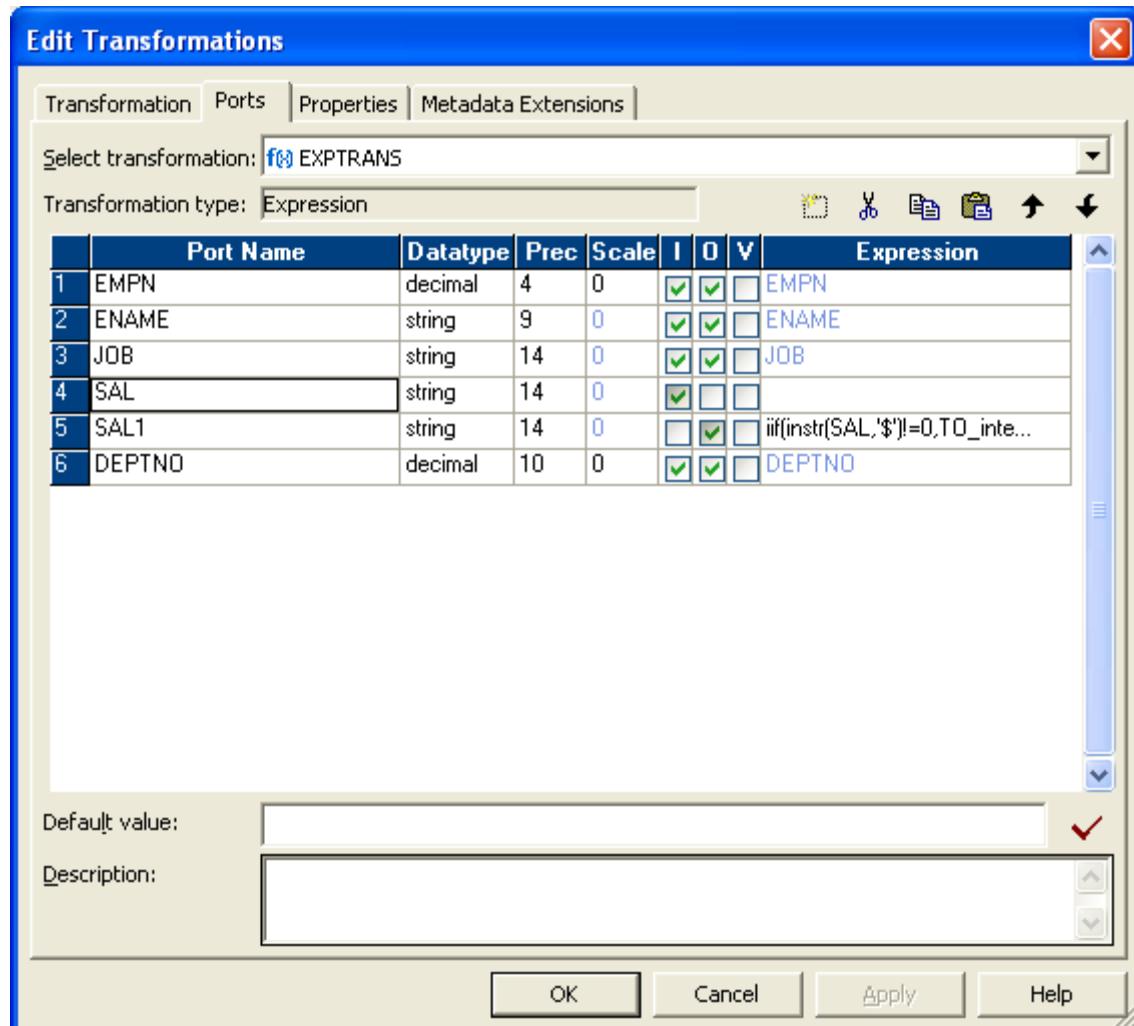
value as bellow



- Then drag the source to mapping area and connect to an expression transformation.

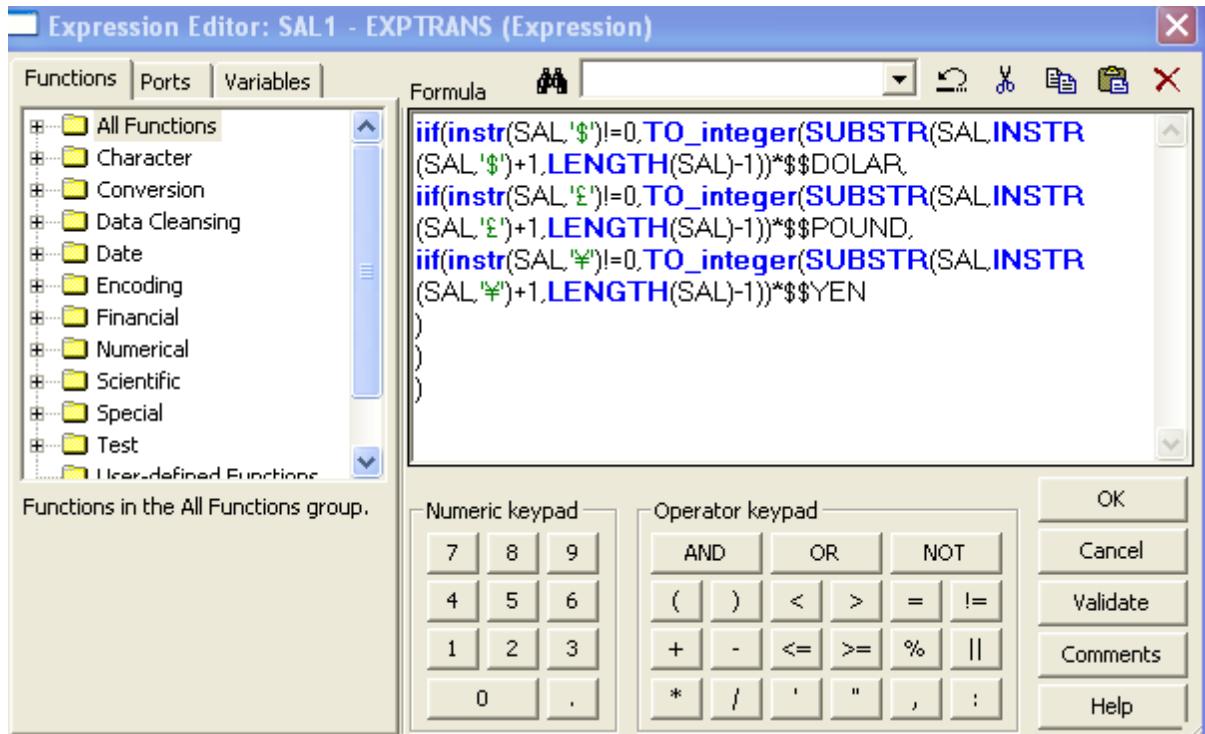


3. In expression create a output port as sal1 and make sal as input only as bellow.



4. In sal1 port write the condition as below

```
iif(instr(SAL,'$')!=0,TO_integer(SUBSTR(SAL,INSTR(SAL,'$')+1,LENGTH(SAL)-1))*$$DOLAR,
iif(instr(SAL,'£')!=0,TO_integer(SUBSTR(SAL,INSTR(SAL,'£')+1,LENGTH(SAL)-1))*$$POUND,
iif(instr(SAL,'¥')!=0,TO_integer(SUBSTR(SAL,INSTR(SAL,'¥')+1,LENGTH(SAL)-1))*$$YEN
)
)
)
```



\$\$DOLLAR, \$\$POUND, \$\$YEN these are mapping parameter . you can multiply price in rupee directly for example dollar price in rupees i.e 48 .

5. Connect required output port from expression to target directly. And run the session.

Consider the following employees table as source

department_no	employee_name
20, R	
10, A	
10, D	
20, P	
10, B	
10, C	
20, Q	
20, S	

Design a mapping to load a target table with the following values from the above source?

department_no, employee_list

10, A
10, A,B
10, A,B,C
10, A,B,C,D
20, A,B,C,D,P

20, A,B,C,D,P,Q

20, A,B,C,D,P,Q,R

20, A,B,C,D,P,Q,R,S

Solution:

Step1: Use a sorter transformation and sort the data using the sort key as department_no and then pass the output to the expression transformation. In the expression transformation, the ports will be

department_no

employee_name

V_employee_list =

IIF(ISNULL(V_employee_list),employee_name,V_employee_list||'|||employee_name)

O_employee_list = V_employee_list

Step2: Now connect the expression transformation to a target table.

Design a mapping to load a target table with the following values from the above source?

department_no, employee_list

10, A

10, A,B

10, A,B,C

10, A,B,C,D

20, P

20, P,Q

20, P,Q,R

20, P,Q,R,S

Solution:

Step1: Use a sorter transformation and sort the data using the sort key as department_no and then pass the output to the expression transformation. In the expression transformation, the ports will be

department_no

employee_name

V_curr_deptno=department_no

V_employee_list = IIF(V_curr_deptno!=

V_prev_deptno,employee_name,V_employee_list||'|||employee_name)

V_prev_deptno=department_no

O_employee_list = V_employee_list

Step2: Now connect the expression transformation to a target table.

Design a mapping to load a target table with the following values from the above source?

department_no, employee_names

10, A,B,C,D

20, P,Q,R,S

Solution:

The first step is same as the above problem. Pass the output of expression to an aggregator transformation and specify the group by as department_no. Now connect the aggregator transformation to a target table.

How to achieve ? Reading a source file with salary prefix \$, in the target the Sal column must store in number .

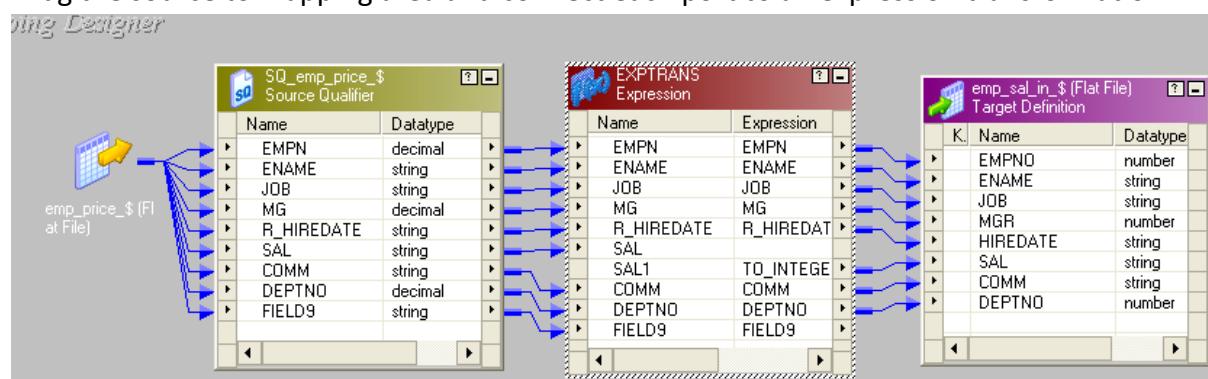
Source

EMPNO	ENAME	JOB	MGR	HIREDATE	SAL	DEPTNO
7369	SMITH	CLERK	7902	17-DEC-80	\$800	20
7499	ALLEN	SALESMAN	7698	20-FEB-81	\$1600	30

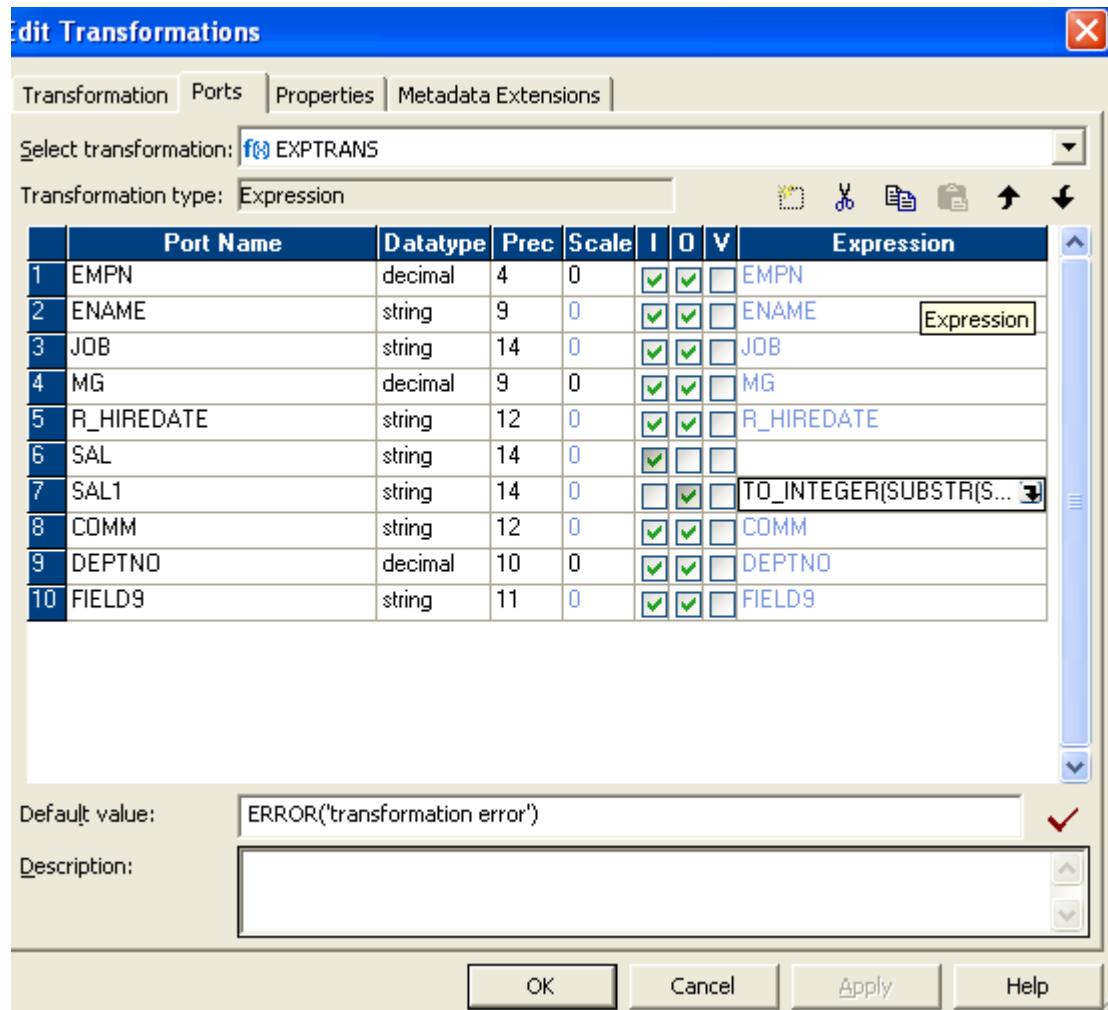
Target

EMPNO	ENAME	JOB	MGR	HIREDATE	SAL	DEPTNO
7369	SMITH	CLERK	7902	17-DEC-80	800	20
7499	ALLEN	SALESMAN	7698	20-FEB-81	1600	30

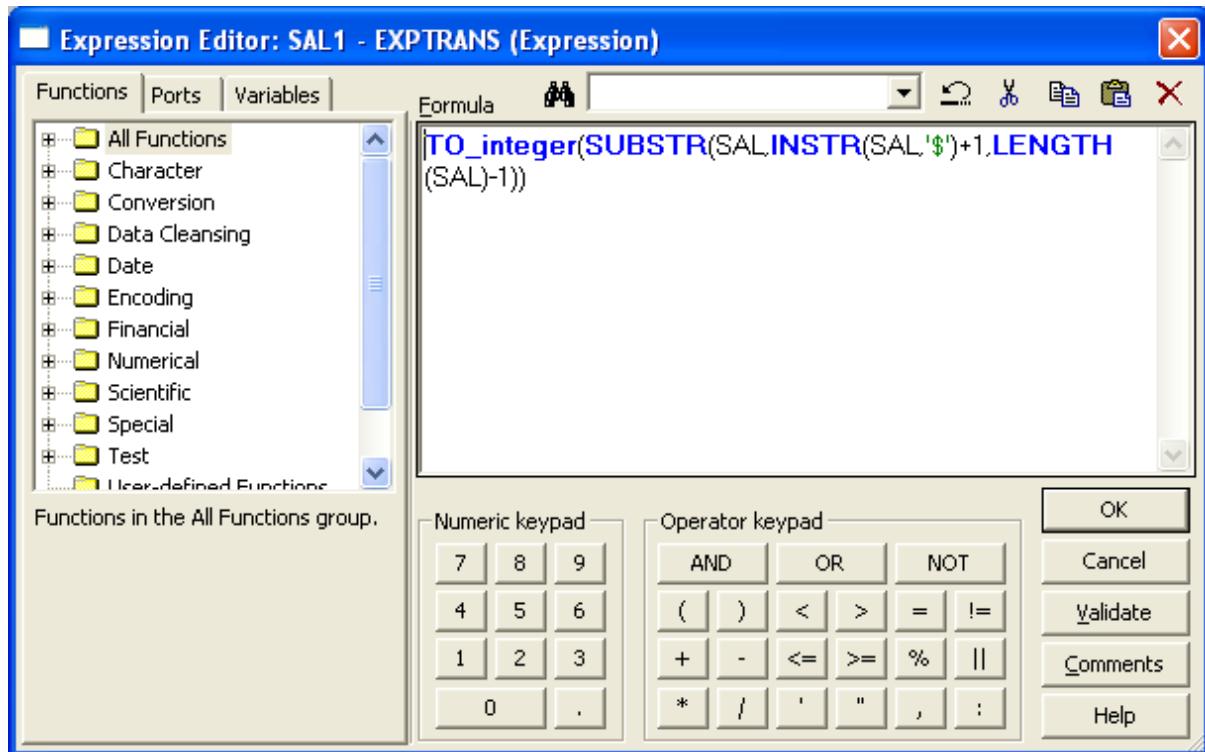
1. Drag the source to mapping area and connect each port to an expression transformation.



2. In expression transformation add a new col sal1 and make it as output and sal as input only as shown in picture.



3. In expression write the condition like this.



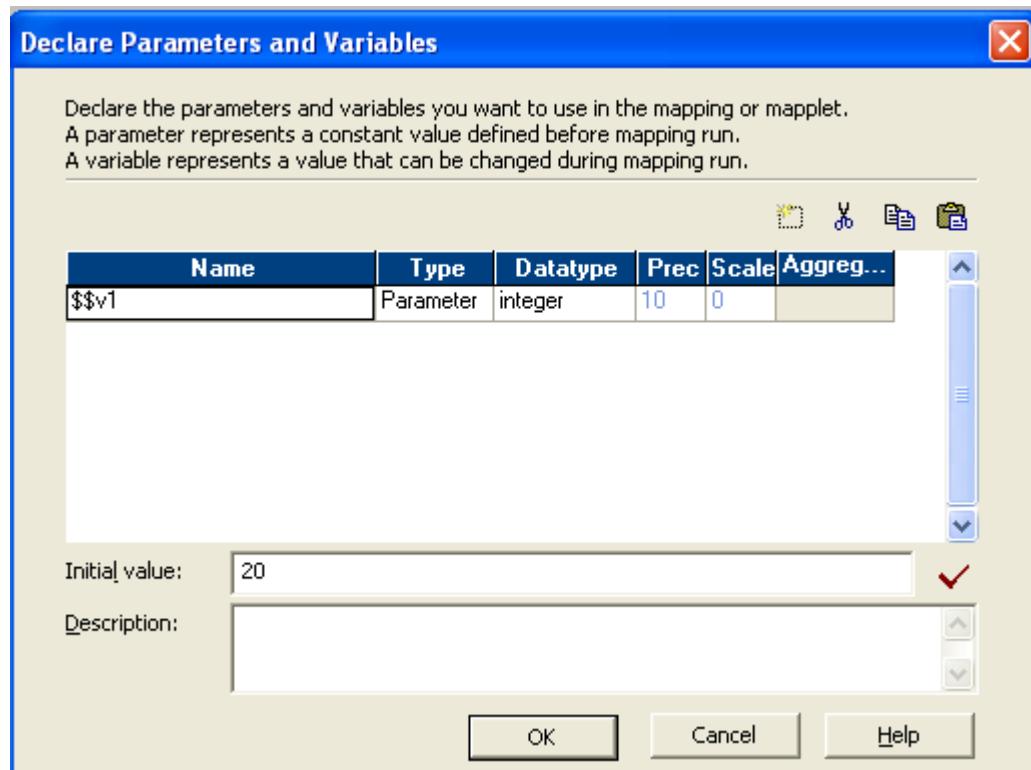
4. connect the required port to target.

How to use mapping parameter and variable in mapping ?

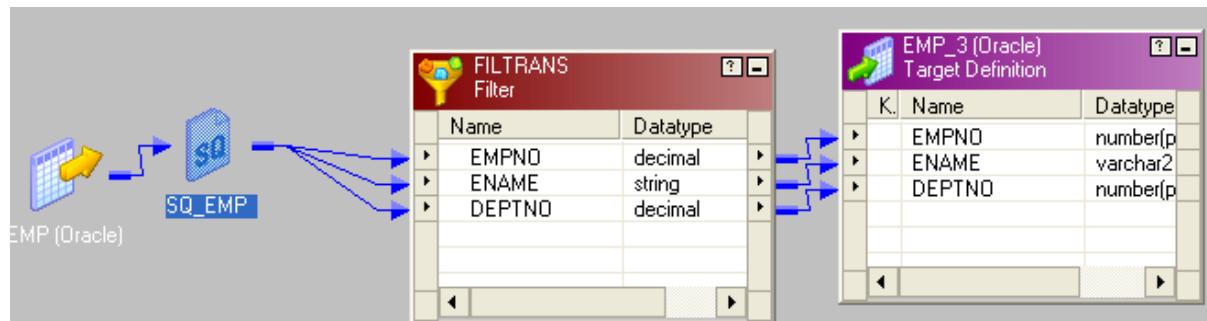
Solution:

1. Go to mapping then parameter and variable tab in the Informatica designer. Give name as \$\$v1, type choose parameter (You can also choose variable), data type as integer and

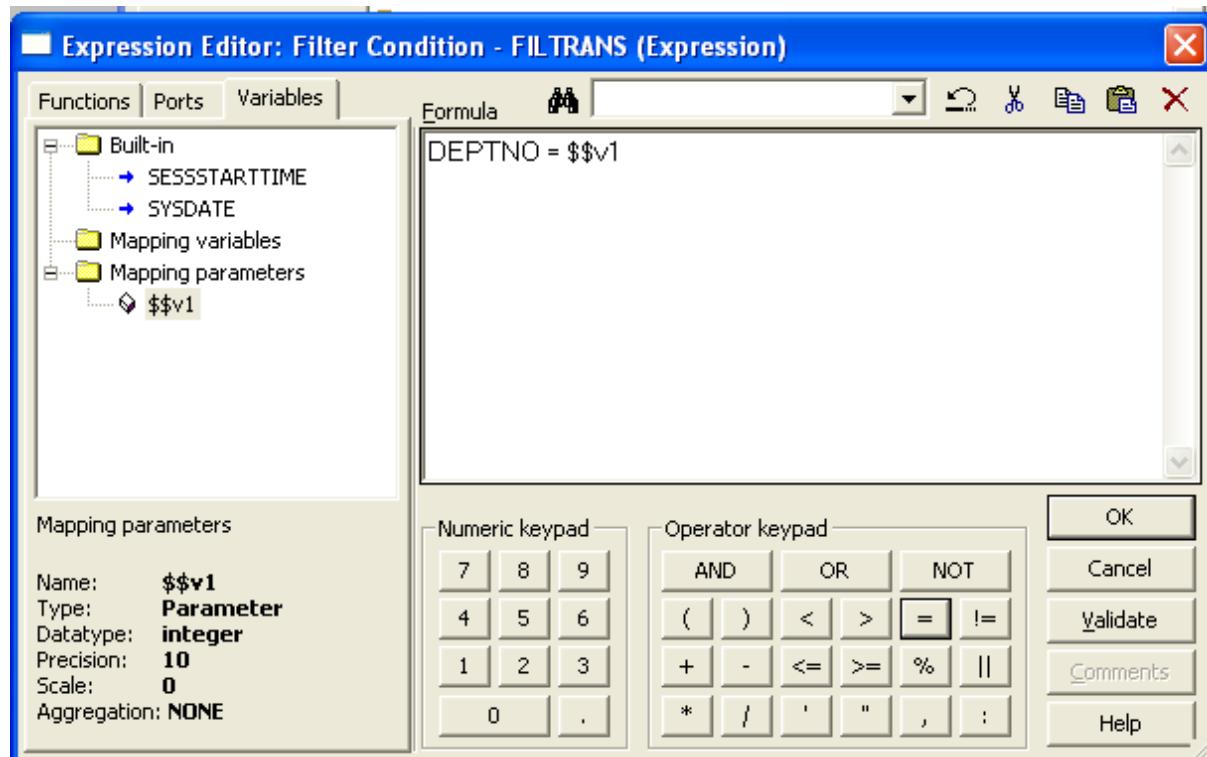
give initial value as 20.



2. Create a mapping as shown in the figure(I have considered a simple scenario where a particular department id will be filtered to the target).



3. In filter set deptno=\$\$v1 (that means only dept no 20 record will go to the target.)



4. Mapping parameter value can't change throughout the session but variable can be changed.
We can change variable value by using text file. I'll show it in next scenario.

Scenario 4

Consider the following product types data as the source.

Product_id, product_type

10, video

10, Audio

20, Audio

30, Audio

40, Audio

50, Audio

10, Movie

20, Movie

30, Movie

40, Movie

50, Movie

60, Movie

Assume that there are only 3 product types are available in the source. The source contains 12 records and you dont know how many products are available in each product type.

4.1 Design a mapping to select 9 products in such a way that 3 products should be selected from video, 3 products should be selected from Audio and the remaining 3 products should be selected from Movie.

Solution:

Step1: Use sorter transformation and sort the data using the key as product_type.

Step2: Connect the sorter transformation to an expression transformation. In the expression transformation, the ports will be

```
product_id  
product_type  
V_curr_prod_type=product_type  
V_count = IIF(V_curr_prod_type = V_prev_prod_type, V_count+1, 1)  
V_prev_prod_type=product_type  
O_count=V_count
```

Step3: Now connect the expression transformation to a filter transformation and specify the filter condition as O_count<=3. Pass the output of filter to a target table.

4.2 In the above problem Q1, if the number of products in a particular product type are less than 3, then you wont get the total 9 records in the target table. For example, see the videos type in the source data. Now design a mapping in such way that even if the number of products in a particular product type are less than 3, then you have to get those less number of records from another product types. For example: If the number of products in videos are 1, then the remaining 2 records should come from audios or movies. So, the total number of records in the target table should always be 9.

Solution:

The first two steps are same as above.

Step3: Connect the expression transformation to a sorter transformation and sort the data using the key as O_count. The ports in soter transformation will be

```
product_id  
product_type
```

O_count (sort key)

Step3: Discard O_count port and connect the sorter transformation to an expression transformation. The ports in expression transformation will be

product_id
product_type
V_count=V_count+1
O_prod_count=V_count

Step4: Connect the expression to a filter transformation and specify the filter condition as O_prod_count<=9. Connect the filter transformation to a target table.

How to achieve ? Design a mapping to convert column data into row data without using the normalizer transformation.

The source data looks like

col1, col2, col3
a, b, c
d, e, f

The target table data should look like

Col
a
b
c
d
e
f

Solution:

Create three expression transformations with one port each. Connect col1 from Source Qualifier to port in first expression transformation. Connect col2 from Source Qualifier to port in second expression transformation. Connect col3 from source qualifier to port in third expression transformation. Create a union transformation with three input groups and each input group should have one port. Now connect the expression transformations to the input groups and connect the union transformation to the target table.

How to achieve ? Design a mapping to convert row data into column data.

The source data looks like

```
id, value  
10, a  
10, b  
10, c  
20, d  
20, e  
20, f
```

The target table data should look like

```
id, col1, col2, col3  
10, a, b, c  
20, d, e, f
```

Solution:

Step1: Use sorter transformation and sort the data using id port as the key. Then connect the sorter transformation to the expression transformation.

Step2: In the expression transformation, create the ports and assign the expressions as mentioned below.

```
id  
value  
V_curr_id=id  
V_count= IIF(v_curr_id=V_prev_id,V_count+1,1)  
V_prev_id=id  
O_col1= IIF(V_count=1,value,NULL)  
O_col2= IIF(V_count=2,value,NULL)  
O_col3= IIF(V_count=3,value,NULL)
```

Step3: Connect the expression transformation to aggregator transformation. In the aggregator transformation, create the ports and assign the expressions as mentioned below.

```
id (specify group by on this port)  
O_col1  
O_col2
```

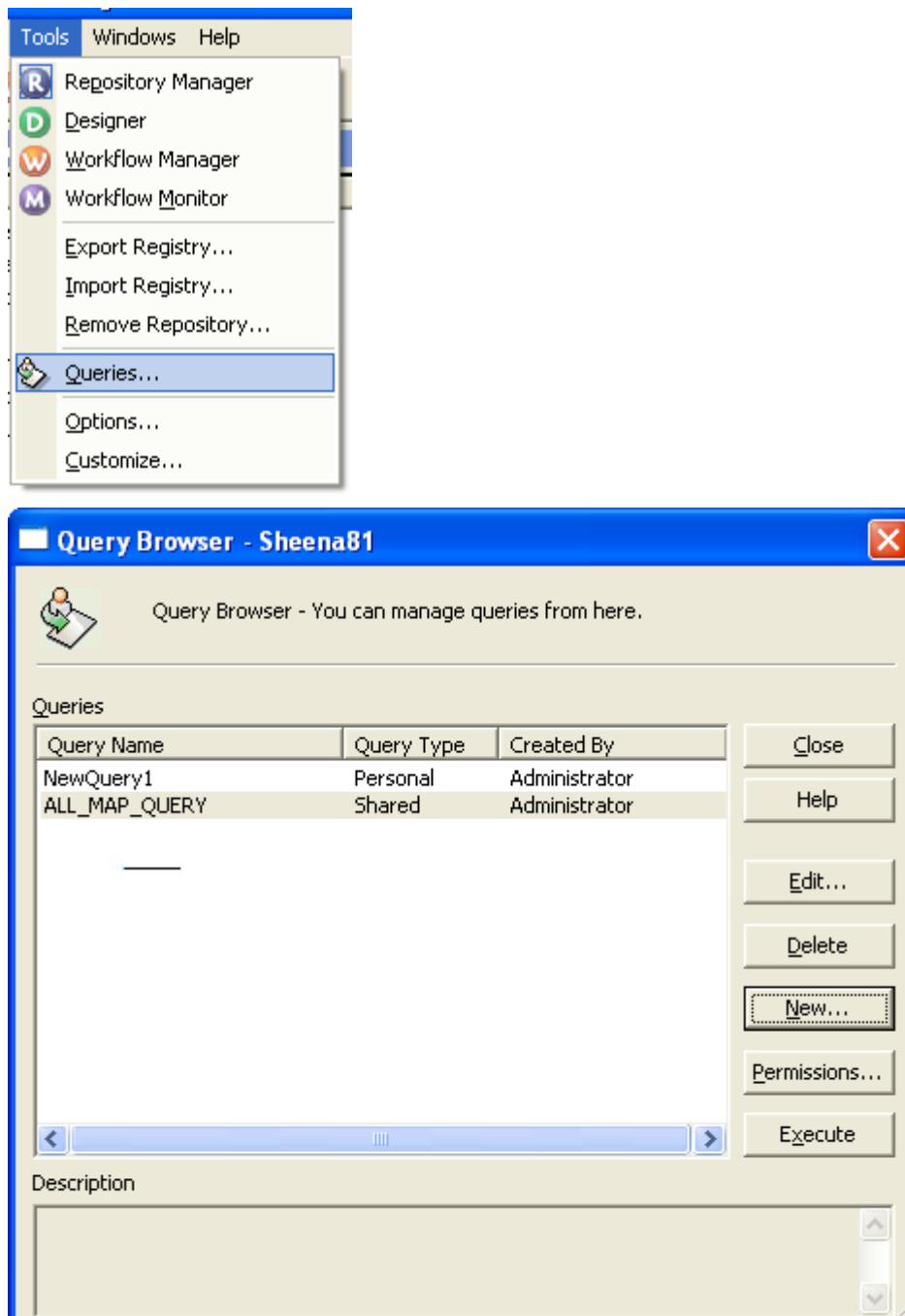
```
O_col3  
col1=MAX(O_col1)  
col2=MAX(O_col2)  
col3=MAX(O_col3)
```

Step4: Now connect the ports id, col1, col2, col3 from aggregator transformation to the target table.

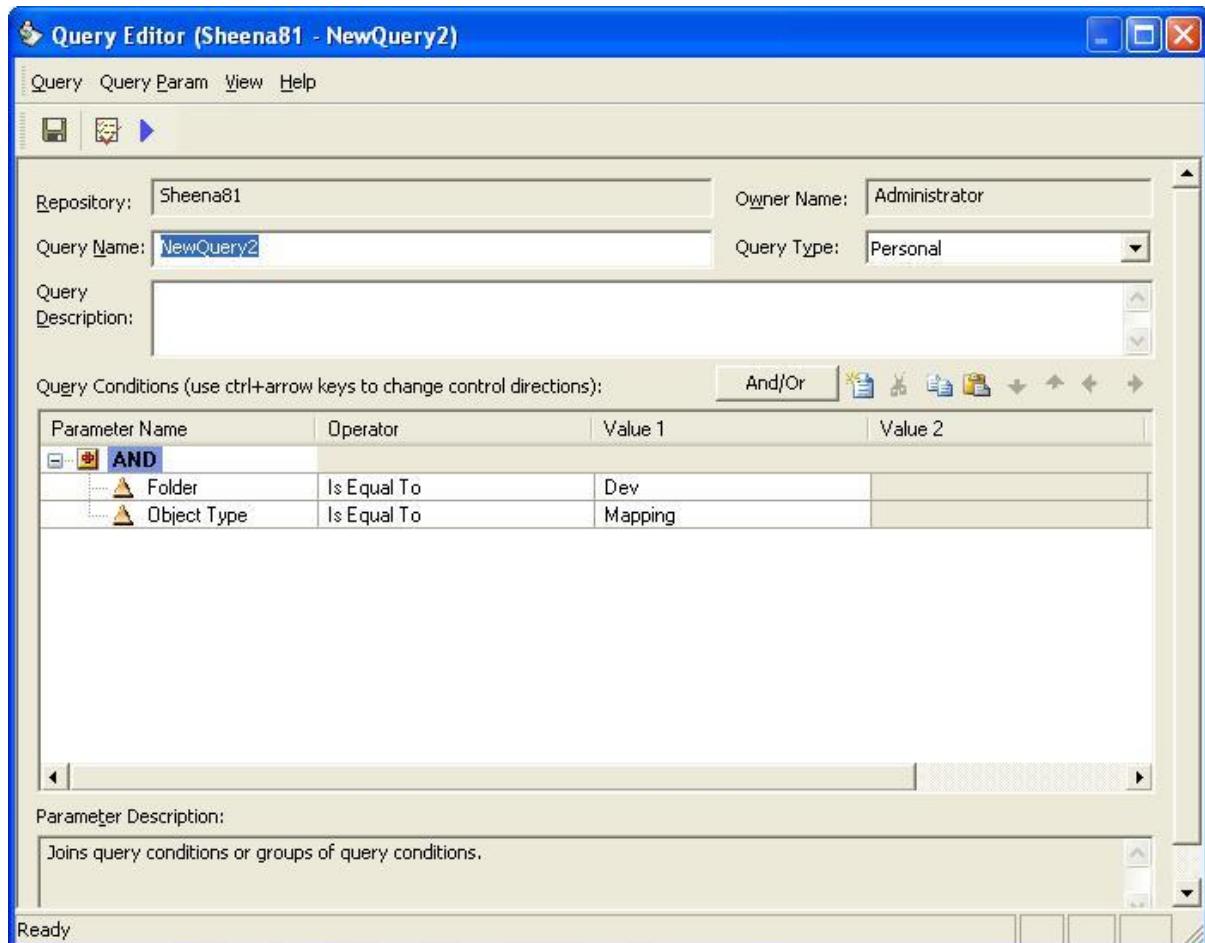
How validate all mapping in repository ?

Solution:

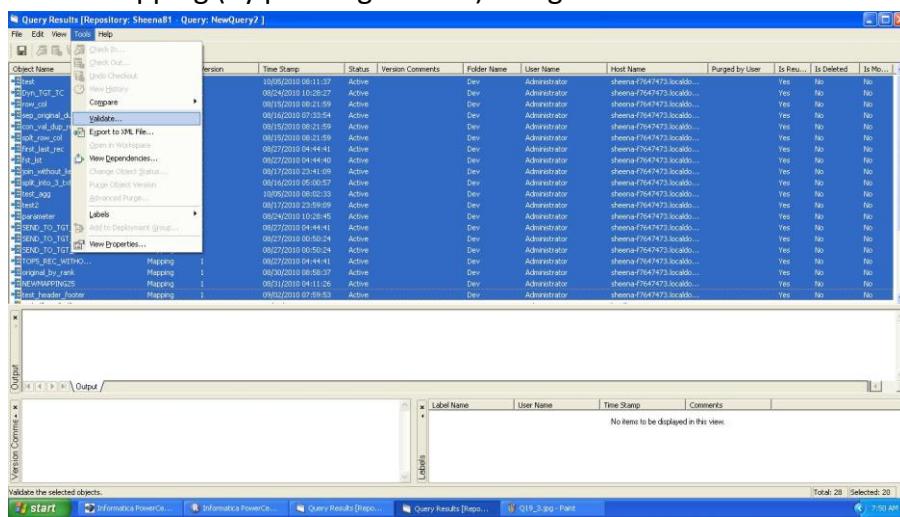
1. In repository go to menu “tool” then “queries”. Query Browser dialog box will appear. Then click on new button.



2. In Query Editor, choose folder name and object type as I have shown in the picture.



3. After that, execute it (by clicking the blue arrow button).
 4. Query results window will appear. You select single mapping (by selecting single one) or whole mapping (by pressing Ctrl + A) and go to "tools" then "validate" option to validate it.



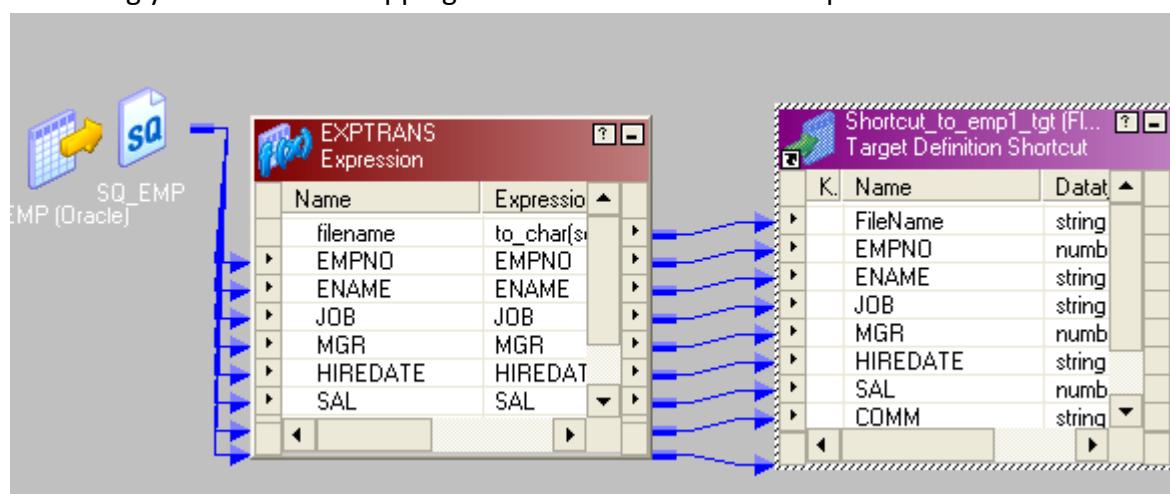
How to generate file name dynamically with name of sys date ?

Solution:

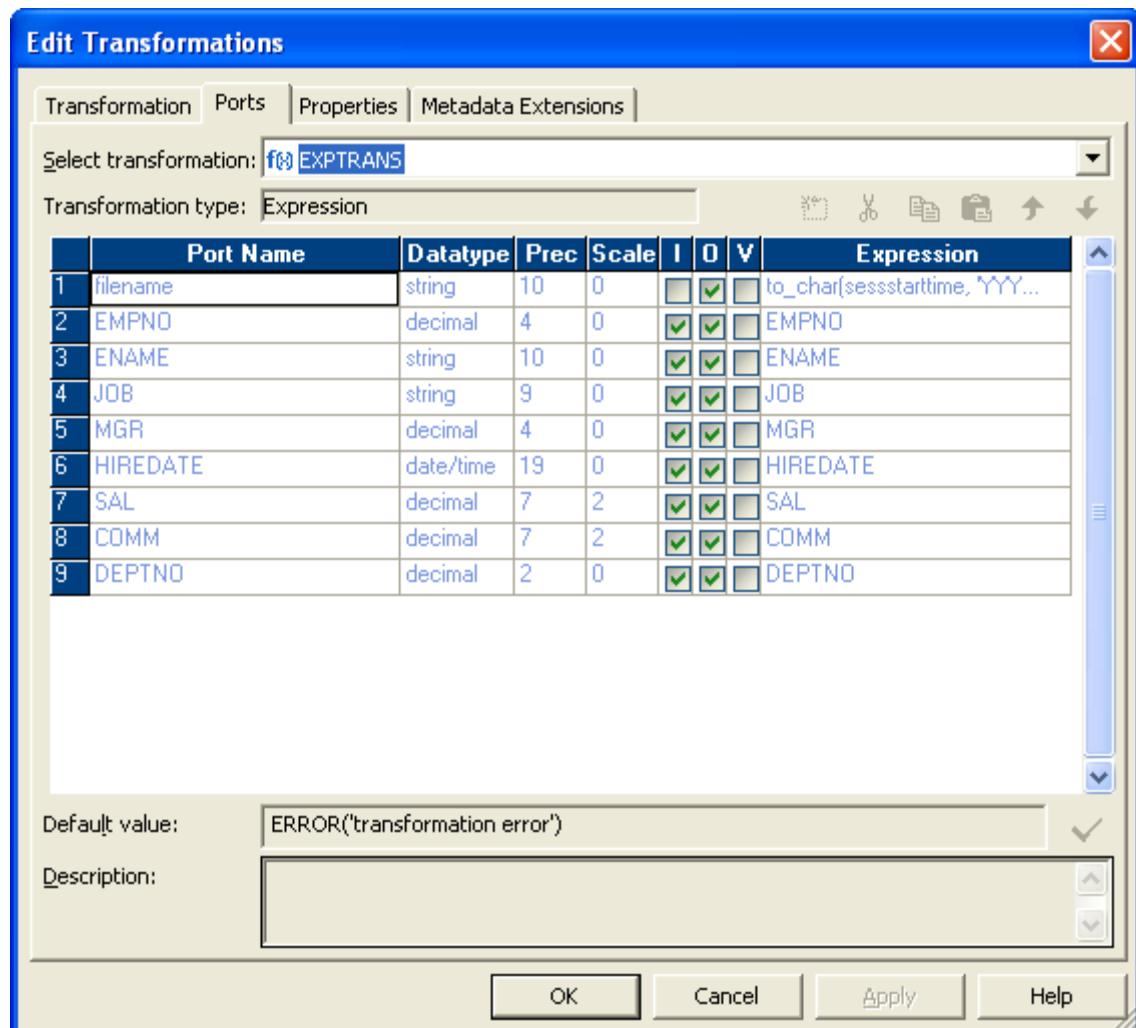
1. Drag your target file to target designer and add a column as show on the picture. It's not a normal column .click on the 'add file name to the table' property. (I have given a red mark there)

	Column Name	Datatype	Prec	Scale	No...	Format	Key Type	Add File
1	EMPN	number	4	0			NOT A KEY	
2	O_ENAME	string	8	0			NOT A KEY	
3	JOB	string	15	0			NOT A KEY	
4	MG	number	9	0			NOT A KEY	
5	R_HIREDATE	string	12	0			NOT A KEY	
6	SAL	number	11	0			NOT A KEY	
7	COMM	string	10	0			NOT A KEY	
8	DEPTNO	number	25	0			NOT A KEY	

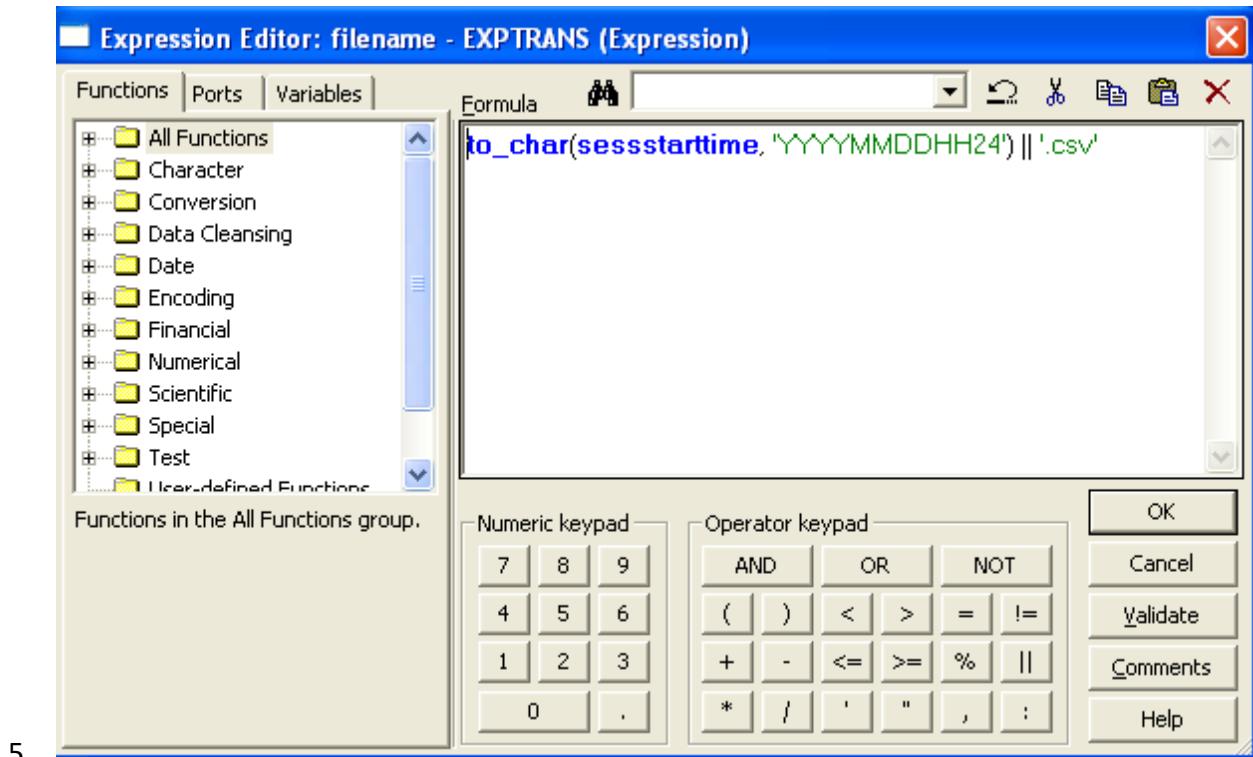
2. Then drag your source to mapping area and connect it to an expression transformation.



3. In expression transformation add a new port as string data type and make it output port.



4. In that output port write the condition like describe as bellow and then map it in to filename port of target. Also send other ports to target. Finally run the session. You will find two file one with sys date and other one is '.out' file which one you can delete.

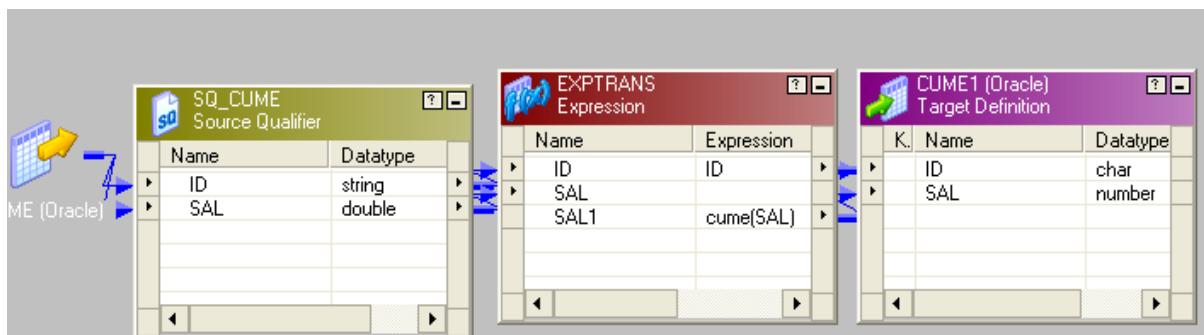


How to produce rows in target table with every row as sum of all previous rows in source table ? See the source and target table to understand the scenario.

SOURCE TABLE	
id	Sal
1	200
2	300
3	500
4	560

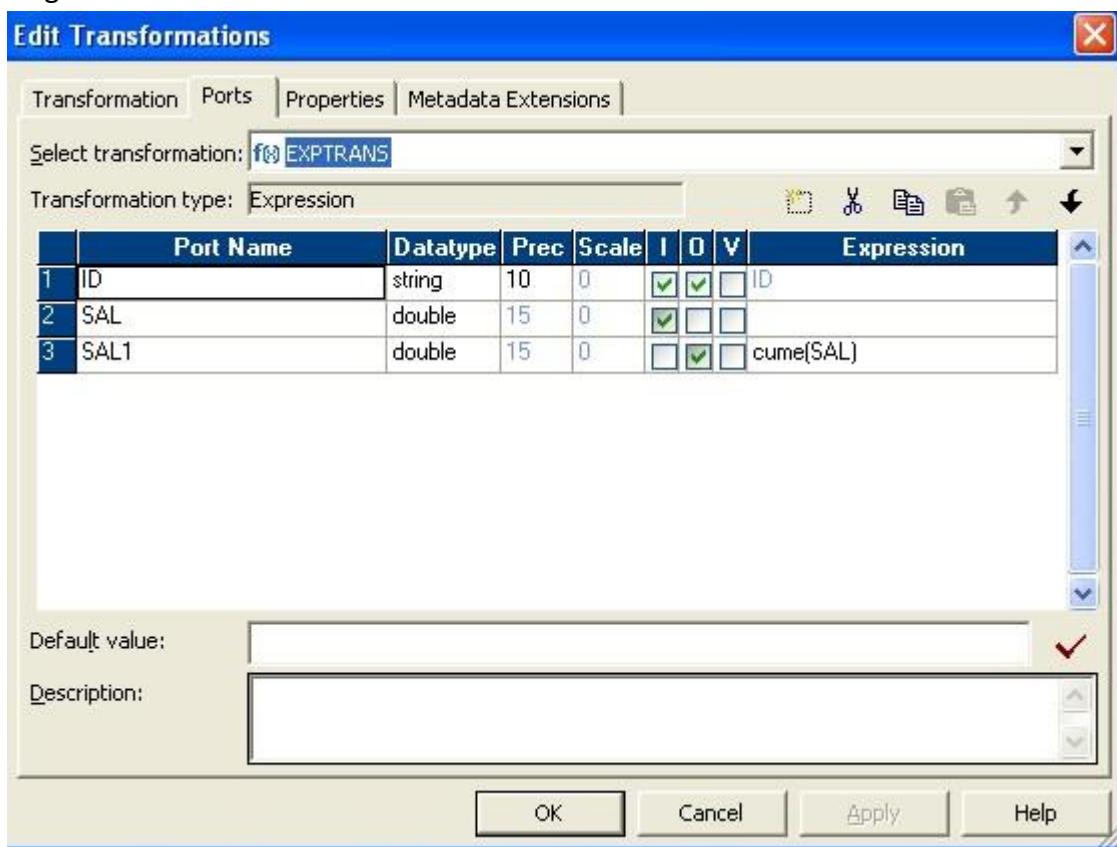
TARGET TABLE	
Id	Sal
1	200
2	500
3	1000
4	1560

1.



2. In expression add one column and make it output(sal1) and sal port as input only.

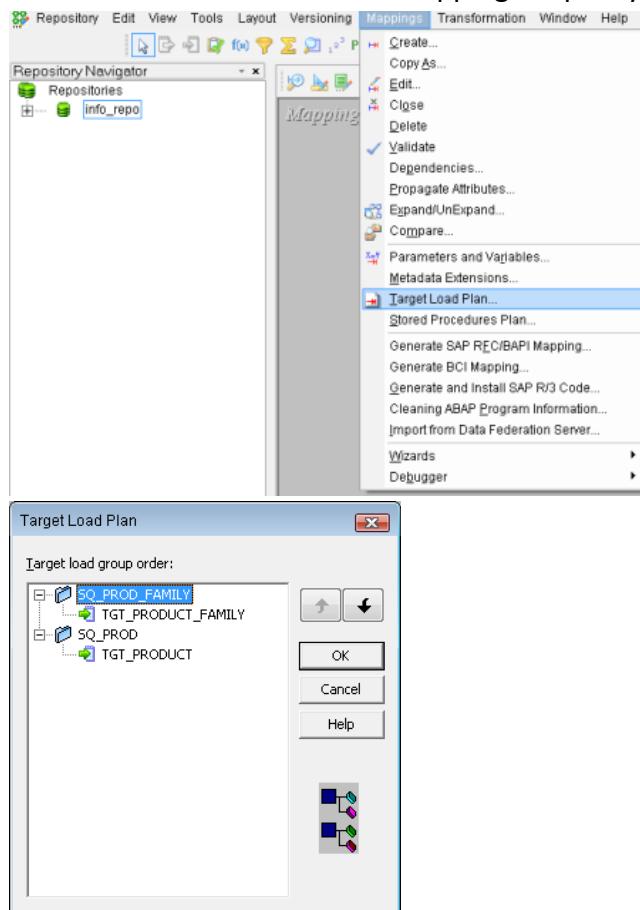
We will make use of a function named cume() to solve our problem, rather using any complex mapping. Write the expression in sal1 as cume(sal) and send the output rows to target.



Suppose we have two Source Qualifier transformations SQ1 and SQ2 connected to Target tables TGT1 and TGT2 respectively. How do you ensure TGT2 is loaded after TGT1?

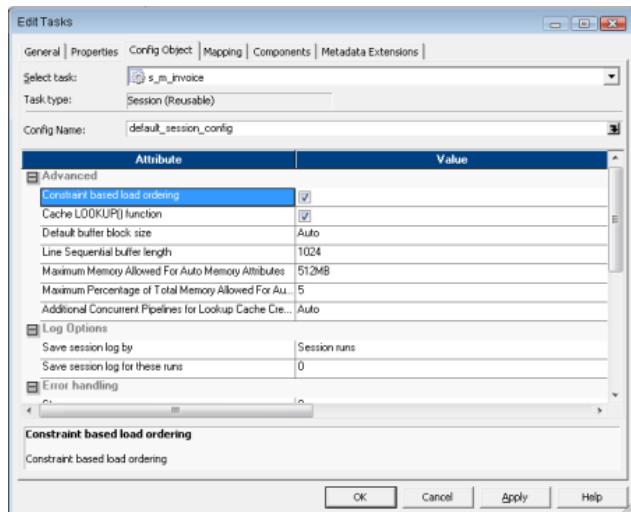
If we have multiple Source Qualifier transformations connected to multiple targets, we can designate the order in which the Integration Service loads data into the targets.

In the Mapping Designer, We need to configure the **Target Load Plan** based on the Source Qualifier transformations in a mapping to specify the required loading order.



Suppose we have a Source Qualifier transformation that populates two target tables. How do you ensure TGT2 is loaded after TGT1?

In the Workflow Manager, we can Configure **Constraint based load ordering** for a session. The Integration Service orders the target load on a row-by-row basis. For every row generated by an active source, the Integration Service loads the corresponding transformed row first to the primary key table, then to the foreign key table. Hence if we have one Source Qualifier transformation that provides data for multiple target tables having primary and foreign key relationships, we will go for Constraint based load ordering.



Scenario: You have two columns in source table T1, in which the col2 may contain duplicate values. All the duplicate values in col2 of will be transformed as comma separated in the column col2 of target table T2.

Source Table: T1

Col1	Col2
a	x
b	y
c	z
a	m
b	n

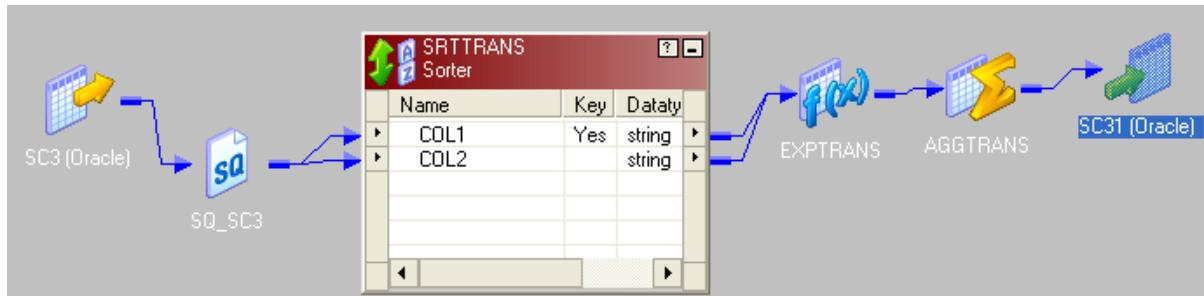
Target Table: T2

col1	col2
a	x,m
b	y,n
c	z

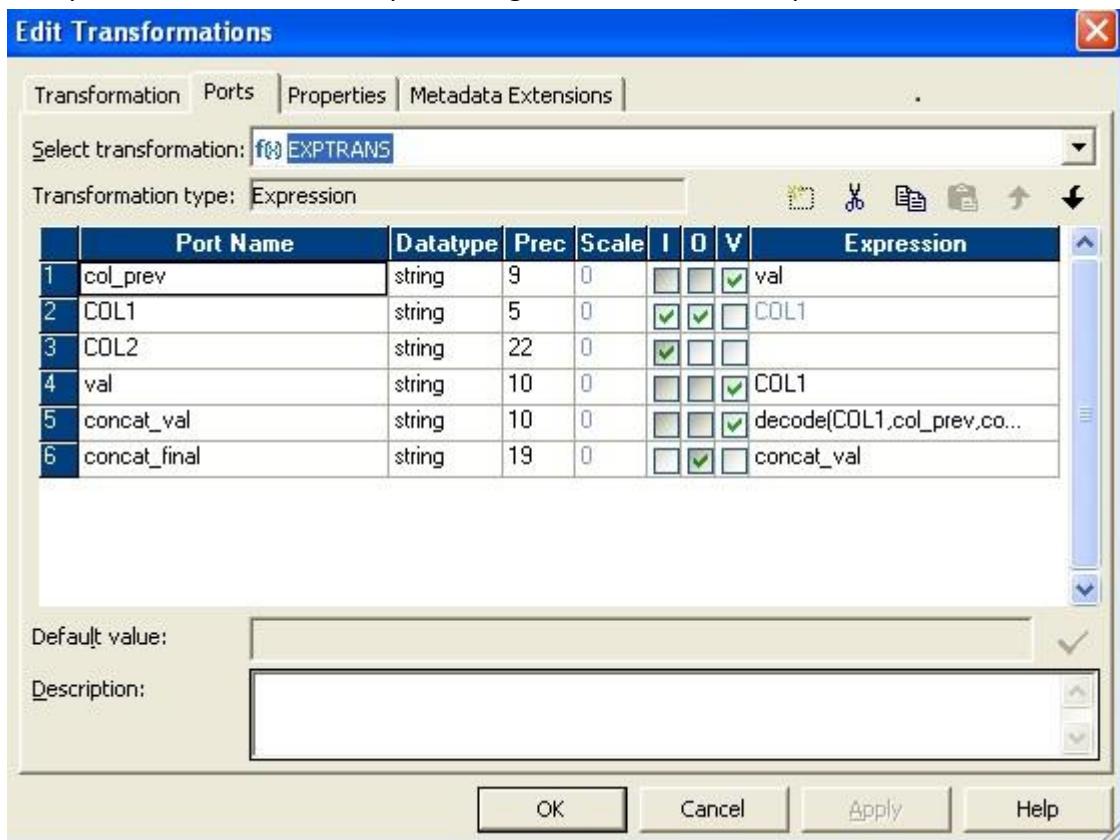
Solution:

- We have to use the following transformation as below.

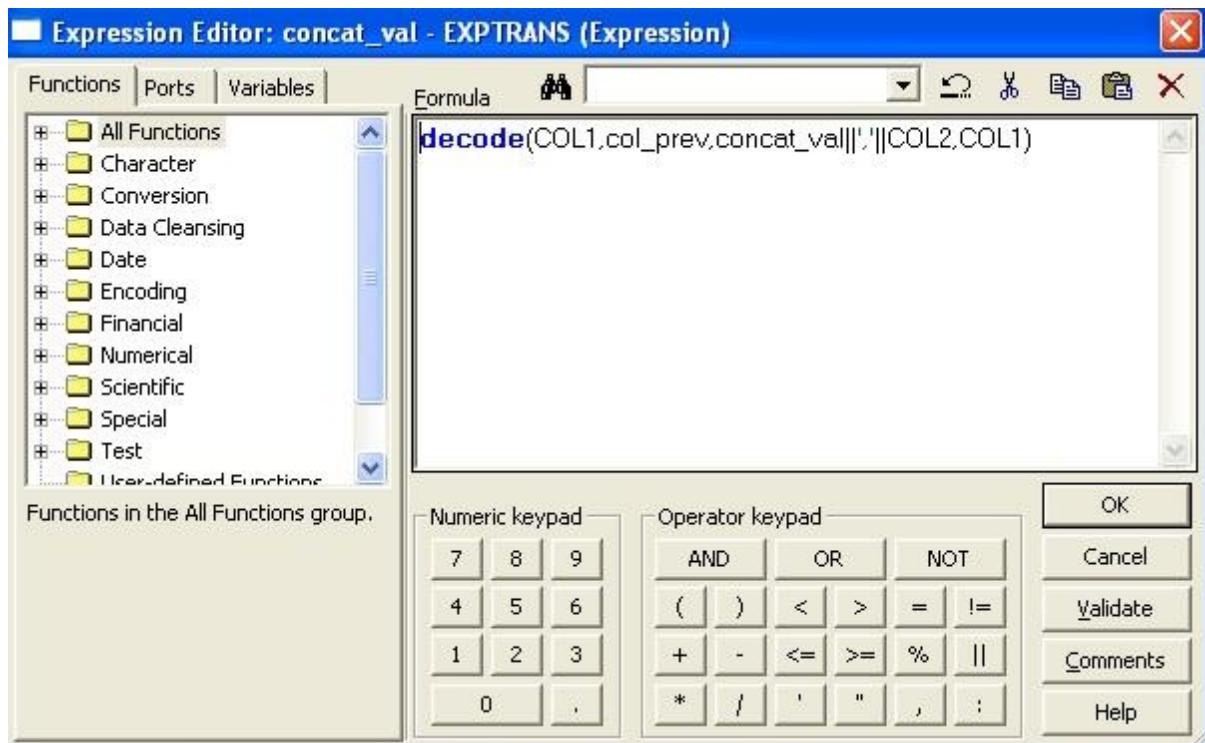
First connect a sorter transformation to source and make col1 as key and its order is ascending. After that connect it to an expression transformation.



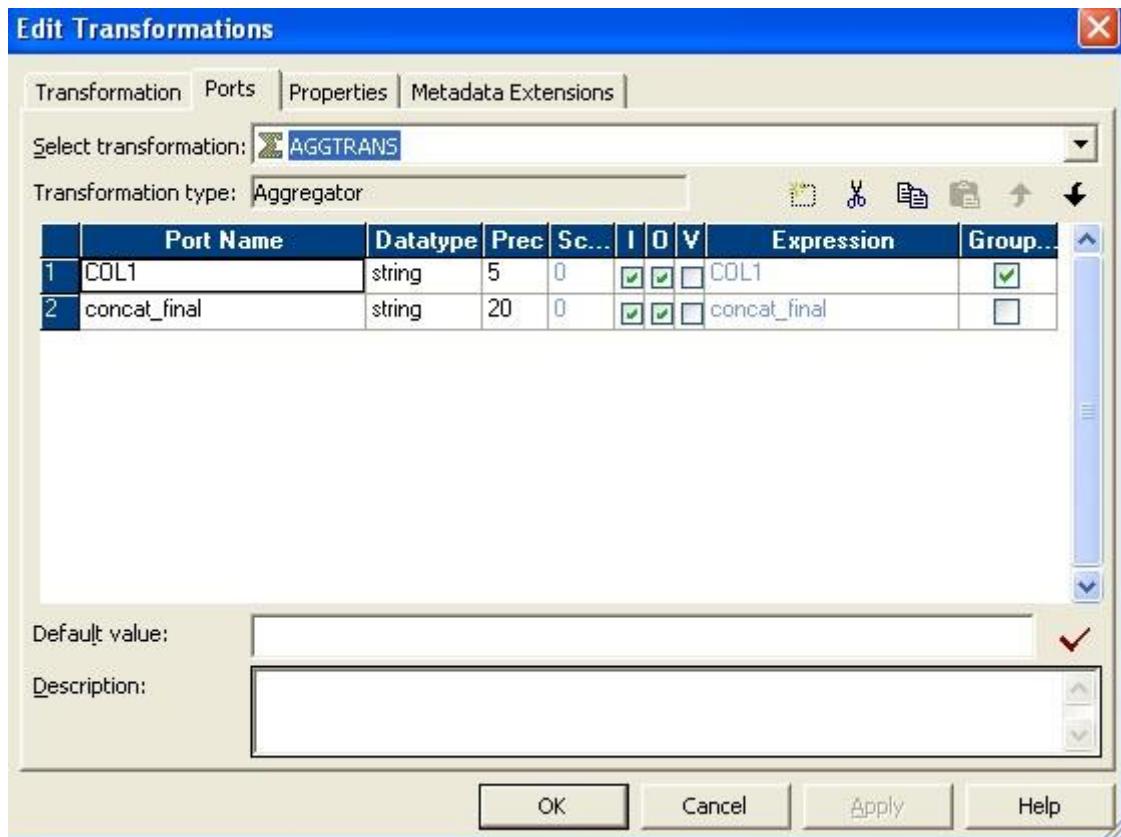
- In Expression make four new port and give them name as in picture below.



3. In concat_val write expression like as describe bellow and send it to an aggregator



4. In aggregator group it by col1 and send it to target



5. Finally run the session.

How to achieve ? The source data contains only column 'id'. It will have sequence numbers from 1 to 1000. The source data looks like as

Id
1
2
3
4
5
6
7
8
....
1000

Create a workflow to load only the Fibonacci numbers in the target table. The target table data should look like as

Id
1
2
3
5
8
13
.....

In Fibonacci series each subsequent number is the sum of previous two numbers. Here assume that the first two numbers of the fibonacci series are 1 and 2.

Solution:

STEP1: Drag the source to the mapping designer and then in the Source Qualifier Transformation properties, set the number of sorted ports to one. This will sort the source data in ascending order. So that we will get the numbers in sequence as 1, 2, 3,1000

STEP2: Connect the Source Qualifier Transformation to the Expression Transformation. In the Expression Transformation, create three variable ports and one output port. Assign the expressions to the ports as shown below.

Ports in Expression Transformation:

id

v_sum = v_prev_val1 + v_prev_val2

v_prev_val1 = IIF(id=1 or id=2,1, IIF(v_sum = id, v_prev_val2, v_prev_val1))

```
v_prev_val2 = IIF(id=1 or id =2, 2, IIF(v_sum=id, v_sum, v_prev_val2) )  
o_flag = IIF(id=1 or id=2,1, IIF( v_sum=id,1,0) )
```

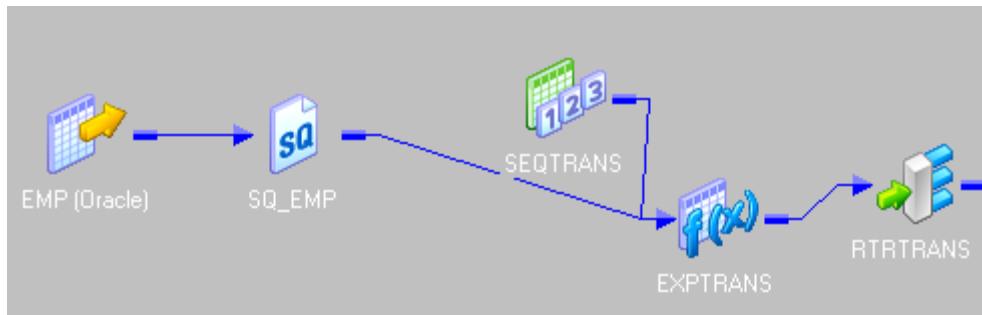
STEP3: Now connect the Expression Transformation to the Filter Transformation and specify the Filter Condition as o_flag=1

STEP4: Connect the Filter Transformation to the Target Table.

Scenario: There is a source table and 3 destination table T1,T2, T3. How to insert first 1 to 10 record in T1, records from 11 to 20 in T2 and 21 to 30 in T3. Then again from 31 to 40 into T1, 41 to 50 in T2 and 51 to 60 in T3 and so on i.e in cyclic order.

Solution:

1. Drag the source and connect to an expression. Connect the next value port of sequence generator to expression.



2. Send the all ports to a router and make three groups as bellow

Group1

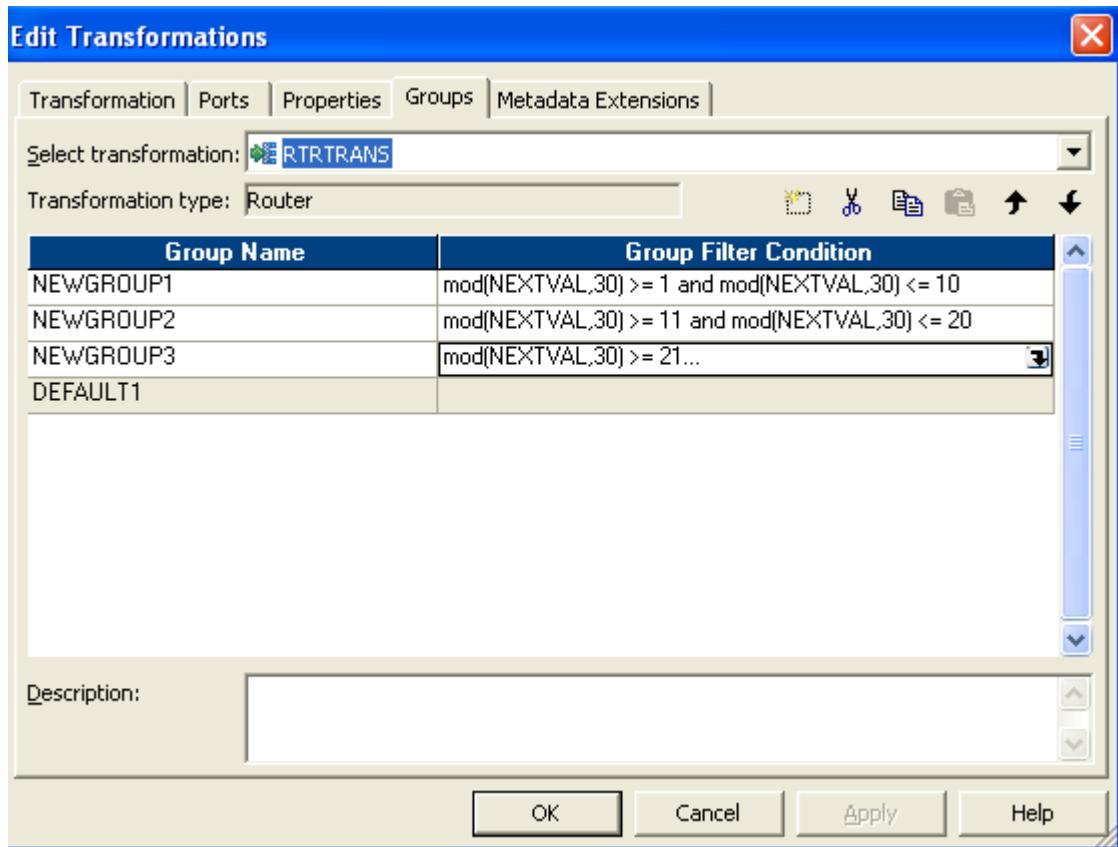
$\text{mod}(\text{NEXTVAL},30) \geq 1 \text{ and } \text{mod}(\text{NEXTVAL},30) \leq 10$

Group2

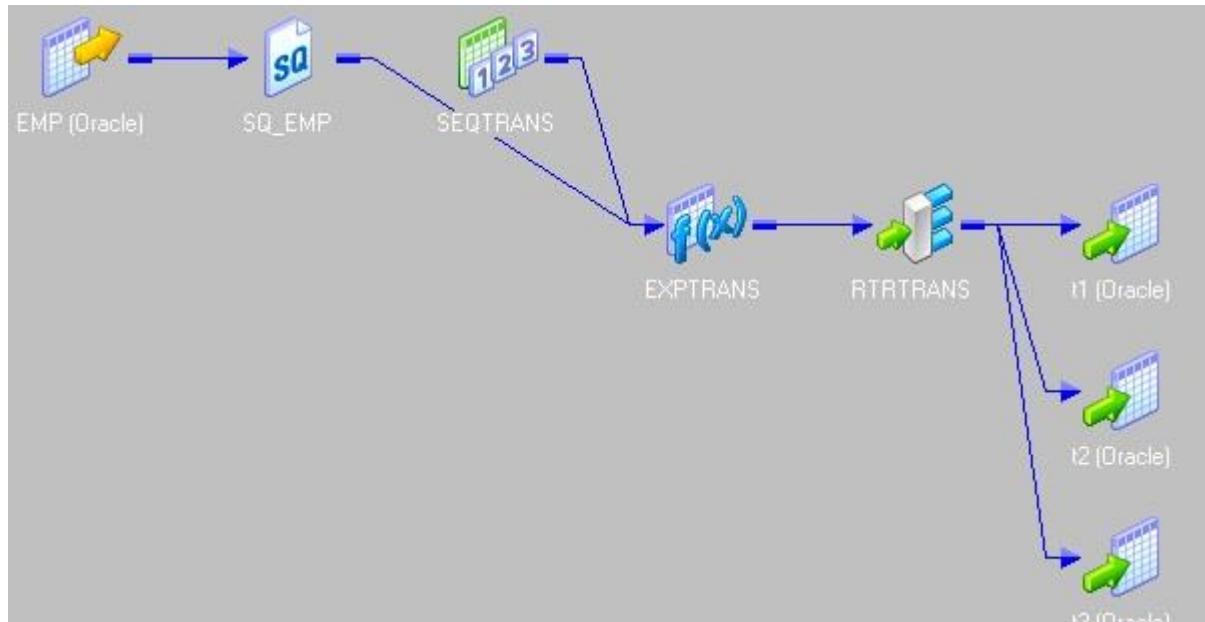
$\text{mod}(\text{NEXTVAL},30) \geq 11 \text{ and } \text{mod}(\text{NEXTVAL},30) \leq 20$

Group3

$\text{mod}(\text{NEXTVAL},30) \geq 21 \text{ and } \text{mod}(\text{NEXTVAL},30) \leq 29 \text{ or } \text{mod}(\text{NEXTVAL},30) = 0$



- Finally connect Group1 to T1, Group2 to T2 and Group3 to T3.



Scenario: The source table contains two columns "id" and "val". The source data looks like

as below

id	val
1	a,b,c
2	pq,m,n
3	asz,ro,liqt

Here the "val" column contains comma delimited data and has three fields in that column.

Create a workflow to split the fields in "val" column to separate rows. The output should look like as below.

id	val
1	a
1	b
1	c
2	pq
2	m
2	n
3	asz
3	ro
3	liqt

Solution:

STEP1: Connect three Source Qualifier transformations to the Source Definition

STEP2: Now connect all the three Source Qualifier transformations to the Union Transformation. Then connect the Union Transformation to the Sorter Transformation. In the sorter transformation sort the data based on Id port in ascending order.

STEP3: Pass the output of Sorter Transformation to the Expression Transformation. The ports in Expression Transformation are:

id (input/output port)

val (input port)

v_currend_id (variable port) = id

v_count (variable port) = IIF(v_current_id!=v_previous_id,1,v_count+1)

v_previous_id (variable port) = id

o_val (output port) = DECODE(v_count, 1,

```
SUBSTR(val, 1, INSTR(val,',',1,1)-1 ),  
2,  
SUBSTR(val, INSTR(val,',',1,1)+1, INSTR(val,',',1,2)-INSTR(val,',',1,1)-1),  
3,  
SUBSTR(val, INSTR(val,',',1,2)+1),  
NULL  
)
```

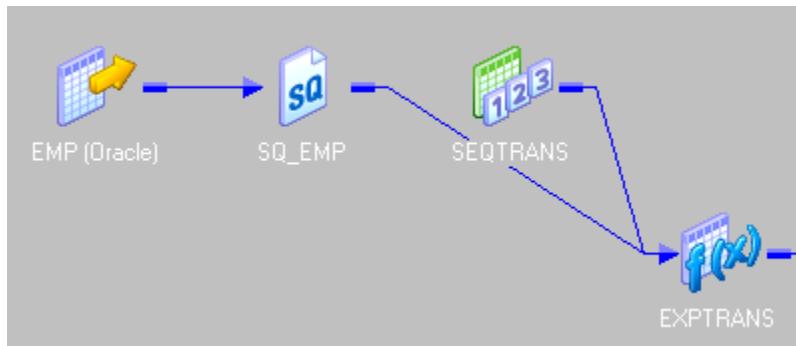
STEP4: Now pass the output of Expression Transformation to the Target definition.

Connect id, o_val ports of Expression Transformation to the id, val ports of Target Definition.

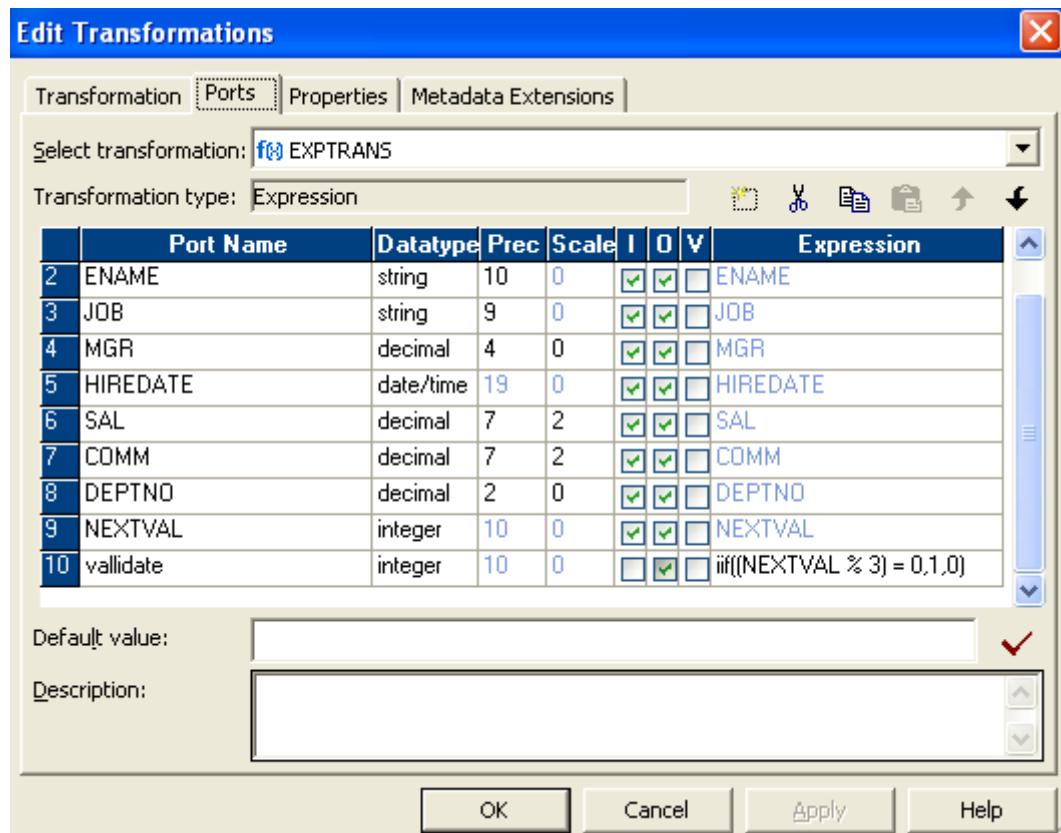
Scenario: How to load every nth row from a Flat file/ relational DB to the target? Suppose n=3, then in above condition the row numbered 3,6,9,12,...so on, This example takes every 3 row to target table.

Solution:

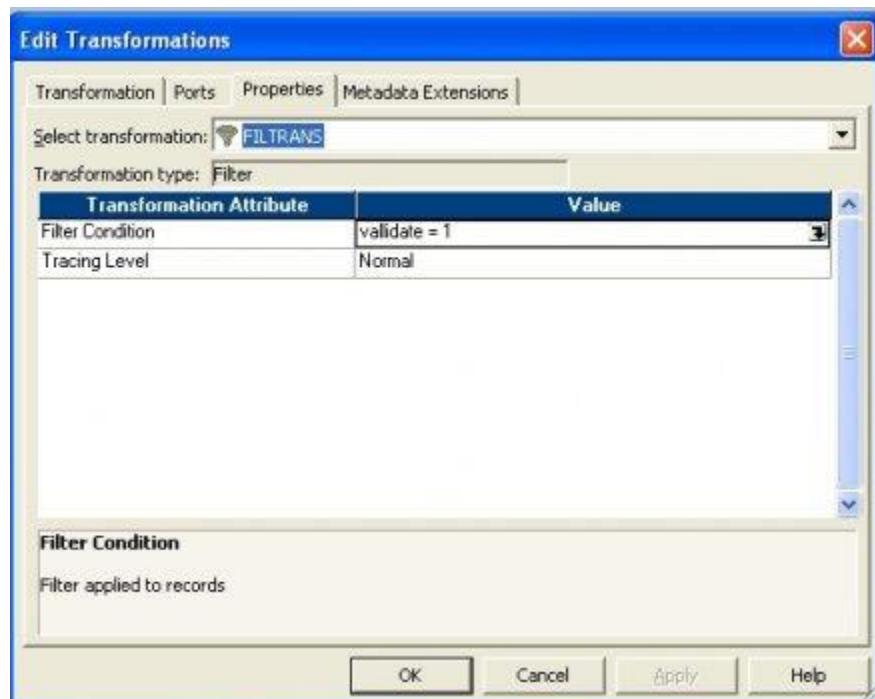
1. Connect an expression transformation after source qualifier.
Add the next value port of sequence generator to expression transformation.



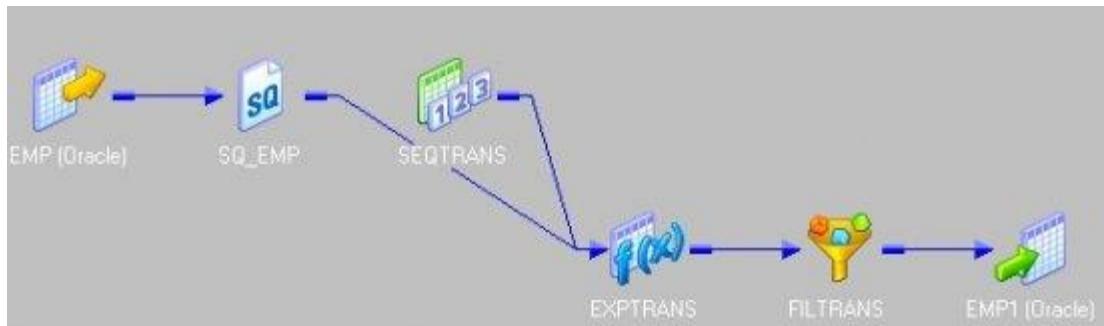
2. In expression create a new port (validate) and write the expression like in the picture below.



3. Connect a filter transformation to expression and write the condition in property like in the picture below.

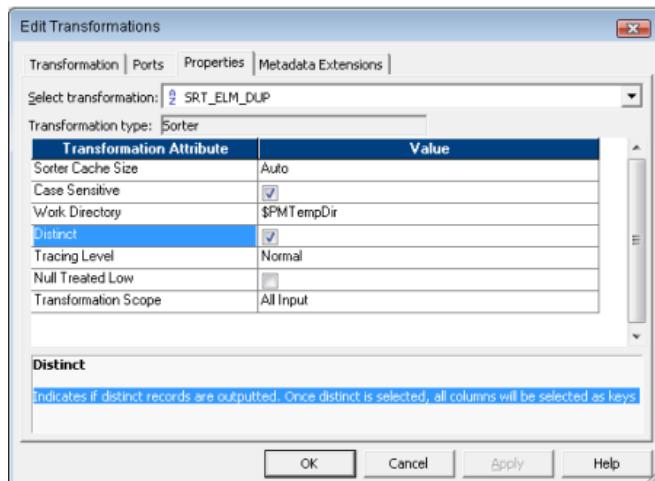


- Finally connect to target.



How to Delete duplicate rows for Flat File sources?

Here since the source system is a **Flat File** you will not be able to select the distinct option in the source qualifier as it will be disabled due to flat file source table. Hence the next approach may be we use a **Sorter Transformation** and check the **Distinct** option. When we select the distinct option all the columns will be selected as keys, in ascending order by default.



How to Delete Duplicate Record Using Informatica Aggregator?

Other ways to handle duplicate records in source batch run is to use an **Aggregator Transformation** and using the **Group By** checkbox on the ports having duplicate occurring data. Here you can have the flexibility to select the last or the first of the duplicate column value records.

There is yet another option to ensure duplicate records are not inserted in the target. That is through Dynamic lookup cache. Using Dynamic Lookup Cache of the target table and associating the input ports with the lookup port and checking the Insert Else Update option

will help to eliminate the duplicate records in source and hence loading unique records in the target.

How to do Loading of Multiple Target Tables?

Scenario

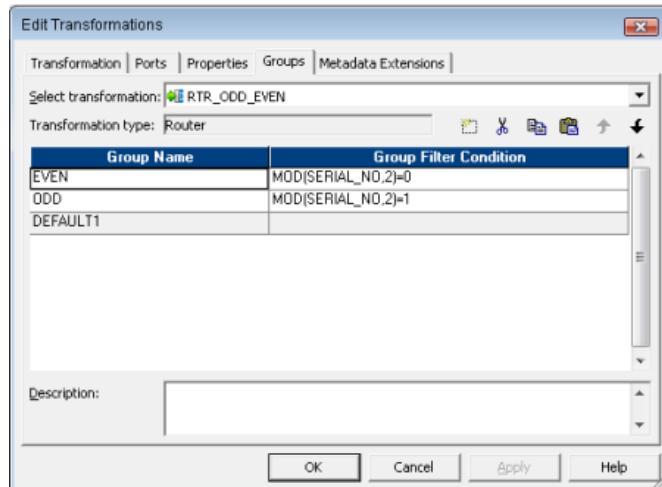
Suppose we have some serial numbers in a flat file source. We want to load the serial numbers in two target files one containing the EVEN serial numbers and the other file having the ODD ones.

Answer

After the Source Qualifier place a **Router Transformation**. Create two **Groups** namely **EVEN** and **ODD**, with filter conditions as:

MOD(SERIAL_NO,2)=0 and MOD(SERIAL_NO,2)=1

... respectively. Then output the two groups into two flat file targets.



How to achieve below scenario using Normalizer transformation?

Scenario 1

Suppose in our Source Table we have data as given below:

Student Name	Maths	Life Science	Physical Science
Sam	100	70	80
John	75	100	85
Tom	80	100	85

We want to load our Target Table as:

Student Name	Subject Name	Marks
Sam	Maths	100
Sam	Life Science	70
Sam	Physical Science	80
John	Maths	75
John	Life Science	100
John	Physical Science	85
Tom	Maths	80
Tom	Life Science	100
Tom	Physical Science	85

Answer

Here to convert the Rows to Columns we have to use the **Normalizer Transformation** followed by an Expression Transformation to Decode the column taken into consideration.

Name the transformations which converts one to many rows i.e increases the i/p:o/p row count. Also what is the name of its reverse transformation?

Normalizer as well as Router Transformations are the Active transformation which can increase the number of input rows to output rows.

Aggregator Transformation performs the reverse action of Normalizer transformation.

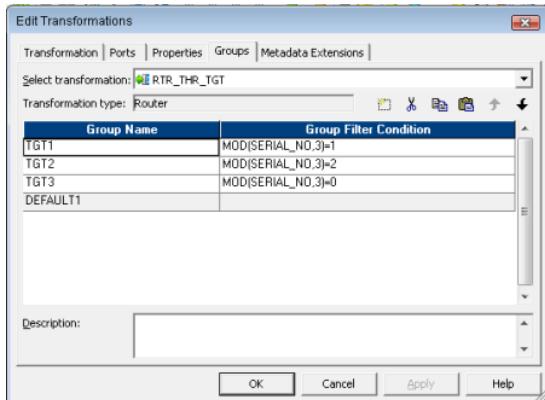
Scenario 2

Suppose we have a source table and we want to load three target tables based on source rows such that first row moves to first target table, second row in second target table, third row in third target table, fourth row again in first target table so on and so forth.

Describe your approach.

We can clearly understand that we need a **Router transformation** to route or filter source data to the three target tables. Now the question is what will be the filter conditions. First of all we need an **Expression Transformation** where we have all the source table columns and along with that we have another i/o port say seq_num, which is gets sequence numbers for each source row from the port **NextVal** of a **Sequence Generator start value 0 and increment by 1**. Now the filter condition for the three router groups will be:

- MOD(SEQ_NUM,3)=1 connected to 1st target table
- MOD(SEQ_NUM,3)=2 connected to 2nd target table
- MOD(SEQ_NUM,3)=0 connected to 3rd target table

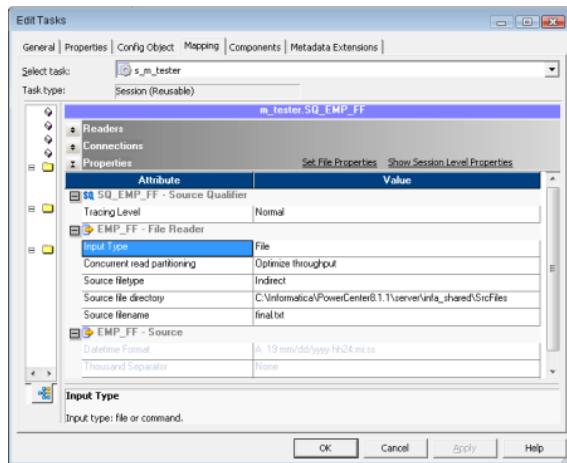


How to Load Multiple Flat Files using one mapping?

Scenario

Suppose we have ten source flat files of same structure. How can we load all the files in target database in a single batch run using a single mapping.

After we create a mapping to load data in target database from flat files, next we move on to the session property of the Source Qualifier. To load a set of source files we need to create a file say final.txt containing the source flat file names, ten files in our case and set the **Source filetype** option as **Indirect**. Next point this flat file final.txt fully qualified through **Source file directory** and **Source filename**.



How can we implement Aggregation operation without using an Aggregator Transformation in Informatica?

We will use the very basic concept of the **Expression Transformation** that at a time we can access the previous row data as well as the currently processed data in an expression transformation. What we need is simple Sorter, Expression and Filter transformation to achieve aggregation at Informatica level.

Scenario

Suppose in our Source Table we have data as given below:

Student Name	Subject Name	Marks
Sam	Maths	100
Tom	Maths	80
Sam	Physical Science	80
John	Maths	75
Sam	Life Science	70
John	Life Science	100
John	Physical Science	85
Tom	Life Science	100

Tom	Physical Science	85
-----	------------------	----

We want to load our Target Table as:

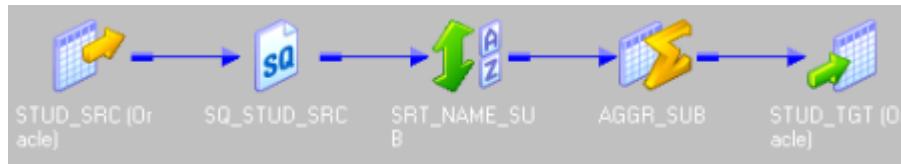
Student Name	Maths	Life Science	Physical Science
Sam	100	70	80
John	75	100	85
Tom	80	100	85

Describe your approach.

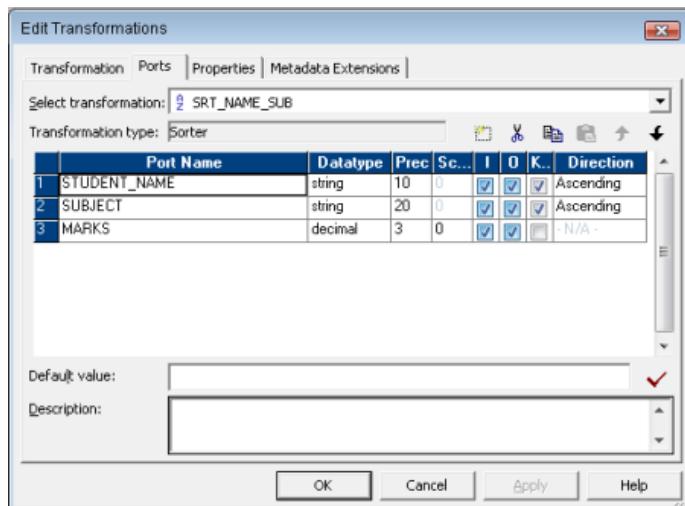
Answer

Here our scenario is to convert many rows to one rows, and the transformation which will help us to achieve this is **Aggregator**.

Our Mapping will look like this:

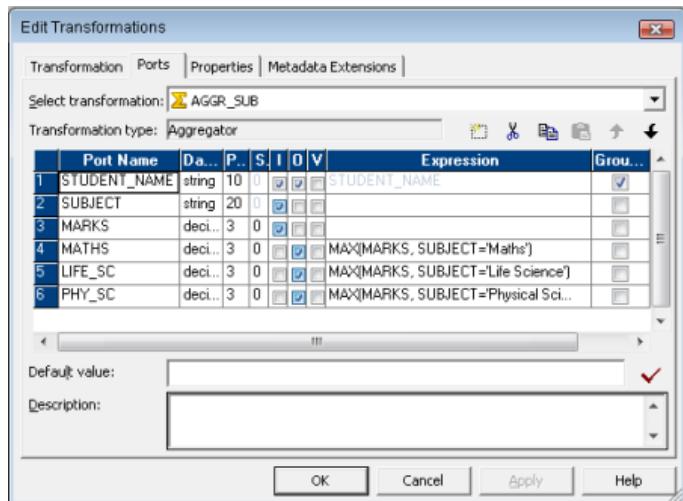


We will sort the source data based on STUDENT_NAME ascending followed by SUBJECT ascending.



Now based on STUDENT_NAME in GROUP BY clause the following output subject columns are populated as

- MATHS: MAX(MARKS, SUBJECT=Maths)
- LIFE_SC: MAX(MARKS, SUBJECT=Life Science)
- PHY_SC: MAX(MARKS, SUBJECT=Physical Science)



What is a Source Qualifier? What are the tasks we can perform using a SQ and why it is an ACTIVE transformation?

A **Source Qualifier** is an Active and Connected Informatica transformation that reads the rows from a relational database or flat file source.

- We can configure the **SQ to join** [Both **INNER** as well as **OUTER JOIN**] data originating from the same source database.
- We can use a source **filter** to reduce the number of rows the Integration Service queries.
- We can specify a number for **sorted ports** and the Integration Service adds an ORDER BY clause to the default SQL query.
- We can choose **Select Distinct** option for relational databases and the Integration Service adds a SELECT DISTINCT clause to the default SQL query.
- Also we can write **Custom/Used Defined SQL** query which will override the default query in the SQ by changing the default settings of the transformation properties.
- Also we have the option to write **Pre** as well as **Post SQL** statements to be executed before and after the SQ query in the source database.

Since the transformation provides us with the property **Select Distinct**, when the Integration Service adds a SELECT DISTINCT clause to the default SQL query, which in turn affects the number of rows returned by the Database to the Integration Service and hence it is an Active transformation.

How to populate Flat file header row, footer row and detail rows to multiple tables ?

Assume that we have a flat file with header row, footer row and detail rows. Now Lets see how to load header row into one table, footer row into other table and detail rows into another table just by using the transformations only.

First pass the data from source qualifier to an expression transformation. In the expression

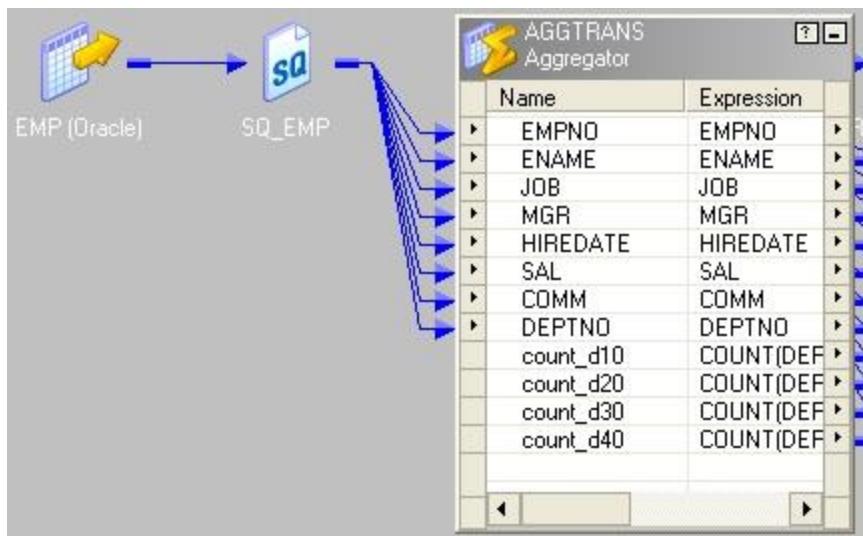
transformation assign unique number to each row (assume exp_count port). After that pass the data from expression to aggregator. In the aggregator transformation don't check any group by port. So that the aggregator will provide last row as the default output (assume agg_count port).

Now pass the data from expression and aggregator to joiner transformation. In the joiner select the ports from aggregator as master and the ports from expression as details. Give the join condition on the count ports and select the join type as master outer join. Pass the joiner output to a router transformation and create two groups in the router. For the first group give the condition as exp_count=1, which gives header row. For the second group give the condition as exp_count=agg_count, which gives the footer row. The default group will give the detail rows.

Scenario 13: There are 4 departments in Emp table. The first one with 100, 2nd with 5, 3rd with 30 and 4th dept has 12 employees. Extract those dept numbers which has more than 5 employees in it, to a target table.

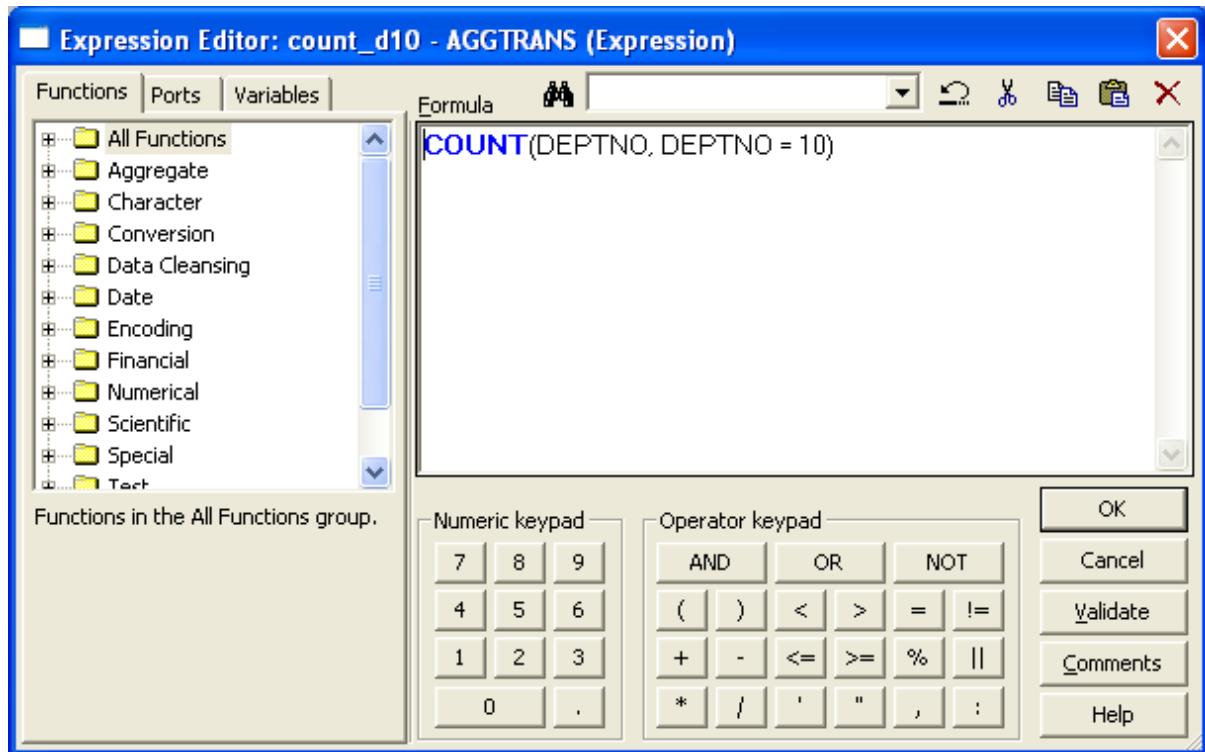
Solution:

1. Put the source to mapping and connect the ports to aggregator transformation.

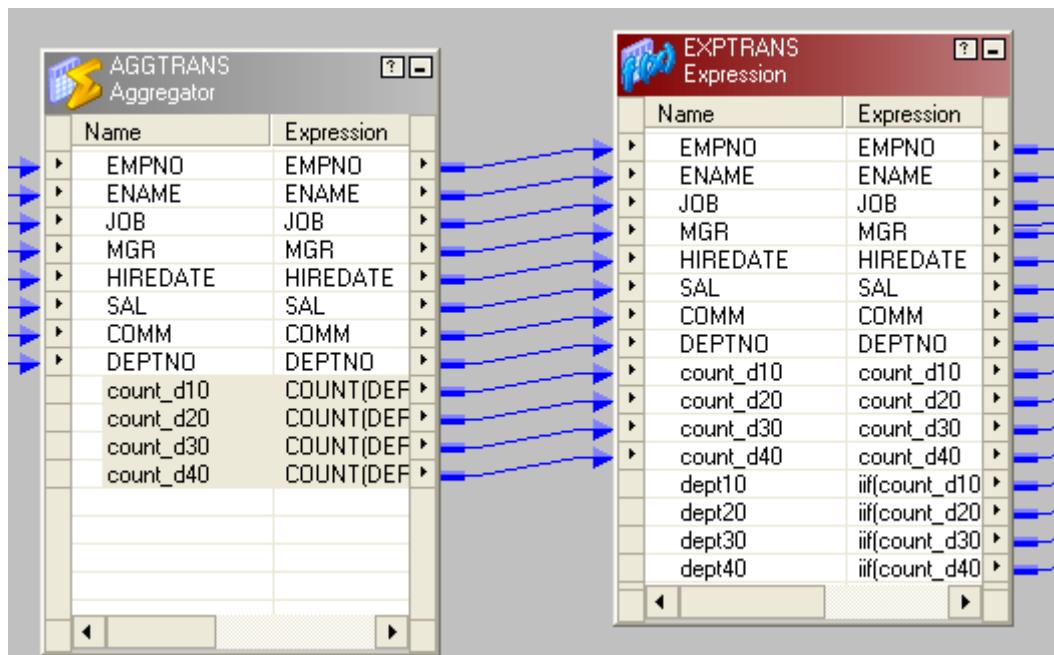


2. Make 4 output ports in aggregator as in the picture above : count_d10, count_d20, count_d30, count_d40.

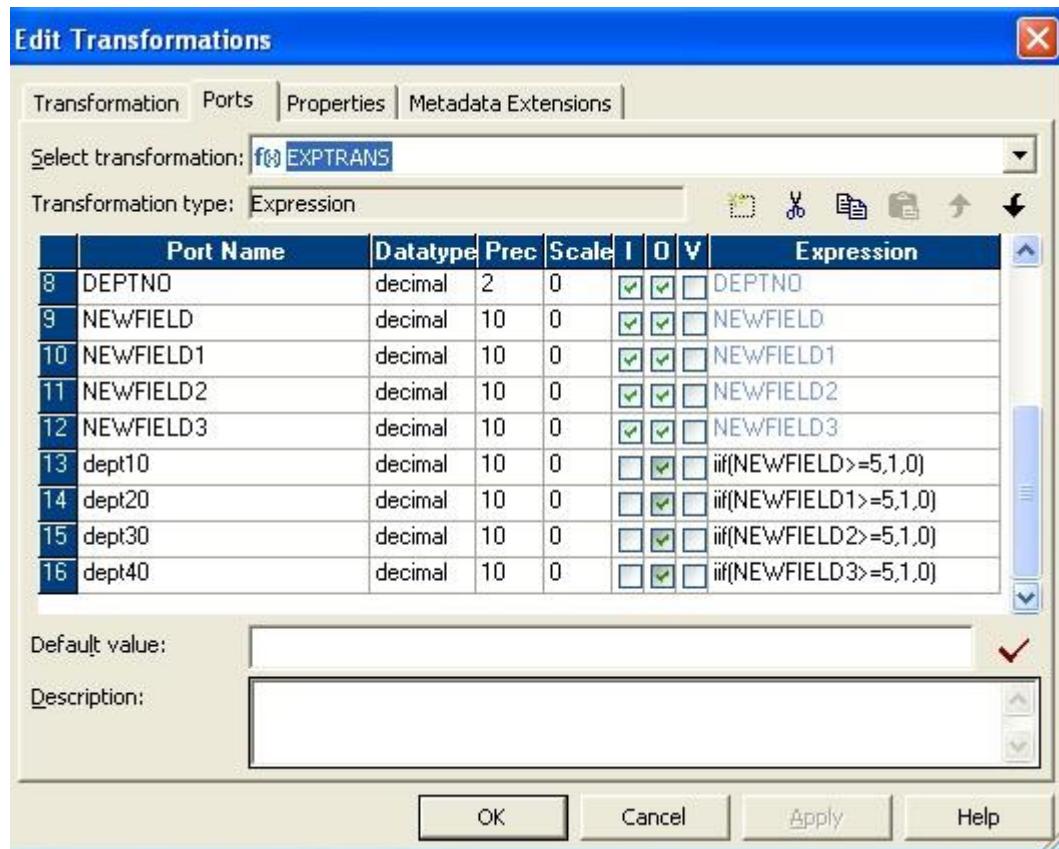
For each port write expression like in the picture below.



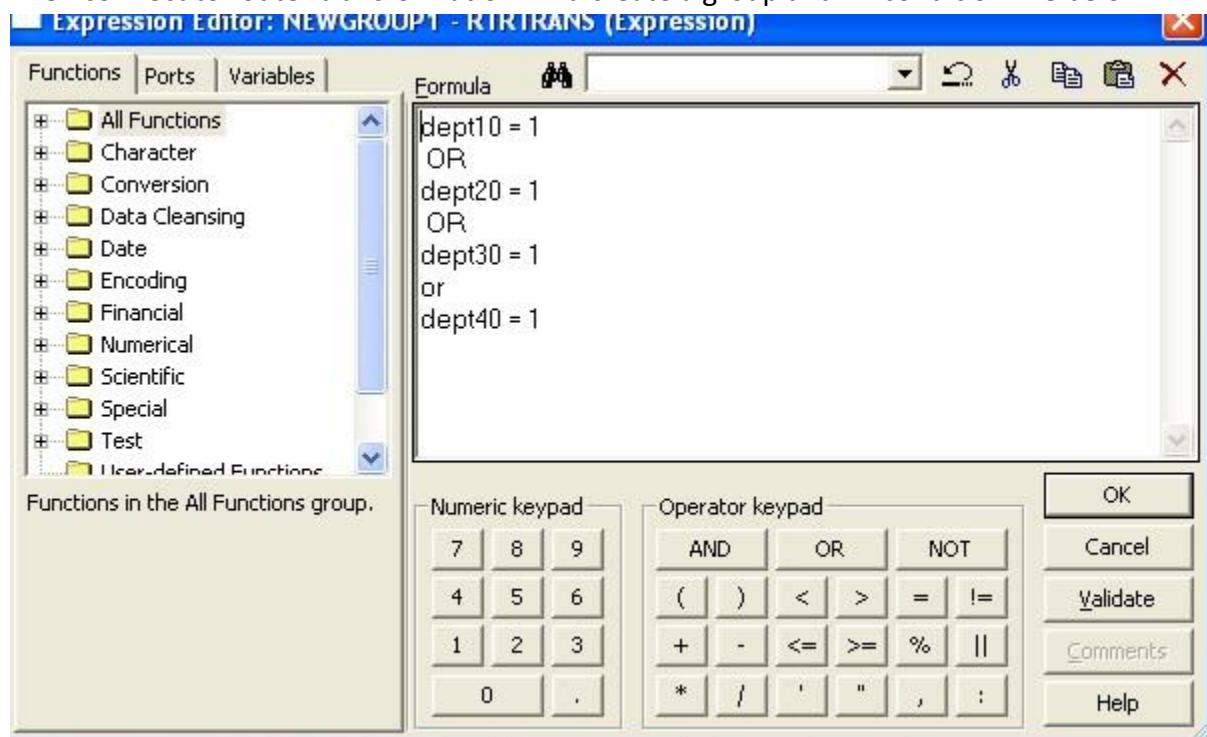
- Then send it to expression transformation.



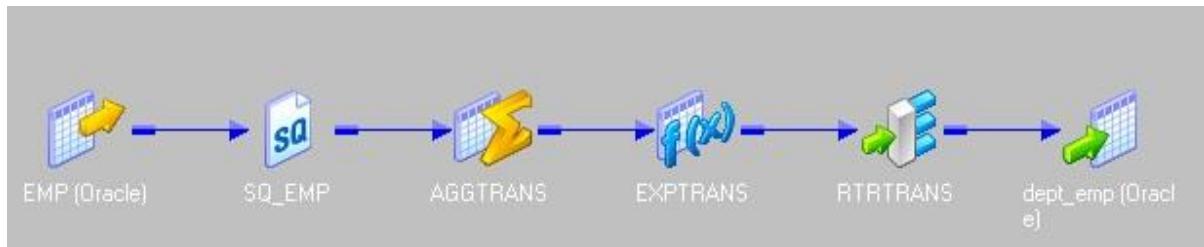
4. In expression make four output ports (dept10, dept20, dept30, dept40) to validate dept no
And provide the expression like in the picture below.



5. Then connect to router transformation. And create a group and fill condition like below.



6. Finally connect to target table having one column that is dept no.



How to Reverse the Contents of Flat File – Informatica?

I have a flat file, want to reverse the contents of the flat file which means the first record should come as last record and last record should come as first record and load into the target file.

As an example consider the source flat file data as

Informatica Enterprise Solution

Informatica Power center

Informatica Power exchange

Informatica Data quality

The target flat file data should look as

Informatica Data quality

Informatica Power exchange

Informatica Power center

Informatica Enterprise Solution

Solution:

Follow the below steps for creating the mapping logic

- Create a new mapping.
- Drag the flat file source into the mapping.

- Create an expression transformation and drag the ports of source qualifier transformation into the expression transformation.
- Create the below additional ports in the expression transformation and assign the corresponding expressions

Variable port: v_count = v_count+1

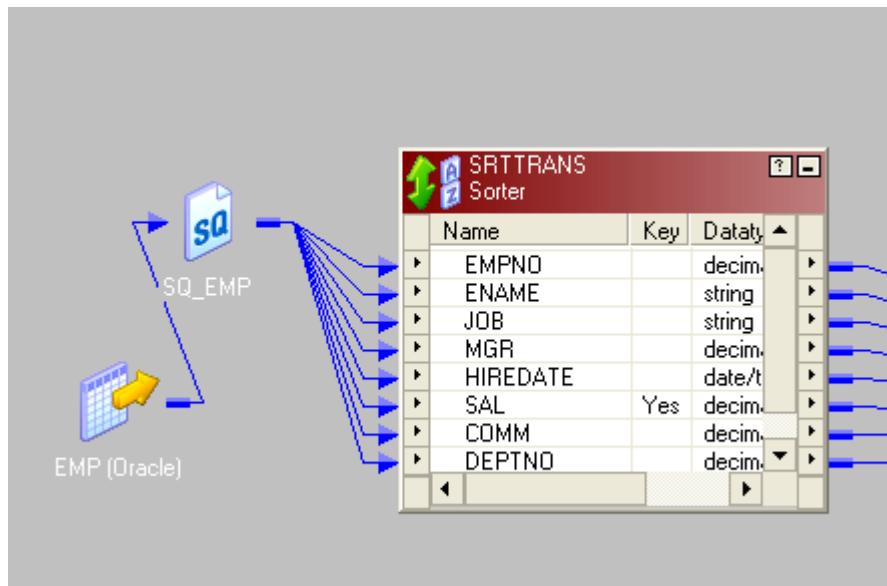
Output port o_count = v_count

- Now create a sorter transformation and drag the ports of expression transformation into it.
- In the sorter transformation specify the sort key as o_count and sort order as DESCENDING.
- Drag the target definition into the mapping and connect the ports of sorter transformation to the target.

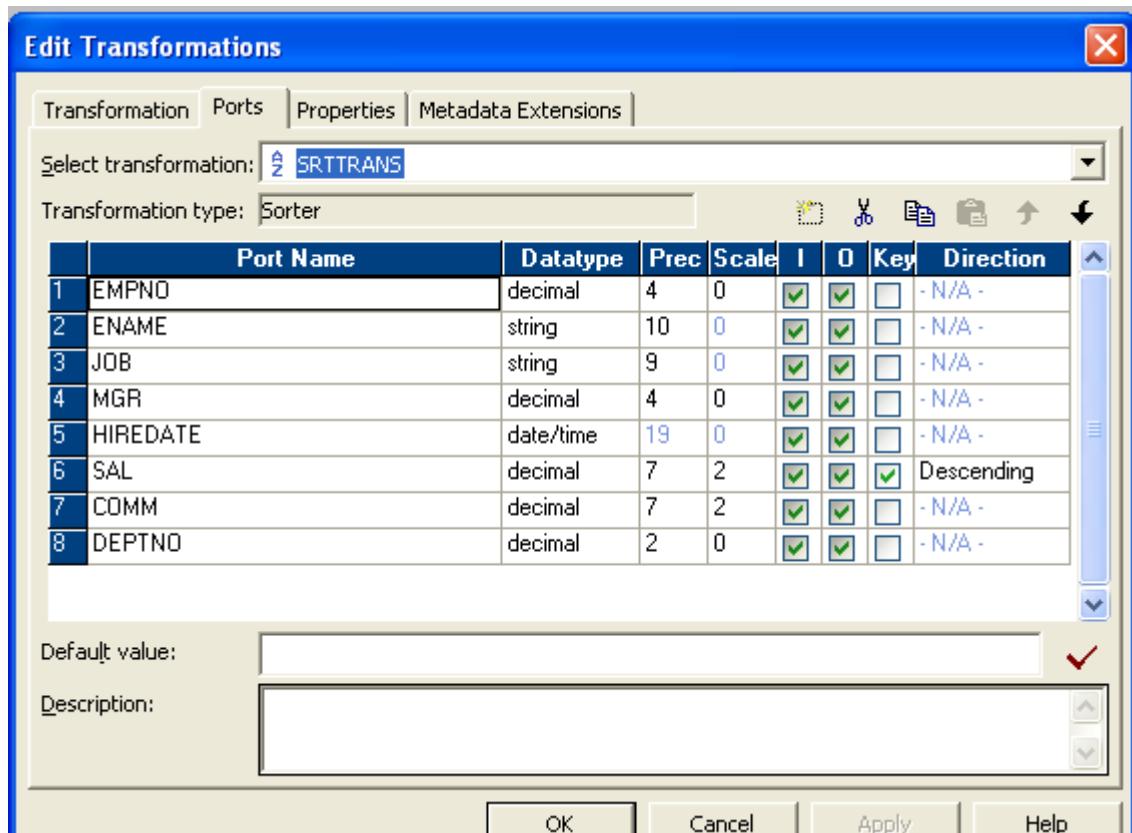
Scenario: How to get top 5 records to target without using rank ?

Solution:

1. Drag the source to mapping and connect it to sorter transformation.

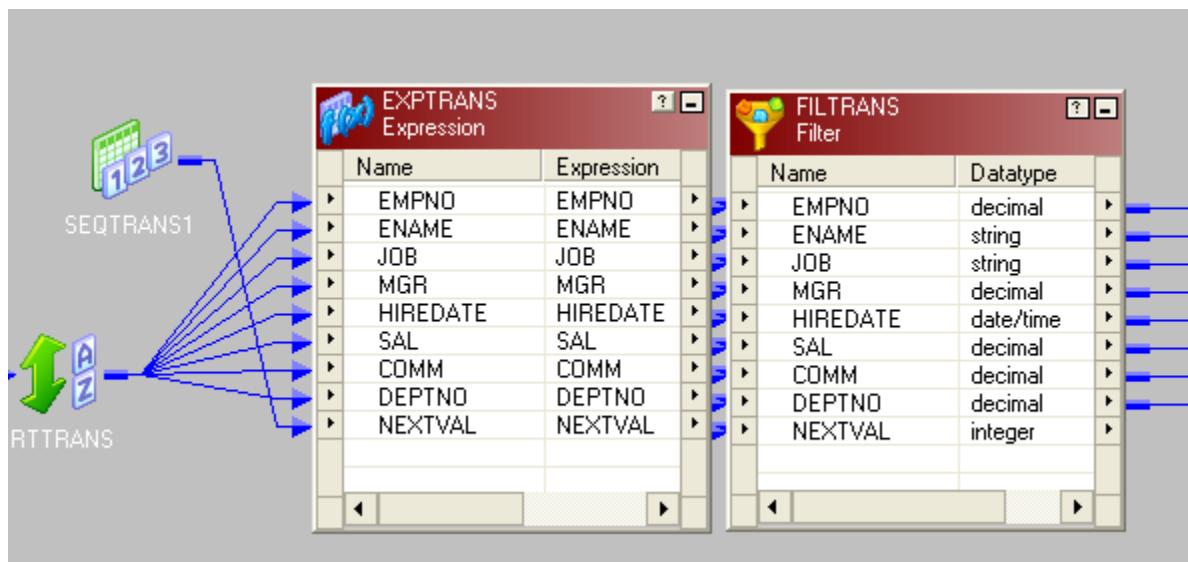


2. Arrange the salary in descending order in sorter as follows and send the record to expression.



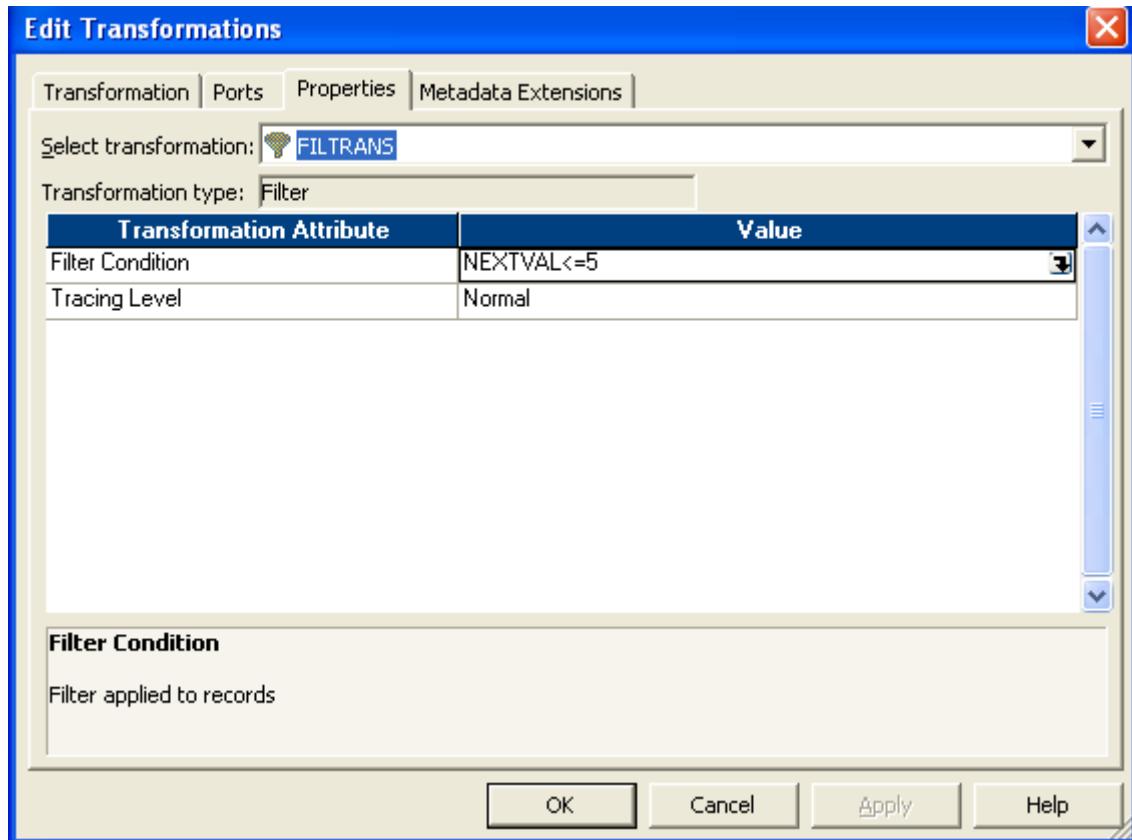
sorter properties

- Add the next value of sequence generator to expression.(start the value from 1 in sequence generator).

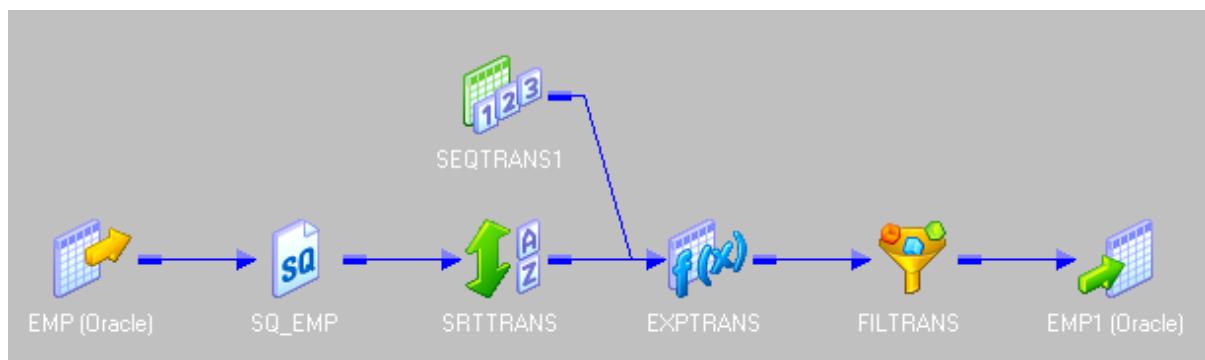


sorter to exp mapping

- Connect the expression transformation to a filter or router. In the property set the condition as follows-



- Finally connect to the target.



Load the header record of the flat file into first target, footer record into second target and the remaining records into the third target.

The solution to this problem I have already posted by using aggregator and joiner. Now we will see how to implement this by reversing the contents of the file.

Solution:

- Connect the source qualifier transformation to the expression transformation. In the expression transformation create the additional ports as mentioned above.

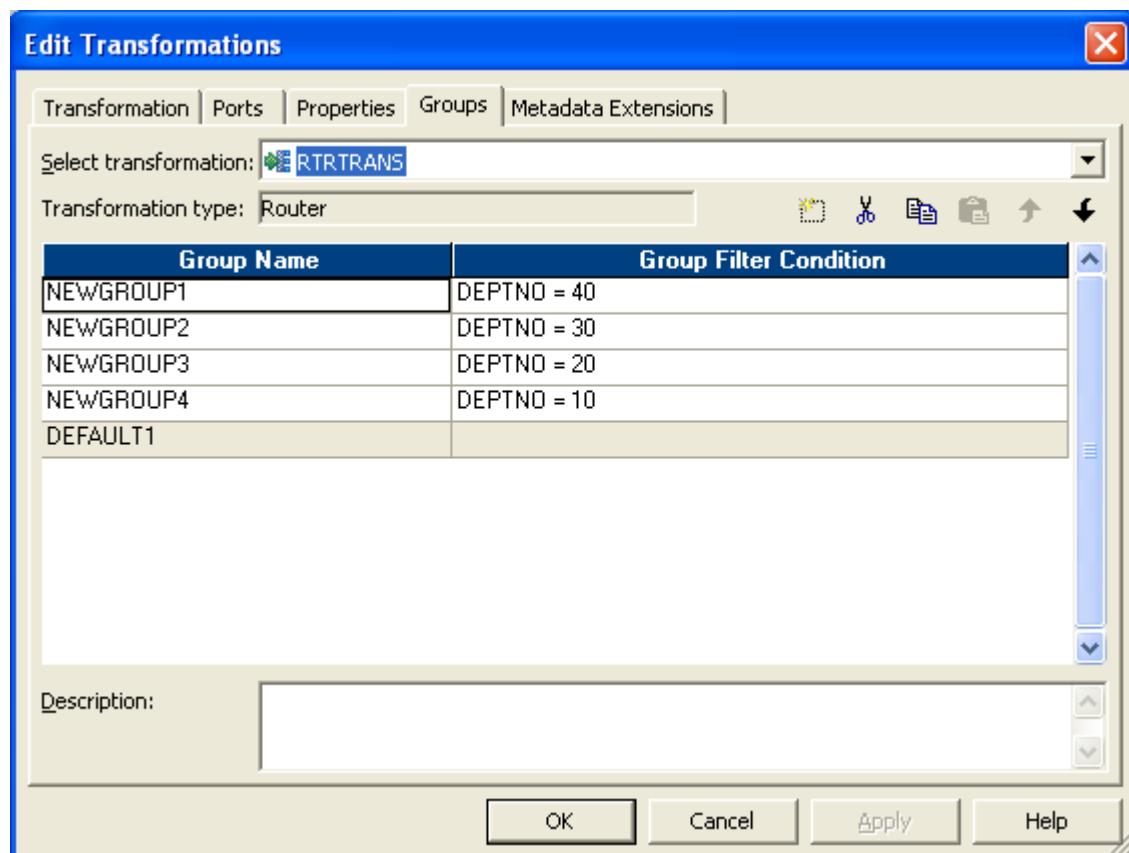
- Connect the expression transformation to a router. In the router transformation create an output group and specify the group condition as o_count=1. Connect this output group to a target and the default group to sorter transformation.
- Sort the data in descending order on o_count port.
- Connect the output of sorter transformation to expression transformation (don't connect o_count port).
- Again in the expression transformation create the same additional ports mentioned above.
- Connect this expression transformation to router and create an output group. In the output group specify the condition as o_count=1 and connect this group to second target. Connect the default group to the third group.

Scenario: In Dept table there are four departments (dept no 40,30,20,10). Separate the record to different target department wise.

Solution:

Step 1: Drag the source to mapping.

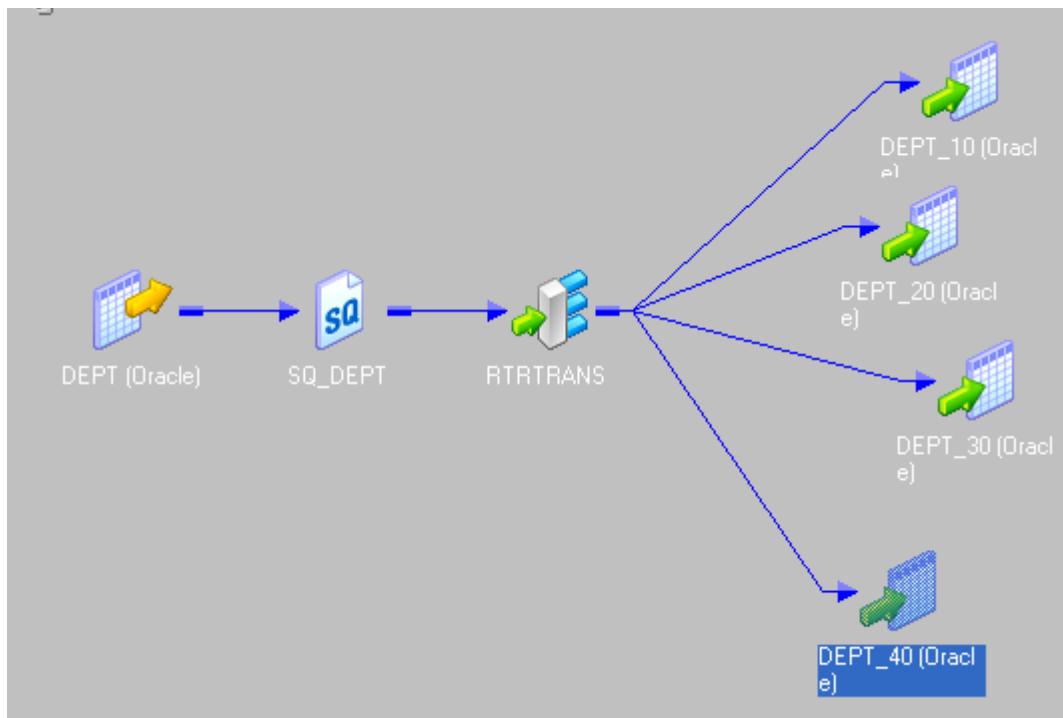
Step 2: Connect the router transformation to source and in router make 4 groups and give condition like below.



router transformation

Step 3: Based on the group map it to different target.

The final mapping looks like below.



router to target

How do you load alternate records into different tables through mapping flow?

The concept is to add a sequence number to the records and then divide the record number by 2. If it is breakable, then move it to one target and if not then move it to another target. Some of the following steps are:

- Drag the source and connect to an expression transformation
- Add the next value of a sequence generator to expression transformation
- In the expression transformation make two ports, one is “odd”, and another one is “even.”
- Write the expression below
- Connect a router transformation to expression
- Make two groups in the router
- Give condition

- Then send the two groups to different targets

Check the Hire-Date is Date or Not?

Scenario: Suppose we have a table with records like this

EMPNO	HIRE_DATE
-------	-----------

1	12-11-87
2	02-04-88;
3	02-2323

empno is number and hire_date is in string format. We have to check the hire_date column, if it is in date format like 'dd-mm-yy', then convert it to date , in the format "mm/dd/yy" and send it to target else send null.

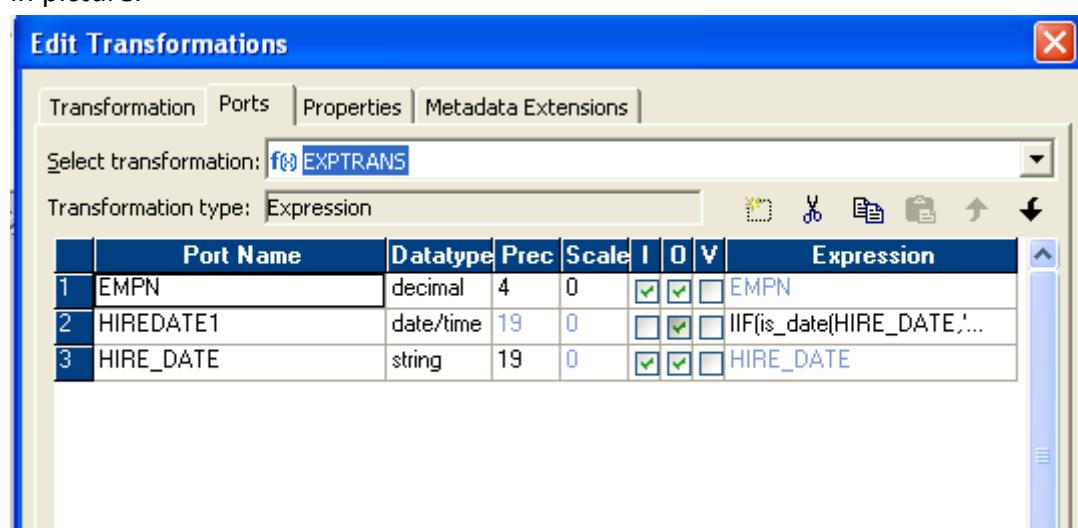
output

EMPNO	HIRE_DATE
-------	-----------

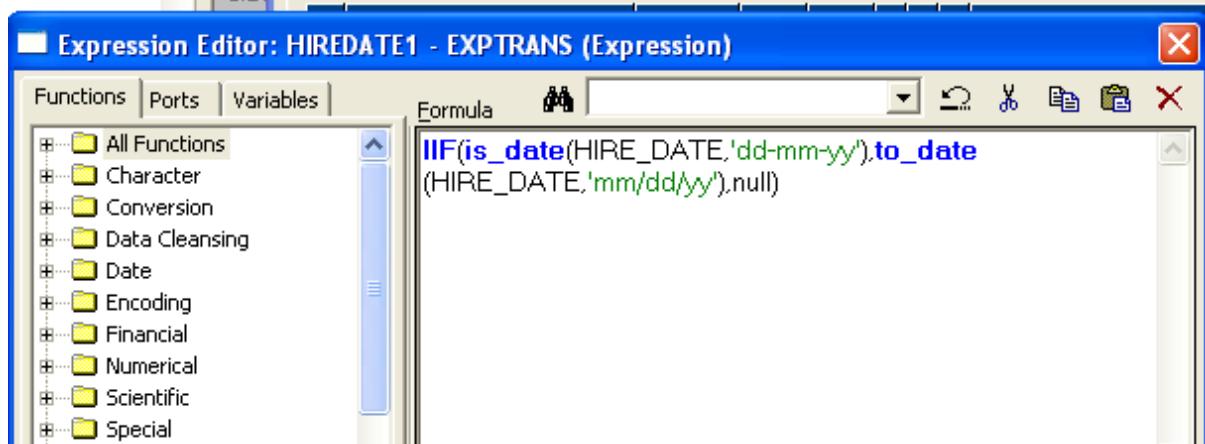
1	11-DEC-87
2	null
3	null

These are the steps for achieving this scenario

1. Connect the ports from SQF to an expression transformation.
2. In expression create another output port hire_date1 and make it to date data-type, shown in picture.



3. In Hire_date1 write the condition like this.



4. Send ports to target.

Check the Date Difference in Hours?

Scenario: There is a order_delivery table having record like this

ORDER_NO	ORDER_DATE	DELIVERY_DATE
2	11-JAN-83	13-JAN-83
3	04-FEB-83	07-FEB-83
1	08-DEC-81	09-DEC-81

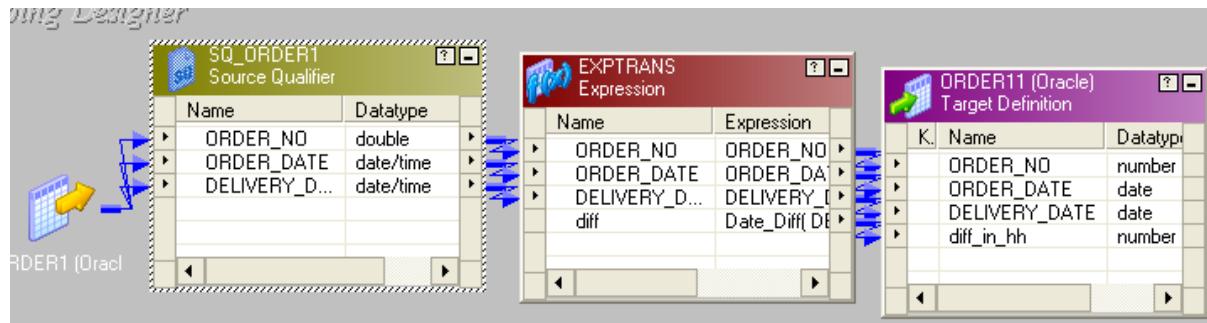
We have to calculate difference between order_date and delivery date in hours and send it to target.

o/p will be

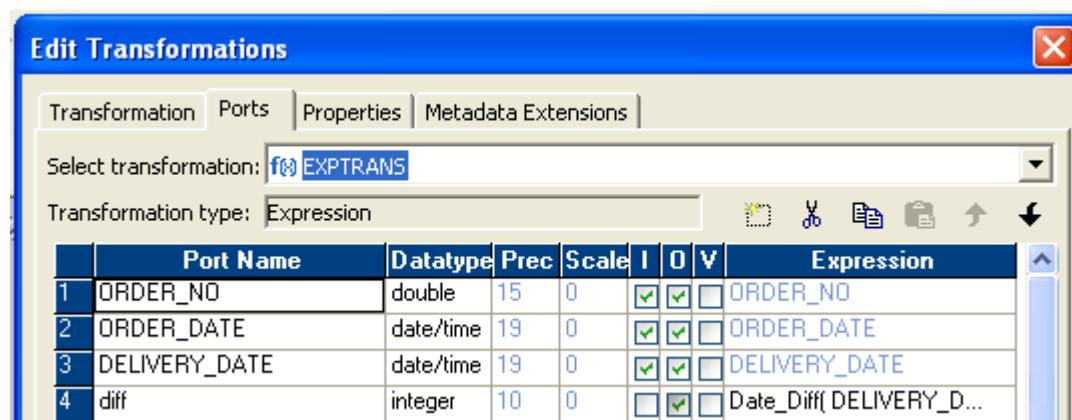
ORDER_NO	ORDER_DATE	DELIVERY_DATE	DIFF_IN_HH
2	11-JAN-83	13-JAN-83	48
3	04-FEB-83	07-FEB-83	72
1	08-DEC-81	09-DEC-81	24

These are the steps for achieving this scenario

1. Connect one expression transformation next to SQF.



2. In expression create one out/put port "diff" and make it integer type.



3. In that port write the condition like this and sent to target.



How to send the data to target with days difference of more than 2 days?

Scenario: From the order_delivery table insert the records to target where , day difference between order_date and delivery_date is greater than 2 days. (Note: see last article , where we discussed finding the time in hour between two dates)

Source

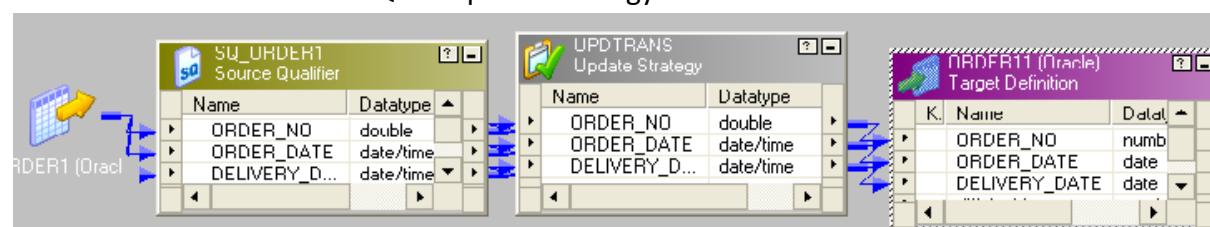
ORDER_NO	ORDER_DATE	DELIVERY_DATE
2	11-JAN-83	13-JAN-83
3	04-FEB-83	07-FEB-83
1	08-DEC-81	09-DEC-81

Target

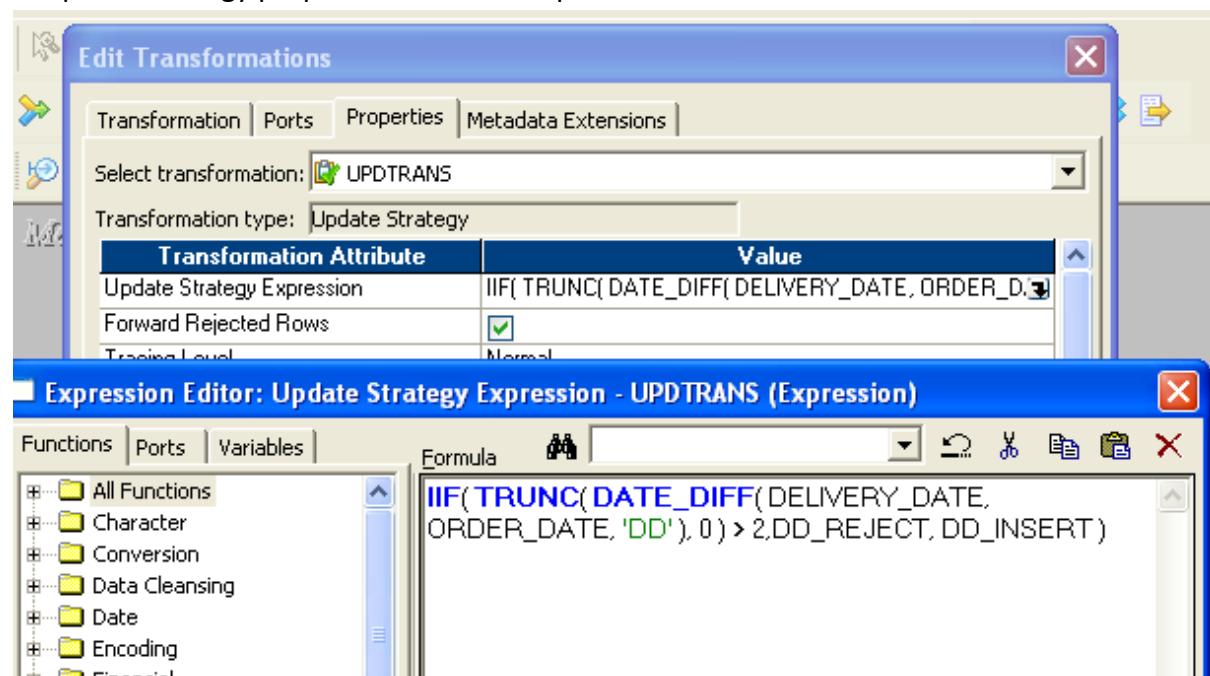
ORDER_NO	ORDER_DATE	DELIVERY_DATE
2	11-JAN-83	13-JAN-83
3	04-FEB-83	07-FEB-83

These are the steps for achieving this scenario

1. Connect all the rows from SQF to update strategy transformation.



2. In update strategy properties write the expression like this



3. Finally send to target.

How to convert Day No. to corresponding month and date of year?

Scenario: Suppose you have a source is like this

Source

E_NO	YEAR	DAYNO
1	01-JAN-07	301
2	01-JAN-08	200

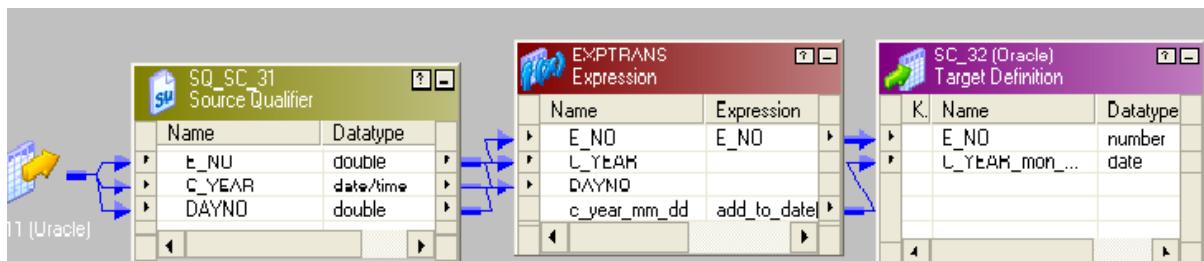
Year column is a date and day no is numeric that represents a day (as in 365 for 31-Dec-Year). Convert the Day no to corresponding year's month and date and then send to target.

Target

E_NO	YEAR_MONTH_DAY
1	29-OCT-07
2	19-JUL-08

These are the basic steps for this scenario

1. Connect SQF with an expression transformation.



2. In expression create one o/p port c_year_mm_dd, make it to date type and in that port write the condition like this.

	Port Name	Datatype	Prec	Scale	I	O	V	Expression
1	E_NO	double	15	0	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	E_NO
2	C_YEAR	date/tme	13	0	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
3	DAYNO	double	15	0	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
4	c_year_mm_dd	date/tme	13	0	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	add_to_date(C_YEAR, 'dd'.DAYNO)

3. Finally send to target

How to get the data of employees who joined in current month?

scenario: Insert the records of those employees who have joined in current month and Reject other rows.

Source

E_NO	JOIN_DATE
1	07-JUL-11
2	05-JUL-11
3	05-MAY-11

If the current month is july ,2011 then target will be like this.

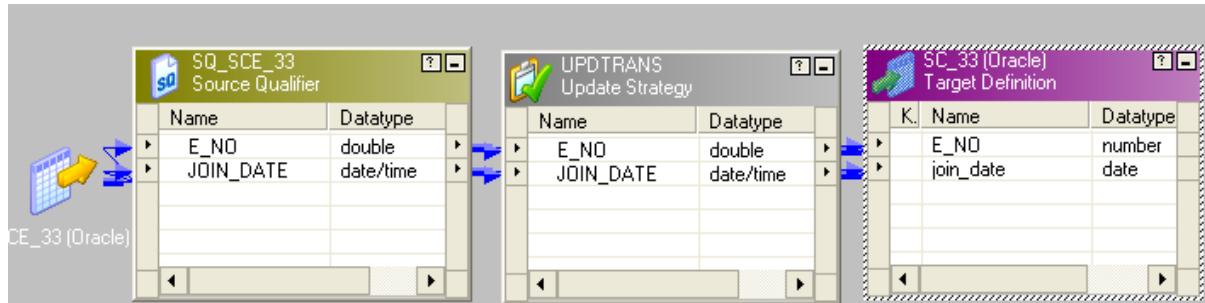
Target

E_NO	JOIN_DATE
1	07-JUL-11

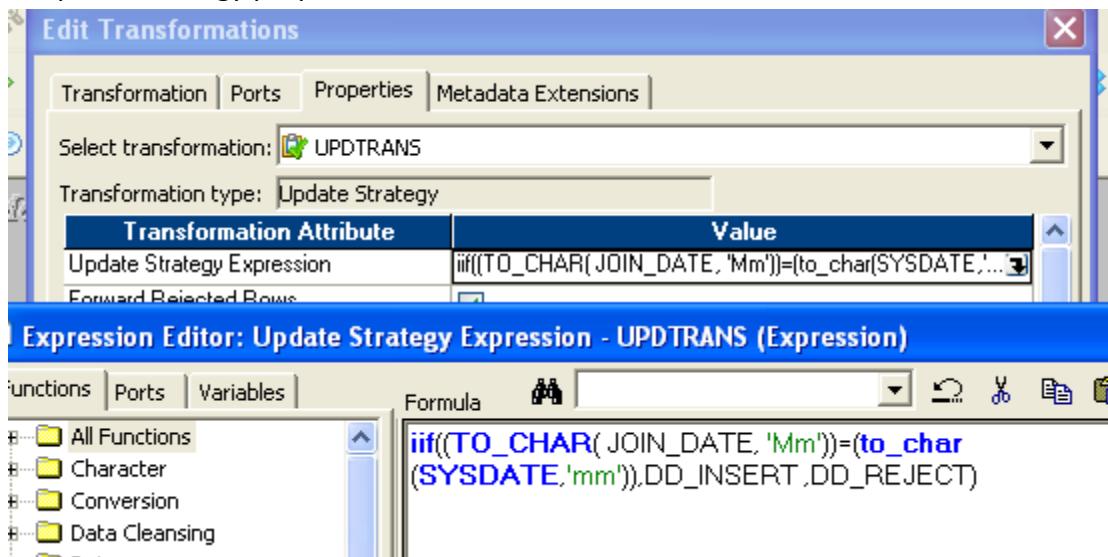
- 1 07-JUL-11
- 2 05-JUL-11

To insert current month records we have to follow these steps

1. Connect one update strategy transformation next to SQF.



2. In update strategy properties write the condition like this



3. Send required ports update strategy to target.

Tell me the process of Extracting first and last name?

Suppose In Ename column there is first name and last name like this

empno ename

- 1 Amit Rao
- 2 Chitra Dash

In target we have to separate ename column to firstname and lastname like this

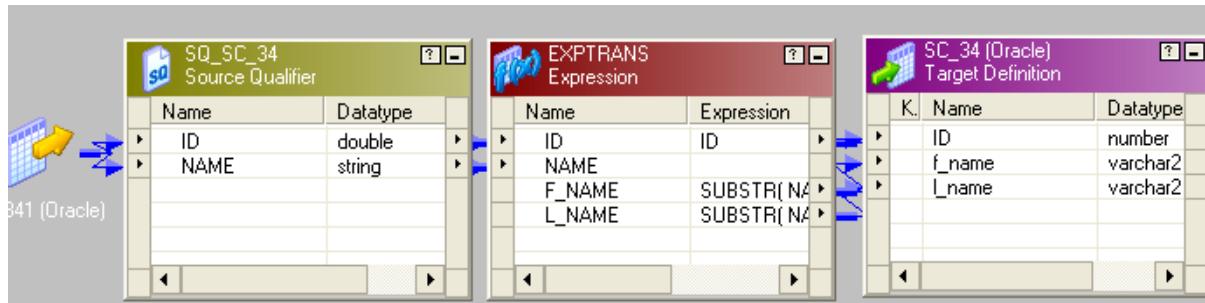
empno firstname Lastname

1 Amit Rao

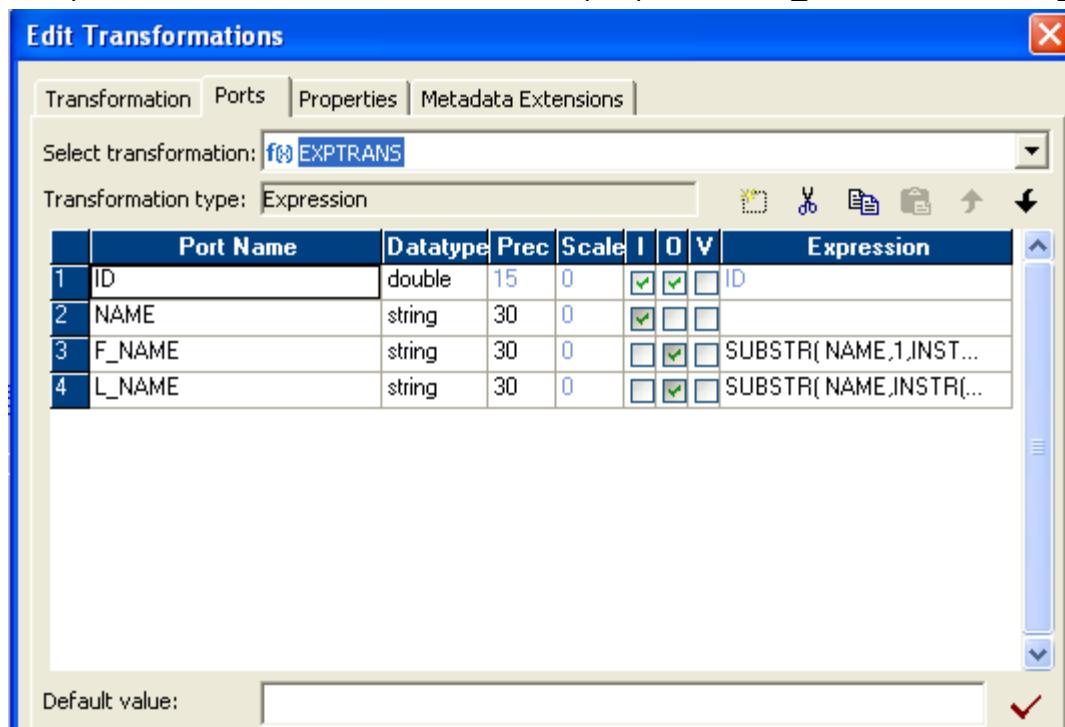
2 Chitra Dash

Steps for solving this scenario

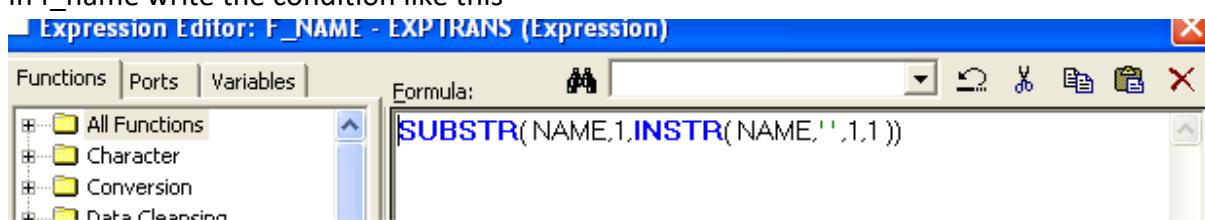
1. Drag the source to mapping area and connect with an expression transformation as shown bellow.



2. In expression transformation create two output port one is f_name and other is l_name.



3. In f_name write the condition like this



4. In l_name write the condition like this



Then connect the target.

How to Separate the original records in target?

Scenario 10: How to separate the original records from source table to separate target table by using rank transformation ?

Source Table

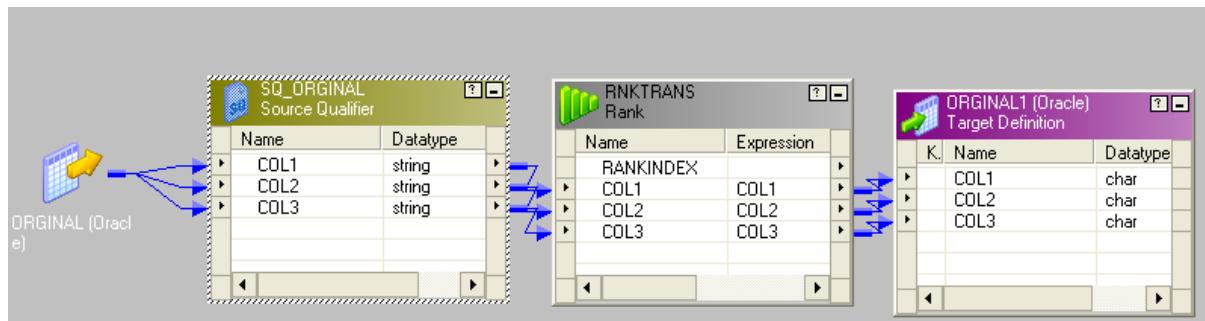
col1	col2	col3
a	b	c
x	y	z
a	b	c
r	f	u
a	b	c
v	f	r
v	f	r

Target Table

Col1	Col2	Col3
a	b	c
x	y	z
r	f	u
v	f	r

Solution:

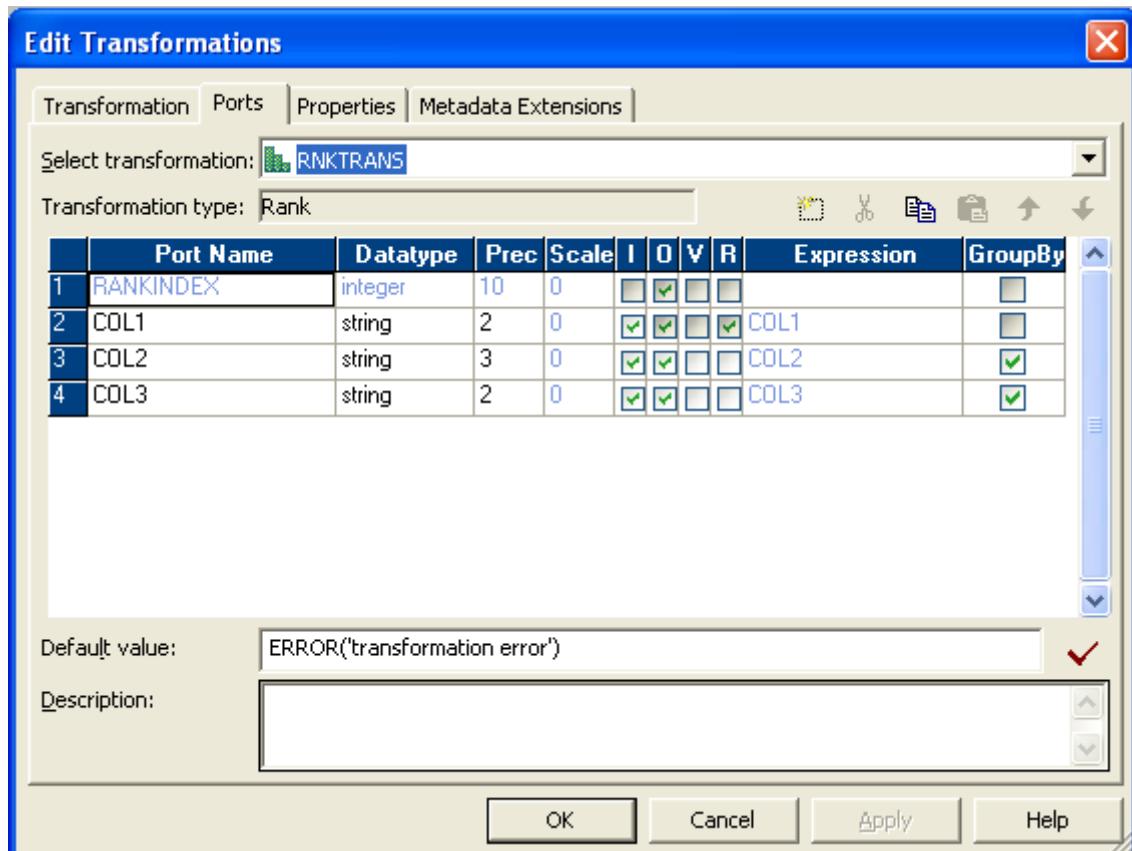
Step 1: Bring the source to mapping.



src to rank mapping

Step 2: Connect the rank to source.

Step 3: In rank, set the property like this.



rank property

Step 4: Then send it to target.

Run the session to see the result.

Tell me the process of Extracting Middle Name From Ename?

Suppose in e_name column is like this

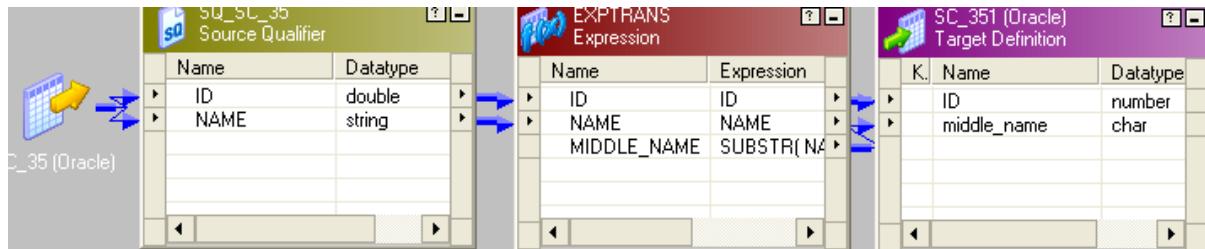
empno	ename
1	Sudhansu Sekher Dash
2	Amiya Prasad Mishra

In target we have to send middle name like this

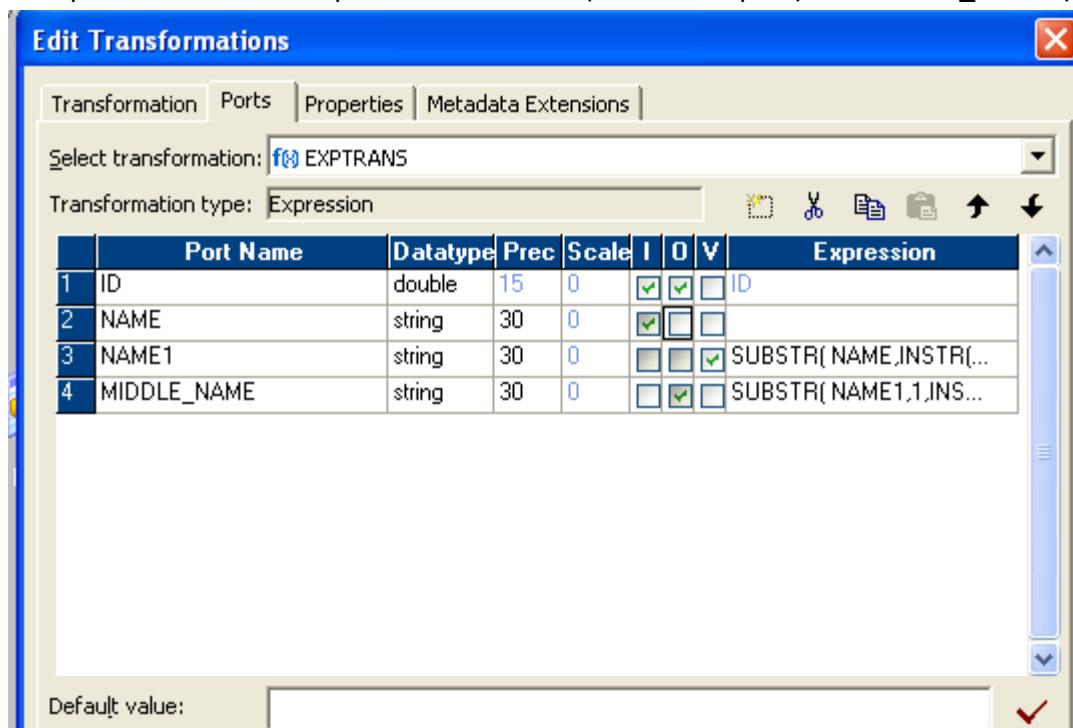
empno	ename
1	Sekher
2	Prasad

These are the steps for achieving this

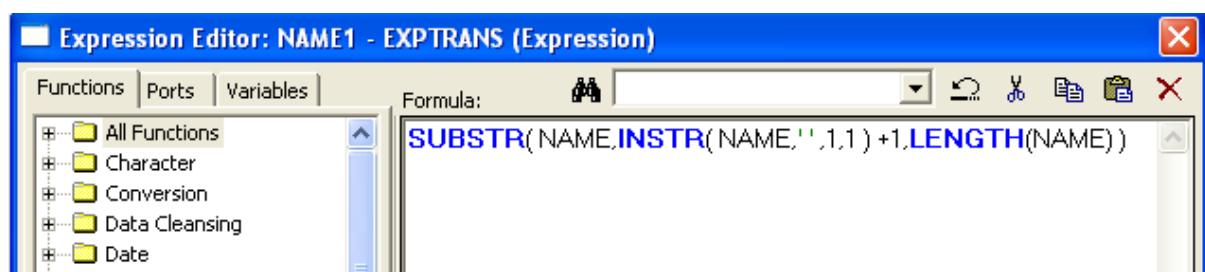
1. Drag the source and connect to an expression transformation



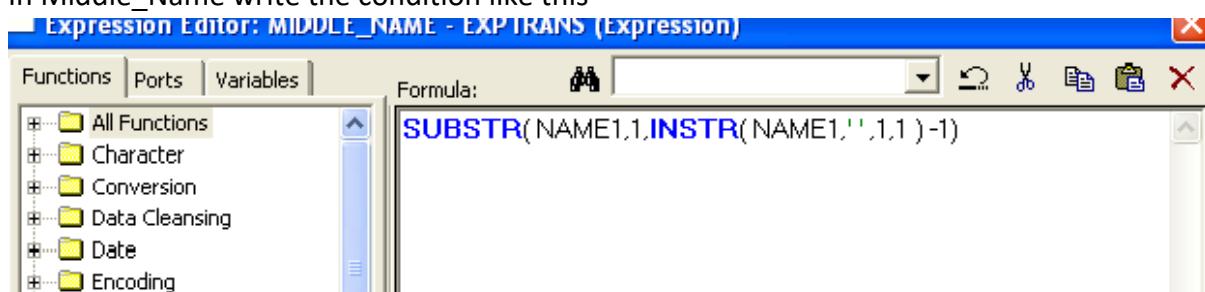
2. In Expression create two ports one is name1(as variable port) and Middle_Name (o/p port)



3. In Name1 write the condition like this



4. In Middle_Name write the condition like this



5. Then send to target.

How to do Alternate Target Loading?

My source is a flat file which contains N number of records. I want to load the source data into two targets such that first five records should loaded into the first target, next five records into the second target table. Again the next source five records into the first target table and so on. How to implement a Informatica mapping logic for this?

Solution:

- Connect the source qualifier transformation to the expression transformation. In the expression transformation, create the below additional ports:
v_cnt (variable port) = v_cnt+1
o_cnt (output port) = v_cnt
- Connect the expression transformation to the router transformation. Create two output groups in the router transformation and specify the following filter conditions:
--Filter condition for first output group
DECODE(substr(o_cnt,-1,1),1,TRUE,2,TRUE,3,TRUE,4,TRUE,5,TRUE,FALSE)
--Filter condition for second output group
DECODE(substr(o_cnt,-1,1),6,TRUE,7,TRUE,8,TRUE,9,TRUE,0,TRUE,FALSE)
- Connect the router transformation output groups to the appropriate targets.

How to Load source data in multiple session run?

I have flat file as a source which contains N number of records. My requirement is to load half of the source data into the target table in the first session run and the remaining half of the records in the second session run. Create Informatica mapping to implement this logic? Assume that the source data does not change between session runs.

Solution:

- Create a mapping to find out the number of records in the source and write the count to a parameter file. Let call this parameter as \$\$SOURCE_COUNT.
- Create another mapping. Go to the mapping parameters and variables, create a mapping variable (\$\$VAR_SESSION_RUNS) with integer data type.
- Connect the source qualifier transformation to the expression transformation. In the expression transformation, create the below additional ports.

v_Count (variable port) = v_Count+1
O_Run_flag (output port) = IIF(\$\$VAR_SESSION_RUNS=0,

```

setvariable($$vAR_SESSION_RUNS,1),
IIF( !ISNULL($$vAR_SESSION_RUNS)
    and v_Count=1,
    2,
    $$vAR_SESSION_RUNS)
)
O_count (output port) = V_Count

```

- Connect the expression transformation to the filter transformation and specify the following filter condition:

```

IIF (O_Run_Flag =1, v_count<= $$SOURCE_COUNT/2,
IIF (O_Run_Flag =2, v_count > $$SOURCE_COUNT/2))

```

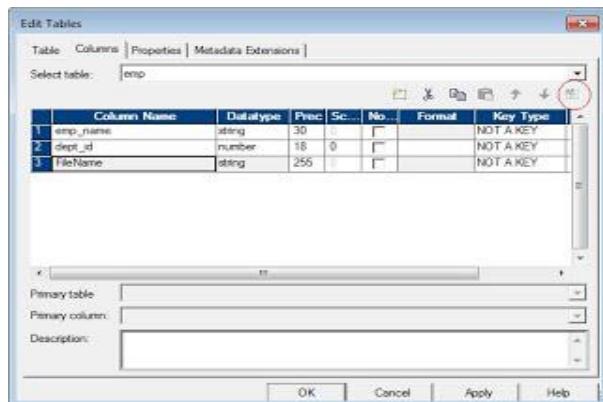
- Connect the filter transformation to the target.
- Here i am assuming that you know how to use a parameter file. That is why I did not specify the complete details.

How to generate Dynamic Target Flat File Name in Informatica?

Informatica 8.x or later versions provides a feature for generating the target files dynamically. This feature allows you to

- Create a new file for every session run
- create a new file for each transaction.

Informatica provides a special port,"FileName" in the Target file definition. This port you have to add explicitly. See the below diagram for adding the "FileName" port.



Go to the Target Designer or Warehouse builder and edit the file definition. You have to click on the button indicated in red color circle to add the special port.

Now we will see some informatica mapping examples for creating the target file name dynamically and load the data.

Tell me the process of Sending alternate record to target?

How to send alternate record to target?

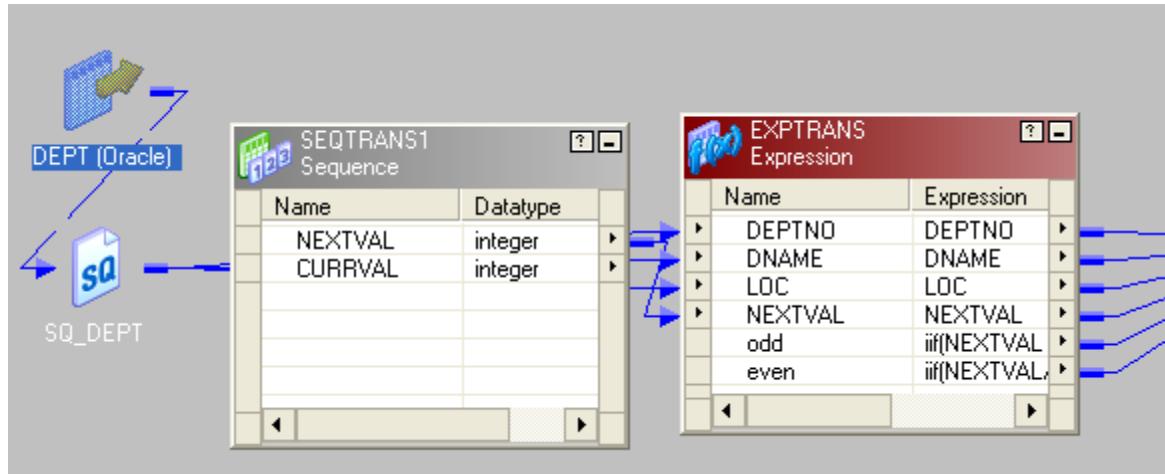
Or

Sending Odd numbered records to one target and even numbered records to another target.

Solution:

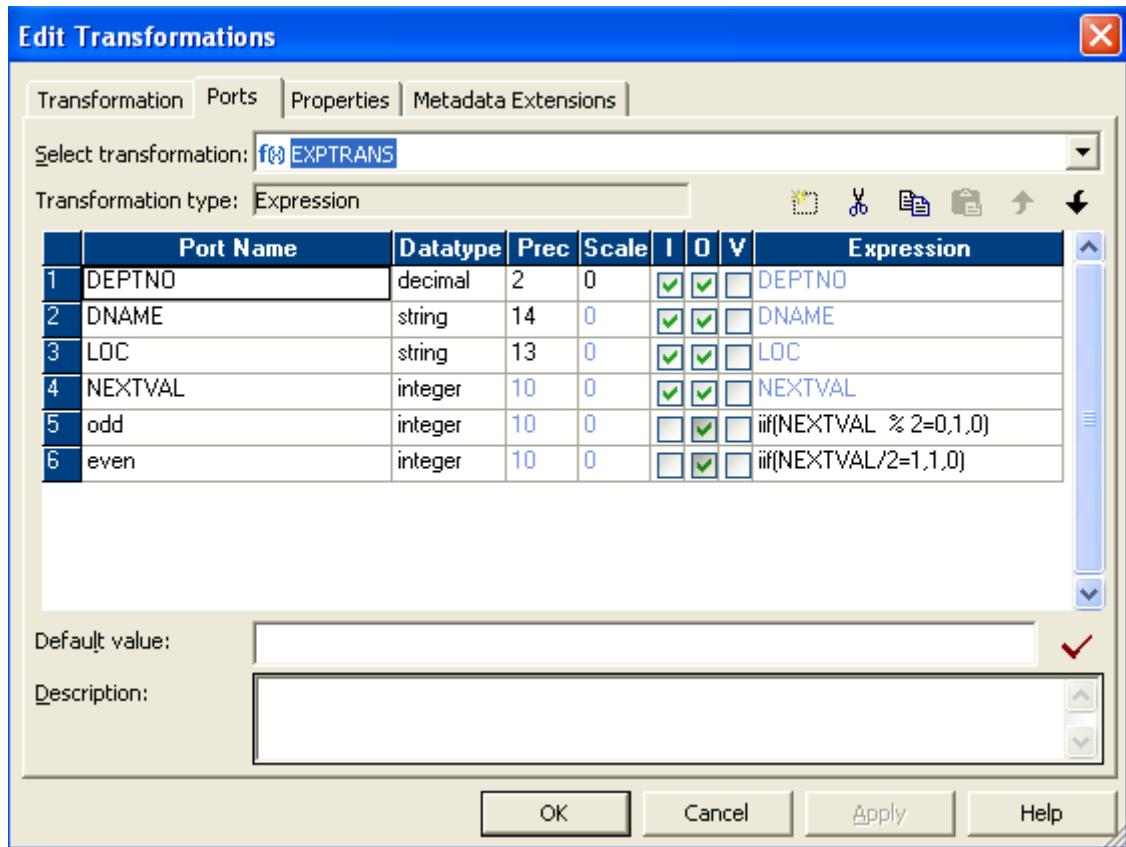
Step 1: Drag the source and connect to an expression transformation.

Step2: Add the next value of a sequence generator to expression transformation.



Step 3: In expression transformation make two port, one is "odd" and another "even".

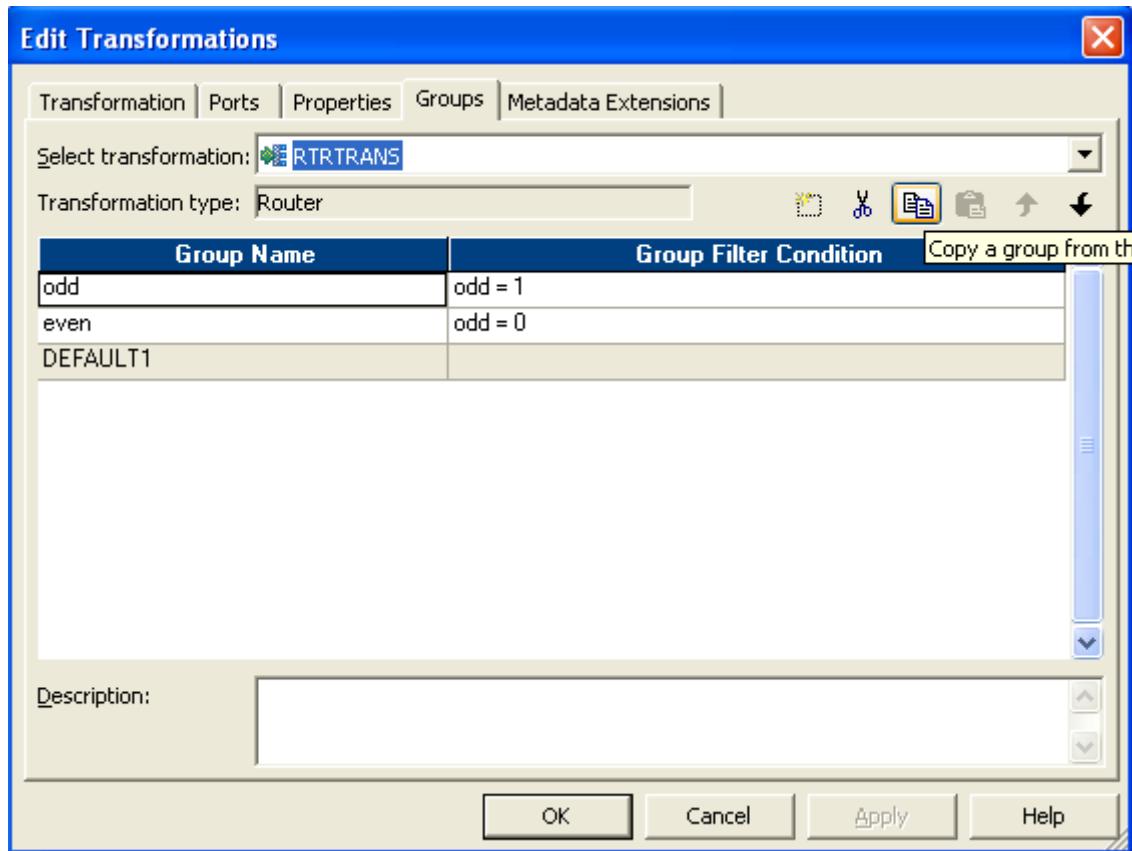
And Write the expression like below



Step 4: Connect a router transformation to expression.

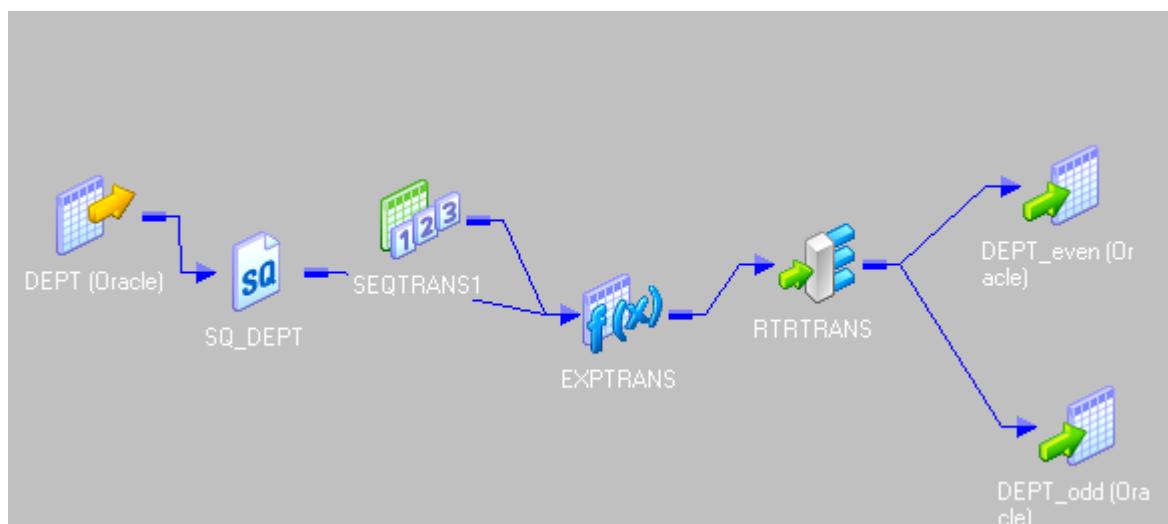
Make two group in router.

And give condition Like below



Step 5: Then send the two group to different targets.

The entire mapping is as below



How to Generate a new file for every session run?

Whenever the session runs you need to create a new file dynamically and load the source data into that file. To do this just follow the below steps:

STEP1: Connect the source qualifier to an expression transformation. In the expression transformation create an output port (call it as File_Name) and assign the expression as 'EMP_''||to_char(sessstarttime, 'YYYYMMDDHH24MISS')||'.dat'

STPE2: Now connect the expression transformation to the target and connect eh File_Name port of expression transformation to the FileName port of the target file definition.

STEP3: Create a workflow and run the workflow.

Here I have used sessstarttime, as it is constant throughout the session run. If you have used sysdate, a new file will be created whenever a new transaction occurs in the session run.

The target file names created would look like EMP_20120101125040.dat.

How to Create a new file once a day?

You can create a new file only once in a day and can run the session multiple times in the day to load the data. You can either overwrite the file or append the new data.

This is similar to the first problem. Just change the expression in expression transformation to 'EMP_''||to_char(sessstarttime, 'YYYYMMDD')||'.dat'. To avoid overwriting the file, use Append If Exists option in the session properties.

How to Create a new file for every session run. The file name should contain suffix as numbers (EMP_n.dat)?

In the above mapping scenario, the target flat file name contains the suffix as 'timestamp.dat'. Here we have to create the suffix as a number. So, the file names should looks as EMP_1.dat, EMP_2.dat and so on. Follow the below steps:

STPE1: Go the mappings parameters and variables -> Create a new variable, \$\$COUNT_VAR and its data type should be Integer

STPE2: Connect the source Qualifier to the expression transformation. In the expression transformation create the following new ports and assign the expressions.

```
v_count (variable port) = v_count+1  
v_file_count (variable port) = IIF(v_count = 1,  
SETVARIABLE($$COUNT_VAR,$$COUNT_VAR+1),$$COUNT_VAR)  
o_file_name (output port) = 'EMP_''||v_file_count||'.dat'
```

STEP3: Now connect the expression transformation to the target and connect the o_file_name port of expression transformation to the FileName port of the target.

How to Create a flat file based on the values in a port?

You can create a new file for each distinct values in a port. As an example consider the employees table as the source. I want to create a file for each department id and load the appropriate data into the files.

STEP1: Sort the data on department_id. You can either use the source qualifier or sorter transformation to sort the data.

STEP2: Connect to the expression transformation. In the expression transformation create the below ports and assign expressions.

v_curr_dept_id (variable port) = dept_id

v_flag (variable port) = IIF(v_curr_dept_id=v_prev_dept_id,0,1)

v_prev_dept_id (variable port) = dept_id

o_flag (output port) = v_flag

o_file_name (output port) = dept_id || '.dat'

STEP4: Now connect the expression transformation to the transaction control transformation and specify the transaction control condition as

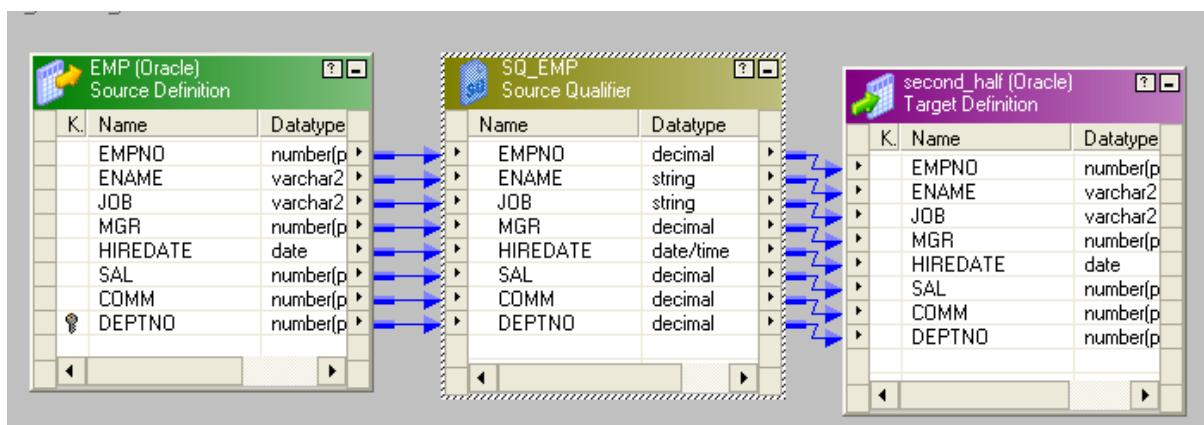
IIF(o_flag = 1, TC_COMMIT_BEFORE, TC_CONTINUE_TRANSACTION)

STEP5: Now connect to the target file definition.

How to send second half record to target?

Solution

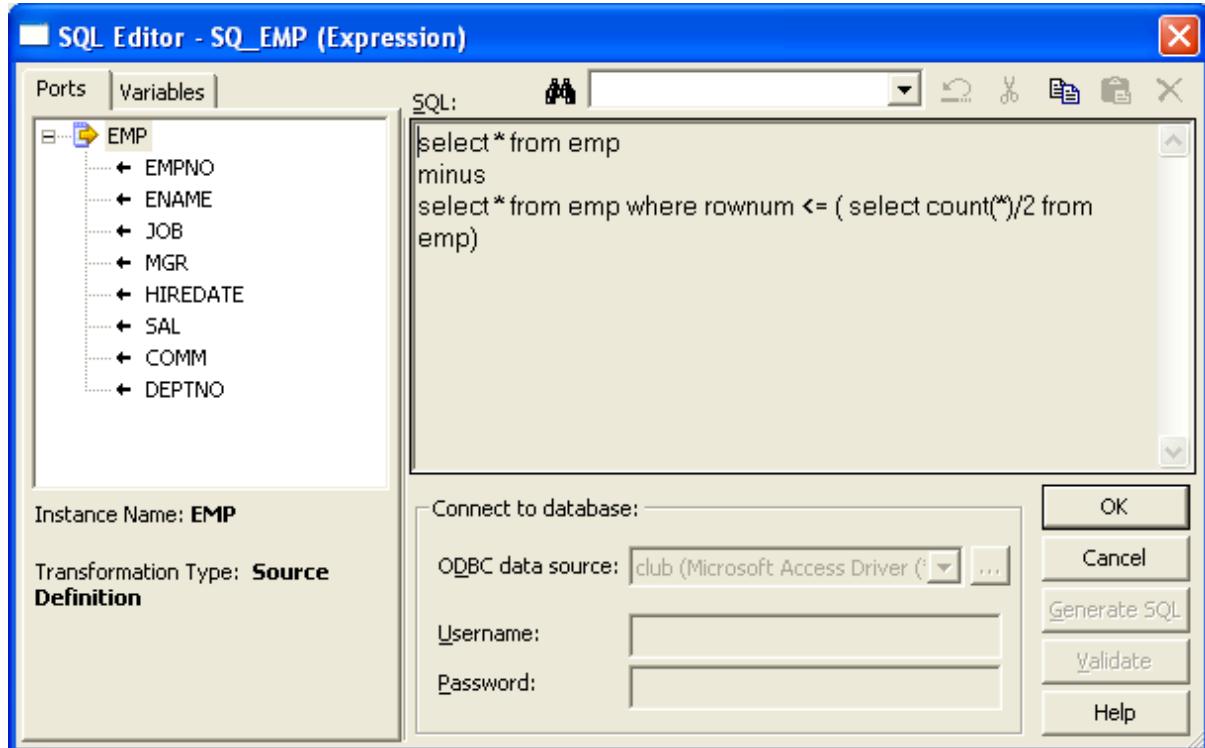
Step 1: Drag and drop the source to mapping.



src to tgt mapping

Step 2: In source-Qualifier , go to property and write the SQL query like

1 select * from emp minus select * from emp where rownum <= (select count(*)/2 from emp))



src qualifier sql query

Step:3 Then connect to target, and run mapping to see the results.

In this problem we will see how to implement the not equal operator, greater than, greater than or equal to, less than and less than or equal to operators when joining two tables in informatica.

Consider the below sales table as an example?

Table name: Sales

product,	prod_quantity,	price ,	Year
A , 10 , 100 , 2010			
B , 15 , 150 , 2010			
A , 8 , 80 , 2011			
B , 26 , 260 , 2011			

Now the problem is to identify the products whose sales is less than in the current year (In

this example: 2011) when compared to the last year.

Here in this example, Product A sold less in 2011 when compared with the sales in 2010.

This problem can be easily implemented with the help of SQL query as shown below

```
SELECT cy.*  
FROM SALES cy,  
      SALES py  
WHERE cy.product = py.product  
AND   cy.year=2011  
AND   py.year=2010  
AND   cy.prod_quantity < py.prod_quantity;
```

In informatica, you can specify only equal to condition in joiner. Now we will see how to implement this problem using informatica.

Solution:

STEP1: Connect two source qualifier transformations to the source definition. Call the first source qualifier transformation as sq_cy (cy means current year) and the other as sq_py (py means previous year).

STEP2: In the sq_cy source qualifier transformation, specify the source filter as price=2011. In the sq_py, specify the source filter as price=2010

STEP3: Now connect these two source qualifier transformations to joiner transformation and make sq_cy as master, sq_py as detail. In the join condition, select the product port from master and detail.

STEP4: Now connect all the master ports and only the prod_quantity port from detail to the filter transformation. In the filter transformation specify the filter condition as prod_quantity < prod_quantity1. Here prod_quantity port is from master port and prod_quantity1 is from detail port.

STEP4: Connect all the ports except the prod_quantity1 of filter transformation to the target definition.

How to implement the not exists operator in informatica which is available in database?

Solution:

Implementing the Not Exists operator is very easy in informatica. For example, we want to

get only the records which are available in table A and not in table B. For this use a joiner transformation with A as master and B as detail. Specify the join condition and in the join type, select detail outer join. This will get all the records from A table and only the matching records from B table.

Connect the joiner to a filter transformation and specify the filter condition as B_port is NULL. This will give the records which are in A and not in B. Then connect the filter to the target definition.

How to generate or load values in to the target table based on a column value using informatica etl tool?

I have the products table as the source and the data of the products table is shown below.

Table Name: Products

Product Quantity

Samsung NULL
Iphone 3
LG 0
Nokia 4

Now i want to duplicate or repeat each product in the source table as many times as the value in the quantity column. The output is

product Quantity

Iphone 3
Iphone 3
Iphone 3
Nokia 4
Nokia 4
Nokia 4
Nokia 4

The Samsung and LG products should not be loaded as their quantity is NULL, 0 respectively.

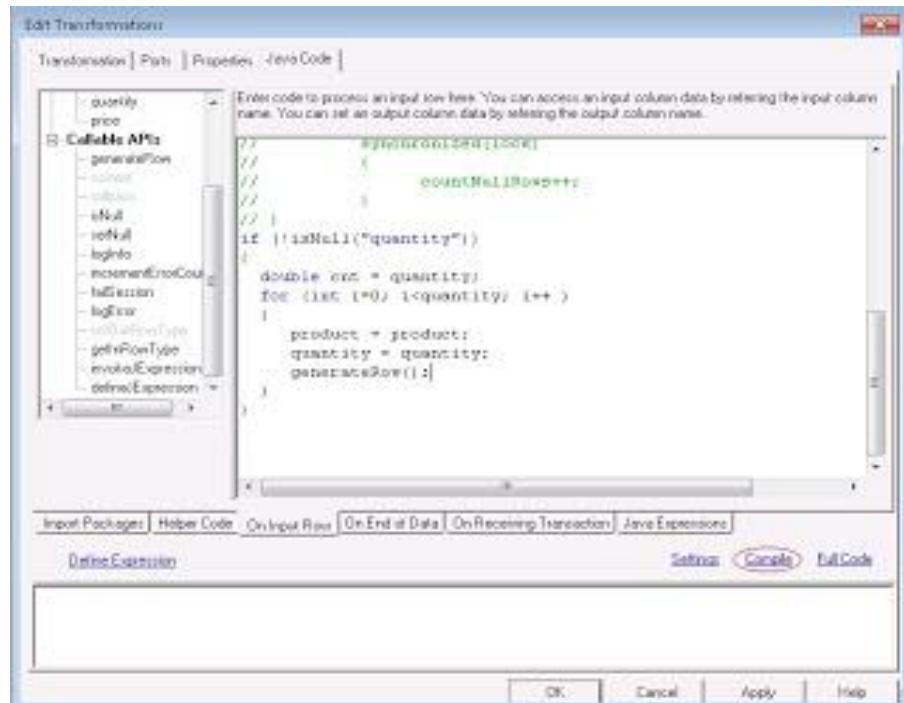
Now create informatica workflow to load the data in to the target table?

Solution:

Follow the below steps

- Create a new mapping in the mapping designer
- Drag the source definition in to the mapping
- Create the java transformation in active mode
- Drag the ports of source qualifier transformation in to the java transformation.
- Now edit the java transformation by double clicking on the title bar of the java transformation and go to the "Java Code" tab.
- Enter the below java code in the "Java Code" tab.

```
if (!isNull("quantity"))
{
    double cnt = quantity;
    for (int i = 1; i <= quantity; i++)
    {
        product = product;
        quantity = quantity;
        generateRow();
    }
}
```



- Now compile the java code. The compile button is shown in red circle in the image.

- Connect the ports of the java transformation to the target.
- Save the mapping, create a workflow and run the workflow.

How to implement SCD TYPE1?

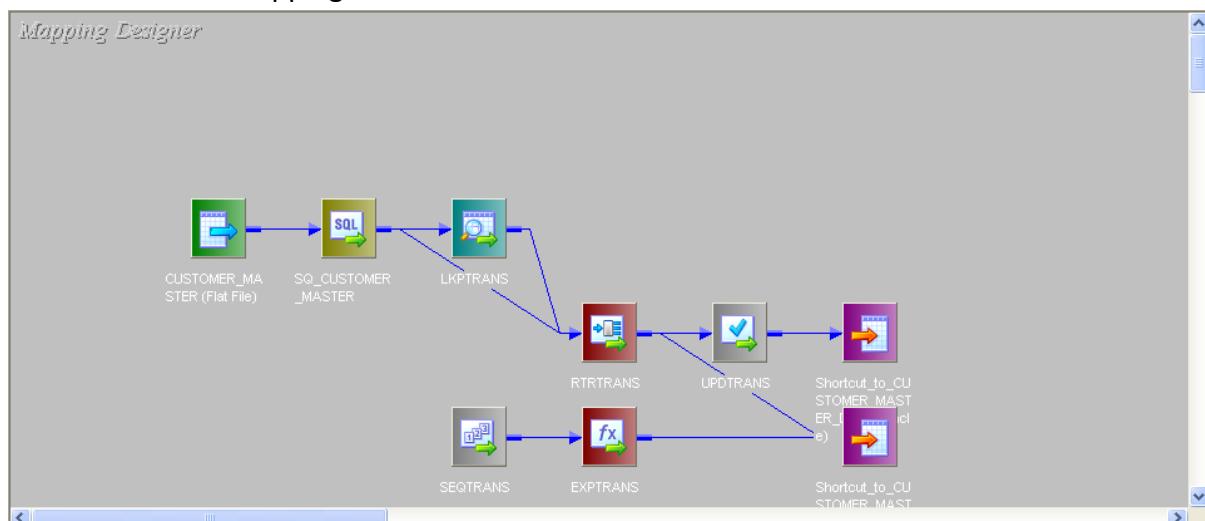
The SCD Type 1 methodology overwrites old data with new data, and therefore does not need to track historical data .

1. Here is the source

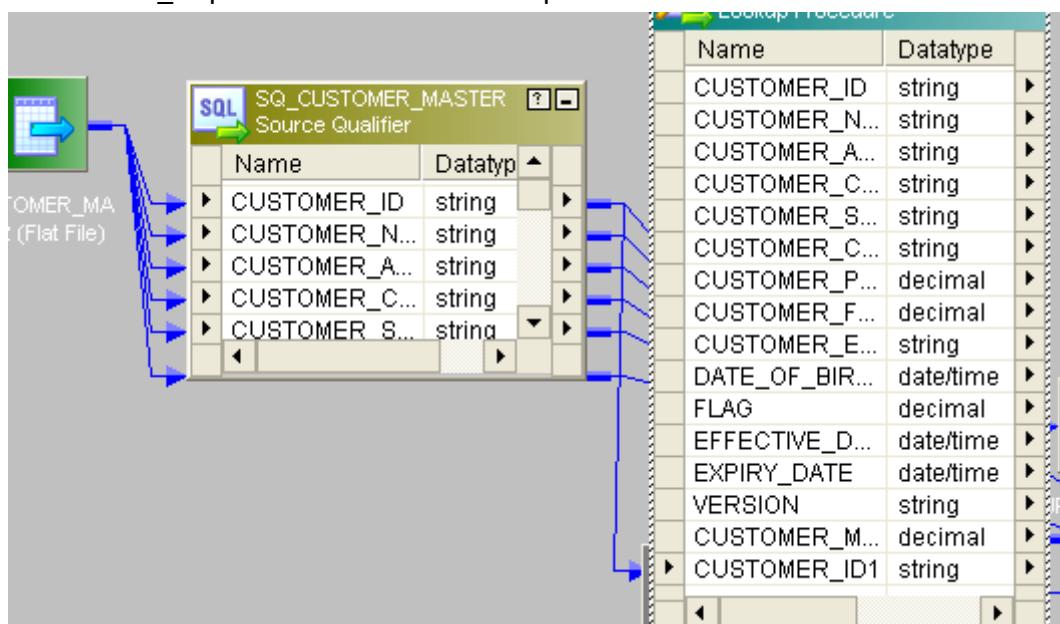
A screenshot of a data integration tool's interface. On the left, there is a green 'MER_MA Flat File' source icon with a blue arrow pointing to a 'SQL Source Qualifier' component. The 'SQL Source Qualifier' has a title bar 'SQ_CUSTOMER_MASTER' and a sub-label 'Source Qualifier'. It contains a table with columns: Name and Datatype. The columns listed are CUSTOMER_ID (string), CUSTOMER_NAME (string), CUSTOMER_ADDRESS (string), CUSTOMER_CITY (string), CUSTOMER_STATE (string), CUSTOMER_COUNTRY (string), CUSTOMER_PHONE (decimal), CUSTOMER_FAX (decimal), CUSTOMER_EMAIL (string), CUSTOMER_TYPE (string), CUSTOMER_INDUSTRY (string), MONTHLY_INCOME_LE... (decimal), DATE_OF_BIRTH (string), GENDER (string), MARITAL_STATUS (string), and ACTION_FLAG (string). Blue arrows connect the source file icon to each column in the table.

Name	Datatype
CUSTOMER_ID	string
CUSTOMER_NAME	string
CUSTOMER_ADDRESS	string
CUSTOMER_CITY	string
CUSTOMER_STATE	string
CUSTOMER_COUNTRY	string
CUSTOMER_PHONE	decimal
CUSTOMER_FAX	decimal
CUSTOMER_EMAIL	string
CUSTOMER_TYPE	string
CUSTOMER_INDUSTRY	string
MONTHLY_INCOME_LE...	decimal
DATE_OF_BIRTH	string
GENDER	string
MARITAL_STATUS	string
ACTION_FLAG	string

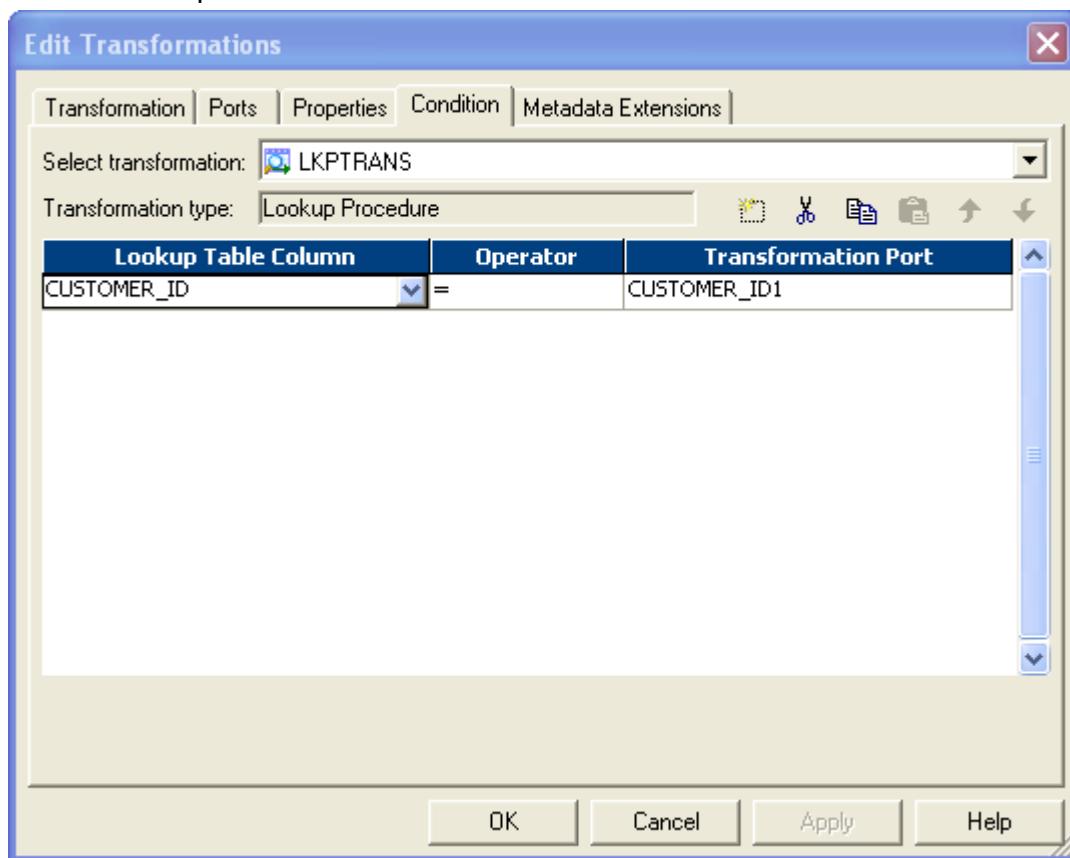
2. We will compare the historical data based on key column CUSTOMER_ID.
3. This is the entire mapping



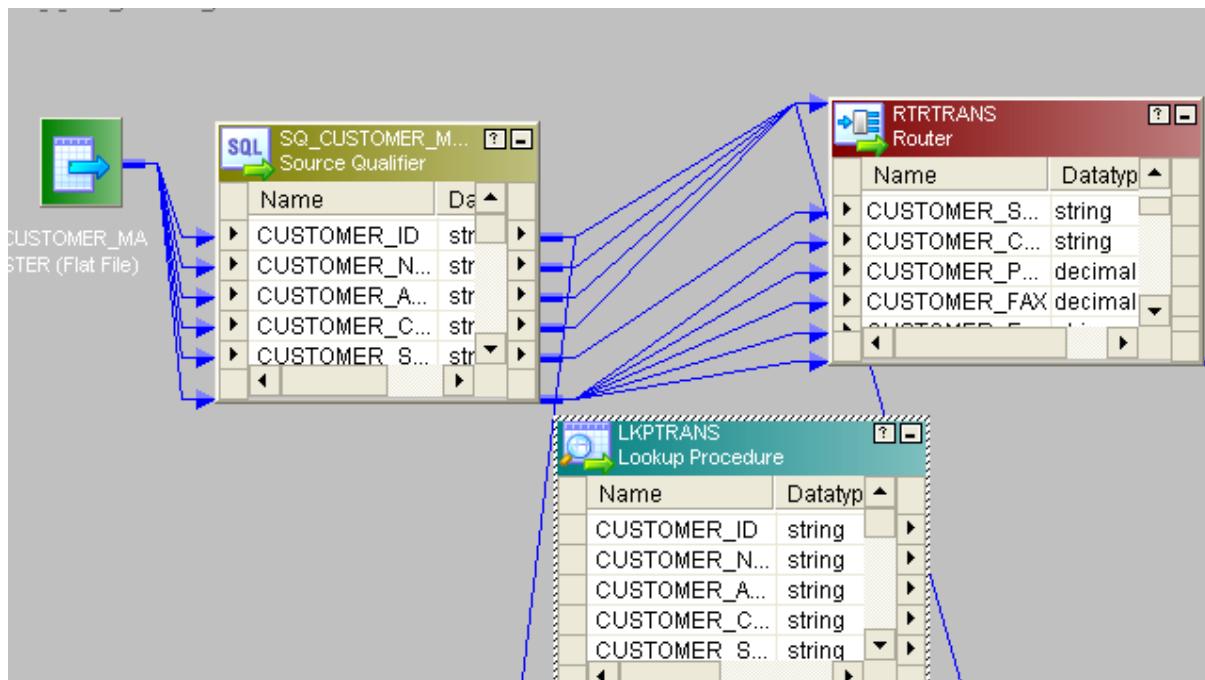
4. Connect lookup to source. In Lookup fetch the data from target table and send only CUSTOMER_ID port from source to lookup



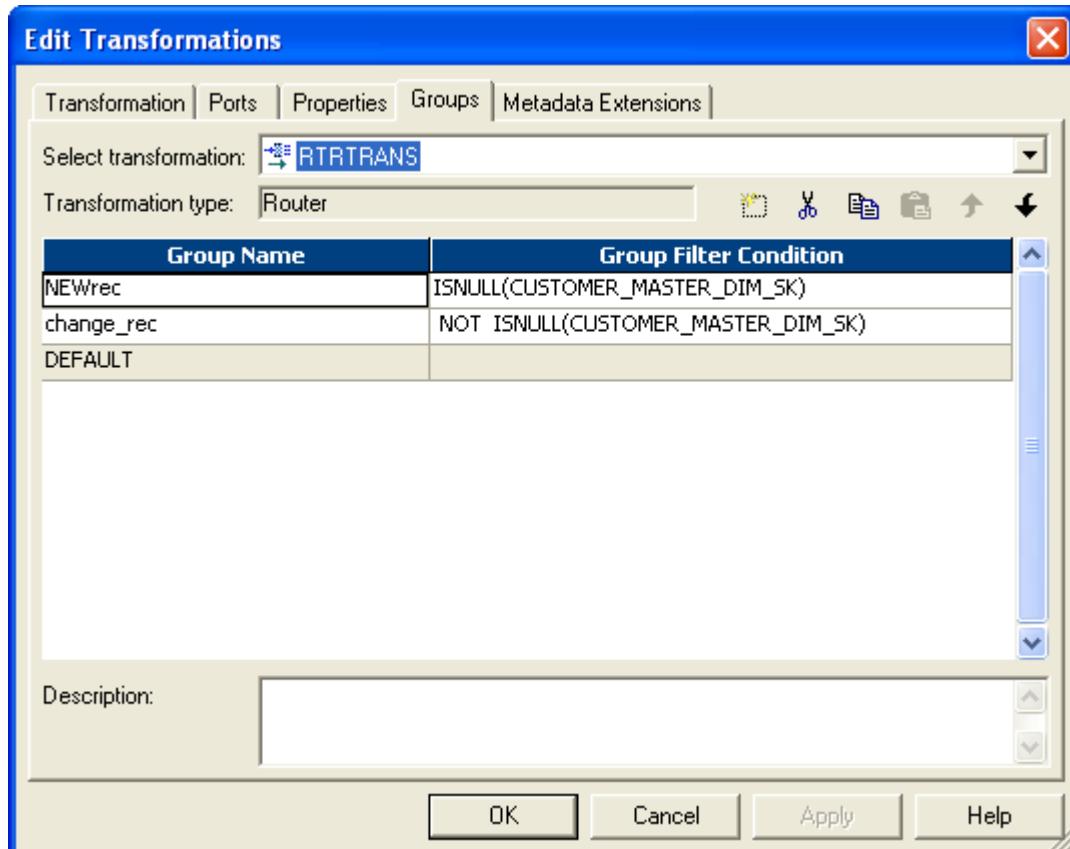
5. Give the lookup condition like this



6. Then rest of the columns from source send to one router transformation

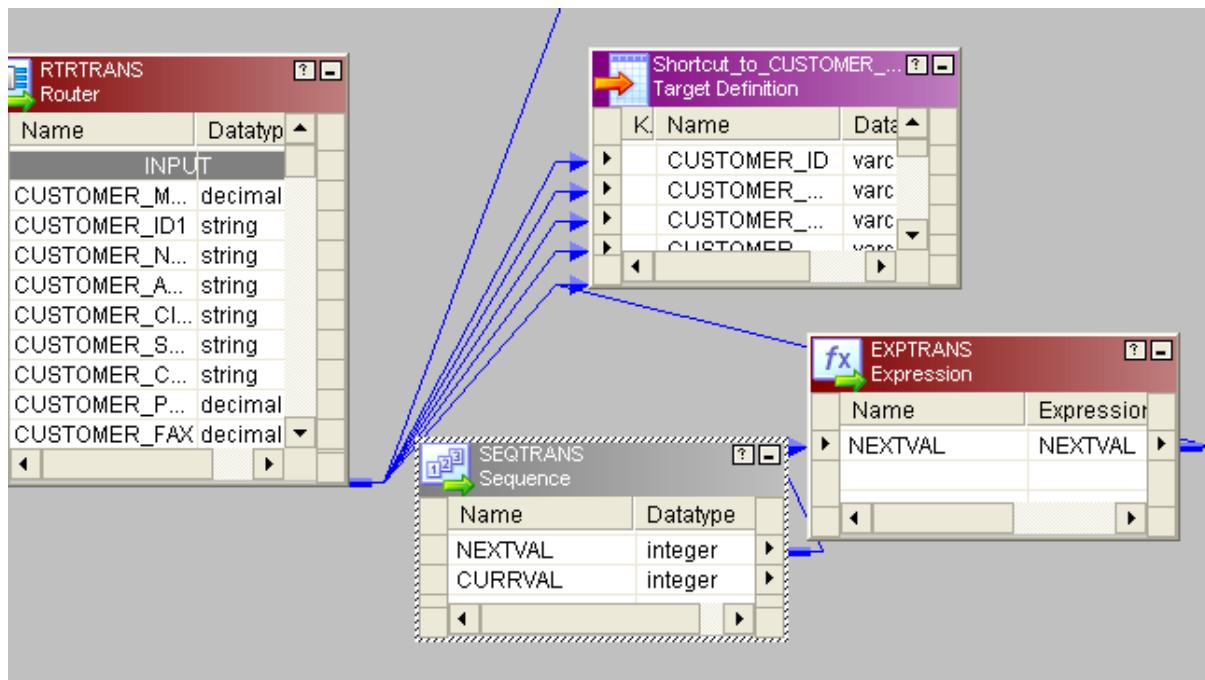


7. In router create two groups and give condition like this

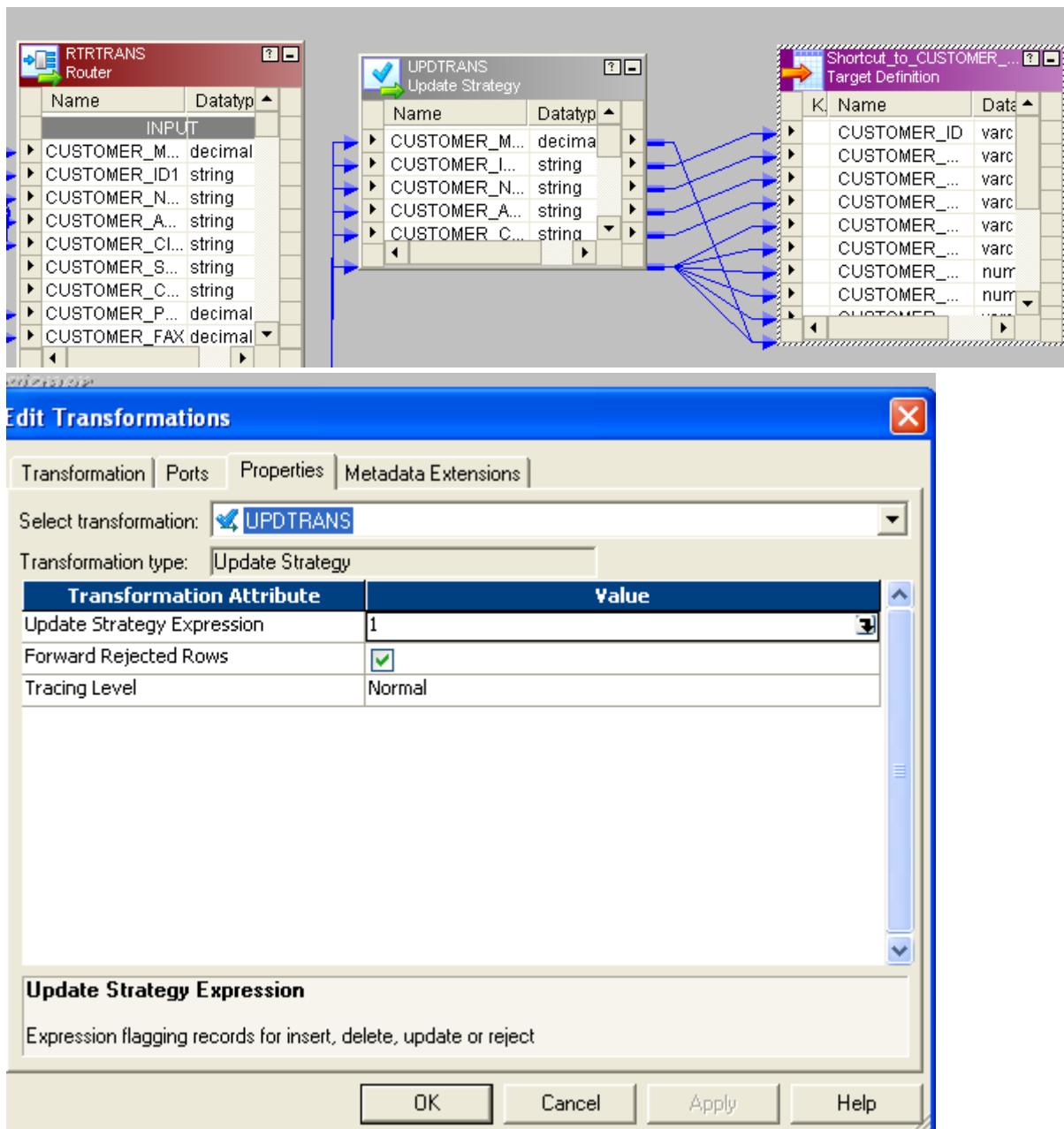


8. For new records we have to generate new customer_id. For that take a sequence generator and connect the next column to expression .New_rec group from router connect to target1(Bring two instances of target to mapping, one for new rec and other for old rec)

.Then connect next_val from expression to customer_id column of target



9. Change_rec group of router bring to one update strategy. and give the condition like this



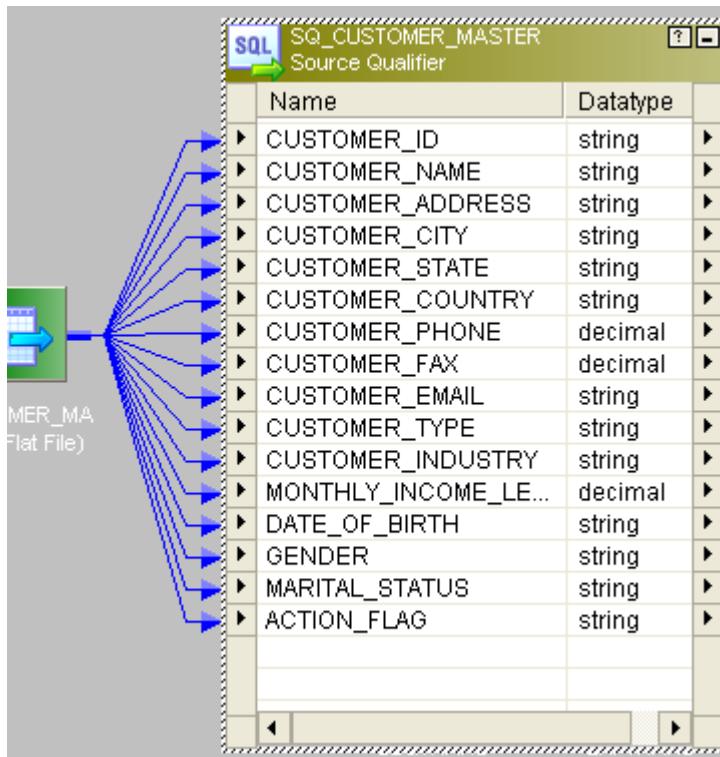
10. Instead of 1 you can give dd_update in update-stratgy. Then connect to target.

How to implement SCD Type2?

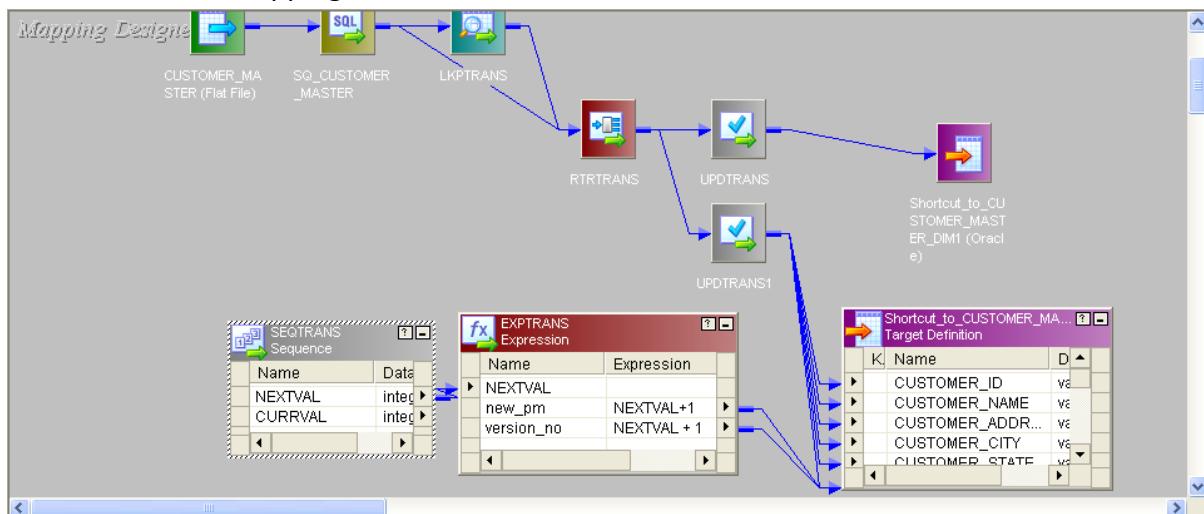
In Type 2 Slowly Changing Dimension, if one new record is added to the existing table with a new information then both the original and the new record will be presented having new records with its own primary key.

- To identifying new_rec we should add one new_pm and one version_no.

2. This is the source.



3. This is the entire mapping

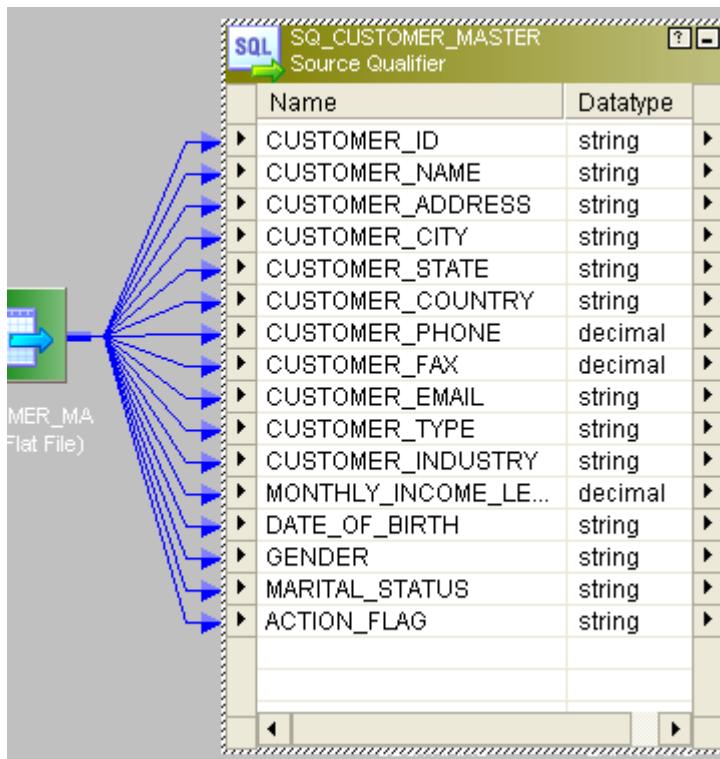


4. All the procedure same as described in SCD TYPE1 mapping. The Only difference is , From router new_rec will come to one update_strategy and condition will be given dd_insert and one new_pm and version_no will be added before sending to target.
5. Old_rec also will come to update_strategy condition will given dd_insert then will send to target.

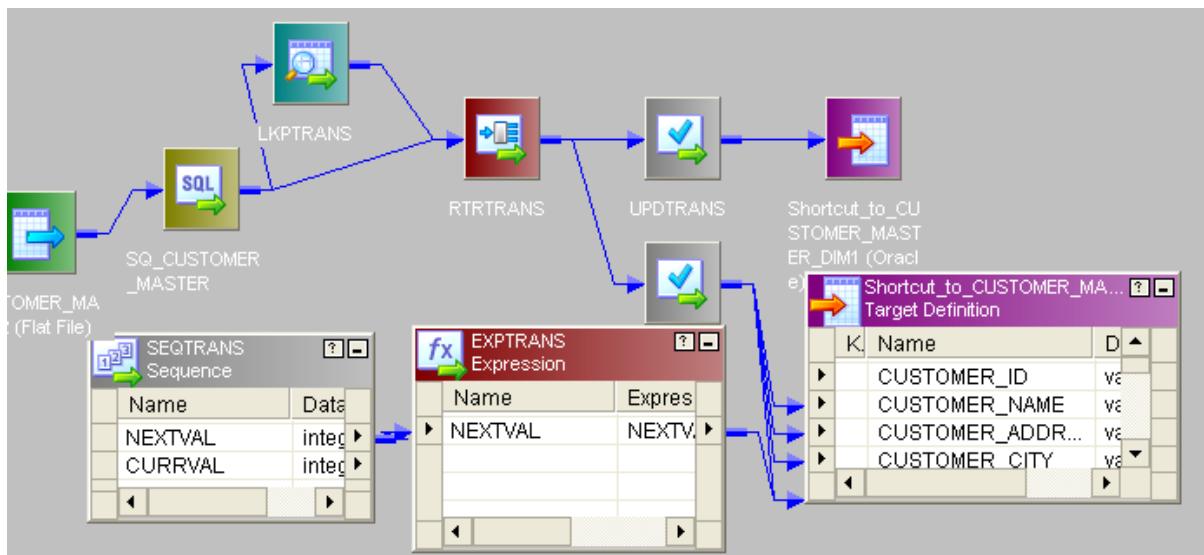
How to implement SCD Type3?

In SCD Type3 ,there should be added two column to identifying a single attribute. It stores one time historical data with current data

1. This is the source



2. This is the entire mapping



3. Up to router transformation ,all the procedure is same as described in Scenario_36 SCD type1.
4. The only difference is after router bring the new_rec to router and give condition dd_insert send to target. Create one new primary key send to target.
5. For old_rec send to update_strategy and set condition dd_insert and send to target.
6. You can create one effective_date column in old_rec table

What needs to be done under Unit Testing?

In unit testing what we need do is something like below

1. Validate source and target
 - Analyse & validate your transformation business rules.
 - We need review field by field from source to target and ensure that the required transformation logic is applied.
 - We generally check the source and target counts for each mapping.
2. Analyse the success and reject rows
 - In this stage we generally customized sql queries to check source and target.
 - Analyse the rejected rows and build the process to handle this rejection.
3. Calculate the load time
 - Run the session and view the statistics
 - We observe how much time is taken by reader and writer .
 - We should look at lesion log and workflow log to view the load statistics
4. Testing performance
 - Source performance
 - Target performance
 - Session performance
 - Network performance
 - Database performance

After unit testing we generally prepare one document as described below

Best Practices in Informatica

We have tried to come up with some of best practices in informatica

- 1) Always try to add expression transformation after source qualifier and before Target. If source or target definition changes it is easier to reconnect the ports
- 2) Always use Cobol File for normaliser in binary format otherwise there are lot of issues specially with comp-3 fields
- 3) Remove unused ports, though unused ports do not have any effect on performance but it is always better to remove them for more visibility
- 4) If possible try to do calculation in Output Ports instead of variable ports as variable ports are need to assign/reassign each time and it can slow down the performance
- 5) Try to avoid complex filter expression instead of that try to evaluate filter expression in upstream expression transformation and pass it to filter transformation. If you use too many complex calculations in filter condition expression it can slow down performance.
- 6) In workflow Source/Target directory Property take advantage of Unix links. Instead of hard coding path in source/target directory specify path with Unix link i.e. suppose in dev environment you are specifying Source directory path as /deserver/team/source and in prod server you specify it as /prod server/team/source .You can get link created in \$PMRootDir in as src_file_dir pointing to /devserver/team/source in

dev server and /prodserver/team/source in prod server and in your source/Target file directory you can put path as \$PMRootDir/src_file_dir In this case there is no need to change Source/Target directory every time you move between production and dev and testing

- 7) In sequence generator do not connect current value port to downstream (unless required) transformation as when we connect current value port from sequence generator transformation Informatica Server processes one row in each block. We can optimize performance by connecting only the NEXTVAL port in a mapping .
- 8) Improve lookup performance by putting all conditions that use the equality operator '=' first in the list of conditions under the condition tab.
- 9) Always remember rule not to cache look tables having more than 550000 rows (Assuming row size 1024) .If your row size is less than or more than 1024 then adjust number of rows accordingly .
- 10) Avoid calculating same value again and again. Instead of that store it in a variable use it several times.

How can we do Performance Tuning in Informatica?

The goal of performance tuning is to optimize session performance so that the sessions run during the available load window for the Informatica Server. Increase the session performance by following.

- **Network:** The performance of the Informatica Server is related to network connections. Data generally moves across a network at less than 1 MB per second, whereas a local disk moves data five to twenty times faster. Thus network connections often affect on session performance. So avoid network connections.
- **Flat files:** If your flat files stored on a machine other than the informatica server, move those files to the machine that consists of informatica server.
- **Less Connections:** Minimize the connections to sources ,targets and informatica server to improve session performance. Moving target database into server system may improve session performance.
- **Staging areas:** If you use staging areas you force informatica server to perform multiple data passes. Removing of staging areas may improve session performance. Use staging area only when its mandatory
- **Informatica Servers:** You can run the multiple informatica servers against the same repository. Distributing the session load to multiple informatica servers may improve session performance.
- Run the informatica server in ASCII data movement mode improves the session performance. Because ASCII data movement mode stores a character value in one byte. Unicode mode takes 2 bytes to store a character.

- **Source qualifier:** If a session joins multiple source tables in one Source Qualifier, optimizing the query may improve performance. Also, single table select statements with an ORDER BY or GROUP BY clause may benefit from optimization such as adding indexes.
- **Drop constraints:** If target consists key constraints and indexes it slows the loading of data. To improve the session performance in this case drop constraints and indexes before we run the session(while loading facts and dimensions) and rebuild them after completion of session.
- **Parallel sessions:** Running a parallel sessions by using concurrent batches will also reduce the time of loading the data. So concurrent batches may also increase the session performance.
- **Partitioning:** the session improves the session performance by creating multiple connections to sources and targets and loads data in parallel pipe lines.
- **Incremental Aggregation:** In some cases if a session contains an aggregator transformation ,you can use incremental aggregation to improve session performance.
- **Transformation Errors:** Avoid transformation errors to improve the session performance. Before saving the mapping validate it and see and if any transformation errors rectify it.
- **Lookup Transformations:** If the session contained lookup transformation you can improve the session performance by enabling the look up cache. The cache improves the speed by saving the previous data and hence no need to load that again
- **Filter Transformations:** If your session contains filter transformation ,create that filter transformation nearer to the sources or you can use filter condition in source qualifier.
- **Group transformations:** Aggregator, Rank and joiner transformation may often decrease the session performance. Because they must group data before processing it. To improve session performance in this case use sorted ports option ie sort the data before using the transformation.
- **Packet size:** We can improve the session performance by configuring the network packet size, which allows data to cross the network at one time. To do this go to server manger, choose server configure database connections.

Crack Your Interview Today



Interview Question Bank

India

[Store](#) [Contact](#)

Cart (0)

Interview Question Bank

For Complete Study Material For Interview Preparations Click On Below Link

https://www.instamojo.com/interview_questions

[Click On Below Links For More Courses](#)

Informatica 650+ Interview Questions & Answers

https://www.instamojo.com/interview_questions/informatica-interview-questions-asked-in-top

SQL 500+ Interview Questions & Answers

https://www.instamojo.com/interview_questions/sql-interview-questions-asked-in-top-it-comp

Unix 500+ Interview Questions & Answers

https://www.instamojo.com/interview_questions/unix-interview-questions-asked-in-top-it-com

SAP BODS 250+ Interview Questions & Answers

https://www.instamojo.com/interview_questions/sap-bods-interview-questions-answers-asked-i

SAP BO 250+ Interview Questions & Answers

https://www.instamojo.com/interview_questions/sap-business-objects-interview-questions-ask

Python 500+ Interview Questions & Answers

https://www.instamojo.com/interview_questions/python-interview-questions-asked-in-top-it-c

HR 150+ Interview Questions & Answers

https://www.instamojo.com/interview_questions/hr-round-interview-questions-answers-asked-i

Data Warehousing and ETL Interview Questions & Answers

https://www.instamojo.com/interview_questions/data-warehousing-and-etl-interview-questions

SCD Types (0,1,2,3,4,5,6,7) & Explanation

https://www.instamojo.com/interview_questions/scd-5a66d

Campus Placement 2000+ Interview Questions & Answers For Freshers

https://www.instamojo.com/interview_questions/campus-placement-interview-questions-with-an

100+ Resume , CV & Cover Letters

https://www.instamojo.com/interview_questions/download-your-favorite-resume-template-and-c

Register Here For Job Referral Program

https://www.instamojo.com/interview_questions/apply-for-referral-program-and-get-an-opport