

Informatica



LEARN IN 1 DAY

KRISHNA RUNGTA

Learn Informatica in 1 Day

By Krishna Rungta

Copyright 2017 - All Rights Reserved – Krishna Rungta

ALL RIGHTS RESERVED. No part of this publication may be reproduced or transmitted in any form whatsoever, electronic, or mechanical, including photocopying, recording, or by any informational storage or retrieval system without express written, dated and signed permission from the author.

Table Of Content

Chapter 1: Introduction

Chapter 2: Architecture

1. [Informatica Domain](#)
2. [PowerCenter Repository](#)
3. [Domain Configuration](#)
4. [Properties of the domain](#)
5. [Powercenter client & Server Connectivity](#)
6. [Repository Service](#)
7. [Integration Service](#)
8. [Sources & Targets](#)

Chapter 3: Configure Clients and Repositories

1. [What is Informatica domain?](#)
2. [How to Open the Administrator Home Page](#)
3. [How to Create Repository Service](#)
4. [How to create repository contents](#)
5. [How to create Integration Service](#)
6. [How to Configuring client and domain](#)
7. [How to create user](#)

Chapter 4: Source Analyzer and Target Designer

1. [How to open Source Analyzer](#)
2. [How to import Source table in Source Analyzer](#)
3. [How to open Target Designer & import target in target designer](#)
4. [How to create folder in Informatica](#)

Chapter 5: Mappings

1. [Why do you need Mapping?](#)
2. [Components of Mapping](#)
3. [What is Stage Mapping?](#)
4. [How to Create a Mapping](#)

5. [Mapping Parameters and Variables](#)
6. [How to Create Mapping Parameter](#)
7. [How to Create Mapping Variable](#)

Chapter 6: Workflows

1. [Open Workflow Manager](#)
2. [Workflow Monitor](#)
3. [How to open Workflow Monitor](#)
4. [Task View](#)
5. [Gantt Chart View](#)

Chapter 7: Debug Mappings & Session Objects

1. [Steps to use debugger in a mapping](#)
2. [Using breakpoint in debugger](#)
3. [Tracing in Informatica](#)
4. [How to set tracing level in a transformation](#)
5. [Session Objects](#)

Chapter 8: Transformations

1. [Transformations & Filter Transformation](#)
2. [Source Qualifier Transformation](#)
3. [Aggregator Transformation](#)
4. [Router Transformation](#)
5. [Joiner Transformation](#)
6. [Rank Transformation](#)
7. [Sequence Generator Transformation](#)
8. [Transaction Control Transformation](#)
9. [Lookup and Re-usable Transformation](#)
10. [Normalizer Transformation](#)
11. [Performance Tuning for Transformation](#)

Chapter 1: Introduction

Informatica is a Software development company, which offers data integration products. It offers products for ETL, data masking, data Quality, data replica, data virtualization, master data management, etc.

Informatica Powercenter ETL/Data Integration tool is a most widely used tool and in the common term when we say Informatica, it refers to the Informatica PowerCenter tool for ETL.

Informatica Powercenter is used for Data integration. It offers the capability to connect & fetch data from different heterogeneous source and processing of data.

For example, you can connect to an SQL Server Database and Oracle Database both and can integrate the data into a third system.

The latest version of Informatica PowerCenter available is 9.6.0. The different editions for the PowerCenter are

- Standard edition
- Advanced edition
- Premium edition

The popular clients using Informatica Powercenter as a data integration tool are U.S Air Force, Allianz, Fannie Mae, ING, Samsung, etc. The popular tools available in the market in competition to Informatica are IBM Datastage, Oracle OWB, Microsoft SSIS and Ab Initio.

Typical use cases for Informatica can be

- An organization migrating from existing legacy system like mainframe to a new database system. So the migration of its existing data into a system can be performed.

- Enterprises setting up their Data Warehouse would require an ETL tool to move data from the Production system to Warehouse.
- Integration of data from various heterogeneous systems like multiple databases and file-based systems can be done using Informatica.
- Informatica can be used as a data cleansing tool.

Informatica is better than its competitors as it offers a wide range of product editions. So the user can opt for a specific edition based upon the requirement. Informatica is constantly featured as Data Integration product leader in the Gartner Magic Quadrant listing.

Informatica is available for all the popular platforms. It offers cloud-based services so that with minimal setup an industry can use this tool. Informatica offers real-time data integration, web services integration, Business to business data integration (B2B), Big data edition, Master Data Management and connectors for social media and Salesforce. Forbes has quoted Informatica as the next Microsoft, this itself reflects the market share Informatica is having over its competitors.

Why do we need Informatica?

Informatica comes to the picture wherever we have a data system available and at the backend we want to perform certain operations on the data. It can be like cleaning up of data, modifying the data, etc. based on certain set of rules or simply loading of bulk data from one system to another.

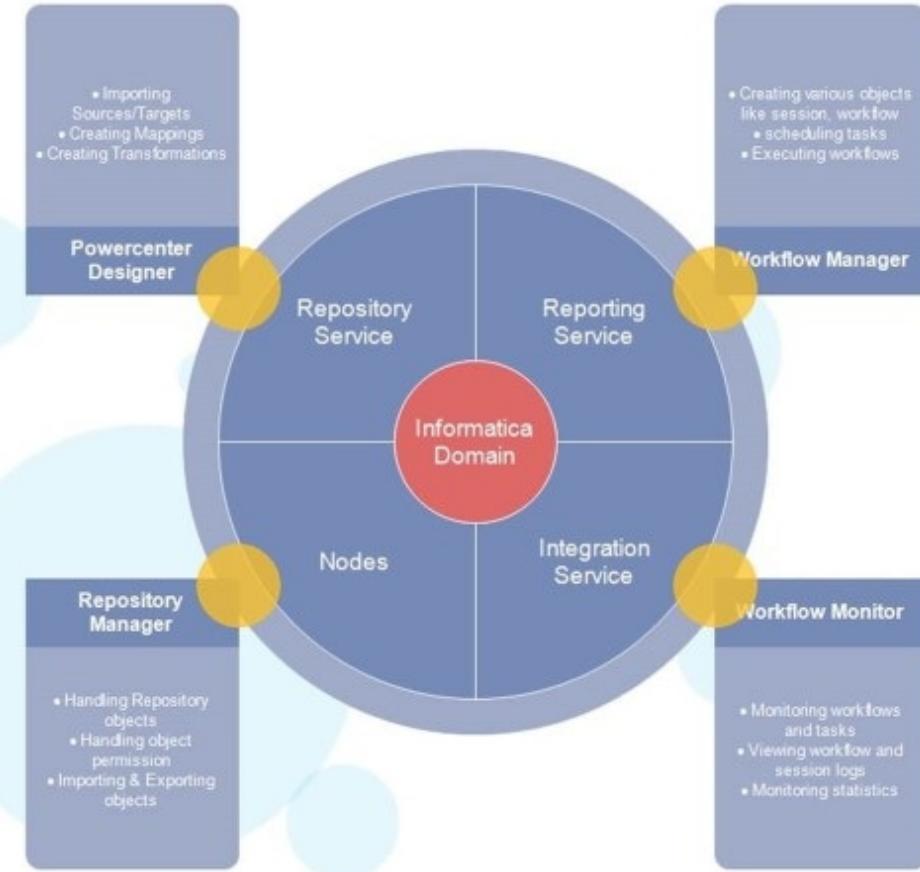
Informatica offers a rich set of features like operations at row level on data, integration of data from multiple structured, semi-structured or unstructured systems, scheduling of data operation. It also has the feature of metadata, so the information about the process and data operations are also preserved.

Chapter 2: Architecture

Before we learn how to use Informatica, we need to understand what are the important components of Informatica and how it works.

Informatica tool consists of following services & components

1. Repository Service – Responsible for maintaining Informatica metadata & providing access of same to other services.
2. Integration Service – Responsible for the movement of data from sources to targets
3. Reporting Service - Enables the generation of reports
4. Nodes – Computing platform where the above services are executed
5. Informatica Designer - Used for creation of mappings between source and target
6. Workflow Manager – Used to create workflows and other task & their execution
7. Workflow Monitor – Used to monitor the execution of workflows
8. Repository Manager – Used to manage objects in repository

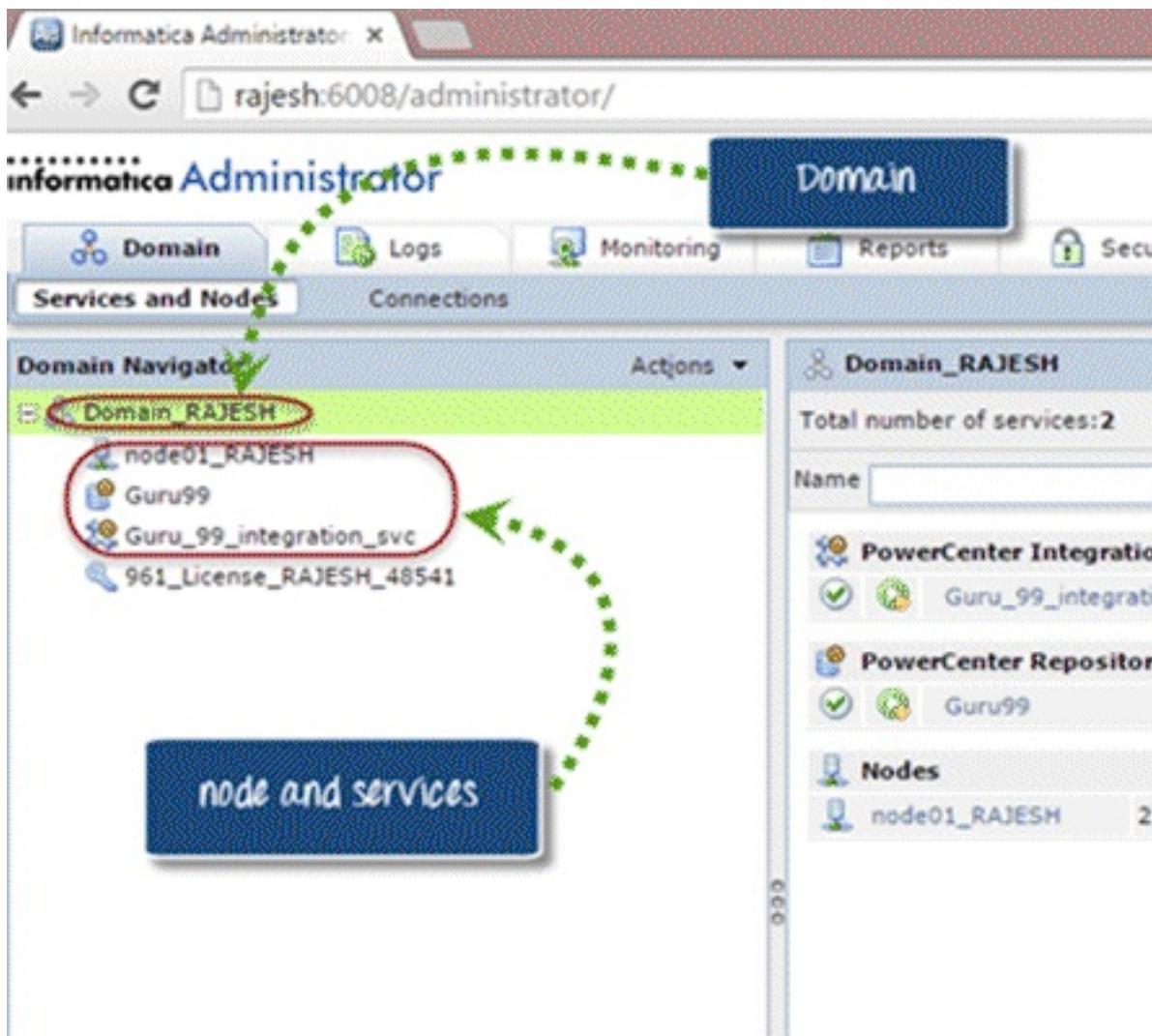


Informatica Domain

The overall architecture of Informatica is Service Oriented Architecture (SOA).

- Informatica Domain is the fundamental administrative unit in Informatica tool
- It is a collection of nodes and services. Further, this nodes and services can be categorized into folders and sub-folders based on the administration requirement.

For example, in the below screenshot, you can see under domain window the folder "Domain_Rajesh" is created under which we have created a node name "node01_rajesh" and services as "guru99 integration services".



Node is a logical representation of a machine inside the domain. Node is required to **run services and processes for Informatica**.

You can have multiple nodes in a domain. In a domain, you will also find a gateway node.

The gateway node is responsible for receiving requests from different client tools and routing those requests to different nodes and services.

There are two types of services in Domain

- **Service Manager:** Service manager manages domain operations like authentication, authorization, and logging. It also runs application services on the nodes as well as manages users and groups.
- **Application Services:** Application service represents the server specific services like integration service, repository service, and reporting service. These services run on different nodes based upon the configuration.

PowerCenter Repository

PowerCenter repository is a relational database like Oracle, Sybase, SQL server and it is managed by repository service. It consists of database tables that store metadata.

There are three Informatica Client tools available in Informatica Powercenter. They are Informatica

- Designer
- Workflow Monitor
- Workflow Manager

These clients can access to the repository using repository service only.

To manage a repository there exists an Informatica service called Repository Service. A single repository service handles exclusively only one repository. Also, a repository service can execute on multiple nodes to increase the performance.

The repository services use locks on the objects, so multiple users cannot modify the same object same time.

You can enable version control in the repository. With the version control feature, you can maintain different versions of the same object.

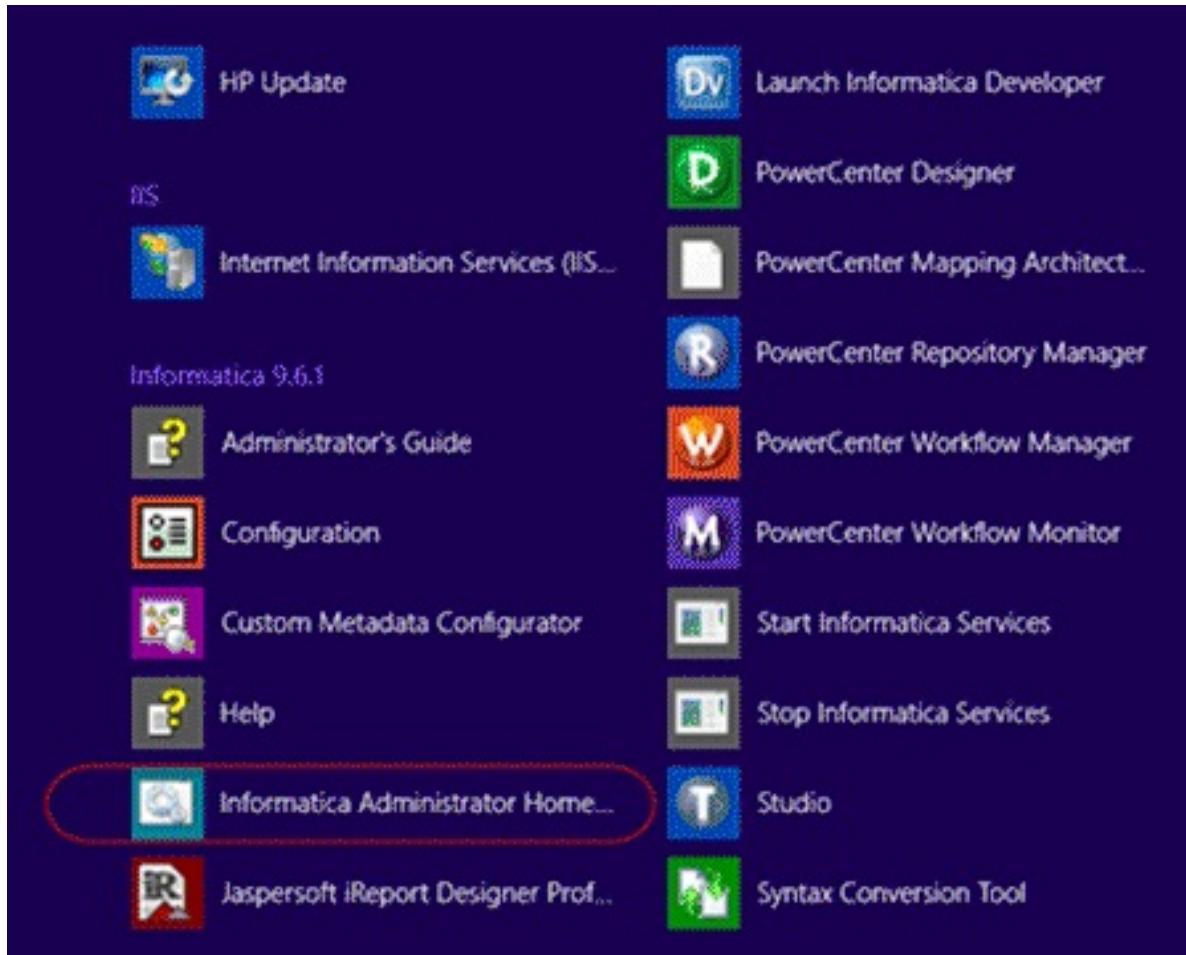
Objects created in the repository can have following three state

- **Valid:** Valid objects are those objects whose syntax is correct according to Informatica. These objects can be used in the execution of workflows.
- **Invalid:** Invalid objects are those who does not adhere to the standard or rules specified. When any object is saved in Informatica, it is checked whether its syntax and properties are valid or not, and the object is marked with the status accordingly.
- **Impacted:** Impacted objects are those whose child objects are invalid. For example in a mapping if you are using a reusable transformation, and this transformation object becomes invalid then the mapping will be marked as impacted.

Domain Configuration

As mentioned earlier, domain is the basic administrative control in Informatica. It is the parent entity which consists of other services like integration service, repository service, and various nodes.

The domain configuration can be done using the Informatica admin console. The console can be launched using web browsers.



Once open in a web browser it prompts for administrator login. The password is set during the Informatica installation.



After login into the Informatica domain, the home page looks something like this.

informatica Administrator

The screenshot shows the 'Domain' tab selected in the top navigation bar. The left pane, titled 'Domain Navigator', lists nodes and services under 'Domain_RAJESH'. The right pane, titled 'Domain_RAJESH', displays service status. A red box highlights the 'Domain_RAJESH' title and the service list. Another red box highlights the 'Nodes' section.

Domain Navigator

- Domain_RAJESH (highlighted)
- node01_RAJESH
- Guru99
- Guru_99_integration_svc
- 961_License_RAJESH_48541

Domain_RAJESH

Total number of services: 2

Name	Type	Count
PowerCenter Integration Service		1
Guru_99_integration_svc		1
PowerCenter Repository Service		1
Guru99		1
Nodes		1
node01_RAJESH	2 service processes	1

In the left pane it shows the existing nodes, repository services, integration services under the domain.

On the main window, it shows the status of those services, whether those are up or down.

Properties of the domain

Click on the properties menu in the admin page to view the properties of the domain.

The screenshot shows the Oracle Database Control interface with the 'Properties' tab selected for the 'Domain_RAJESH' domain. The 'General Properties' section is expanded, showing configuration like Name (Domain_RAJESH), Resilience Timeout (30), and Dispatch Mode (RoundRobin). The 'Database Properties' section is also expanded, showing details such as Database Type (ORACLE), Database Host (localhost), and Database Port (1521). A table at the bottom lists nodes with their status and gateway configuration.

Node Name	Status	Gateway
node01_RAJESH	Available	Yes

Key properties of the domain are

Resilience timeout – If any of the integration service or repository services goes down then resilience timeout is the no of seconds the application service tries to connect to those services.

Restart Period – It is the maximum number of seconds the domain spends to restart a service.

Dispatch Mode – It is the policy used by the load balancer to dispatch tasks to various nodes.

Database type – The type of database on which domain is configured.

Database host – Hostname of the machine on which domain is configured.

Database port & name – It is the database port and the database instance name for the domain.

These properties can be modified based upon requirement.

Powercenter client & Server Connectivity

PowerCenter client tools are development tools which are installed on the client machines. Powercenter designer, workflow manager, a repository manager, and workflow monitor are the main client tools.

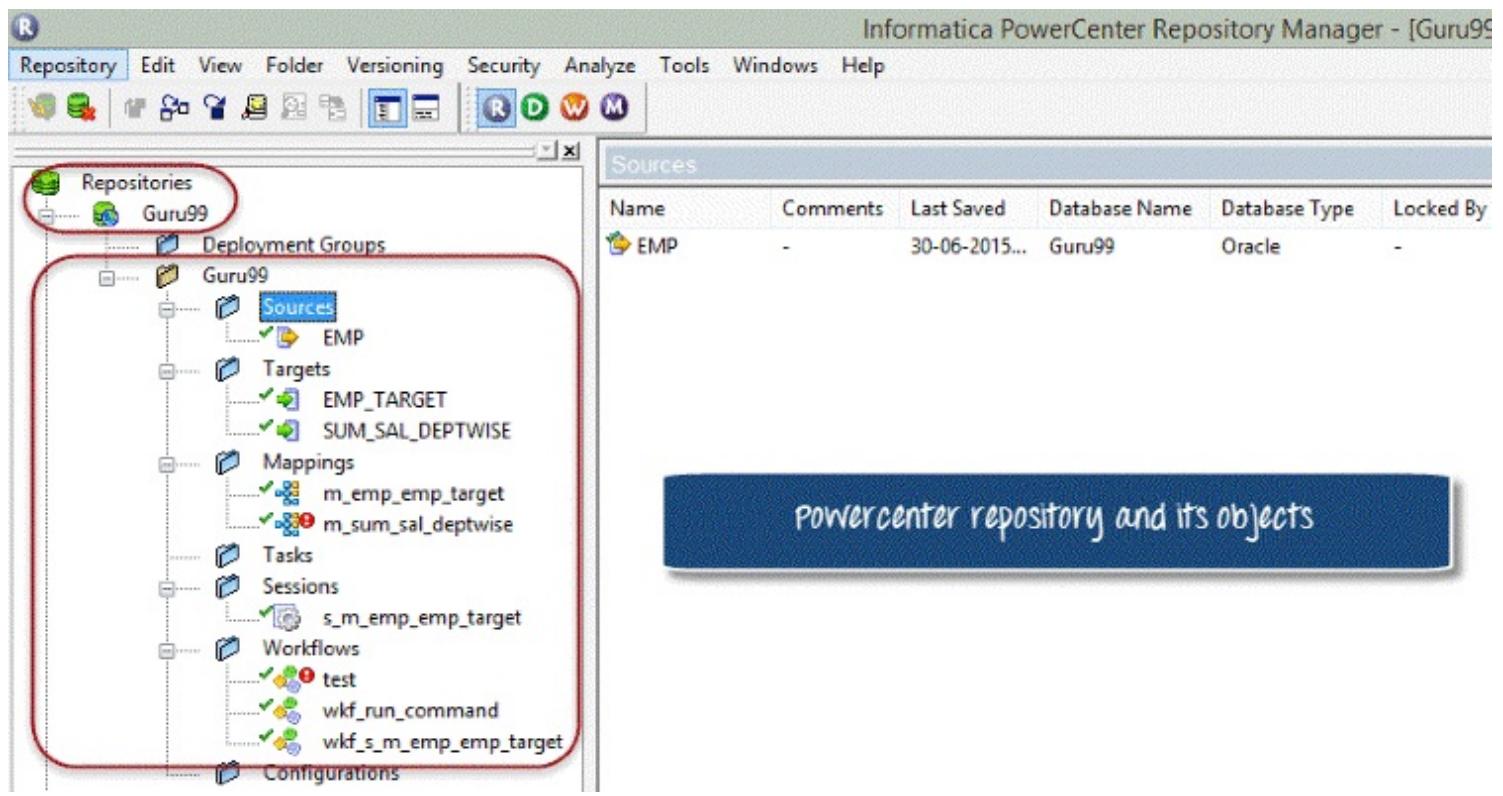
The mappings and objects that we create in these client tools are saved in the Informatica repository which resides on the Informatica server. So the client tools must have network connectivity to the server.

On the other hand, PowerCenter client connects to the sources and targets to import the metadata and source/target structure definitions. So it also must have connectivity to the source/target systems.

- To connect to the integration service and repository service, PowerCenter client uses TCP/IP protocols and
- To connect to the sources/targets PowerCenter client uses ODBC drivers.

Repository Service

The repository service maintains the connections from Powercenter clients to the PowerCenter repository. It is a separate multi-threaded process, and it fetches, inserts and updates the metadata inside the repository. It is also responsible for maintaining consistency inside the repository metadata.



Integration Service

Integration service is the executing engine for the Informatica, in other words, this is the entity which executes the tasks that we create in Informatica. This is how it works

- A user executes a workflow
- Informatica instructs the integration service to execute the workflow
- The integration service reads workflow details from the repository
- Integration service starts execution of the tasks inside the workflow
- Once execution is complete, the status of the task is updated i.e. failed, succeeded or aborted.
- After completion of execution, session log and workflow log is generated.
- This service is responsible for loading data into the target systems
- The integration service also combines data from different sources

For example, it can combine data from an oracle table and a flat file source.

So, in summary, Informatica integration service is a process residing on the Informatica server waiting for tasks to be assigned for the execution. When we execute a workflow, the integration service receives a notification to execute the workflow. Then the integration service reads the workflow to know the details like

which tasks it has to execute like mappings & at what timings. Then the service reads the task details from the repository and proceeds with the execution.

Sources & Targets

Informatica being an ETL and Data integration tool, you would be always handling and transforming some form of data. The input to our mappings in Informatica is called source system. We import source definitions from the source and then connect to it to fetch the source data in our mappings. There can be different types of sources and can be located at multiple locations. Based upon your requirement the target system can be a relational or flat file system. Flat file targets are generated on the Informatica server machine, which can be transferred later on using ftp.

Relational– these types of sources are database system tables. These database systems are generally owned by other applications which create and maintain this data. It can be a Customer Relationship Management Database, Human Resource Database, etc. for using such sources in Informatica we either get a replica of these datasets, or we get select privileges on these systems.

Flat Files - Flat files are most common data sources after relational databases in Informatica. A flat file can be a comma separated file, a tab delimited file or fixed width file. Informatica supports any of the code pages like ascii or Unicode. To use the flat file in Informatica, its definitions must be imported similar to as we do for relational tables.

Chapter 3: Configure Clients and Repositories

After installing Informatica server and client, Informatica server needs to be configured.

The Informatica architecture is a Service Oriented Architecture (SOA). Therefore, before using Informatica tools, server and client components needs to be installed and configured.

Prerequisite to configure clients, integration service and repository service.

- Informatica server and client should be installed.
- Informatica service should be up and running.
- Oracle or other compatible databases must be installed, and database service must be up.
- Database users must be created for repository service.
- Domain administrator console credentials must be available (it is configured during server installation)

The following components needs to be set up before you can start working in Informatica PowerCenter.

1. **Informatica domain**
2. **creating repository service**
3. **creating integration service**
4. **Creating users**

What is Informatica domain?

Informatica Domain is a basic administrative unit for various services and hosts.

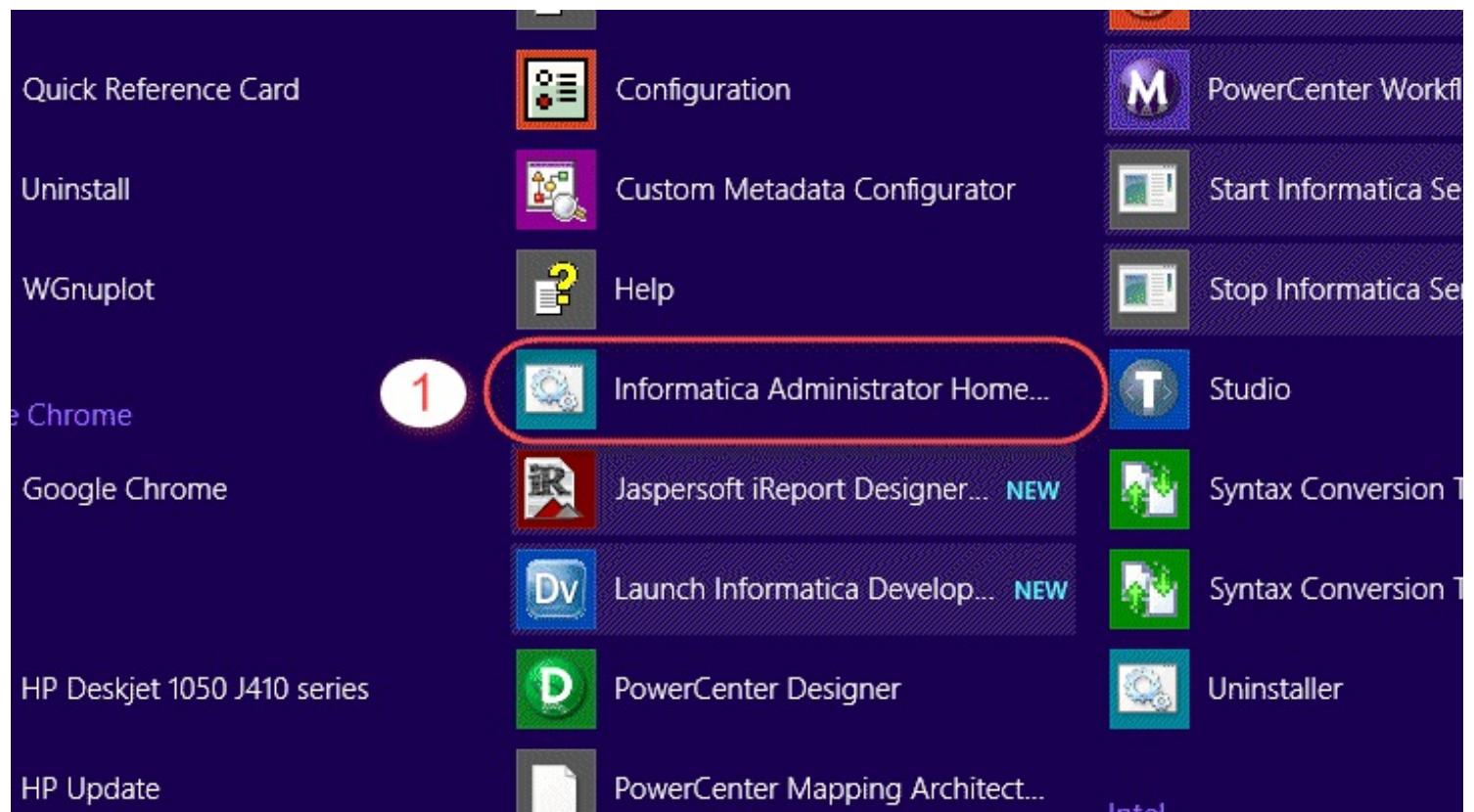
Using domain console you can perform various administrative tasks like adding a node, creating repository service, creating integration service, creating reporting service, managing & creating users, monitoring services, etc.

During installation of Informatica server, system prompts for setting up the domain credentials and database connection.

You need to use the same credentials when you access the domain console.

How to Open the Administrator Home Page

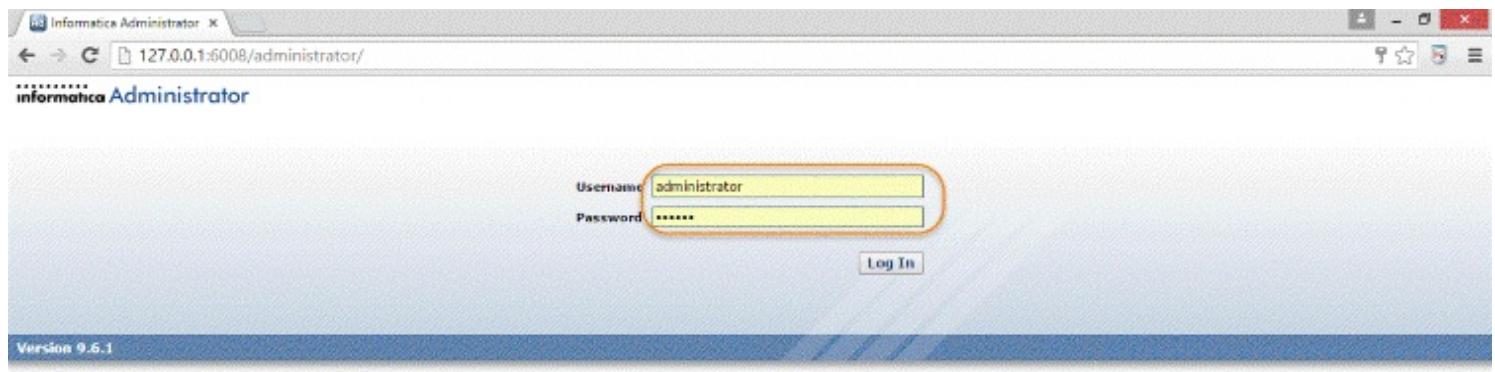
Step 1) From the program menu open the Informatica administrator home page



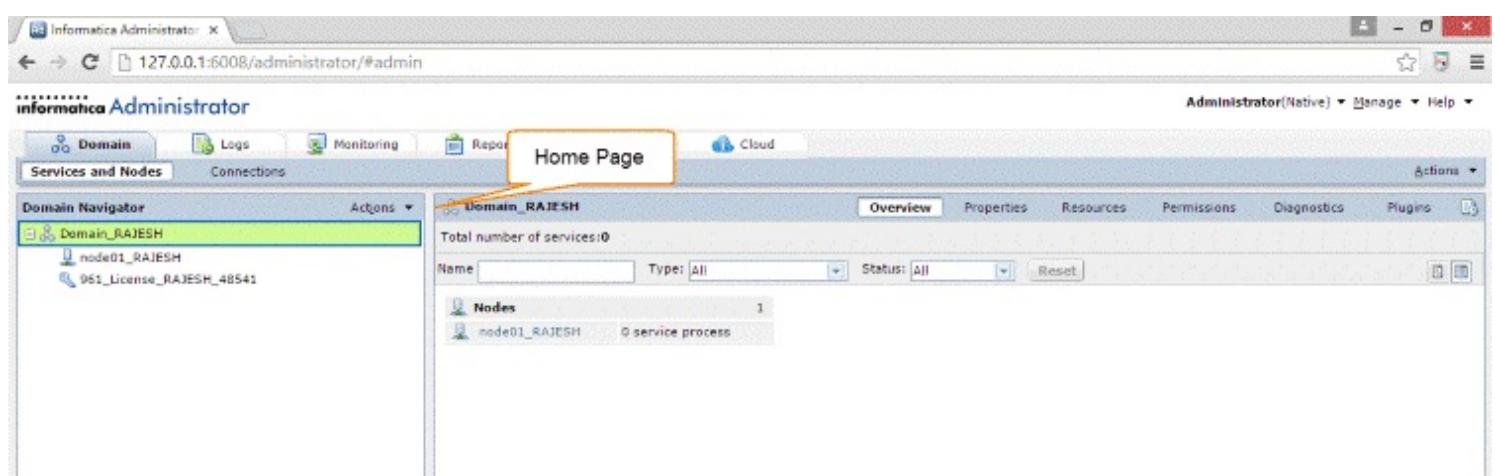
This will open up domain home page in the web browser.

Troubleshooting- if the home page doesn't open up, check whether Informatica server process and database service are up and running. If not, start the same.

Step 2) Enter the domain credentials (these credentials are generated during server installation).



Step 3) On successful login the Informatica administrator home page will appear.



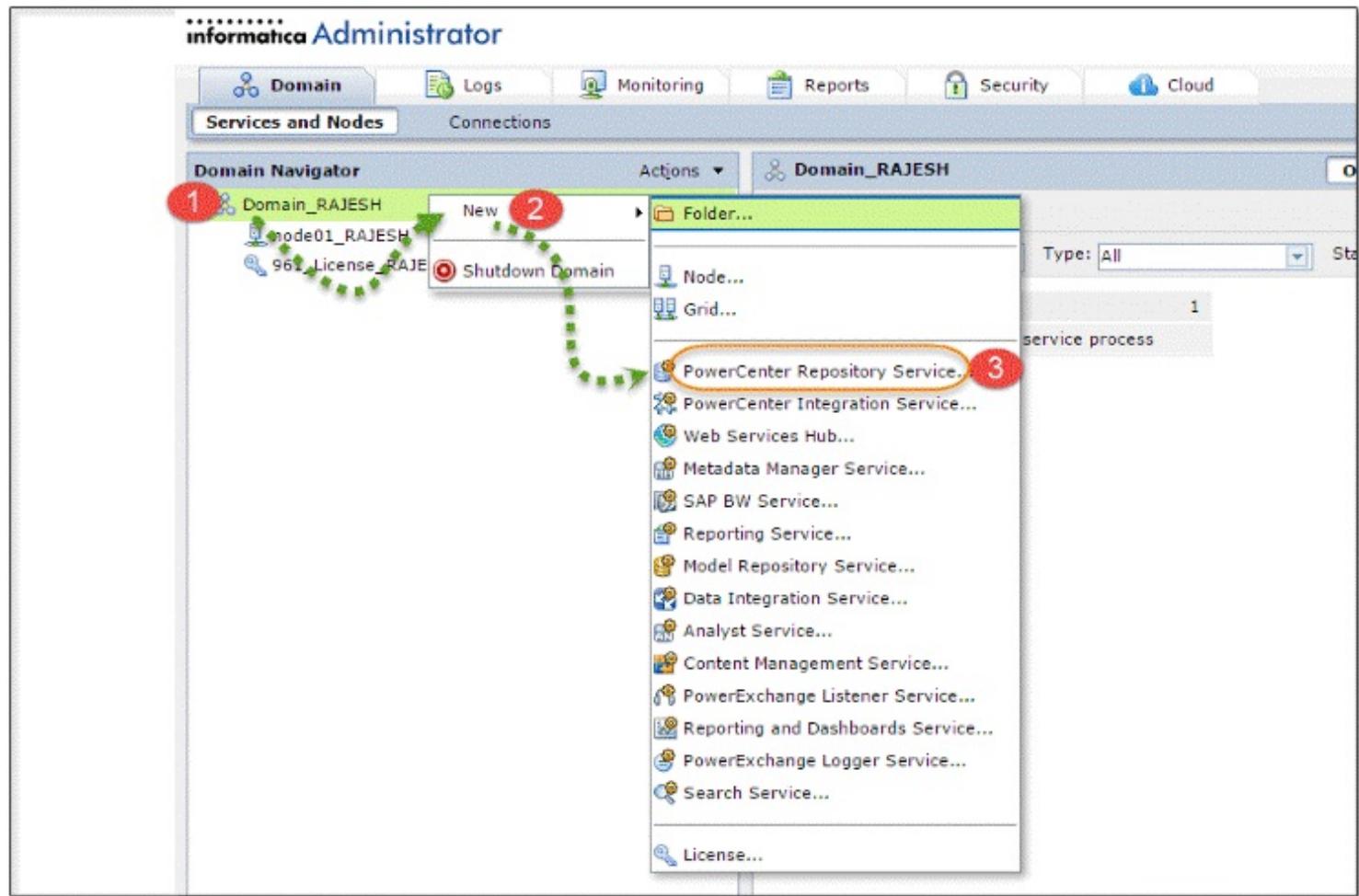
How to Create Repository Service

It is the first service that needs to be created in the domain console. The pre-requisite is to have a database user created in the database along with basic privileges (connect, resource), as this service maintains repository data in the database. You can create versioned objects in the repository using the enable version control option.

You can select oracle, DB2, Sybase or SQL Server databases to create a repository

Step 1)

1. Right-click on the domain name
2. Select new option
3. Select PowerCenter repository service.



Step 2) In the next screen

1. Enter Name for the repository
2. Select License
3. Select Node
4. Click on Next Button

New PowerCenter Repository Service - Step 1 of 2

Fields marked with an asterisk (*) are required.

Specify the properties for this new PowerCenter Repository Service.

Name *	<input type="text" value="Guru99"/> 1
Description	<input type="text"/>
Location *	<input type="text" value="Domain_RAJESH"/> 2
License	<input type="text" value="961_License_RAJESH_48541"/> 3
Node *	<input type="text" value="node01_RAJESH"/>

4

Next >

Step 3) In the next screen,

1. Enter the database properties.
2. Select the option – enable version control, if version control feature is required for Informatica development.
3. Select finish button.

New PowerCenter Repository Service - Step 2 of 2

Fields marked with an asterisk (*) are required.

Specify the database properties for this new PowerCenter Repository Service.

Database Type *

1

Oracle

Username *

Guru99

Password *

Connection String *

Guru99_Oracle

Code Page *

MS Windows Latin 1 (ANSI), superset of Latin1

Tablespace Name

Specify the creation options for the new PowerCenter Repository Service.

- Content exists under specified connection string. Do not create new content.
 No content exists under specified connection string. Create new content.

Create as Global Repository (May not be reverted to local)

2 Enable version control (A versioned repository cannot be unversioned)

Enable the PowerCenter Repository Service

3

< Back

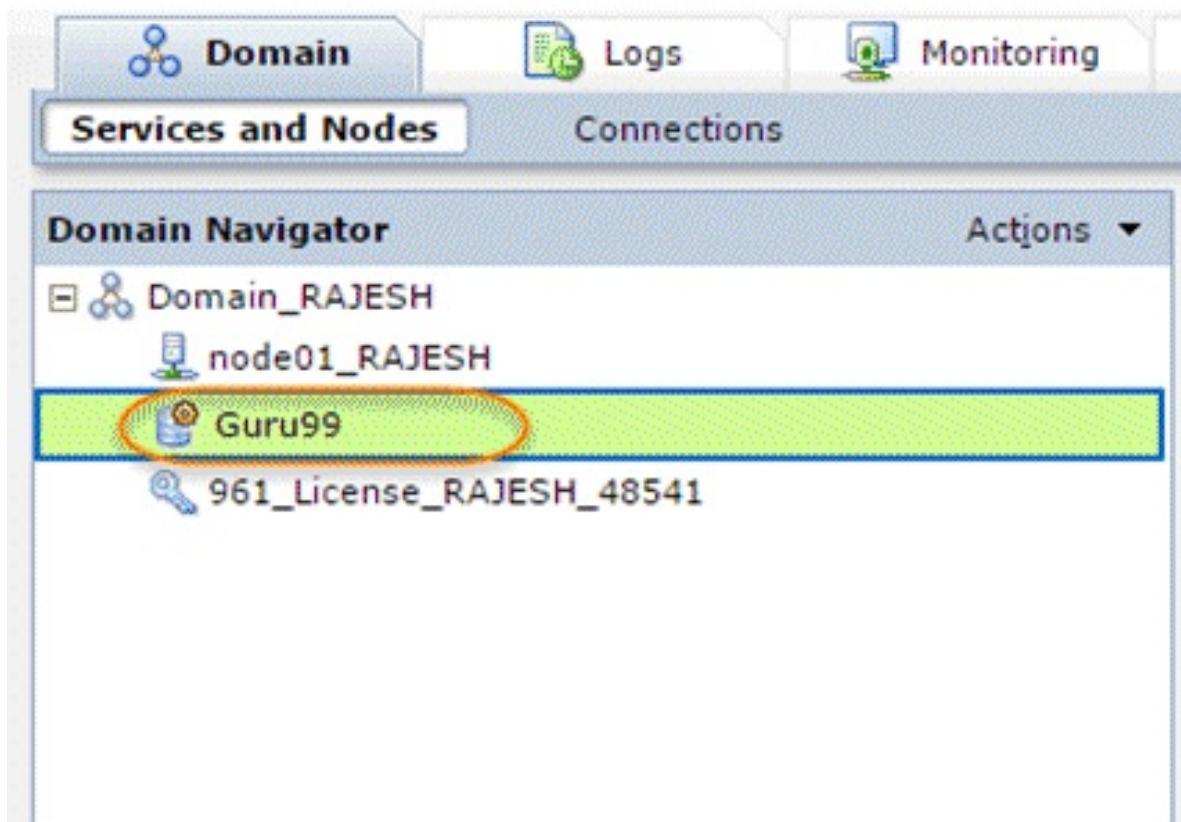
Finish

Cancel

Note – the connection string name is the name of database instance service name.

Step 4) The Repository service will be created, and will be listed under the domain tree.

***** informatica Administrator



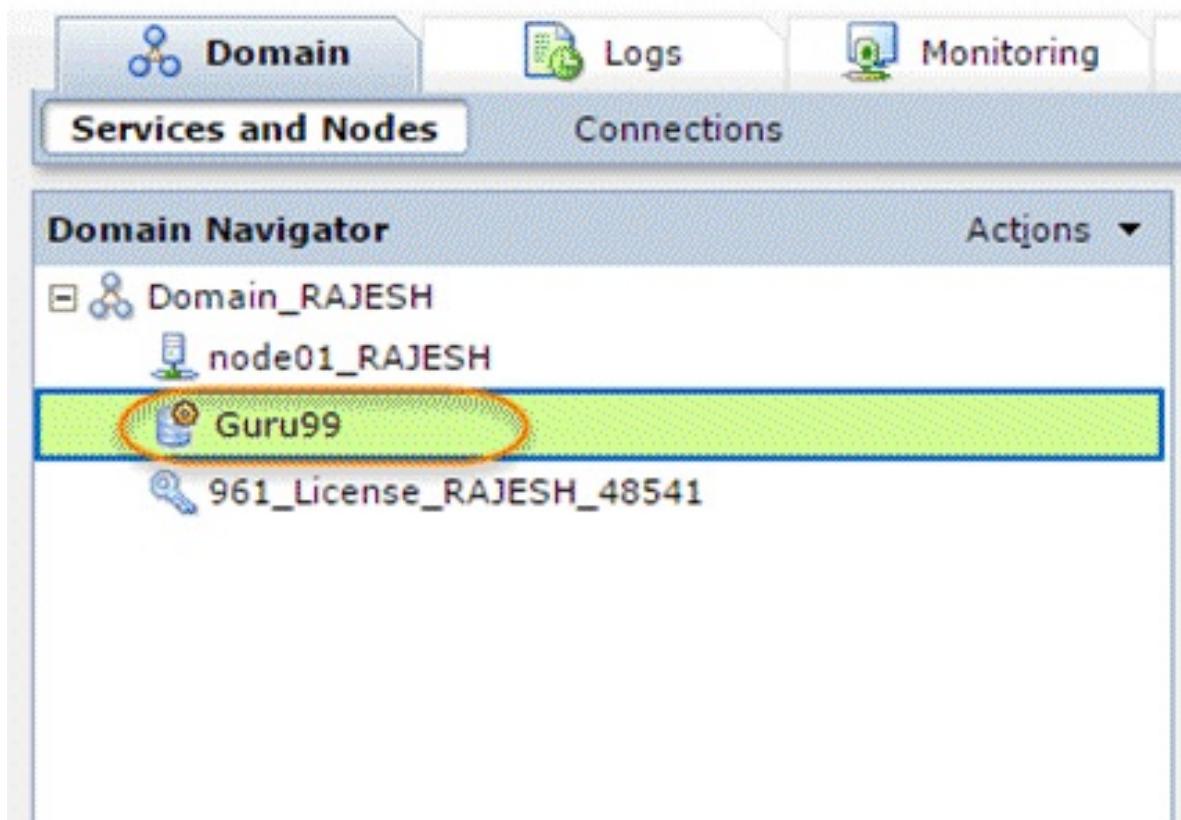
Troubleshooting – if there is an error during service creation, check whether the correct database credentials are provided, and connection string is correct. After any changes made to repository properties, recycle the repository service to make the changes effective.

How to create repository contents

After the repository service is created, **repository contents** needs to be created (this is one time activity for a repository). Without repository contents, you cannot start the repository service. Repository content is basically database tables which contains the details of objects created in Informatica.

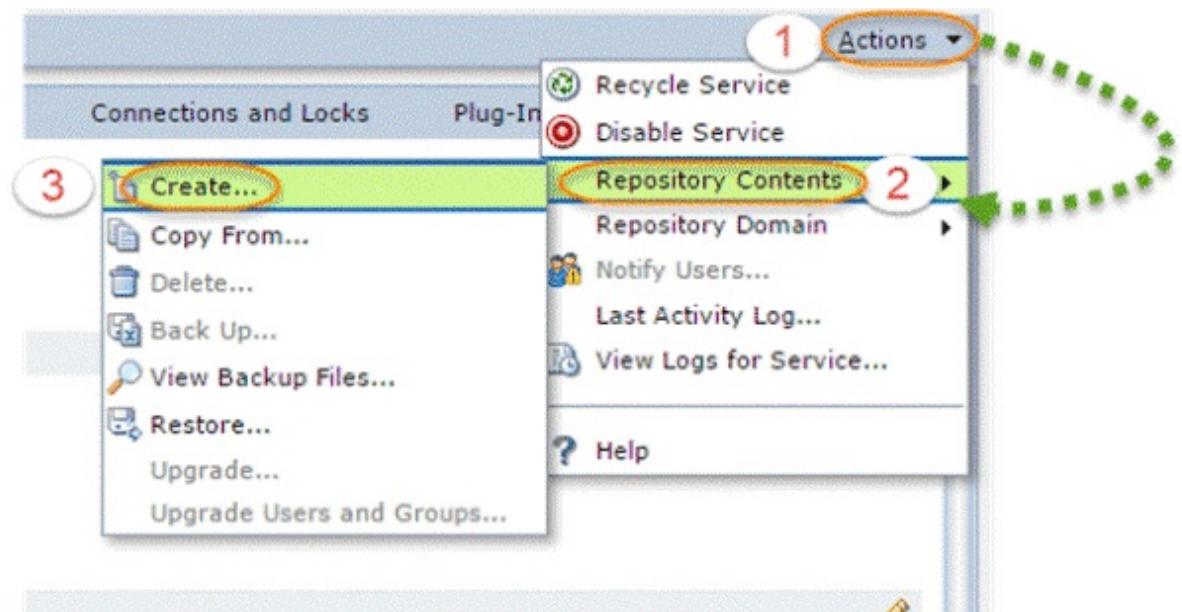
Step 1) Select the repository service under the domain tree.

informatica Administrator



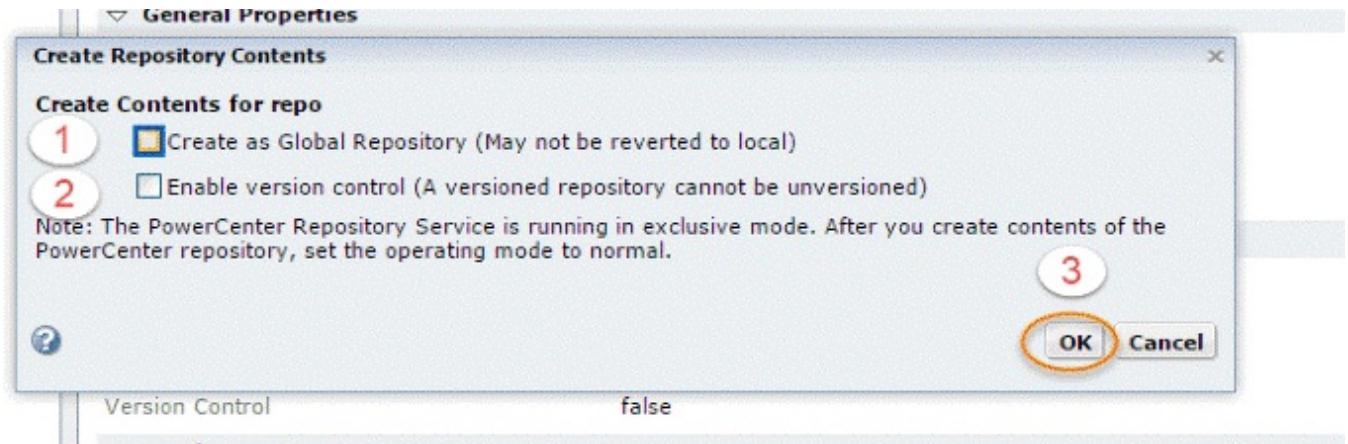
Step 2) After selecting the repository service.

1. Select the action menu on right side of page
2. Select repository contents option
3. Select create option



Step 3) In the next screen

1. Select the options of local/global repository (we are keeping it unchecked to create local repository)
2. Select version control (if required, we are keeping it unchecked to skip version control feature)
3. Click ok.



Step 4) Once repository content is created, change the repository service from exclusive to normal mode

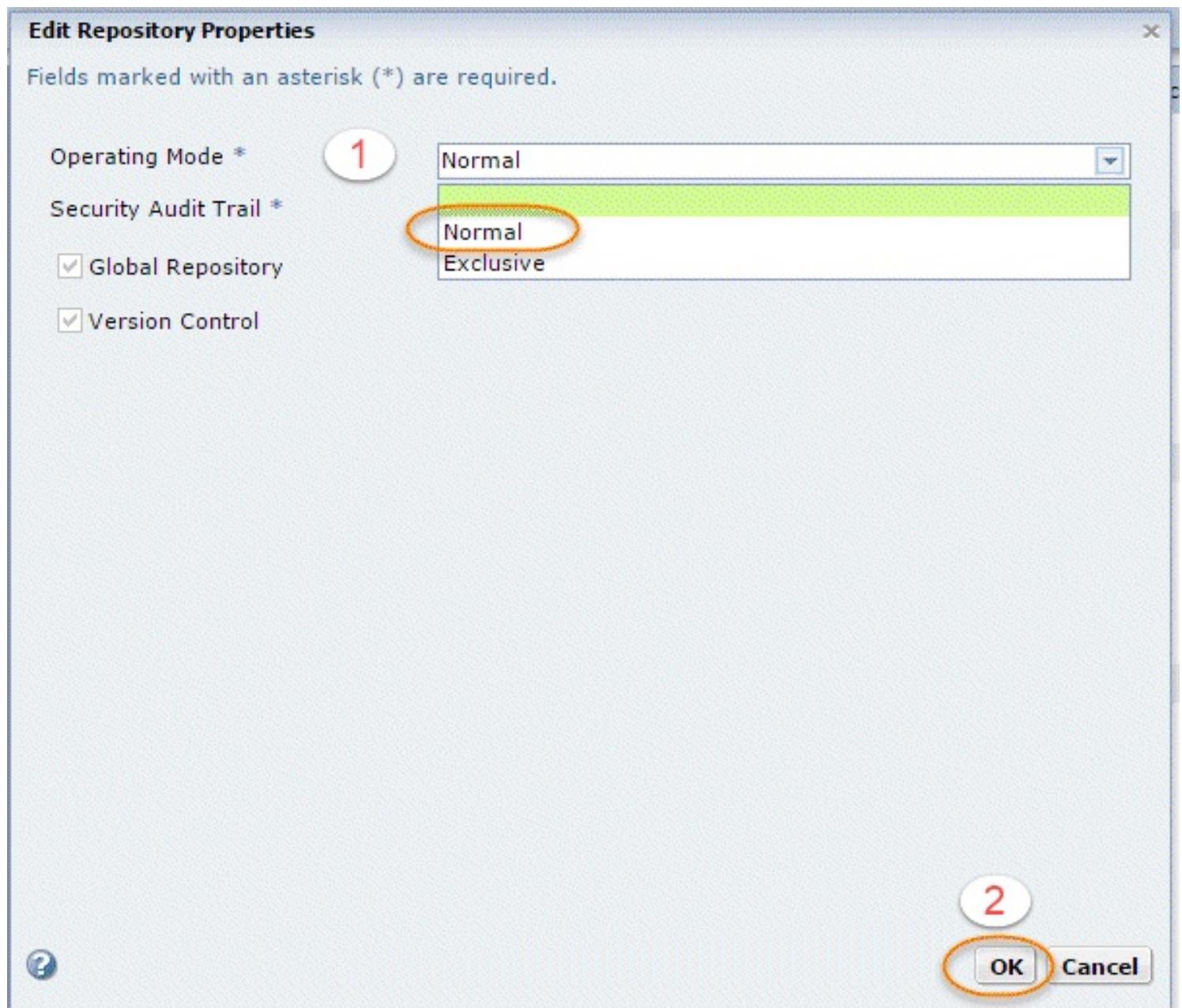
Note – If the repository service is in exclusive mode, users would not be able to connect to the repository

Select the edit property option of repository



Step 5)

1. Select the mode from exclusive to normal
2. Select the ok button.



Once done, the repository service will be running in normal mode.

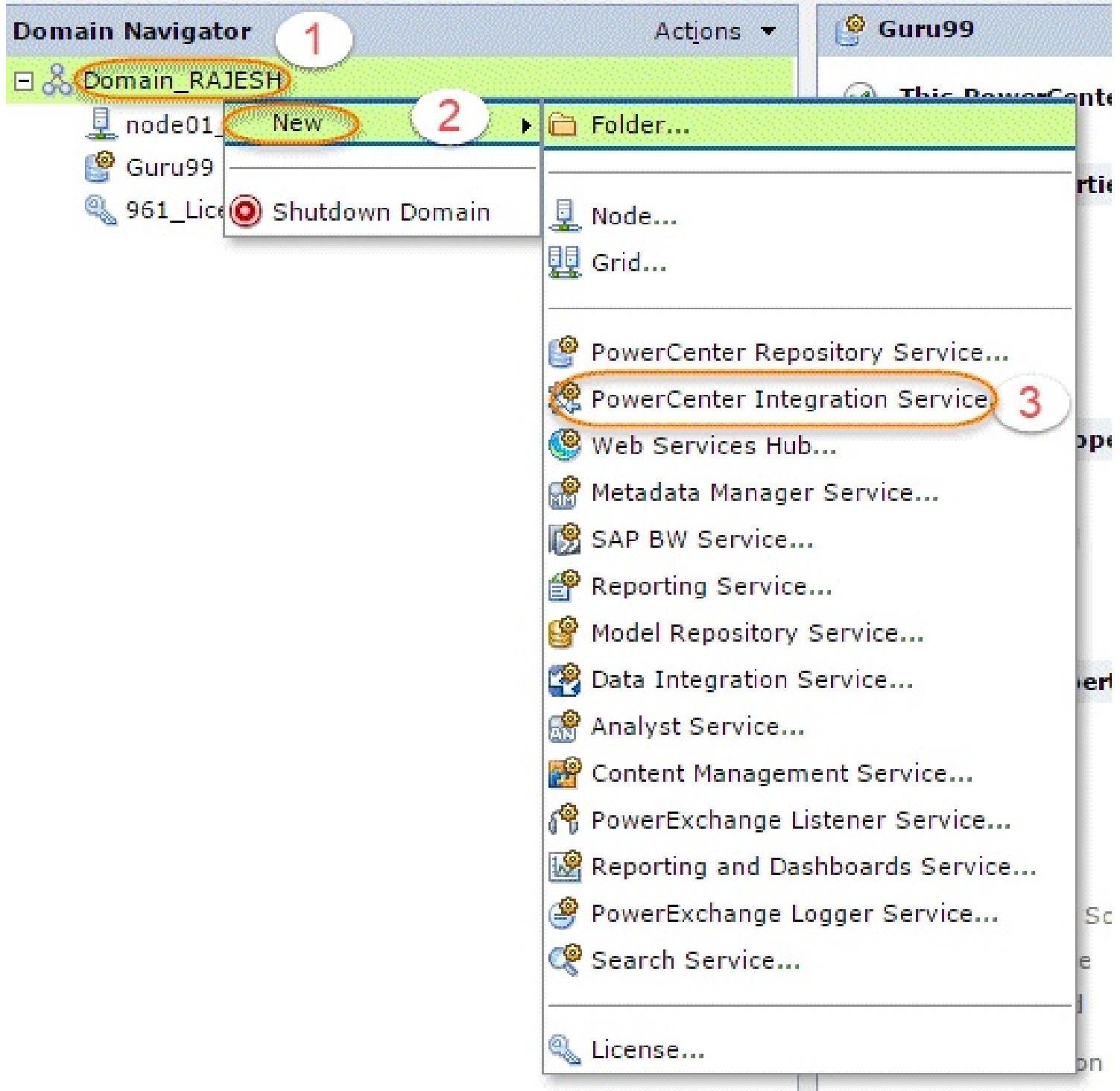
How to create Integration Service

The integration service is responsible for the execution of workflows. Whenever, any workflow is executed from workflow manager, scheduler or command prompt, the integration service receives a request. This service locks the respective workflow, fetches the details about the workflow, mapping & session from the Informatica repository and executes the workflow. This service is also responsible for creating workflow logs, reading parameter files, creating process threads to execute workflows, execute any other task created in workflow manager, etc.

To Create Integration Service

Step 1)

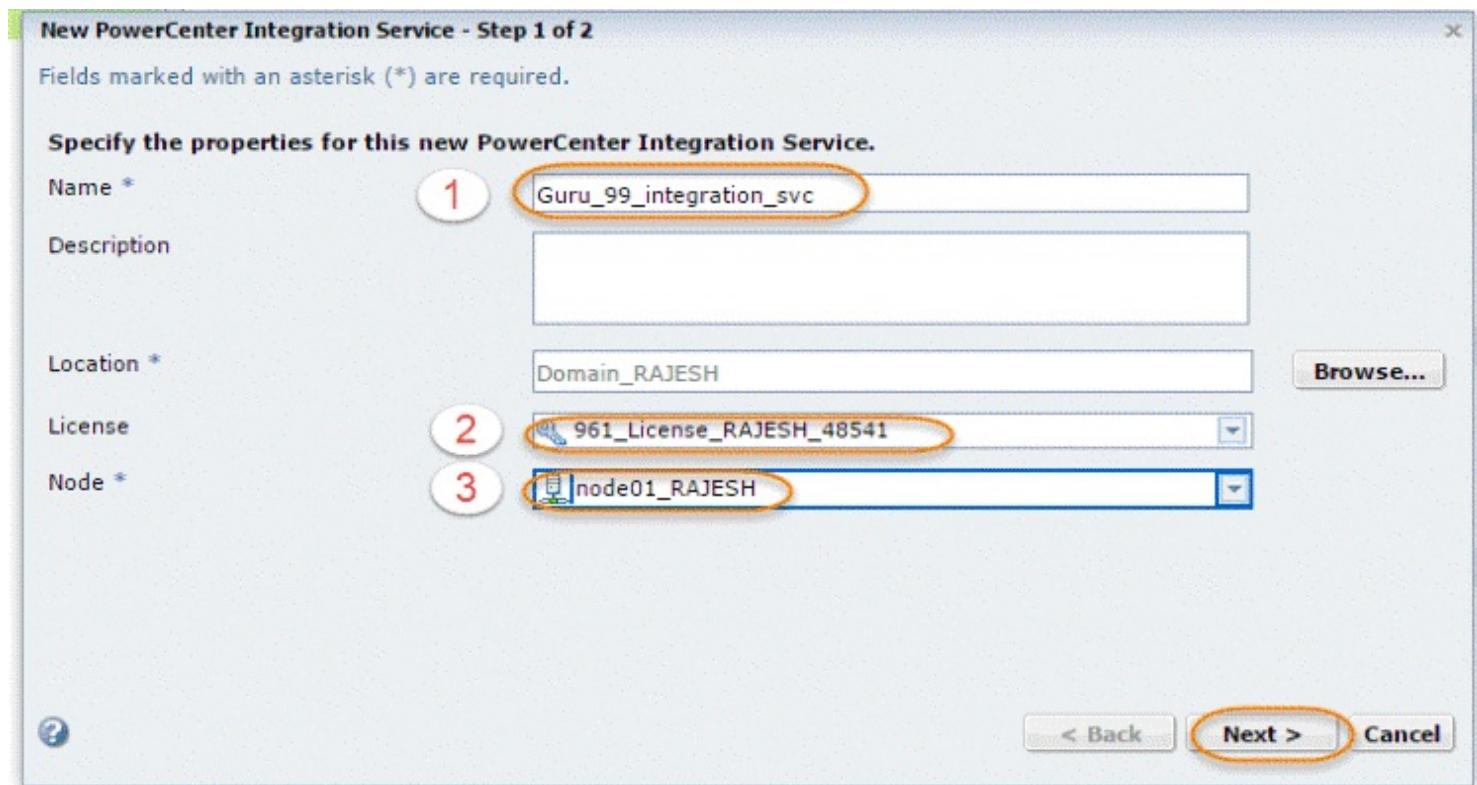
1. Right click on the **domain** navigator tree
2. Select new option
3. Select PowerCenter integration service



Step 2) In this step we will,

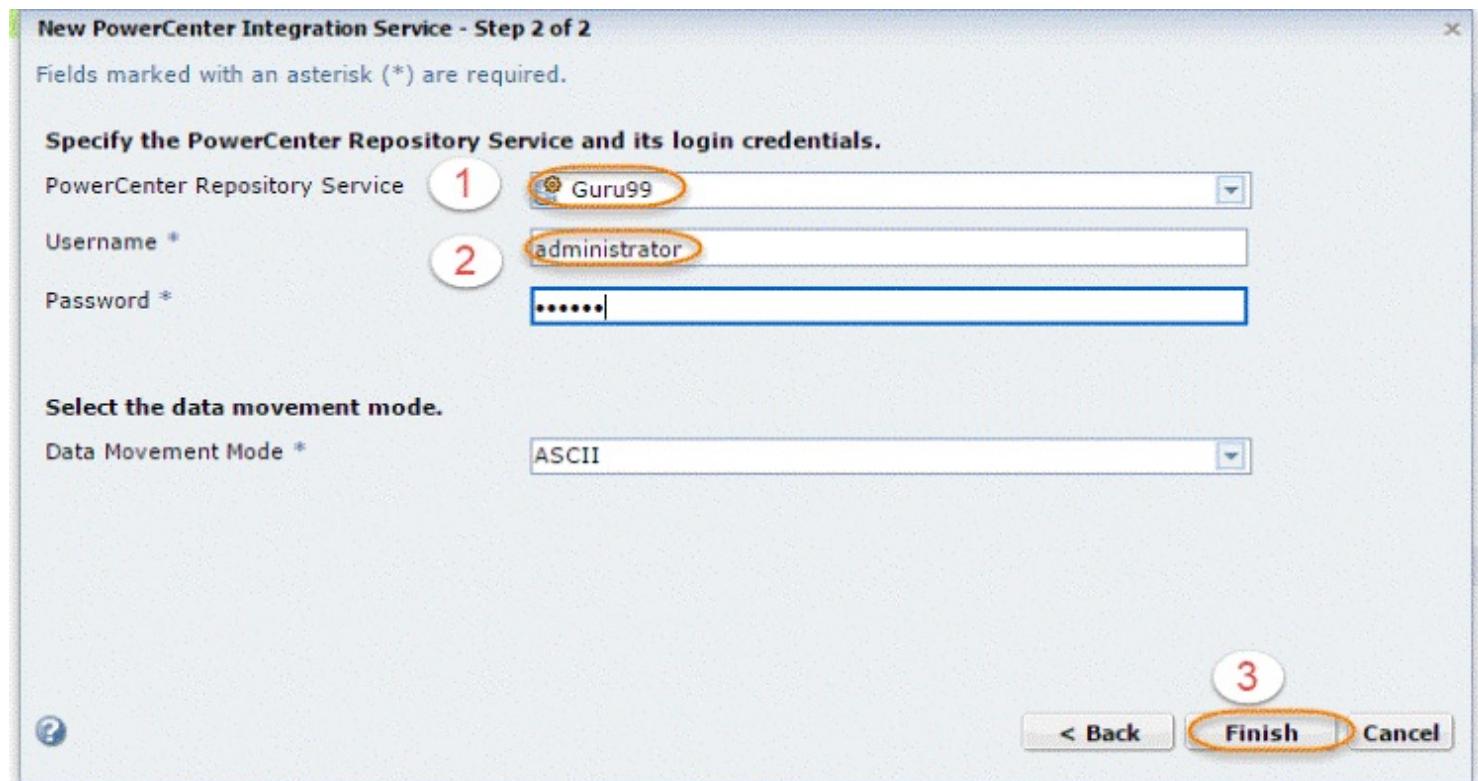
1. Enter integration service name
2. Select the license

3. Select a node and click next.



Step 3) On the next page

1. Select the repository service
2. Enter Informatica Administrator username and password.
3. Select Finish.



Step 4) To enable Integration service.

1. Click on the Integration service in domain tree and Select Actions option.
2. Select the Enable service option

1

Actions ▾

Recycle Service

Enable Service

View Logs for Service...

Help

2

ciated Repository

Pr

When both integration service and repository services are created and enabled, they should be in running status.

Total number of services: 2

Name	Type:	Status:
PowerCenter Integration Service	All	All
Guru_99_integration_svc	All	All
PowerCenter Repository Service	All	All
Guru99	All	All
Nodes	All	All
node01_RAJESH	All	All
	2 service processes	

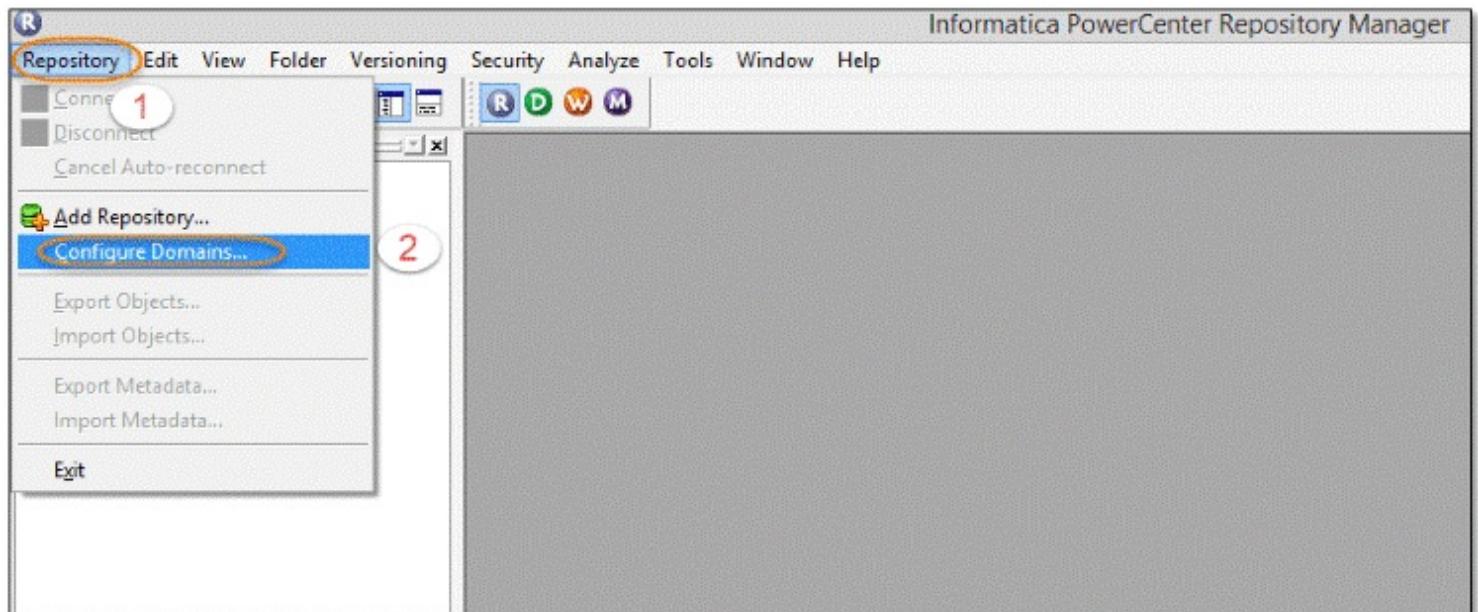
green check box
represents running service

How to Configuring client and domain

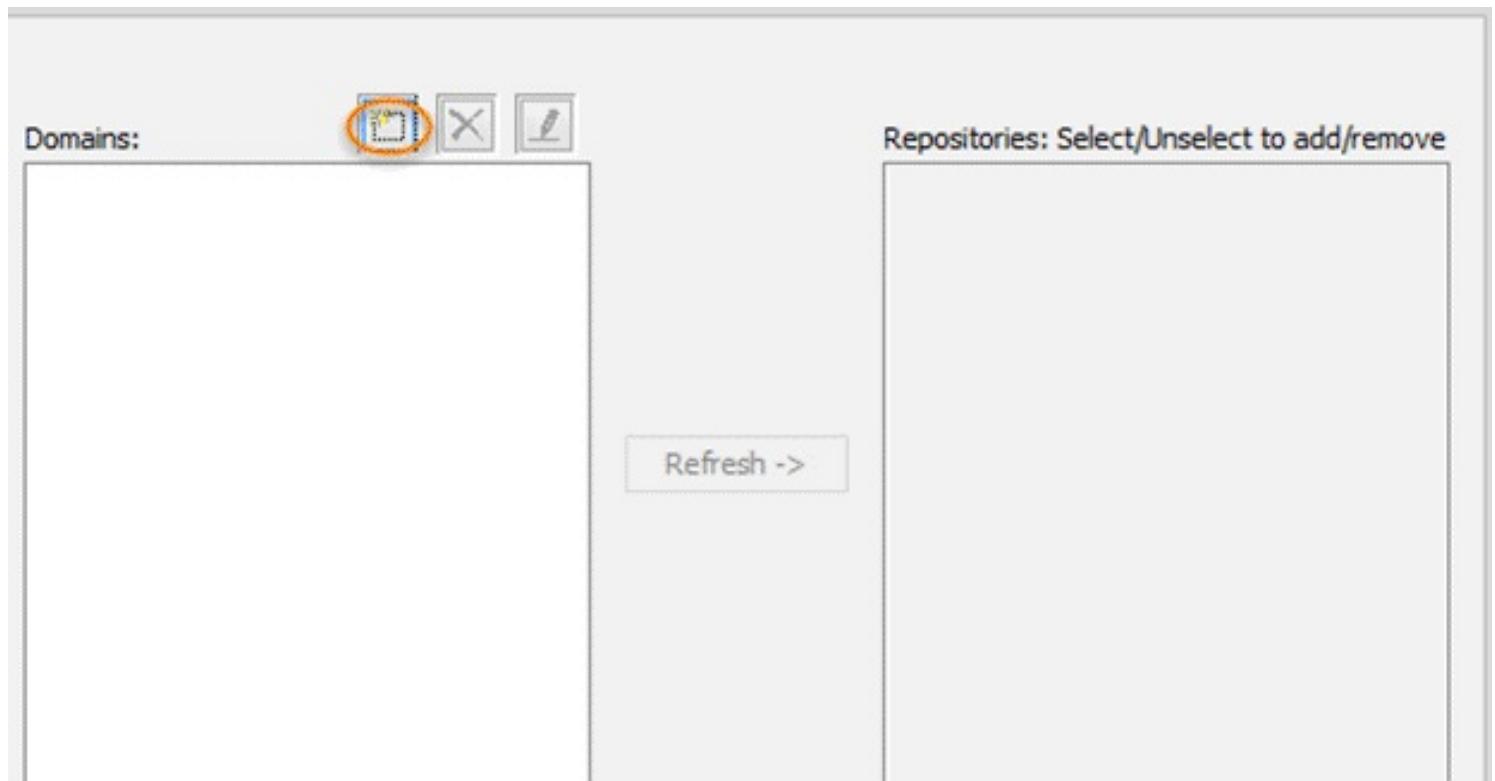
Before you start using the Informatica integration and repository service of the Informatica server, you have to configure the Informatica domain at the client. To add the domain at client following steps are required.

Step 1) Open the Informatica repository designer

1. Select repository menu
2. Select configure domains

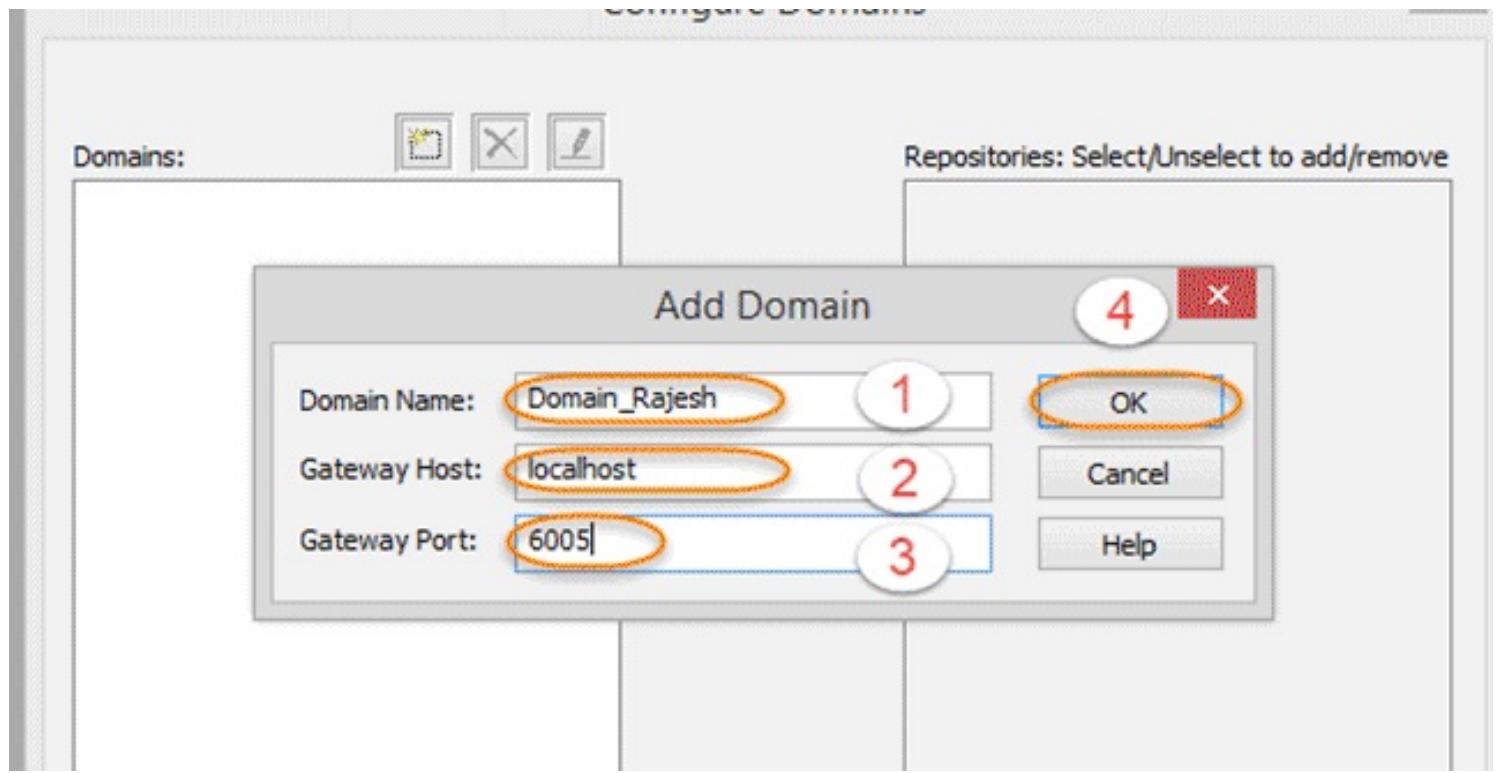


Step 2) Select the add new domain option

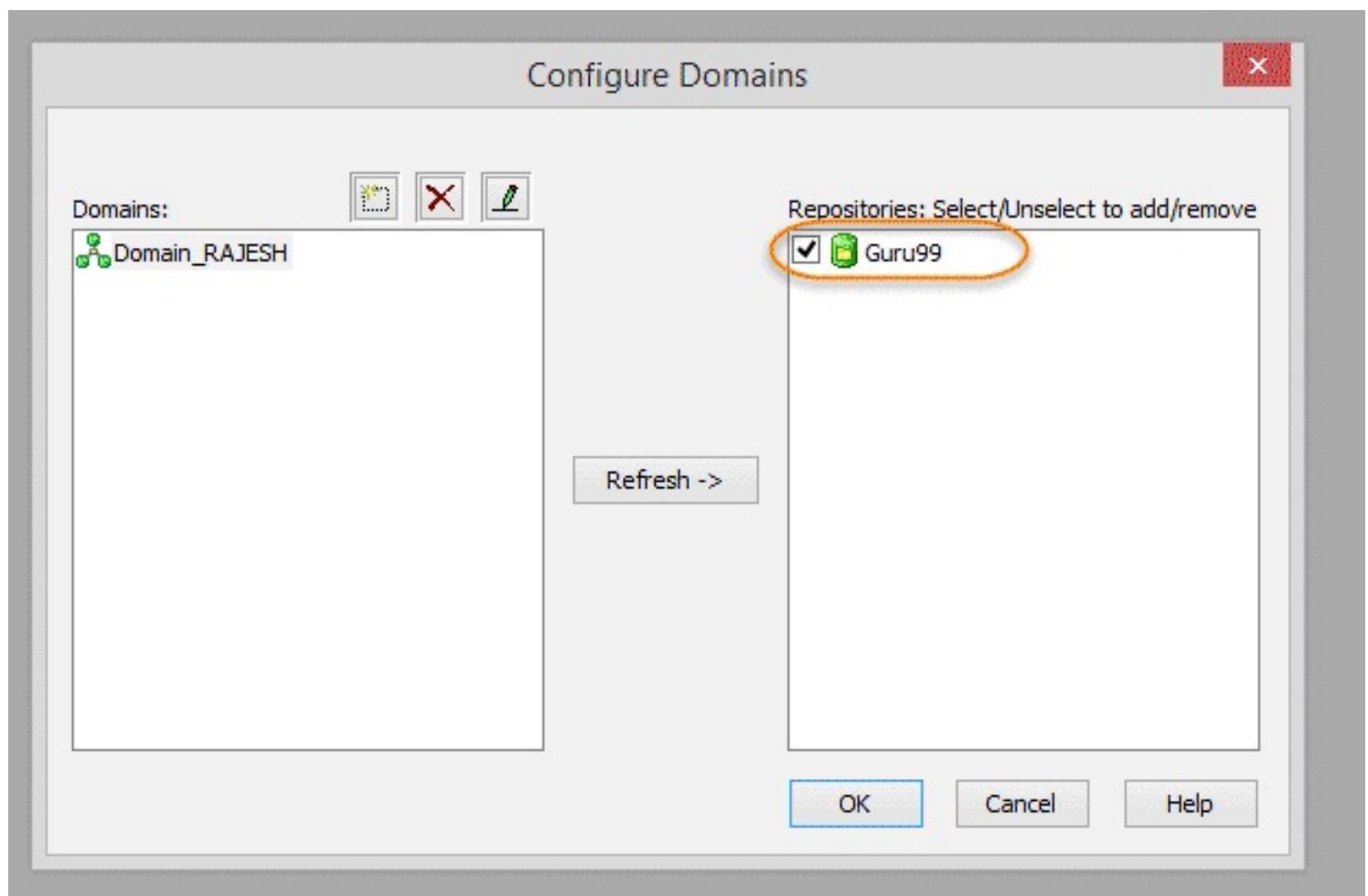


Step 3) In the add domain window

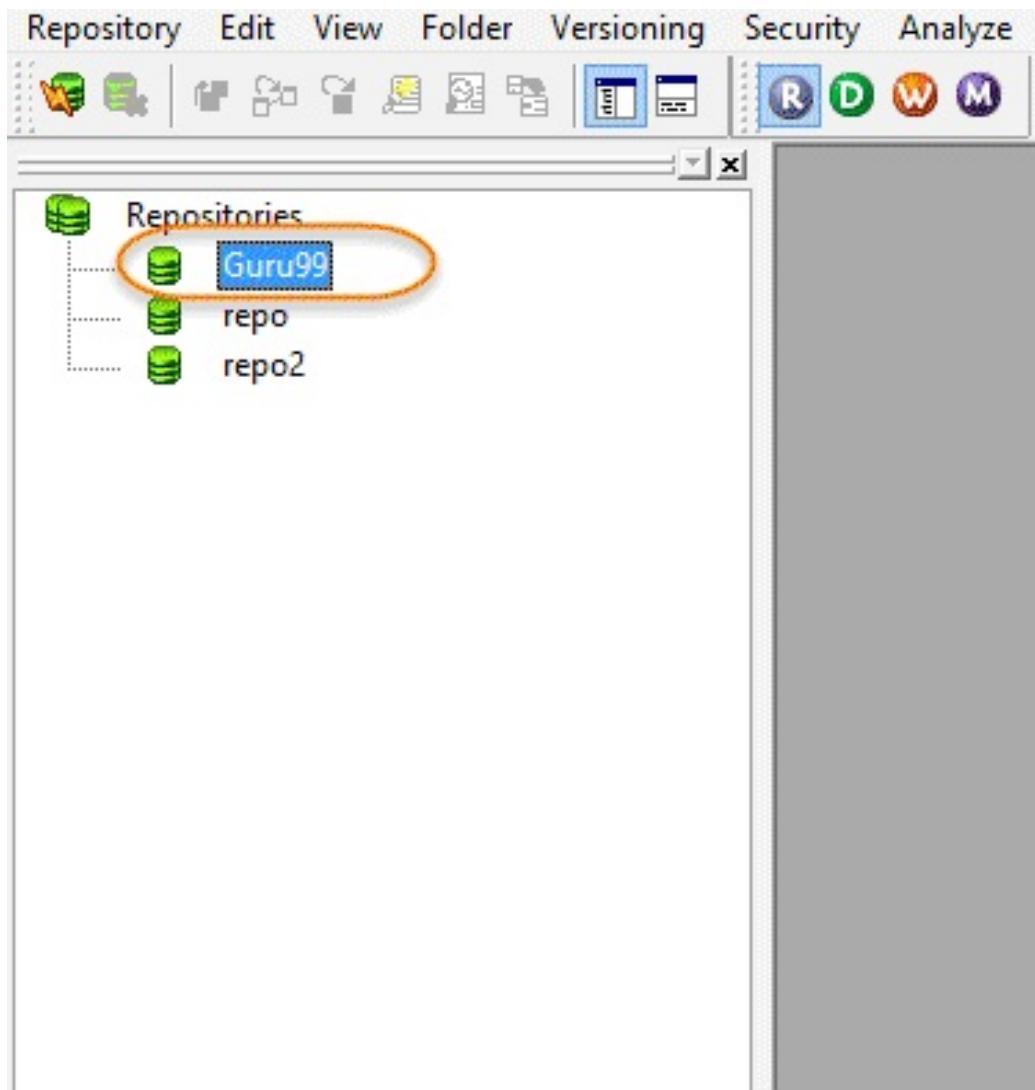
1. Enter Domain name
2. Enter Gateway host
3. Enter Gateway port (the port no is assigned during the Informatica server installation process)
4. Select ok



Step 4) The repository will be listed on right side of the window. Select the checkbox and Select ok.



The repository will be listed in the repository manager navigation tree.

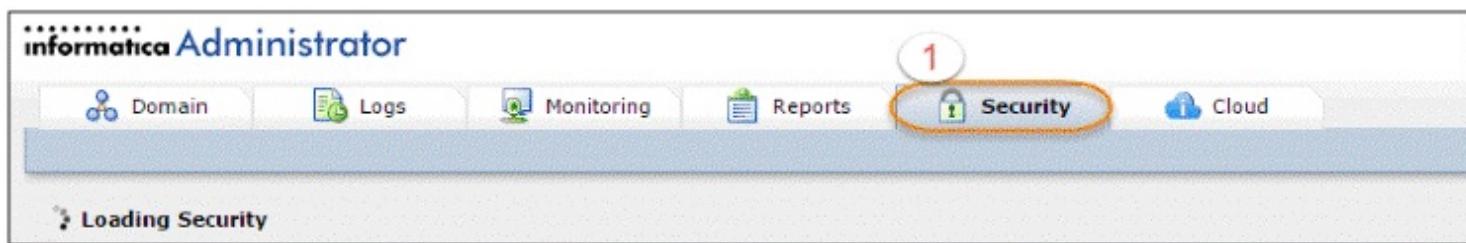


How to create user

To create mappings in Informatica designer and to perform various tasks like, executing workflows, managing repository folders, monitoring workflows, etc. a valid Informatica user is required, and the user must have privileges. A user can be administrator user or developer user based on the roles assigned to it. You can create any no. of users based upon requirement.

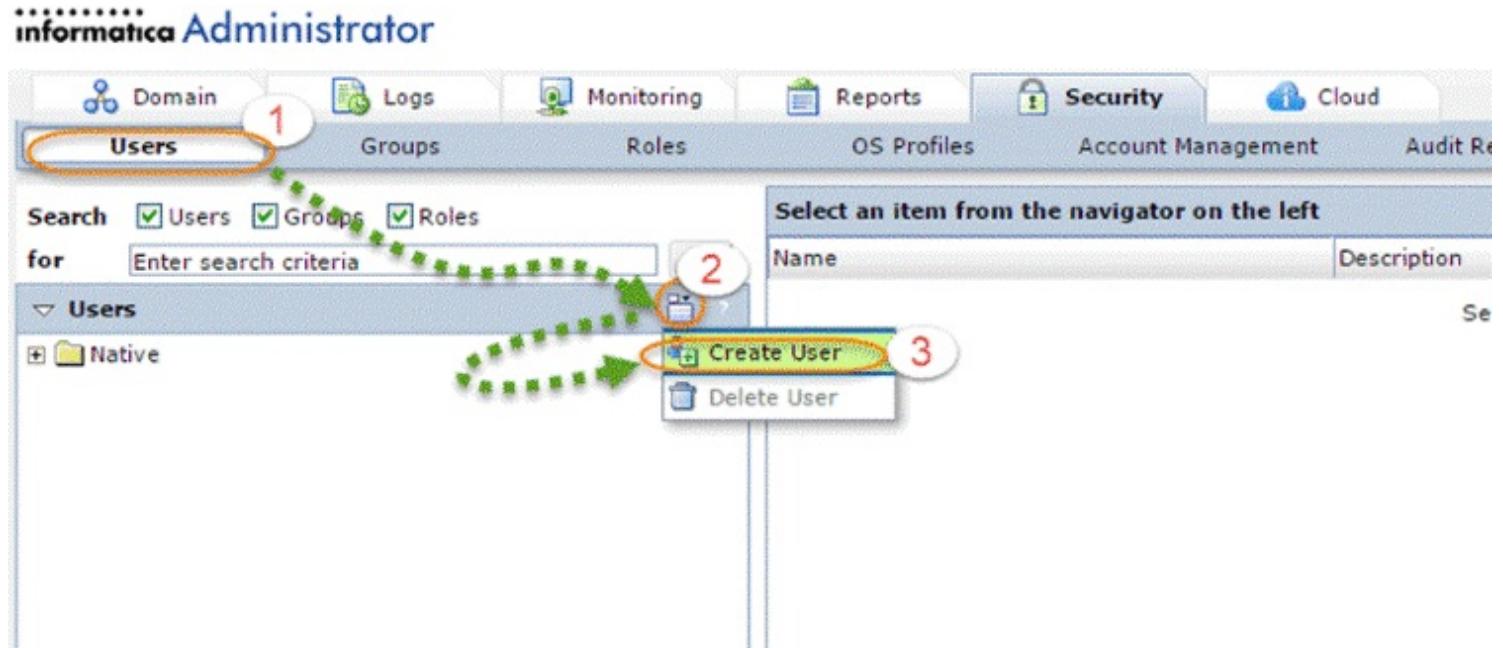
To create Informatica user

Step 1) In the Informatica Administrator home page Select security tab



Step 2) In this step we will,

1. Select user tab
2. Select action menu
3. Select create user option



Step 3) On create user page

1. Enter desired login name
2. Enter desired password
3. Confirm password
4. Select ok

Create User

Fields marked with an asterisk (*) are required.

Login Name *	Guru99	1
Password *	*****	2
Confirm Password *	*****	3
Security Domain	Native	
Full Name *	Guru99 user	
Description		
Email		
Phone		
<input type="button" value="OK"/> 4 <input type="button" value="Cancel"/>		

Step 4)The user must have privilege to domain and repository, to assign the privilege

1. Click on username
2. Select privileges tab
3. Select the domain.

Domain Logs Monitoring Reports Security Cloud

Users Groups Roles OS Profiles Account Management Audit Reports

Search Users Groups Roles
for Go

Overview Privileges Permissions

Guru99 Edit

Select the domain or service to view the assigned privileges. The domain or services that you do not select will be listed below.

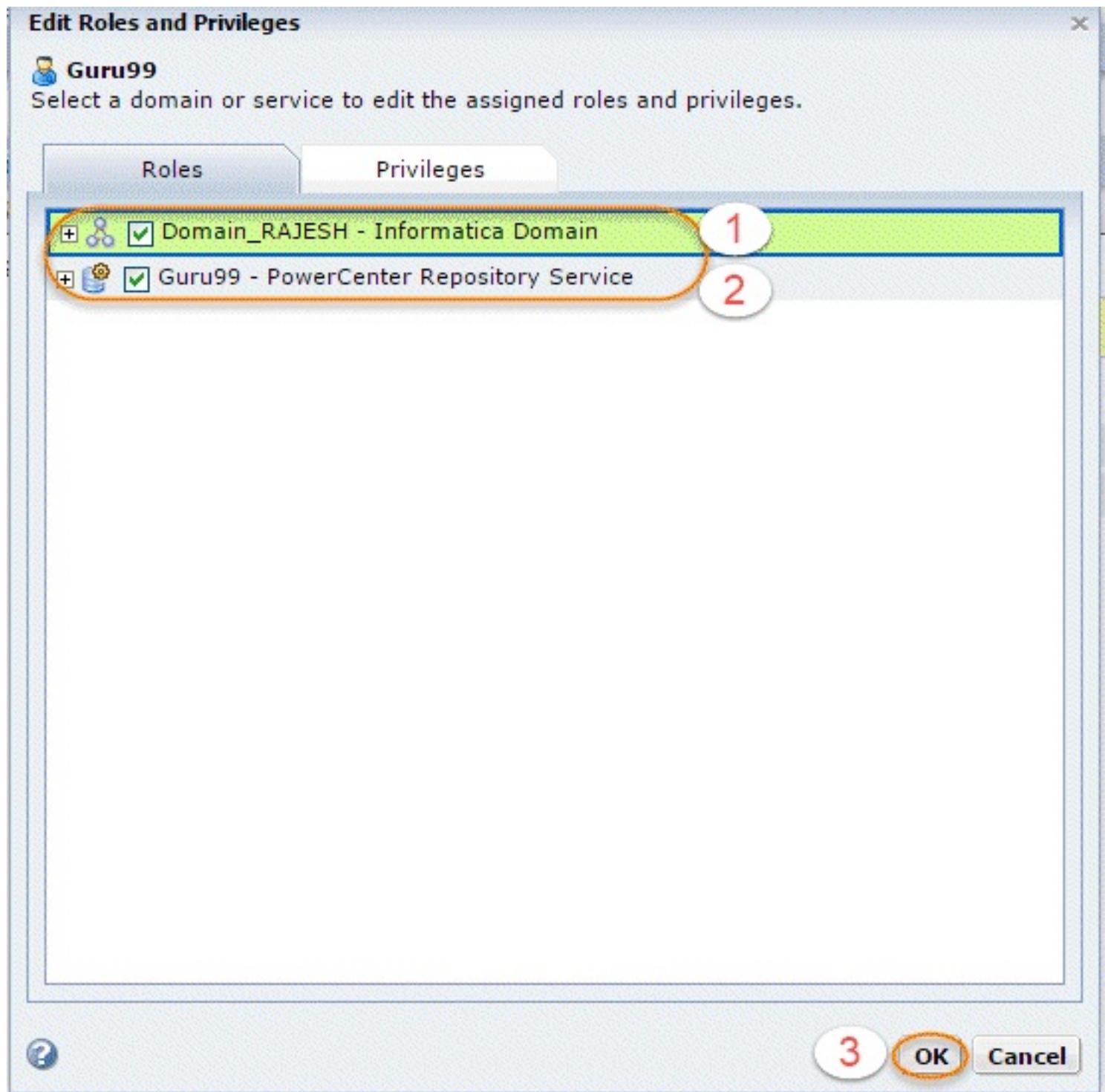
This user has no privileges assigned.

3 Domain_RAJESH - Informatica Domain (No privileges assigned)
Guru99 - PowerCenter Repository Service (No privileges assigned)

Step 5) In the edit role window

1. Select check box of domain
2. Select check box for repository service
3. Select ok.

This will assign repository and domain privilege to the user.



Now you will be able to connect to the repository using client tools.

This completes the exercise of repository service creation, integration service creation, domain configuration, user creation and adding a repository in the client.

Chapter 4: Source Analyzer and Target Designer

In ETL/Data Warehouse, you will encounter different sources and targets.

A Source in ETL is an entity from where you pull the records, and then you store these records in temporary tables (staging tables) or Informatica transformation caches. On the basis of mapping design document/business requirement you make a change in these data records (transform the data) and then you load the transformed data in another tables structures, called target tables.

In every Informatica mapping, there will always be a source and a target. To manage different sources and targets in Informatica, you have to use source analyzer and target designer. These tools are integrated into Powercenter designer tool and can be launched from there itself.

With the help of Source analyzer, you can create or import different types of sources in Informatica. Similarly by using target designer you can create or import different types of targets. Informatica provides you the feature to either create source/target from scratch by designing its structure or you can create source/target by importing the definitions. When you Import Source from the database, the following metadata gets imported.

- Source (Table) name
- Database location
- Name of Columns
- Columns Data types
- Constraints

You can also define the key relationships in the tables, which is applicable only at Informatica level and is stored in the repository.

The source or target created/imported in Informatica can be reused any no of times in different mappings. Every mapping must have at least one loadable target. Otherwise mapping will be invalid.

Note- When we create source/target in source analyzer/target designer, structures are only created in Informatica. At the database level, there is no object created. So, you have to create Database objects having the same structure as you have created in Informatica.

You can import following type of sources using source analyzer

- Relation tables (database tables), views and synonyms
- Flat files
- Cobol files
- XML files

Performance tip – To improve the performance of Relational Source tables, use indexes on the source database tables. On the target, tables disable or remove constraints and indexes for performance.

How to open Source Analyzer

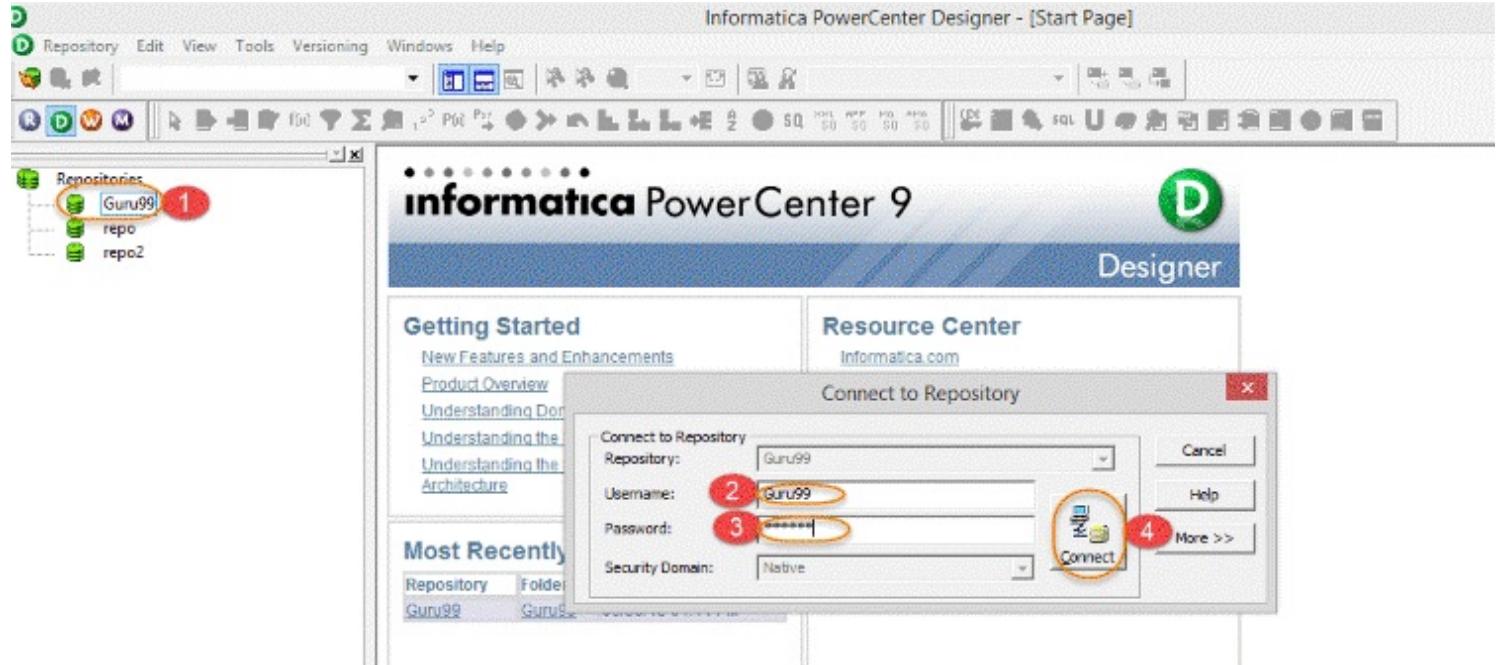
Step 1 - Open Informatica PowerCenter designer tool

	Custom Metadata Configurator	09-06-2015 22:57	Shortcut	3 KB
	PowerCenter Designer	09-06-2015 22:57	Shortcut	2 KB
	PowerCenter Mapping Architect for Visio	09-06-2015 22:57	Shortcut	2 KB
	PowerCenter Repository Manager	09-06-2015 22:57	Shortcut	2 KB
	PowerCenter Workflow Manager	09-06-2015 22:57	Shortcut	2 KB
	PowerCenter Workflow Monitor	09-06-2015 22:57	Shortcut	2 KB

Step 2 – In next screen

1. Double click on the repository which you want to connect

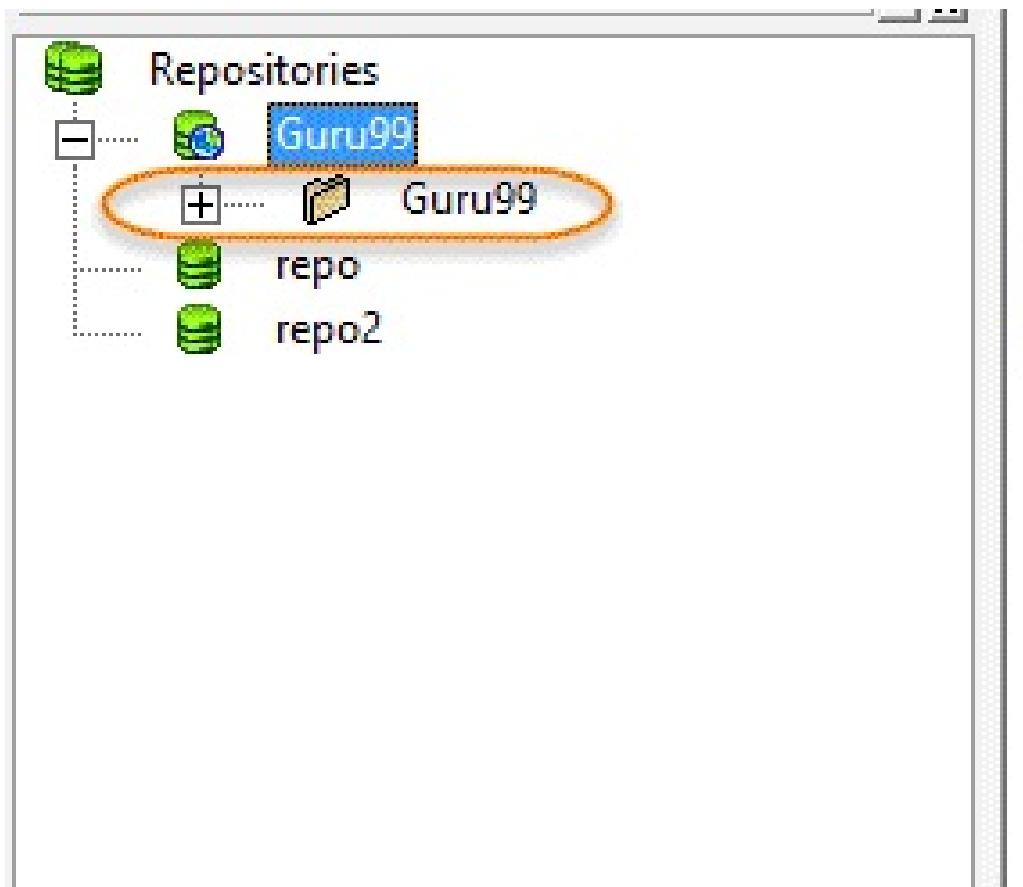
2. Enter username
3. Enter Password
4. Click on Connect button



After Successful login, the folders of the user will be listed under the repository name

Note –

1. If login is not successful, check if the user has privileges of connecting to repository
2. If no folder is visible under repository name, check if the folder is created.
3. If a folder was created recently then disconnect from the repository and reconnect.

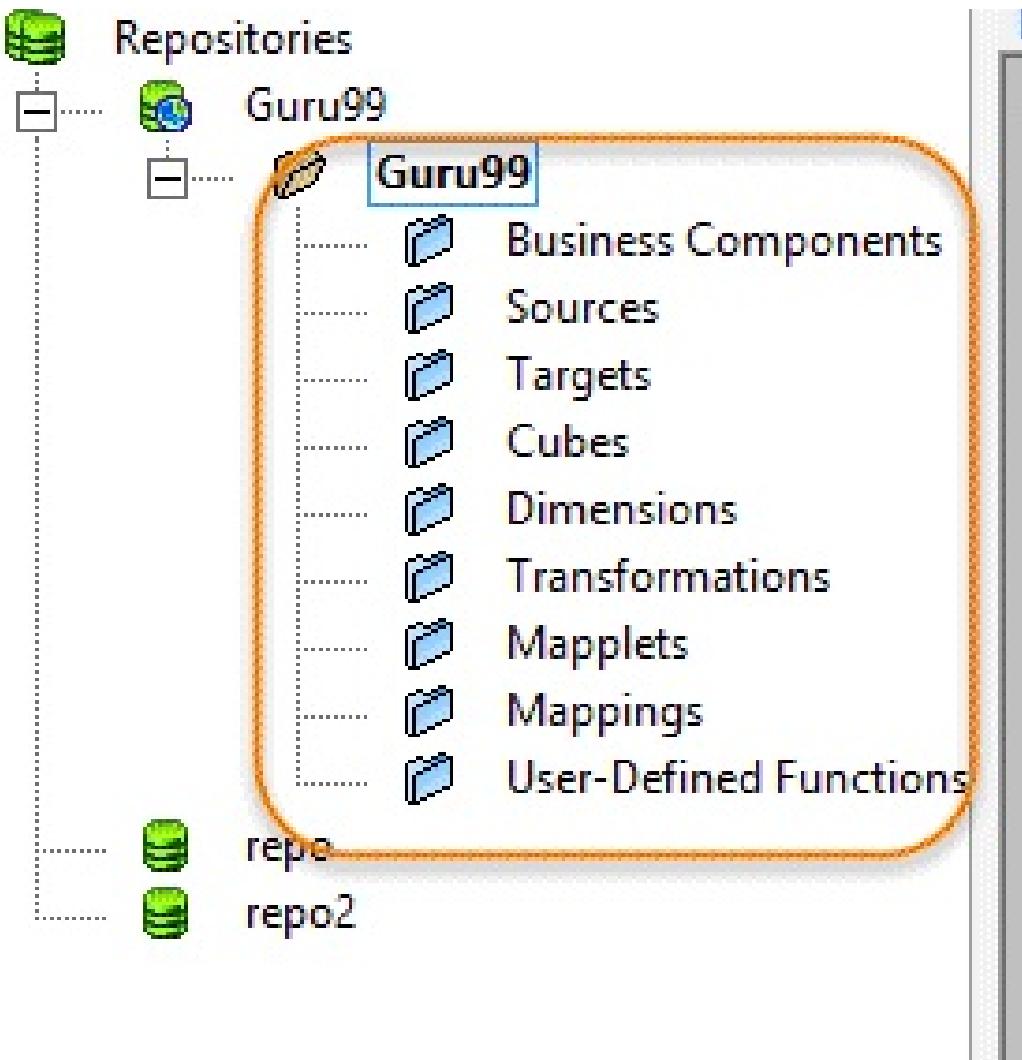


Step 3 – In the next step

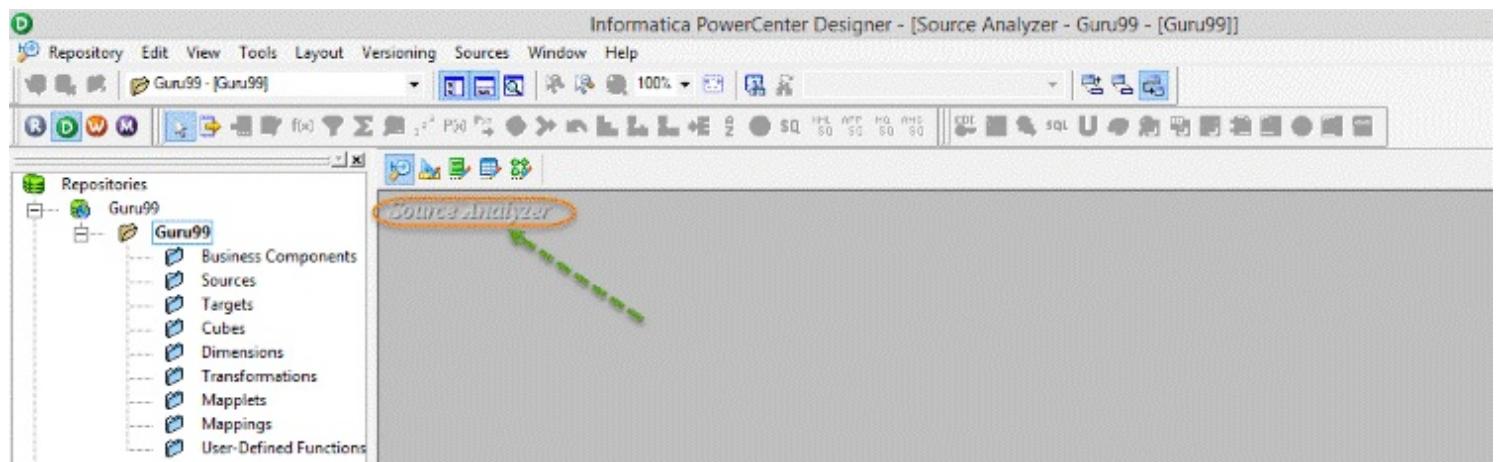
1. Right click on the folder
2. Select open option

The screenshot shows the Informatica PowerCenter Designer application. On the left, there's a toolbar with icons for Refresh (R), Data Flow (D), Workflow (W), and Metadata (M). Below the toolbar is a 'Repositories' tree view. The 'Guru99' repository is expanded, and its first child folder, also named 'Guru99', is selected and highlighted with a blue border. A context menu is open at this node, with the 'Open' option circled in red and labeled with a red number '2'. Other options in the menu include 'Allow Docking' (circled with a red number '1'), 'Hide', 'Connect', 'Disconnect', 'Find Checkouts...', and 'Properties'. The main workspace is titled 'informatica PowerCenter 9 Designer'. It features a 'Getting Started' section with links to 'New Features and Enhancements', 'Product Overview', 'Understanding Domains', 'Understanding the Repository', 'Understanding the Integration Service Architecture', and a 'Resource Center' section with links to 'Informatica.com', 'mysupport.informatica.com', 'Developer Network', and 'Informatica Events'. At the bottom, there's a 'Most Recently Opened' section showing a single entry for 'Guru99' last opened on '06/28/15 07:57 PM'. The bottom right corner features the 'INFORMATICA® The Data Integration Company™' logo.

When a folder is open, subfolders of that folder will be listed.



Step 4 – Click on the Source analyzer menu, as shown in the figure.

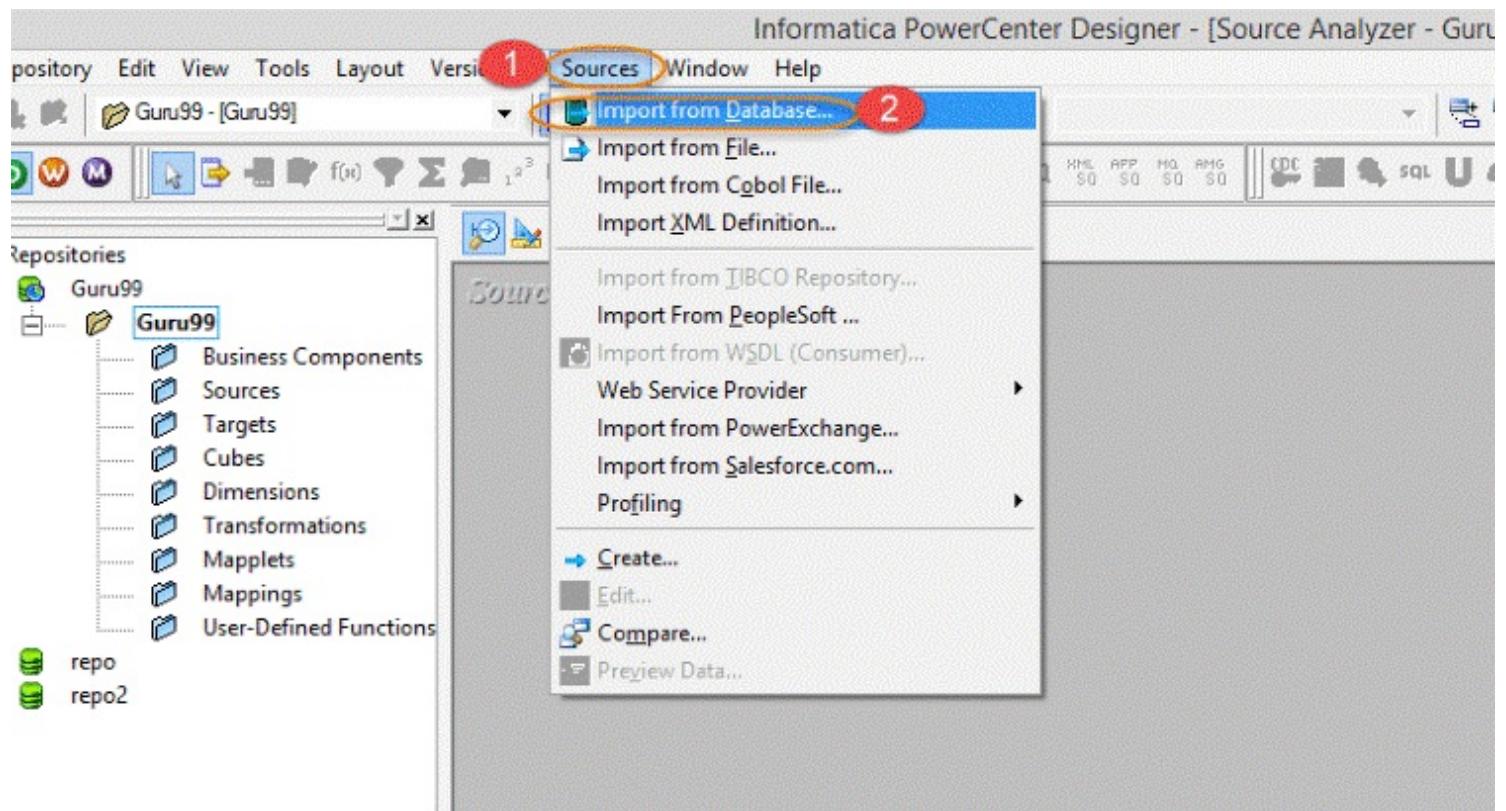


How to import Source table in Source Analyzer

Step 1 - In source analyzer

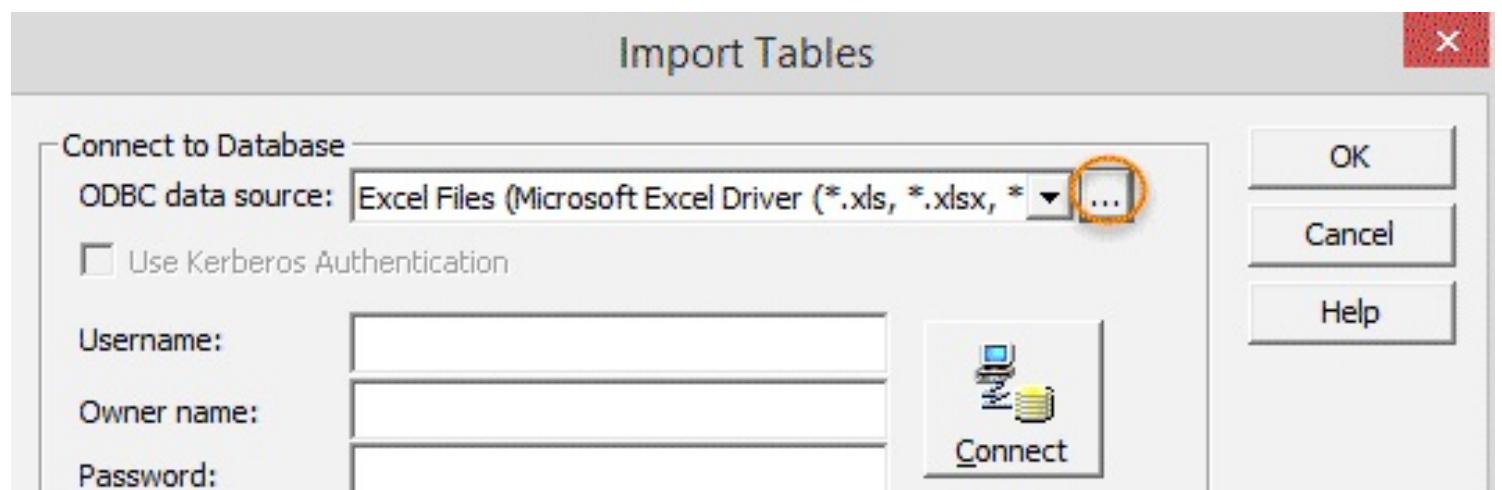
1. Click on tab "sources" from the main menu

2. Select import from database option, after this ODBC Connection box will open.

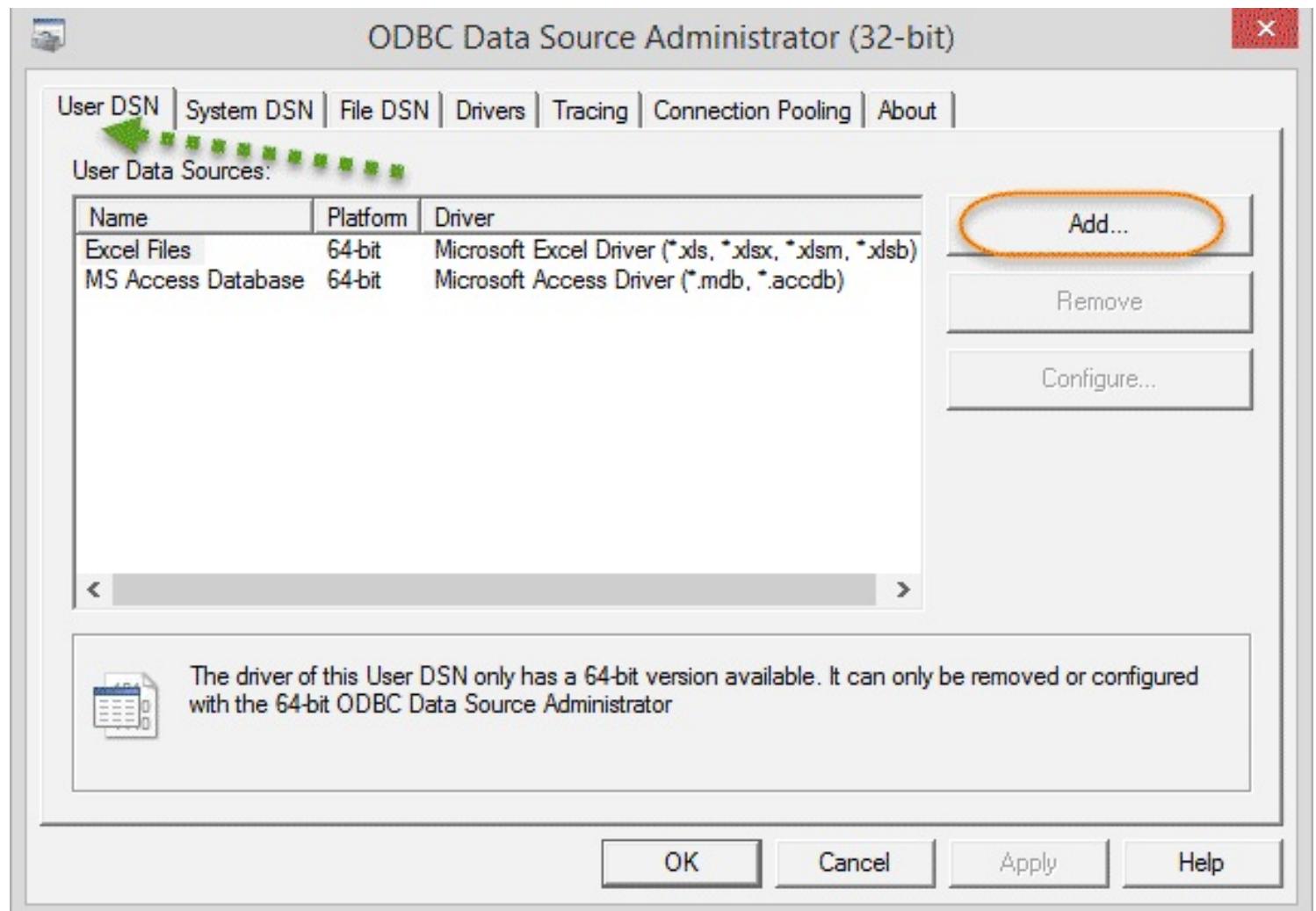


Step 2 – We will now create ODBC connection (If you already have ODBC connection, move to step 3)

1. Click on the button next to ODBC data Source.



2. On the next page, Select user DSN tab and click Add button.



A DSN (Data Source Name) is basically a data structure, and it contains information about specific details (database type, location, user details, etc.). This information is required by ODBC driver so that it can connect to that database.

- When you click on the add button, you will see a list of drivers for various databases (Oracle, SQL Server, Sybase, Microsoft Access, Excel, etc.) The driver which you select depends on what database you want to connect. Select the driver for the data source. Here in guru99 you will be using oracle database tables. Select Oracle wire protocol driver.

Create New Data Source

Select a driver for which you want to set up a data source.

Name	V
DataDirect 7.1 MySQL Wire Protocol	7
DataDirect 7.1 New SQL Server Wire Protocol	7
DataDirect 7.1 Oracle Wire Protocol	7
DataDirect 7.1 SQL Server Wire Protocol	7
DataDirect 7.1 Sybase Wire Protocol	7
Driver da Microsoft para arquivos texto (*.txt; *.csv)	6
Driver do Microsoft Access (*.mdb)	6
Driver do Microsoft dBase (*.dbf)	6
Driver do Microsoft Excel (*.xls)	6
Driver do Microsoft Powerpoint (*.dhx)	6

< Back

Finish

Cancel

4. On the next page, select the general tab and enter database details. Then Click on test connect.

ODBC Oracle Wire Protocol Driver Setup

[?](#)

Pooling | Bulk | Client Monitoring | Advanced Security | About
General | Advanced | Security | Performance | Failover

Data Source Name: Help

Description:

Standard Connection

Host:

Port Number:

SID:

Service Name:

TNSNames Connection

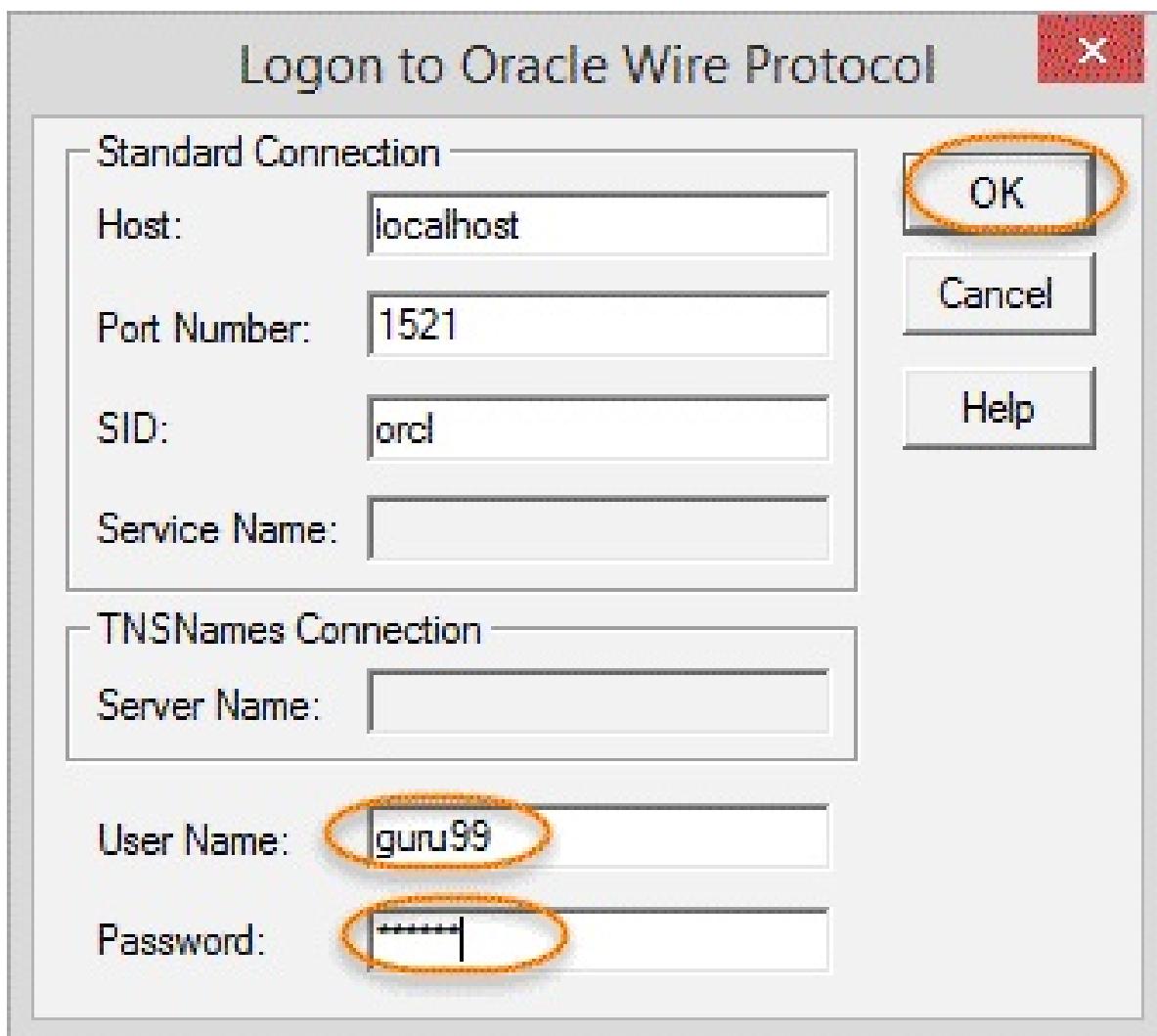
Server Name:

TNSNames File

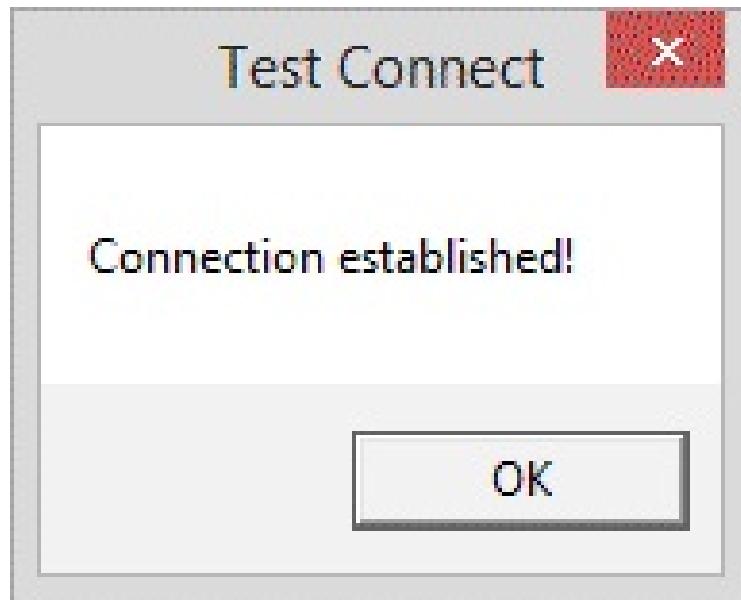
Edition Name:

Note – Database details are specific to your installation depending on hostname, port and SID you selected during database installation. You can get this details by opening **tnsnames.ora** file located in your oracle installation folder.

- Once you made a Test Connect, enter database username, password and then select "OK" button



- If the connection is successful, it will show a message box.



7. Select OK for test connect window and ODBC Driver setup window. The Created ODBC Data Source will be added to user data sources. Now we are set with ODBC Data source.

Step 3 - In guru99, you will be using oracle's Scott/Tiger schema tables. If you do not have these tables in the database, create those using this script.

Step 4 – In the import tables window.

1. Select ODBC data source for the oracle database.
2. Enter database username
3. Enter database password
4. Click on connect/reconnect button. This will show tables for the database user.
5. Expand the tree under tables folder and select EMP table
6. Select OK button.

Import Tables

6

OK

Cancel

Help

Connect to Database

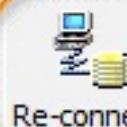
ODBC data source: 1 Guru99 (DataDirect 7.1 Oracle Wire Protocol)

 Use Kerberos Authentication

Username: 2 guru99

Owner name: GURU99

Password: 3 *****



Re-connect

4

Select tables

GURU99

TABLES

DEPT

EMP 5

OPB_PCIS_SERVICELEASE

OPB_PCIS_SESSIONLEASE

VIEWS

Show owners:

Default

All

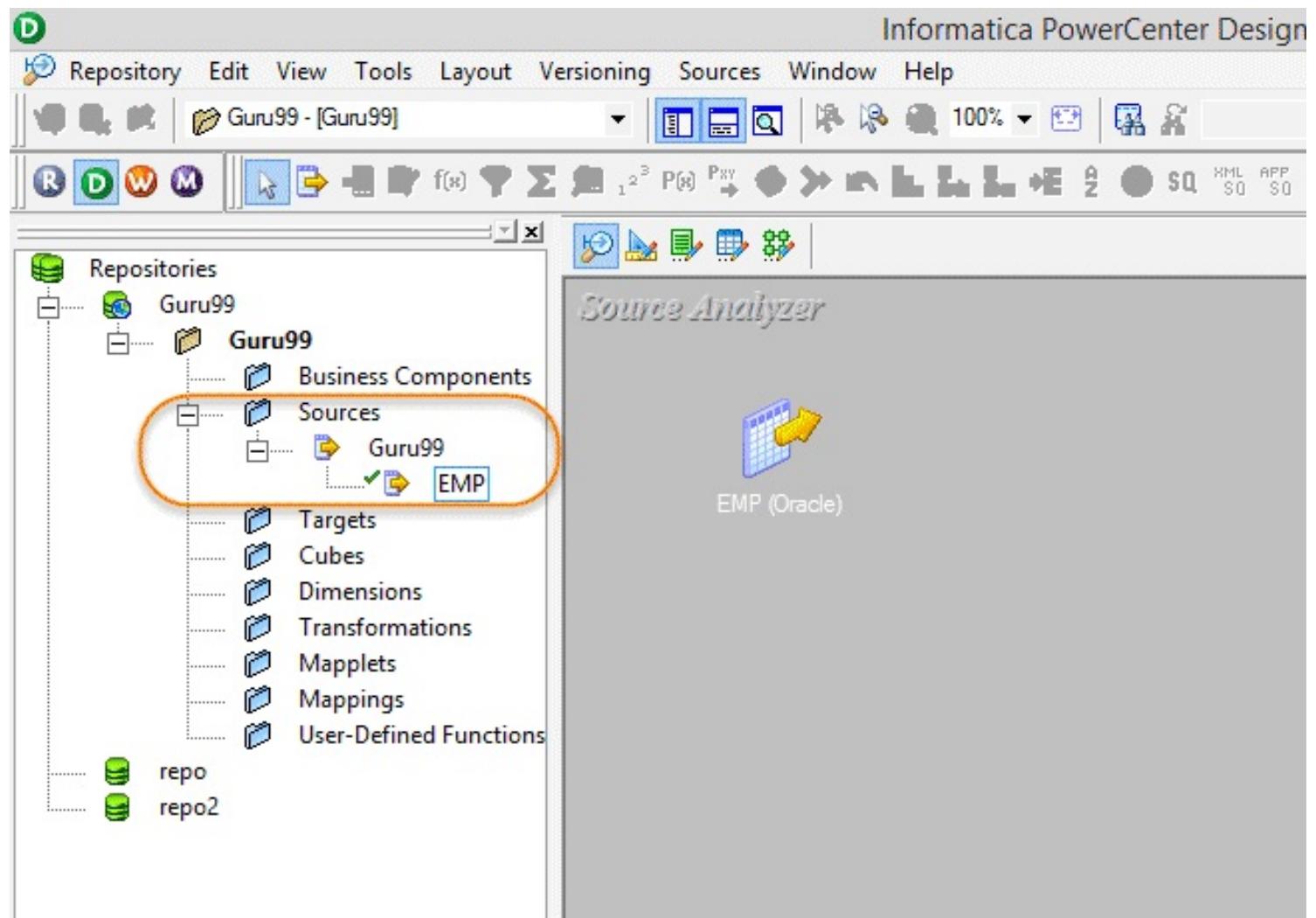
Search for tables named:

Search

Select all

Select none

Step 5- The table will be imported in the Informatica Source Analyzer. Use "ctrl+s" keys to save changes to the repository.



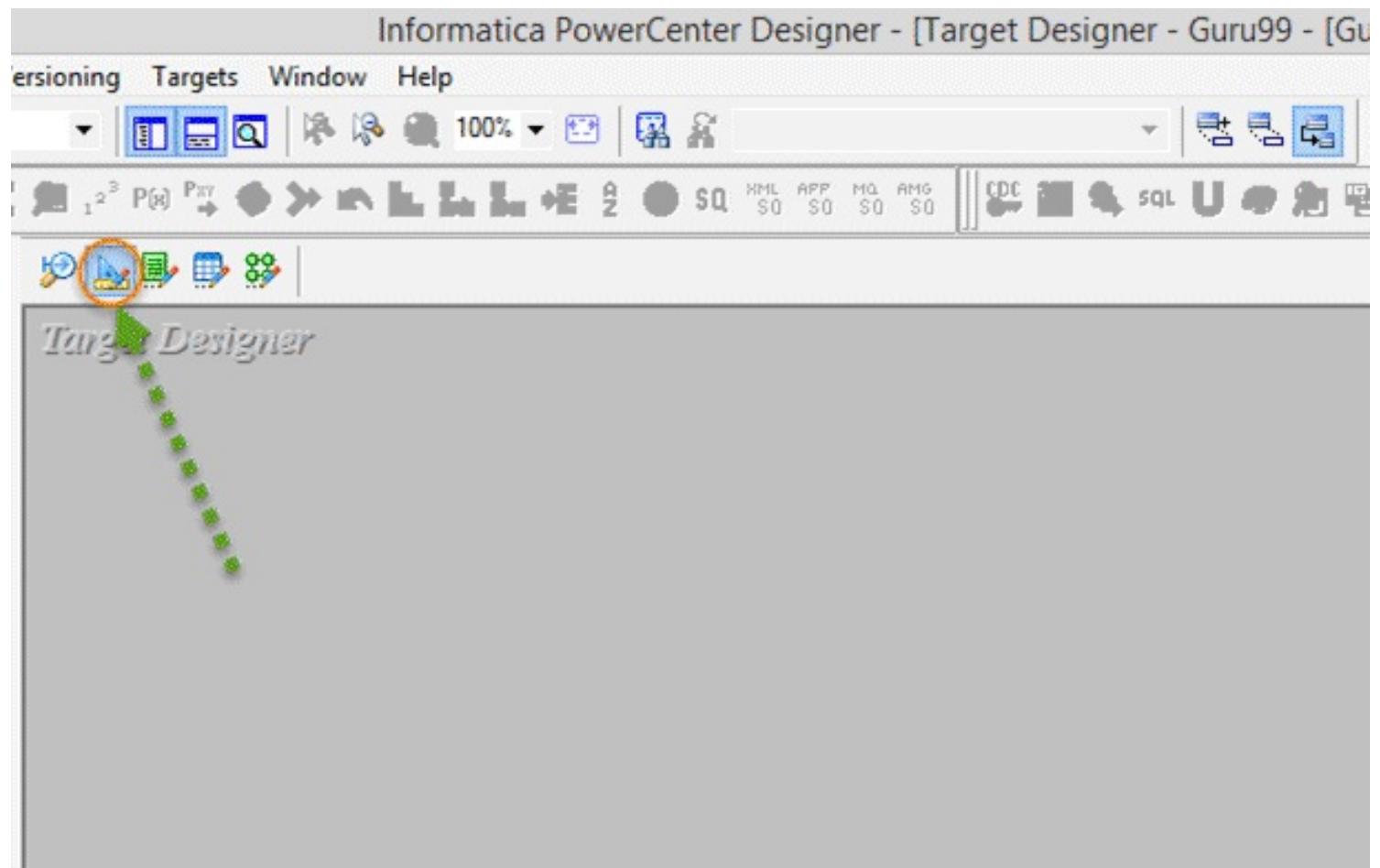
Similarly, you can import any other database tables in source analyzer.

How to import target in target designer

In this tutorial, you will create target table emp_target, which will be having the identical structure of the source EMP table. To import target table, the target table must be present in the database. To create target table use the below script. Types of targets available in Informatica are relational, XML and flat files.

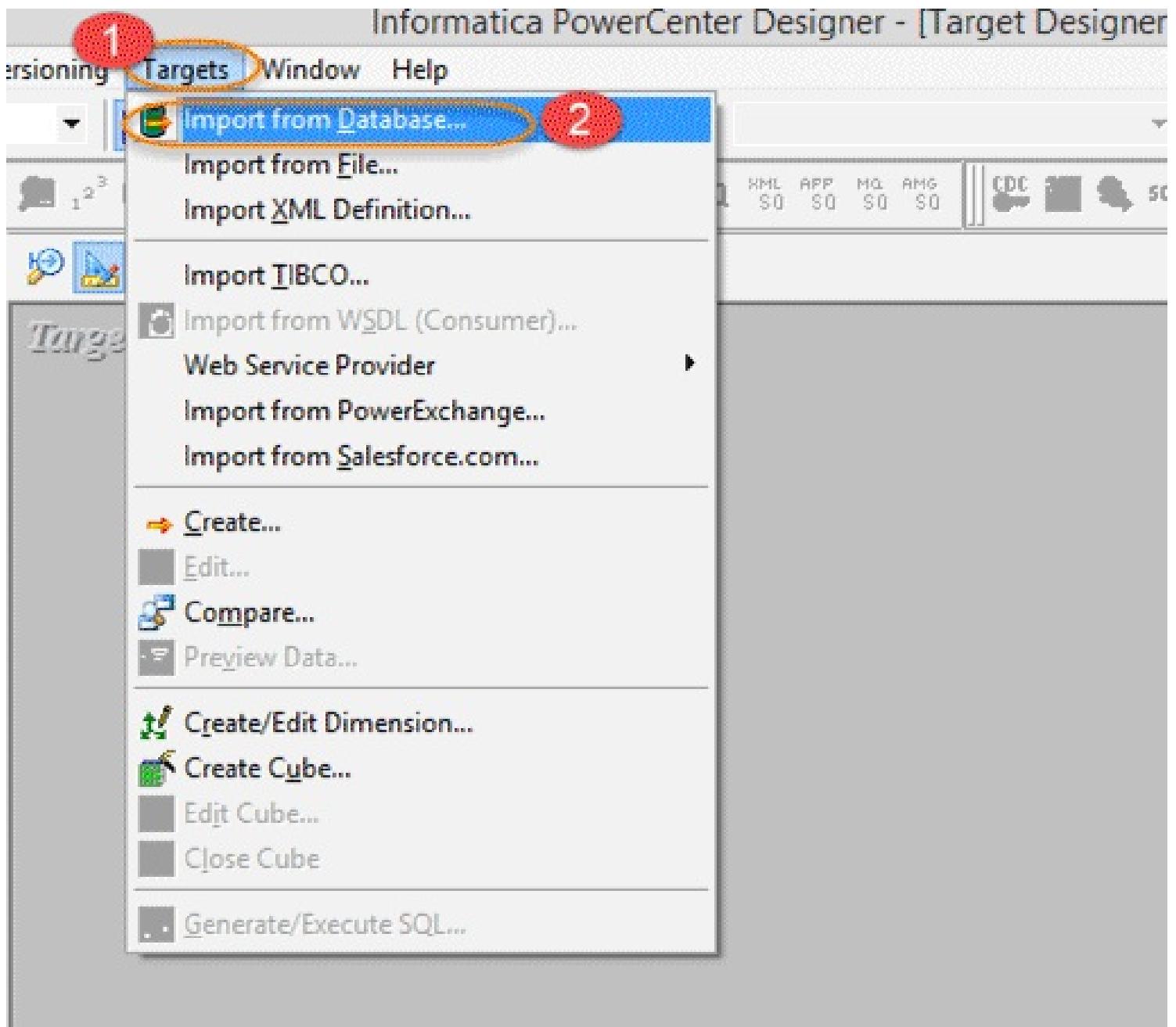
Note – here in this tutorial, both source and target table are present in the same database schema (guru99). However, you can also create a target in a different schema.

Step 1- In the Informatica Designer, click target designer icon to launch target, designer.



Step 2- In next step

1. Select option "Targets" from the main menu
2. Select "import" from database option



Step 3 - In the import table window.

1. Select ODBC data source for the oracle database.
2. Enter database username
3. Enter database password
4. Click on connect/reconnect button. This will show tables for the database user.
5. Expand the tree under tables folder and select EMP_TARGET table
6. Select OK button

Import Tables

Connect to Database

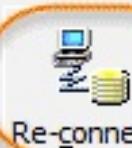
ODBC data: 1 Guru99 (DataDirect 7.1 Oracle Wire Protocol) ...

Use Kerberos Authentication

Username: 2 guru99

Owner name: GURU99

Password: 3 *****



Re-connect 4

OK 6

Cancel

Help

Select tables

GURU99

TABLES

DEPT

EMP

5 EMP_TARGET

OPB_PCIS_SERVICELEASE

OPB_PCIS_SESSIONLEASE

VIEWS

Show owners:

Default

All

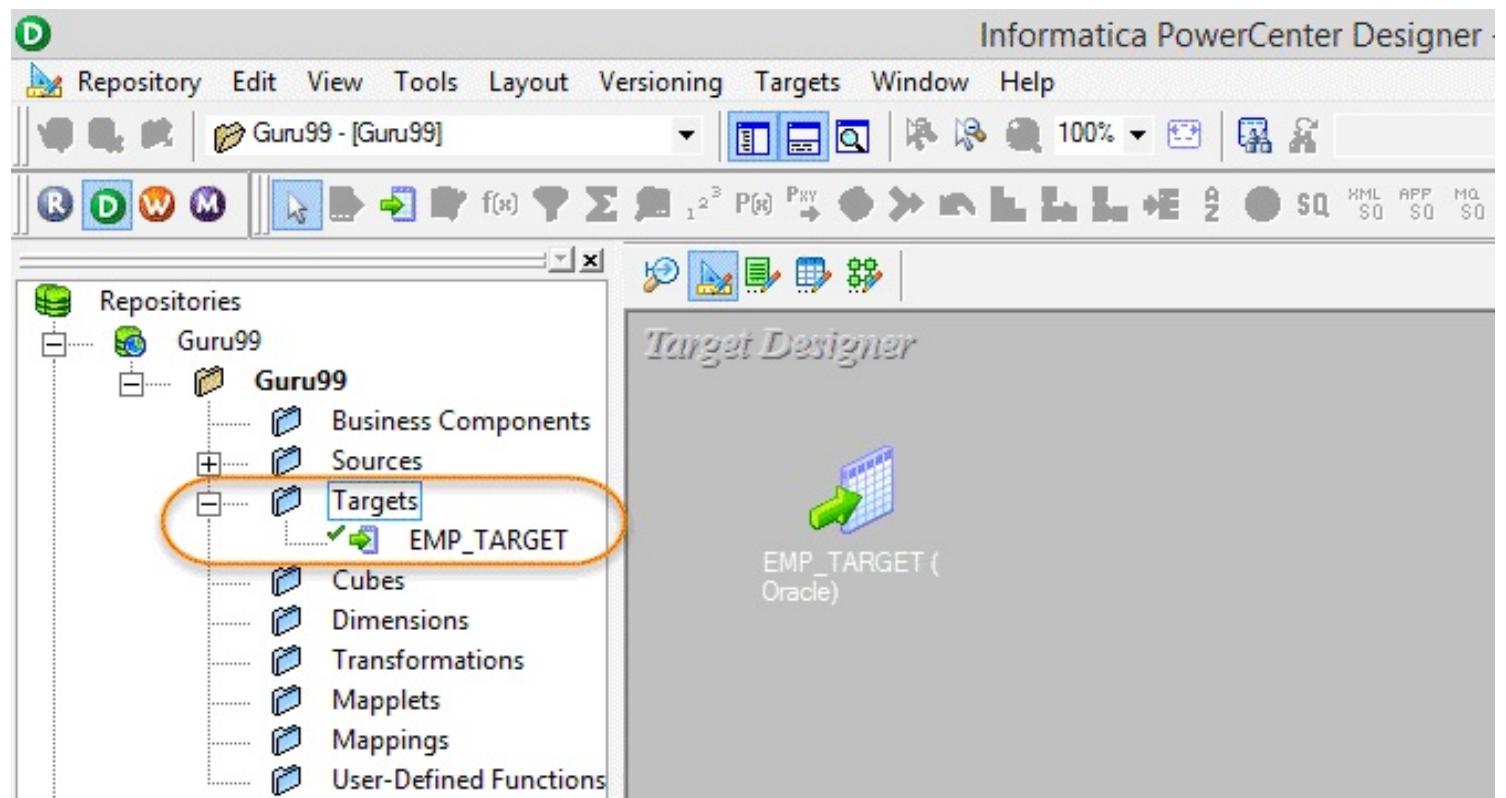
Search for tables named:

Search

Select all

Select none

This will import target in Target Designer of Informatica Powercenter. Use "ctrl+s" key to save the changes.



In this tutorial, you have imported sources and targets. Now, you are all set to create your first mapping.

How to create folder in Informatica

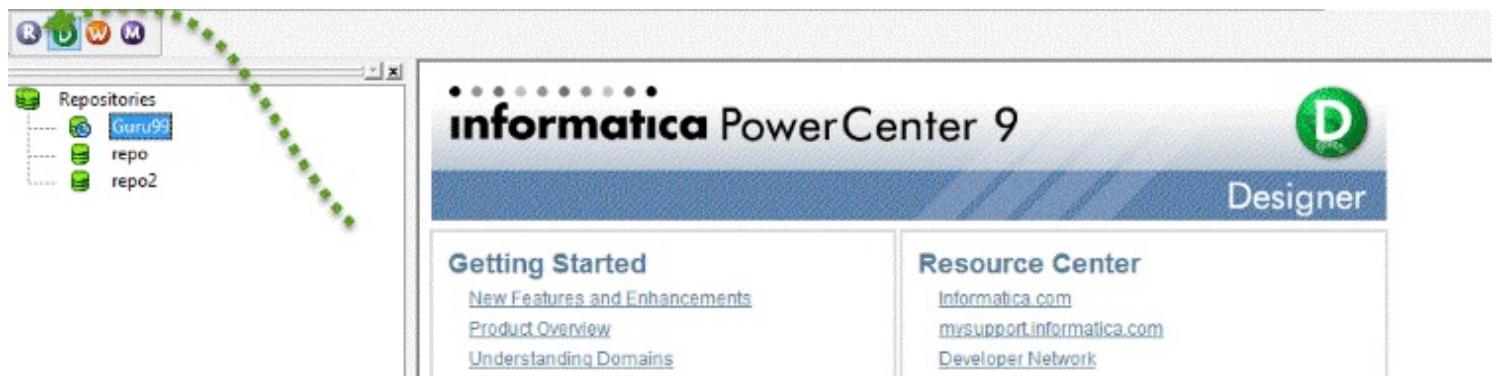
Informatica folders are logical containers, which keeps the objects that you create in Informatica designer/ Workflow manager (mappings, mapplets, sources, targets, etc.)

To create folder

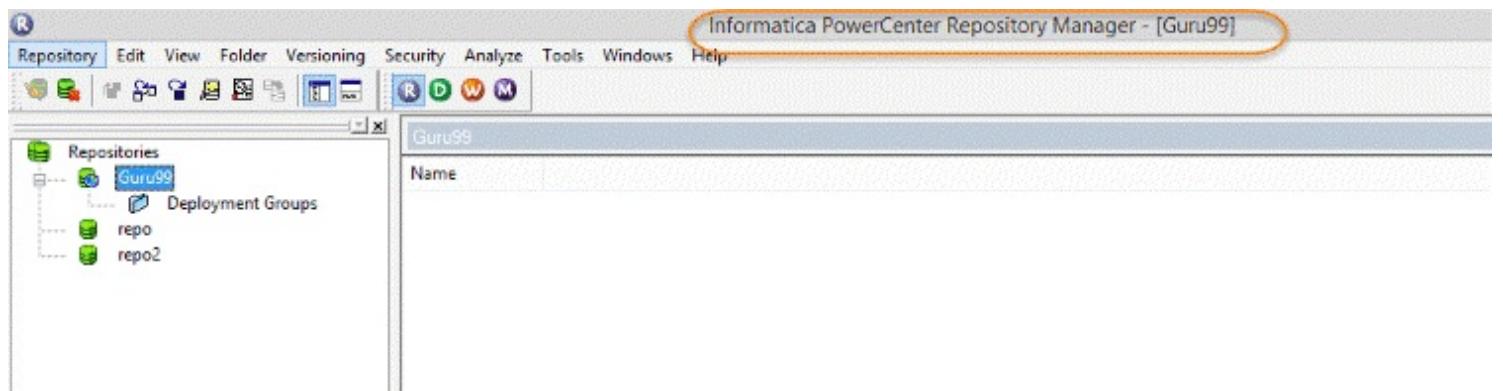
Step 1 – Open Repository Manager

	Custom Metadata Configurator	09-06-2015 22:57	Shortcut
	PowerCenter Designer	09-06-2015 22:57	Shortcut
	PowerCenter Mapping Architect for Visio	09-06-2015 22:57	Shortcut
	PowerCenter Repository Manager	09-06-2015 22:57	Shortcut
	PowerCenter Workflow Manager	09-06-2015 22:57	Shortcut
	PowerCenter Workflow Monitor	09-06-2015 22:57	Shortcut

Note - If Informatica Designer is already open, then repository manager can be opened using the shortcuts available in the toolbox.

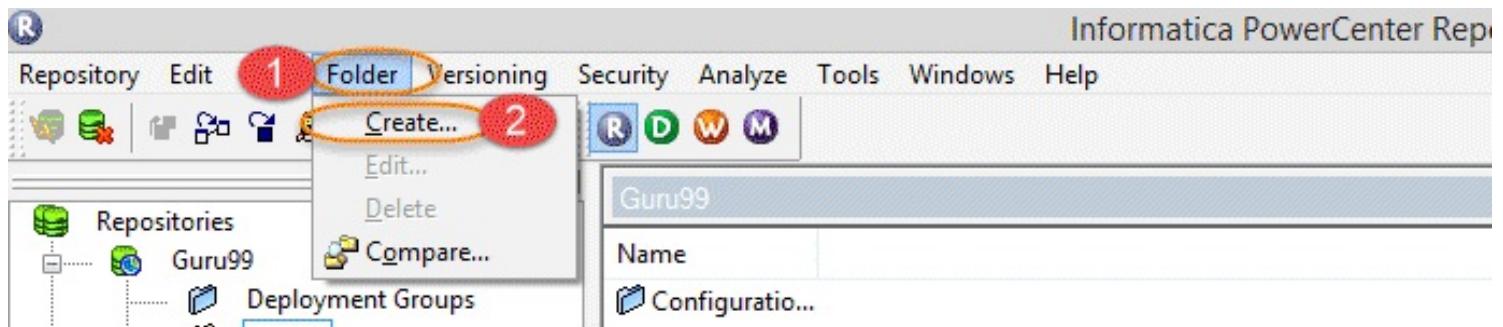


This will open Informatica Repository Manager in a separate window.



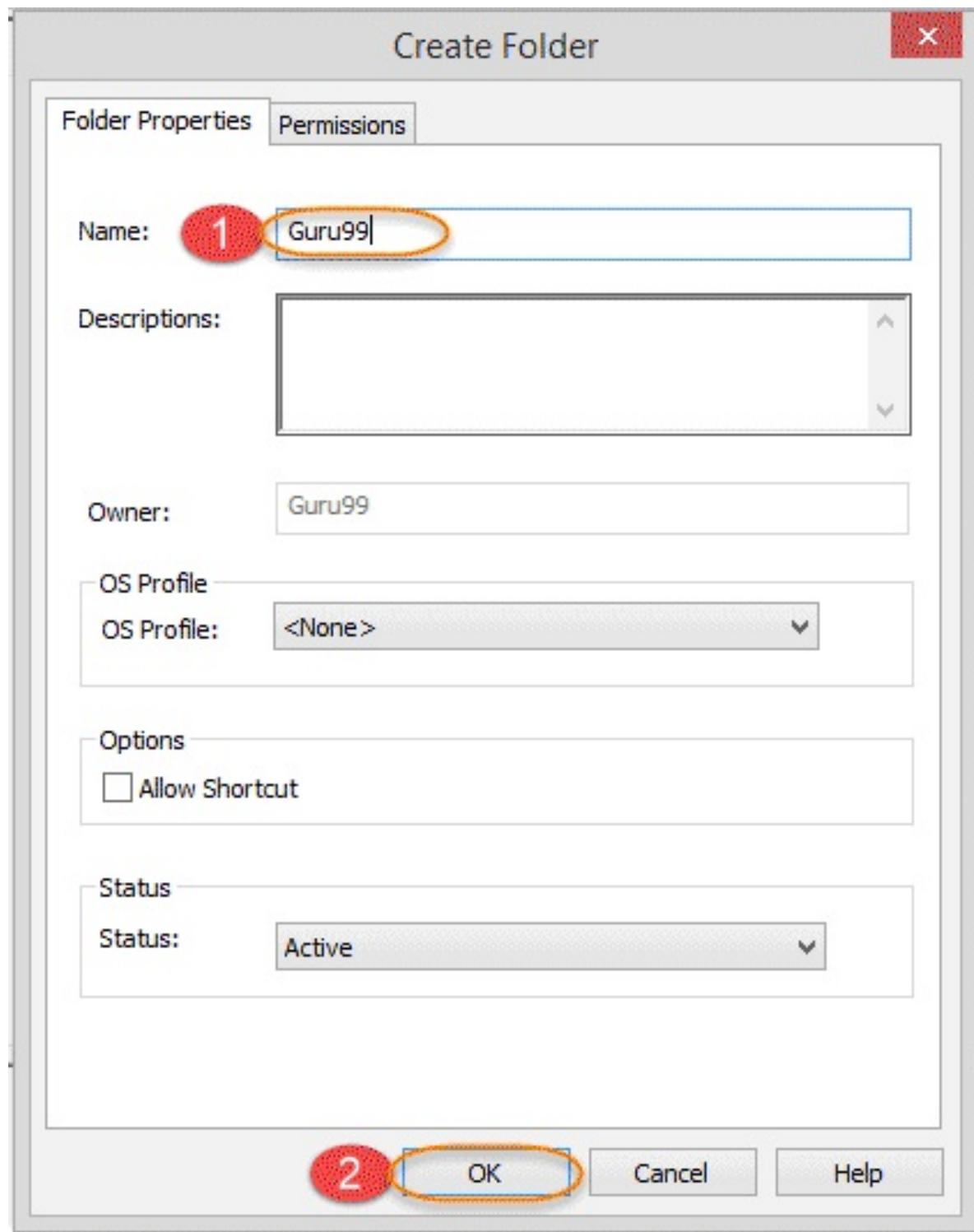
Step 2 - Follow these steps

1. Click on folder menu
2. Select Create option

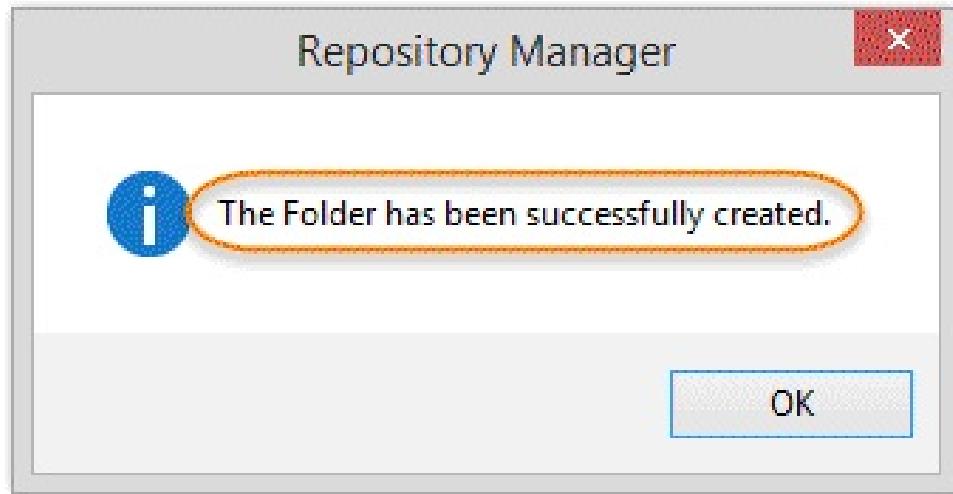


Step 3 – In next step

1. Enter Desired folder name
2. Select ok



This will create a new folder and a message will appear saying "The folder has been successfully created."



The newly created folder will be listed under the respective repository.



This completes the tutorial. See you in the next one!

Chapter 5: Mappings

A mapping is a collection of source and target objects linked together by a set of transformations. These transformations consist of a set of rules, which define the data flow and how the data is loaded into the targets.

A mapping consists of following set of objects

- **Source Definition** - Source definition defines the structure and characteristic of the source, its underlying data types, type of the data source, etc.
- **Transformation** - Transformation objects define how the source data is transformed, and various functions can be applied during the process
- **Target Definition** - Target definition defines the final target where the data will be loaded.
- **Links** - Links connect the source definition to different transformations and target tables. It defines how the data flows from source to target and the transformations.

Why do you need Mapping?

Mapping is an object in Informatica with the help of which you can define how the source data is modified before it reaches the destination or target object. Like if you have employee name as "Bill Clinton" in your source system and in the target system the requirement is to have employee name in the format as "Clinton Bill", such operations can be designed at the mapping level. In basic terms, what you do with the source data is defined at the mapping level.

Mapping is the basic Informatica object with the help of which we can define the data transformation details and source/target object characteristics. Mappings help us to define the data transformation at the individual column levels for each row. Even in a single mapping you can handle multiple sources and targets.

Components of Mapping

Basic components of a mapping are

- **Source tables**
- **Mapping parameters and variables**
- **Target objects**
- **Mapping transformations**

There are various objects that constitute a mapping. A mapping can consist of **sources, targets, mapping parameter and variables, mapplets, various transformations, and user-defined functions**.

- **Mapping Source:** Mapping sources are the objects from where you fetch the source data. It can be a database table, flat file, XML source or COBOL file source
- **Mapping target:** Mapping target is our destination objects where final processed data gets loaded. Mapping target can be a relational table of a database, a flat file or XML file. Sources and targets are mandatory in any mapping, their type can differ
- **Mapping Parameters and Variables:** Mapping parameters and variables helps you to create temporary variable objects which will help you to define and store temporary values while mapping data processing. Mapping parameters and variables are optional users defined data types, which can be created for a mapping and can be referenced and updated for a specific requirement. We will learn more about mapping parameters and variables in this section
- **Mapplets:** They are objects which consist of a set of transformation, source or targets. Mapplets are generally created to reuse the existing functionality of a set of transformations. It can be used in any no of mappings.

What is Stage Mapping?

A stage mapping is a mapping in where we create the replica of the source table. For Example, in a production system if you have an "employee" table then you can create an identical table "employee_stage" in ETL schema.

Having a local stage table offers various advantages, like production downtime, won't affect your ETL system because you have your own "employee_stage" table, instead of referring to production "employee" table. In a Production system, there can be other operations and processes which affect the performance. However, when you have

replica staging table, only ETL processes will access it. This offers performance benefits.

In Stage Mappings,

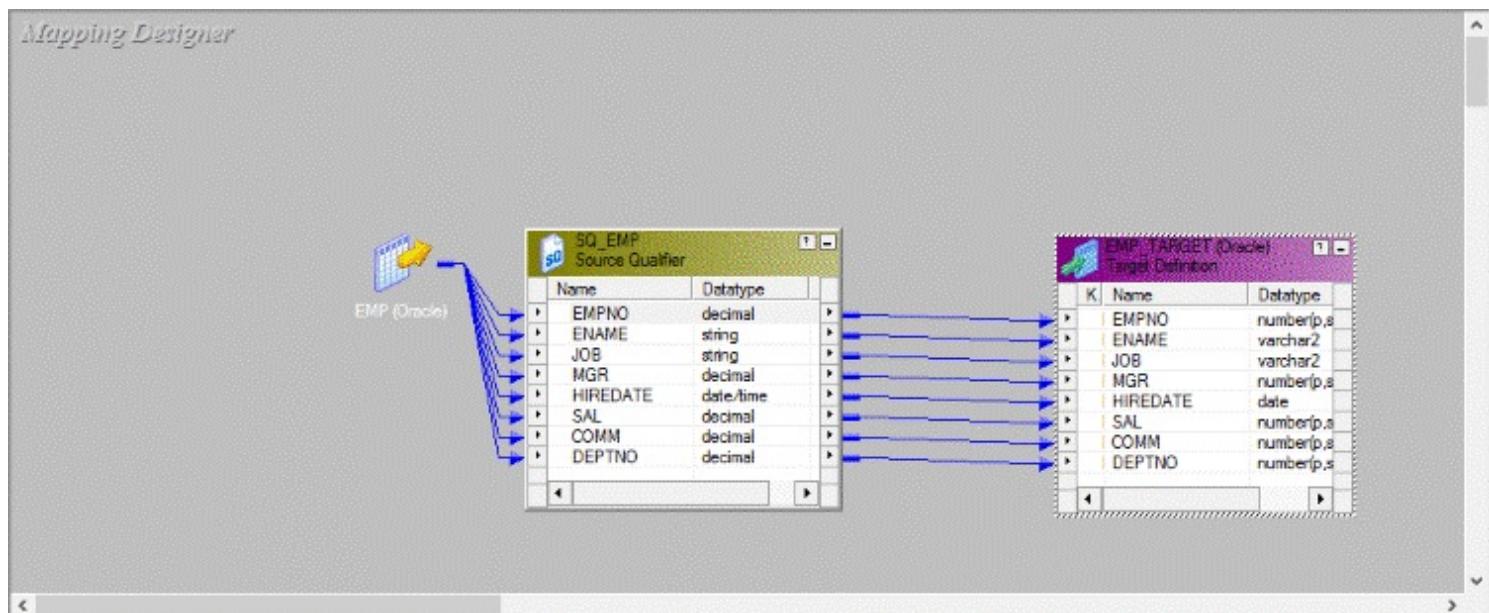
- Source and Target tables have identical structures
- The data in the target table is a replica of source table data or
- Data in stage (target) table is a subset of source data.

For example, if your source table contains employee details of deptno 10, 20, 30, and 40. The staging table can be a table having employee records of deptno 10 & 30 only.

The purpose of creating stage tables in Data warehouse is to make the process of data transformation efficient by fetching only those data which is relevant to us and also to minimize the dependency of ETL/Data Warehouse from the real-time operational system.

How to Create a Mapping

In this exercise, we will create a stage mapping, in which source will be "emp table" and the target will be "emptgt".

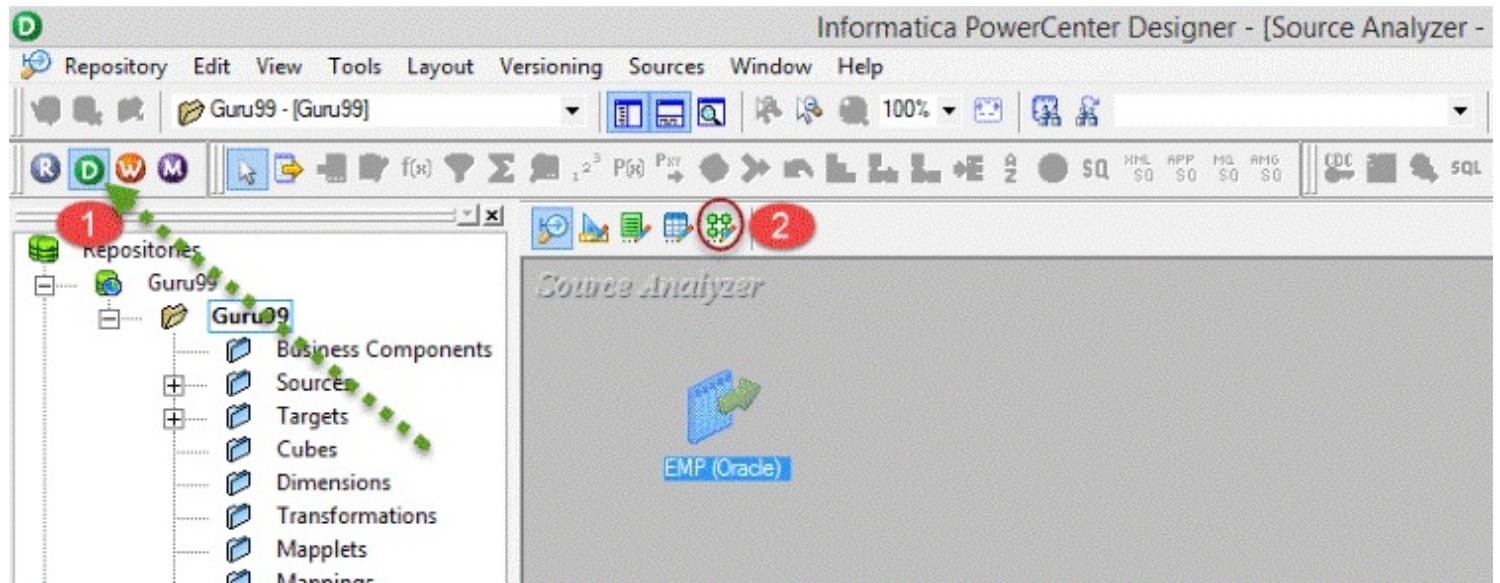


Naming Convention – mapping names are prefixed with 'm_' followed by source and target table names separated by underscore sign.

Example – if you are loading emp_target table from the emp table, then mapping name can be 'm_emp_emp_target'.

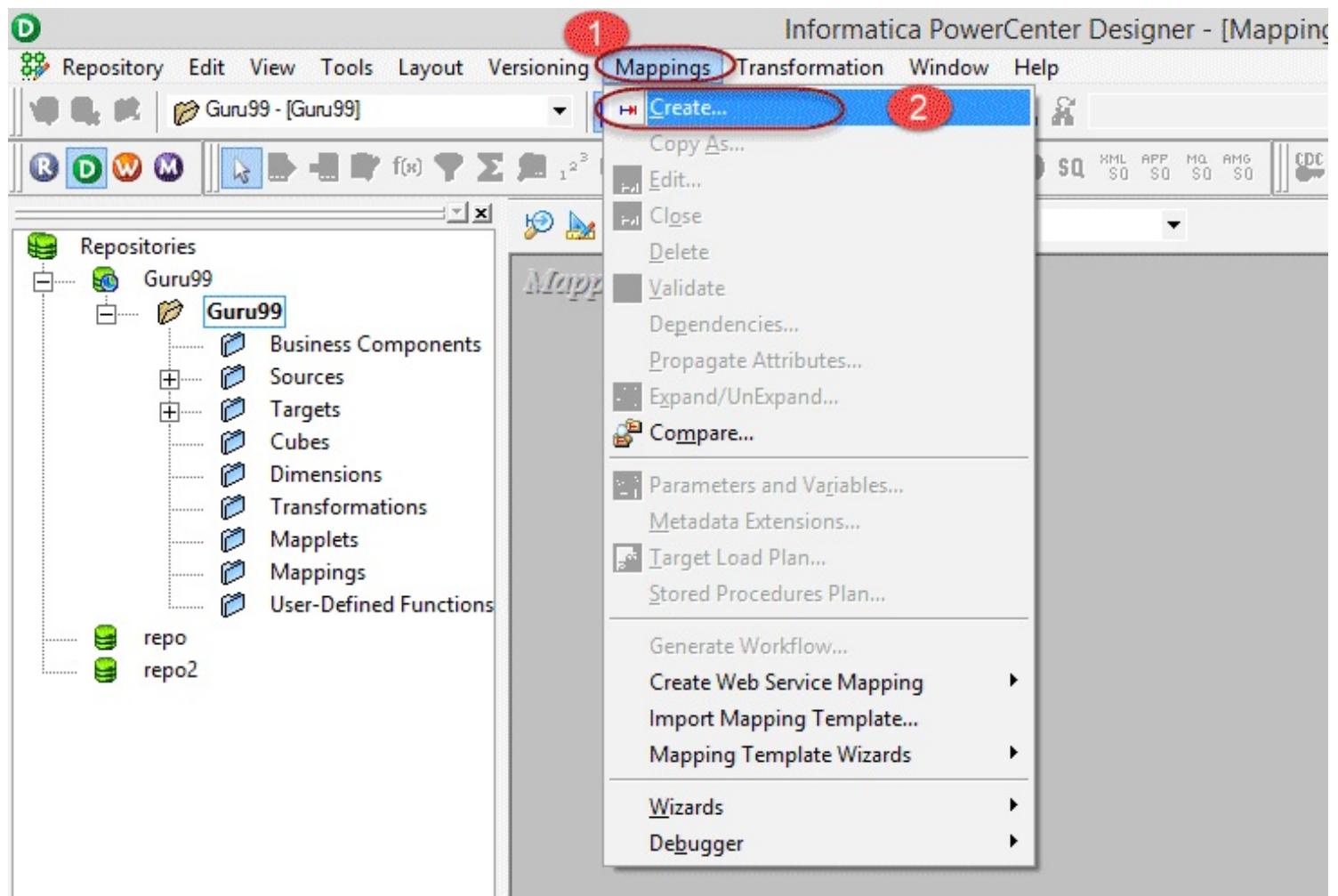
Step 1 – Launching Mapping Designer

1. Open Informatica Designer Tool
2. Click on Mapping Designer Icon to launch Mapping Designer

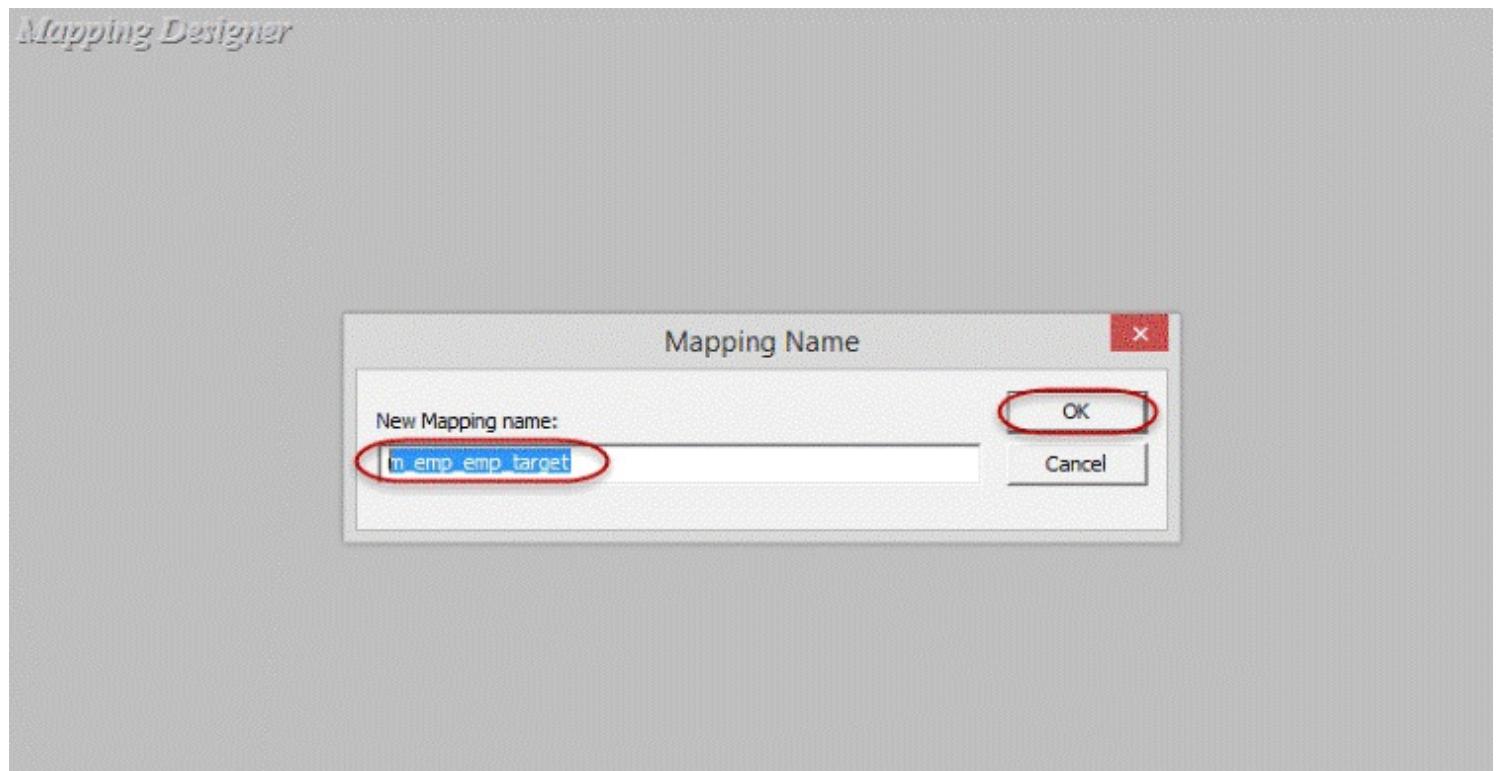


Step 2 – In Mapping Designer

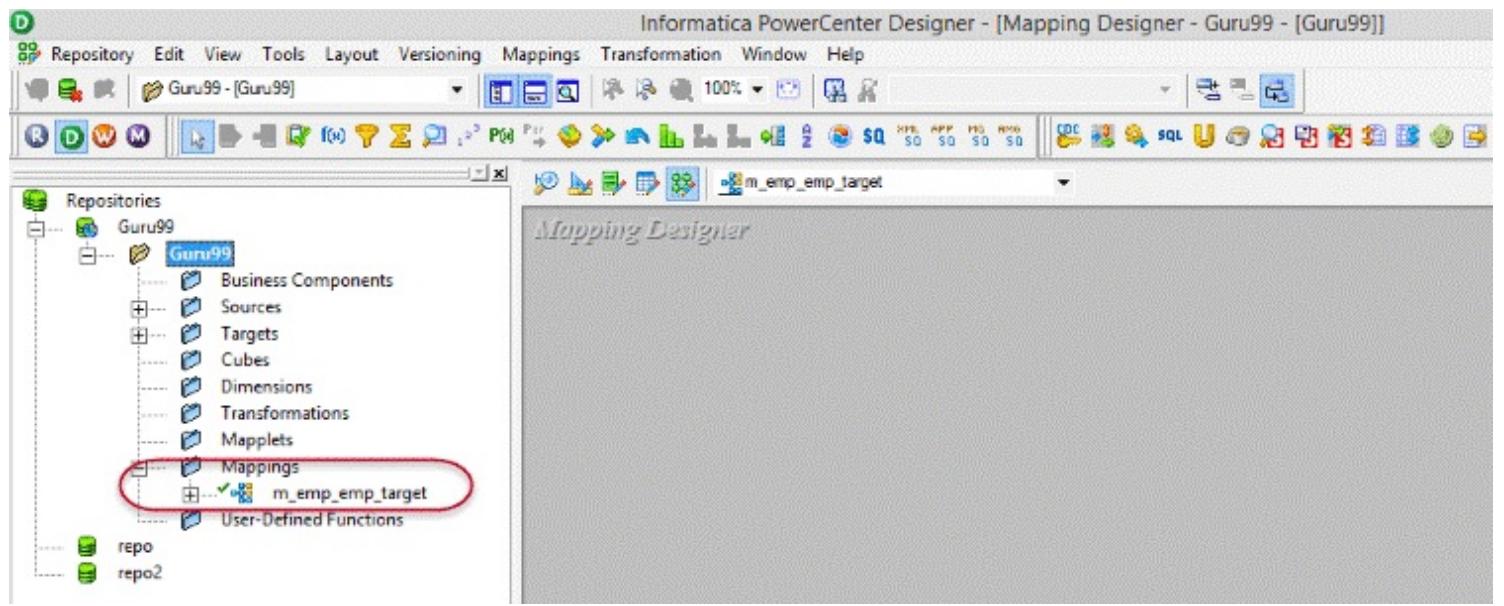
1. Click on Mappings Menu
2. Select Create Option



Step 3 – Enter Mapping name as 'm_emp_emp_target' and select OK Button.



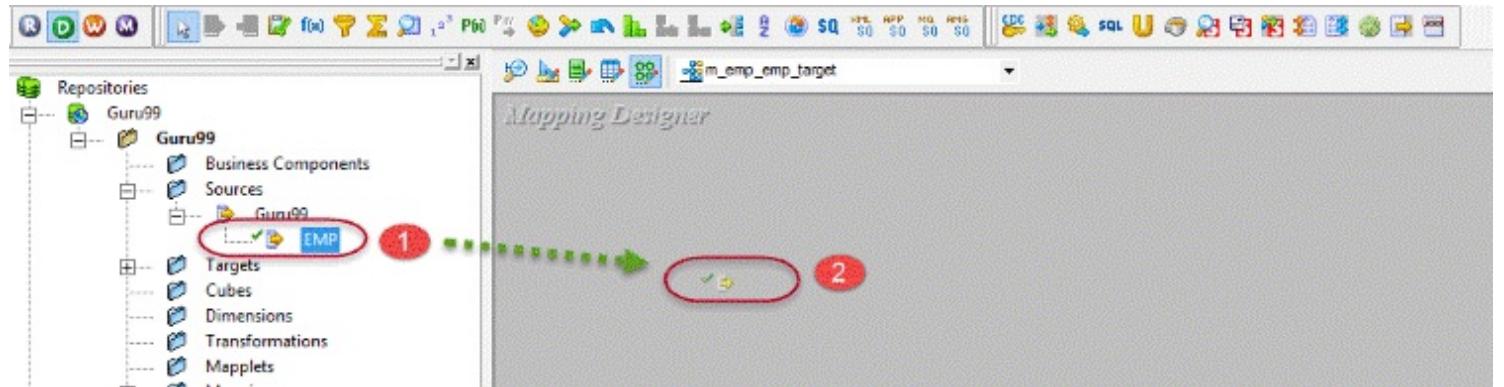
Mapping will be created and listed under mappings folder.



A Mapping must have at least a **source** and a **target**, you will add sources and targets to the mapping.

Step 4 – In this step we will,

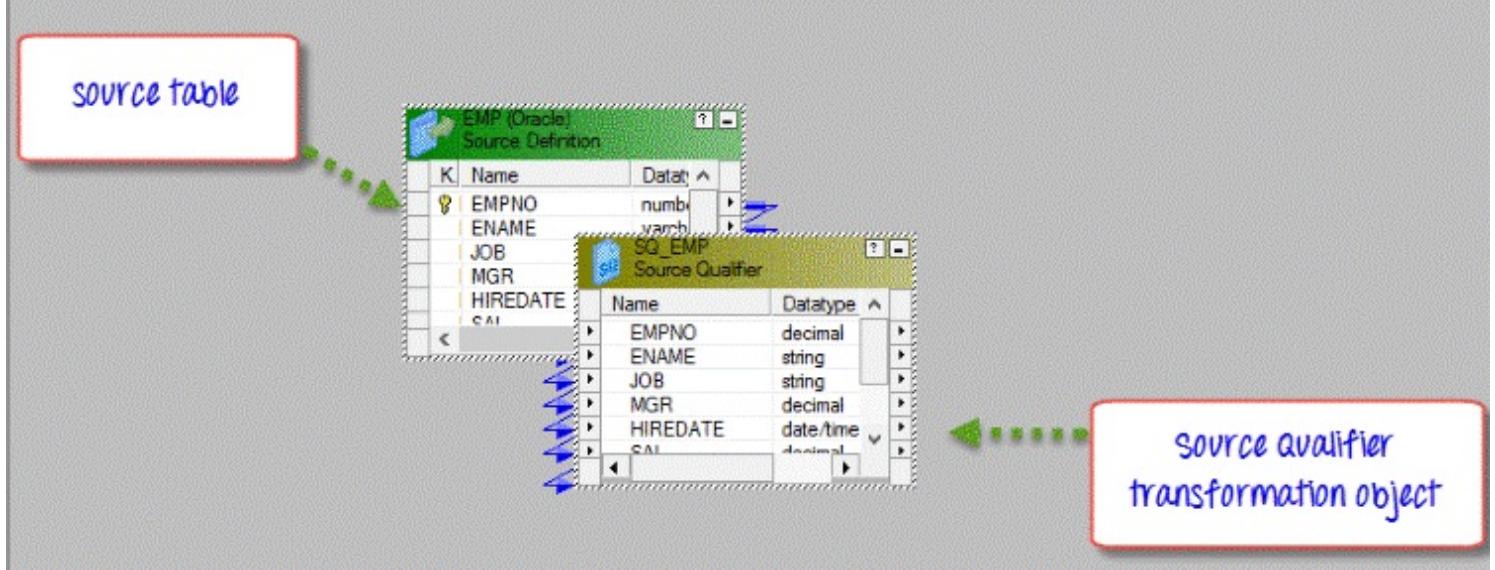
1. Select "emp" source table under sources folder.
2. Drag and drop "emp" table into mapping designer.



In mapping designer, imported source table will be shown.

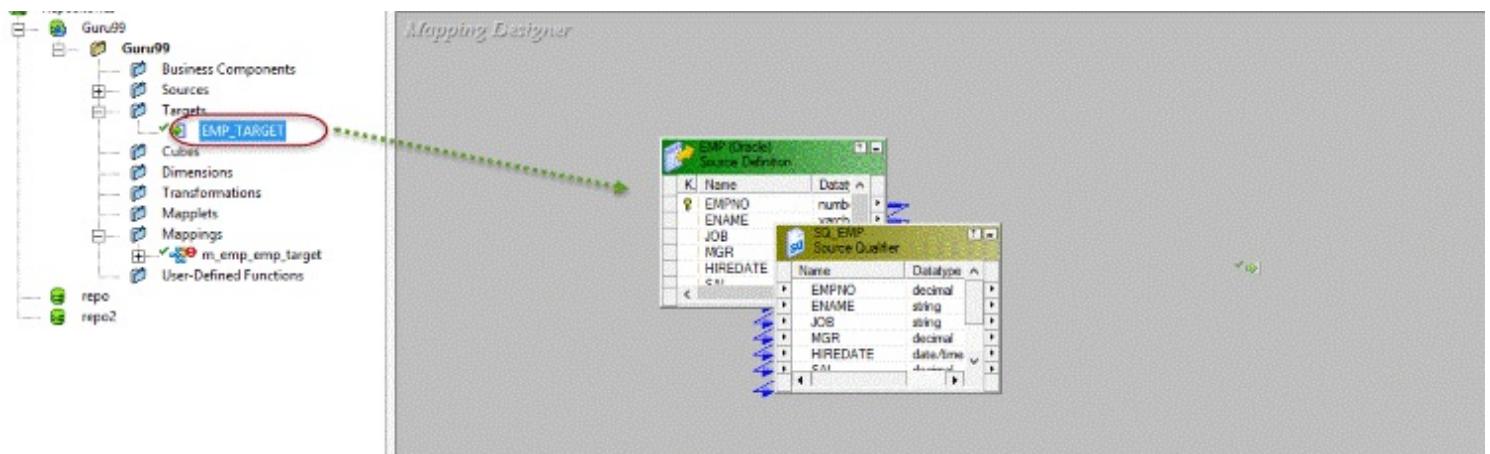
Note – When you import any relational (database) table in a mapping, an additional object of source qualifier type will also be created. This source qualifier transformation is necessary and helps Informatica integration service to identify the source database table and its properties. Whenever you import a source table, source qualifier transformation will also be created. You should never delete a source qualifier object in a mapping.

Mapping Designer

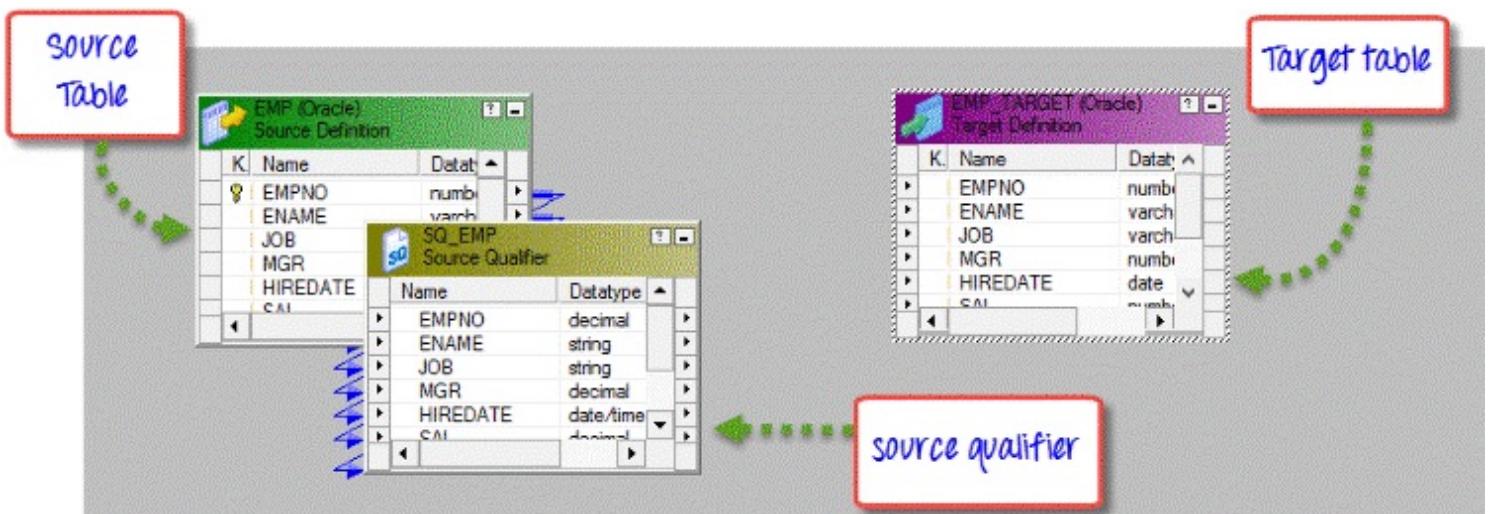


Step 5 – In this step we will,

1. Select "emp_target" source table under Targets folder.
2. Drag and drop "emp_target" table into mapping designer

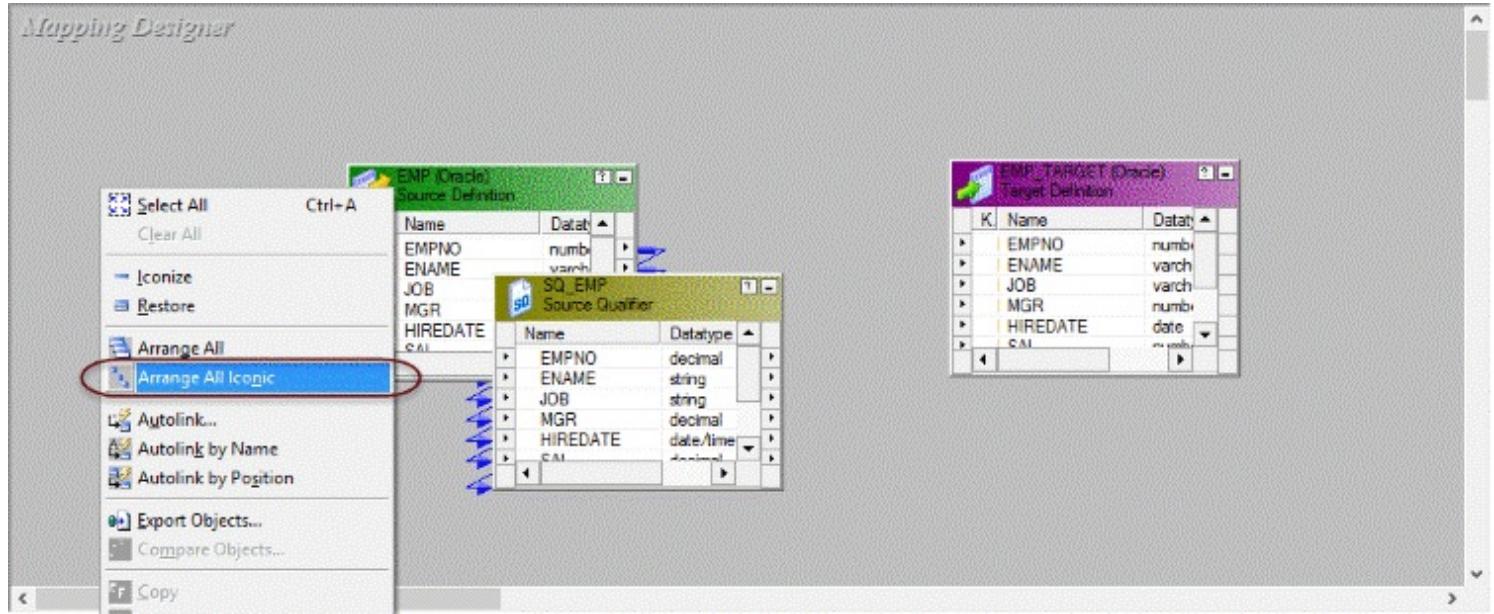


In mapping designer, "target table" will be imported and shown.

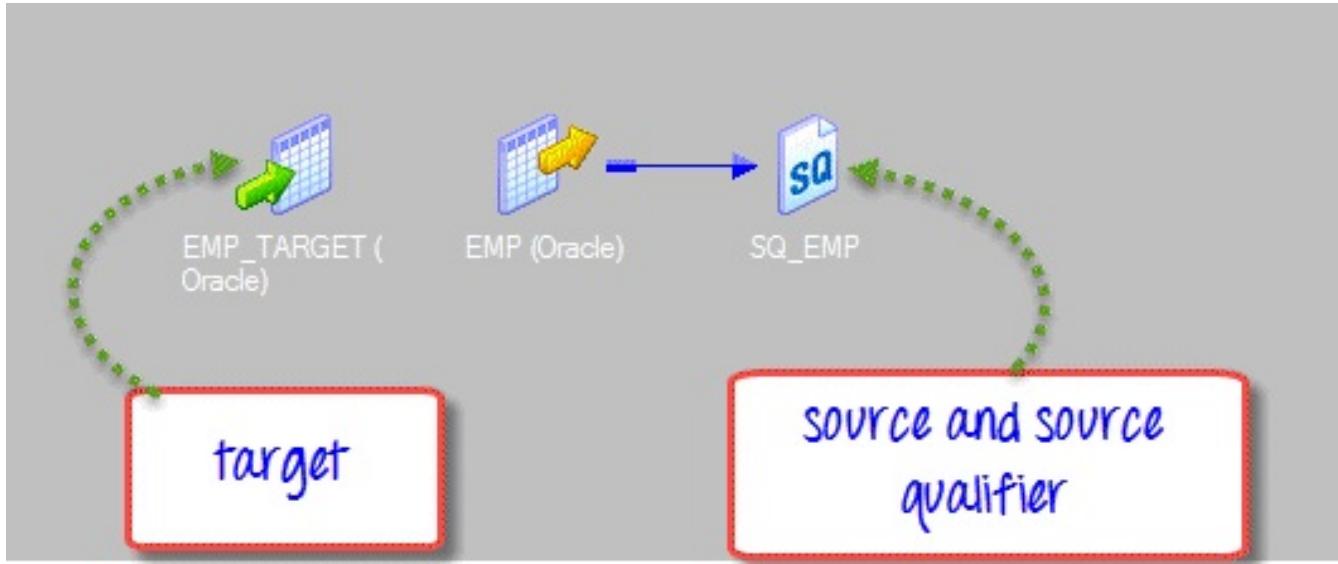


To manage the view space, you can iconize these objects in the mapping.

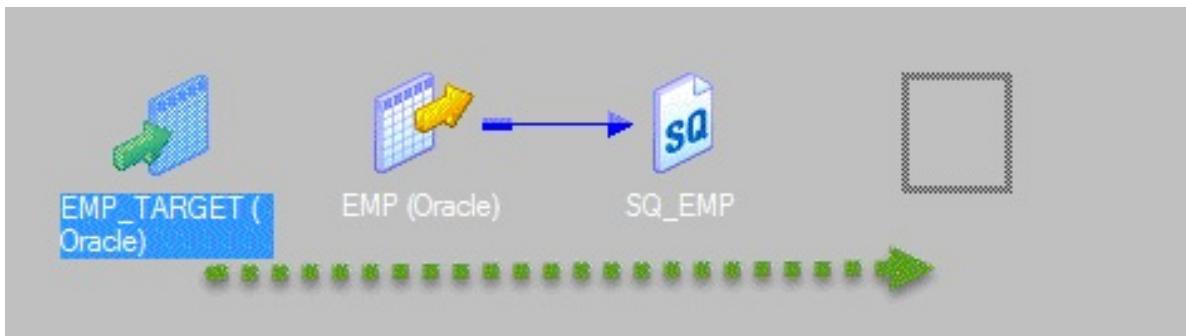
Step 6 – Right click anywhere in the mapping designer empty workspace and select option – Arrange all iconic.



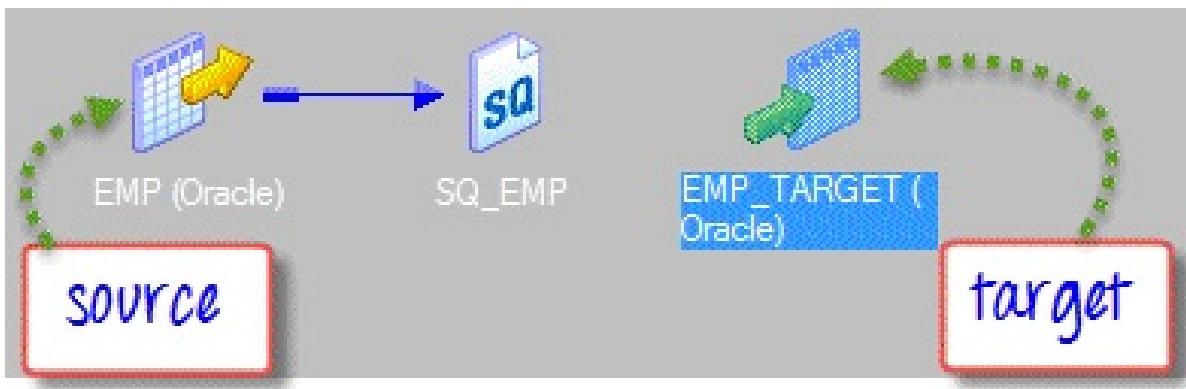
After selecting the option "Arrange all Iconic", the workspace will look like this.



Step 7 - In Informatica, We design with the flow from left to right. So, source tables should be at the left side, and target tables should be at right. To arrange tables in our workspace, Select the "emp_target" table then drag and drop it to the right side of the emp table.



After this rearrangement, the workspace will look like this.



Note – Periodically use "ctrl+s" shortcut to save changes to the repository.

Step 8 - Now you have source and target tables in your mapping, but the mapping is not yet complete. The source and target tables should be linked to complete a mapping.

To Link source and targets

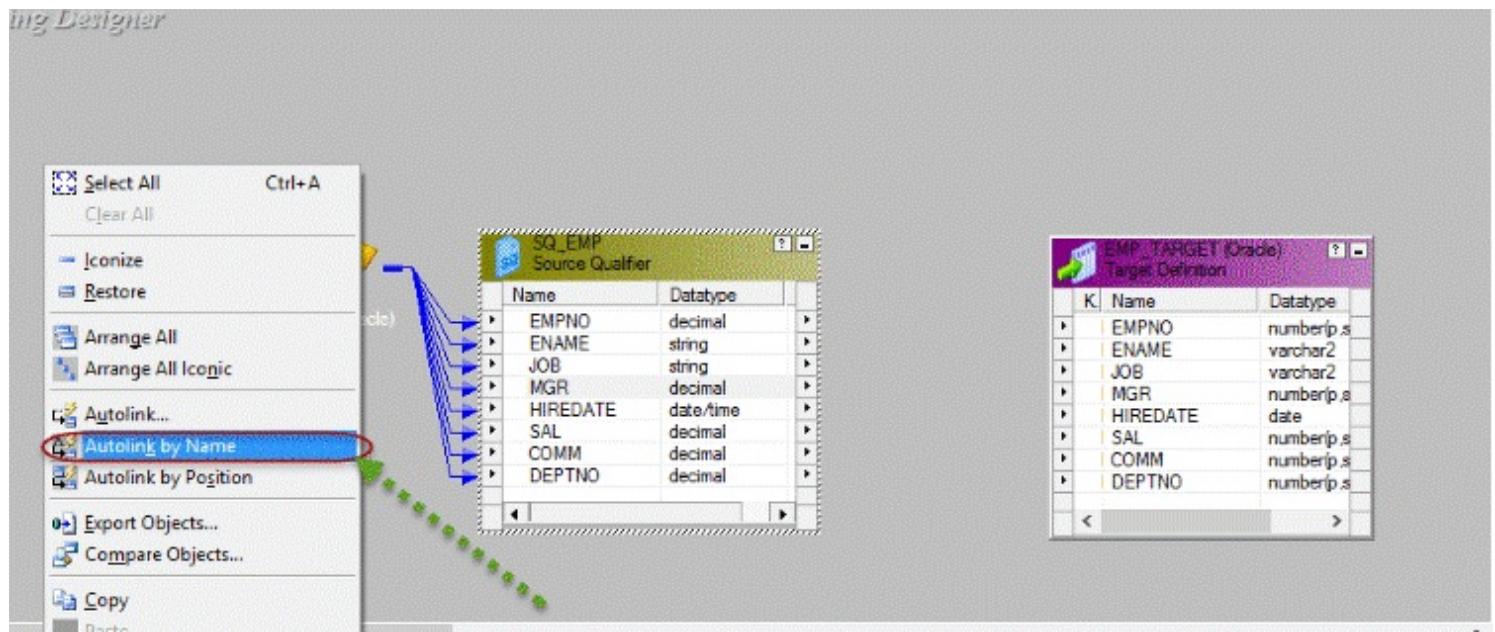
Double click on SQ_EMP table and EMP_TARGET table to change the view of them from iconic to the graphic.

Double click on objects to turn them from iconic view to graphic view

Name	Datatype
EMPNO	decimal
ENAME	string
JOB	string
MGR	decimal
HIREDATE	date/time
SAL	decimal
COMM	decimal
DEPTNO	decimal

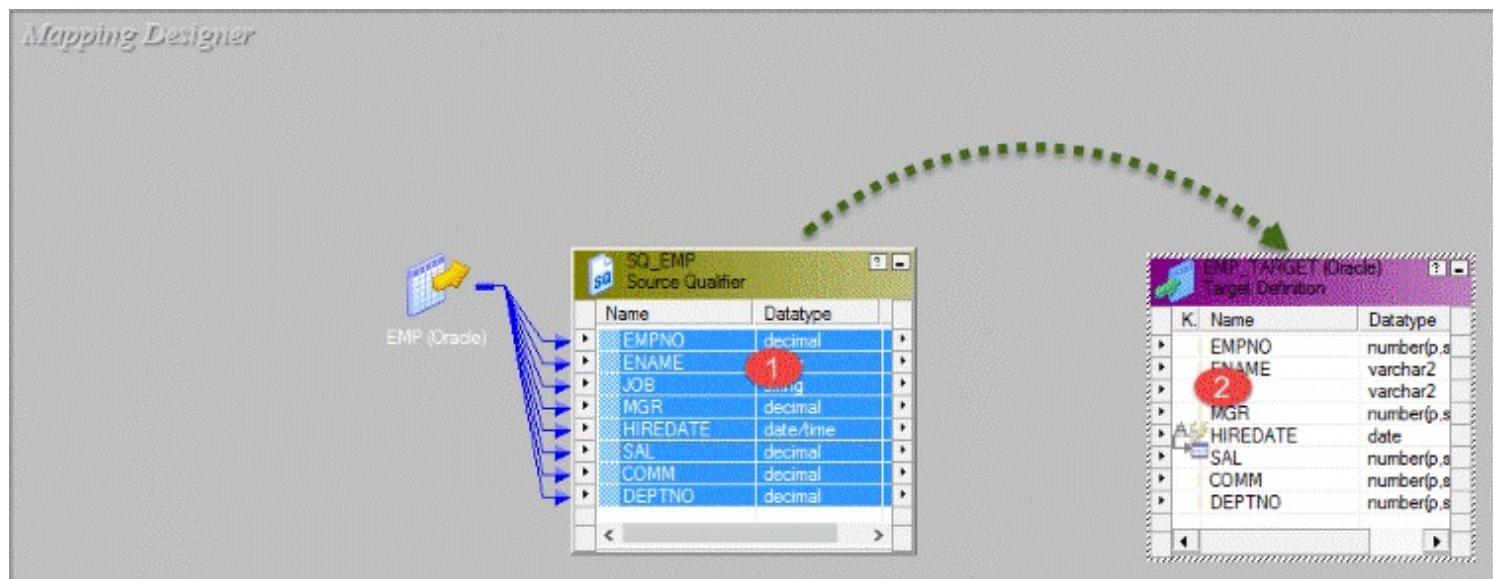
K.	Name	Datatype
	EMPNO	number(p,s)
	ENAME	varchar2
	JOB	varchar2
	MGR	number(p,s)
	HIREDATE	date
	SAL	number(p,s)
	COMM	number(p,s)
	DEPTNO	number(p,s)

Step 9 – Right Click on mapping designer workspace and select "Autolink" by name option.



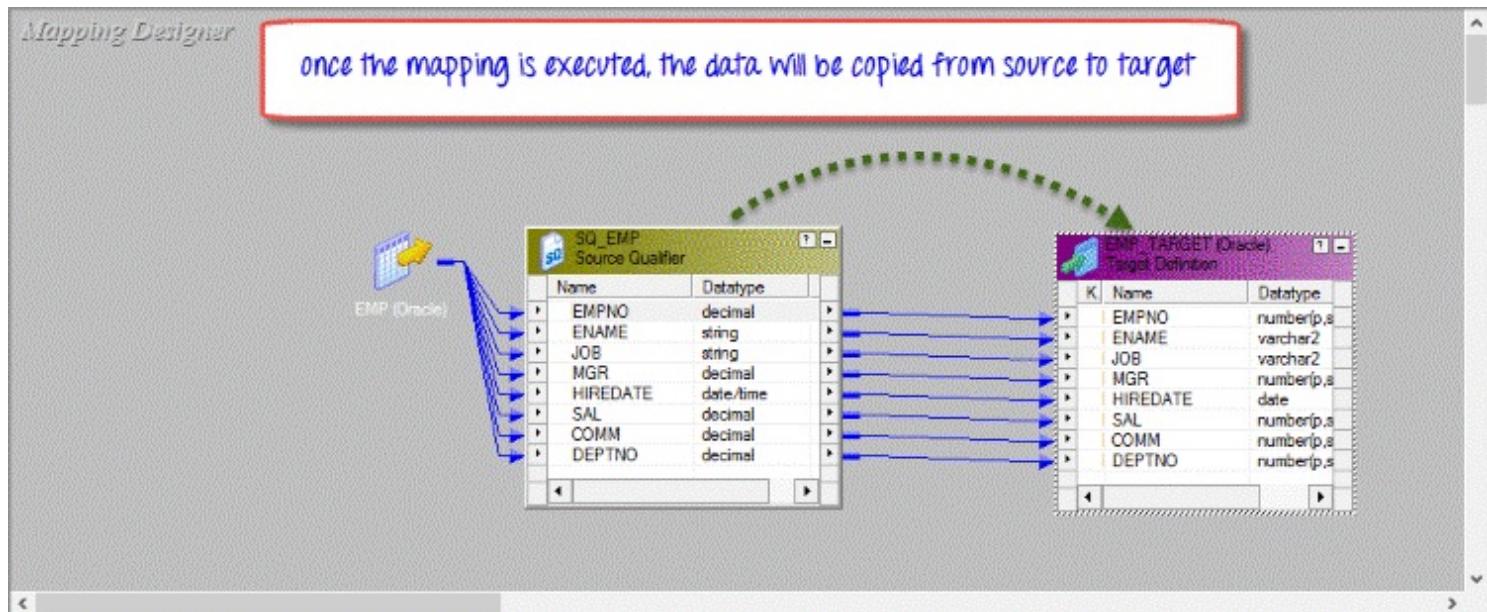
Step 10 – To link source with target table

1. Select Source table columns.
2. Drag and drop columns to the target table.



The Source and the Target tables will be linked, and connecting arrows will appear from source to the target table.

once the mapping is executed, the data will be copied from source to target



Note – Here you have linked all source columns to the respective target table column. It means, for every source record fetched, all the columns of the target will get loaded. If you want to exclude any specific column from getting loaded, click on that column link and press delete key from the keyboard. The link will get removed, and the target column will not get loaded.

Step 11 - Use shortcut "ctrl+s" to save changes to your mapping. On the output window, you can see the message of mapping validation/parsing. It confirms that your mapping is valid. Also, there will be a temporary green tick mark next to the mapping name in mapping folder tree indicating mapping is done successfully.

Repositories

- Guru99
 - Guru99
 - Business Components
 - Sources
 - Targets
 - Cubes
 - Dimensions
 - Transformations
 - Mapplets
 - Mappings
 - + m_emp_emp_target
 - User-Defined Functions
- repo
- repo2

Mapping Designer

EMP (Oracle)

...data flow validation completed with no errors.
Parsing mapping m_emp_emp_target...
...parsing completed with no errors.
***** Mapping m_emp_emp_target is VALID *****
mapping m_emp_emp_target updated.

In mappings there can be a requirement, where we need to pass variable to the mapping or there can be a scenario where we need to calculate temporary variables and further required to store them for the next session run of the mapping. For these purposes, we create mapping parameters and variables.

Mapping Parameters and Variables

Like every programming language, Informatica has its own way of defining

parameters and variables. But unlike other programming languages, Informatica isn't a code based language. To create parameters and variables in Informatica, you have to follow the predefined syntax and navigation.

Difference between parameters and variables –

Mapping Parameters	Mapping Variables
Mapping parameters are those data types whose value once assigned remains constant throughout the mapping run. Like if you have created a mapping parameter deptno=20, then the value 20 will be constant for the whole mapping run. The parameter wherever it will be referenced will always return value 20 for that instance of mapping run. For a new mapping instance, the parameter value can be redefined.	Mapping variables are objects which can be referenced throughout the mapping run (to access their values) and their values can be reassigned. For example, a mapping variable of total_salary can be used in a mapping, and its value can be updated based on salaries.

The mapping parameters and variables are specific to a mapping only. They cannot be referenced in another mapping.

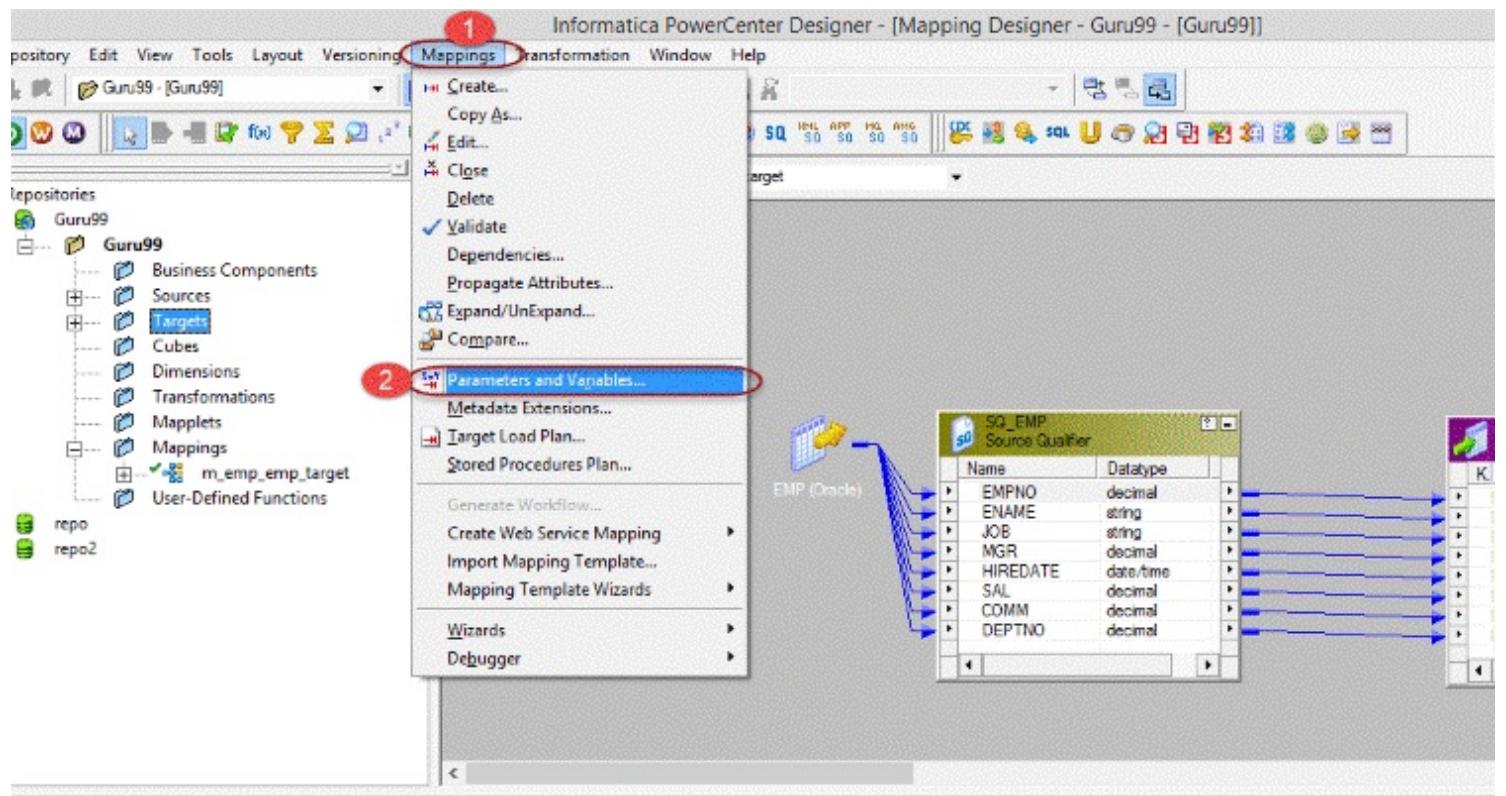
How to Create Mapping Parameter

When you create a mapping parameter, during execution of mapping Integration service looks for its assigned value. This values can be assigned to following places.

- Inside parameter file
- In pre-session variable assignment
- Initial value in repository
- Default value assigned during variable creation

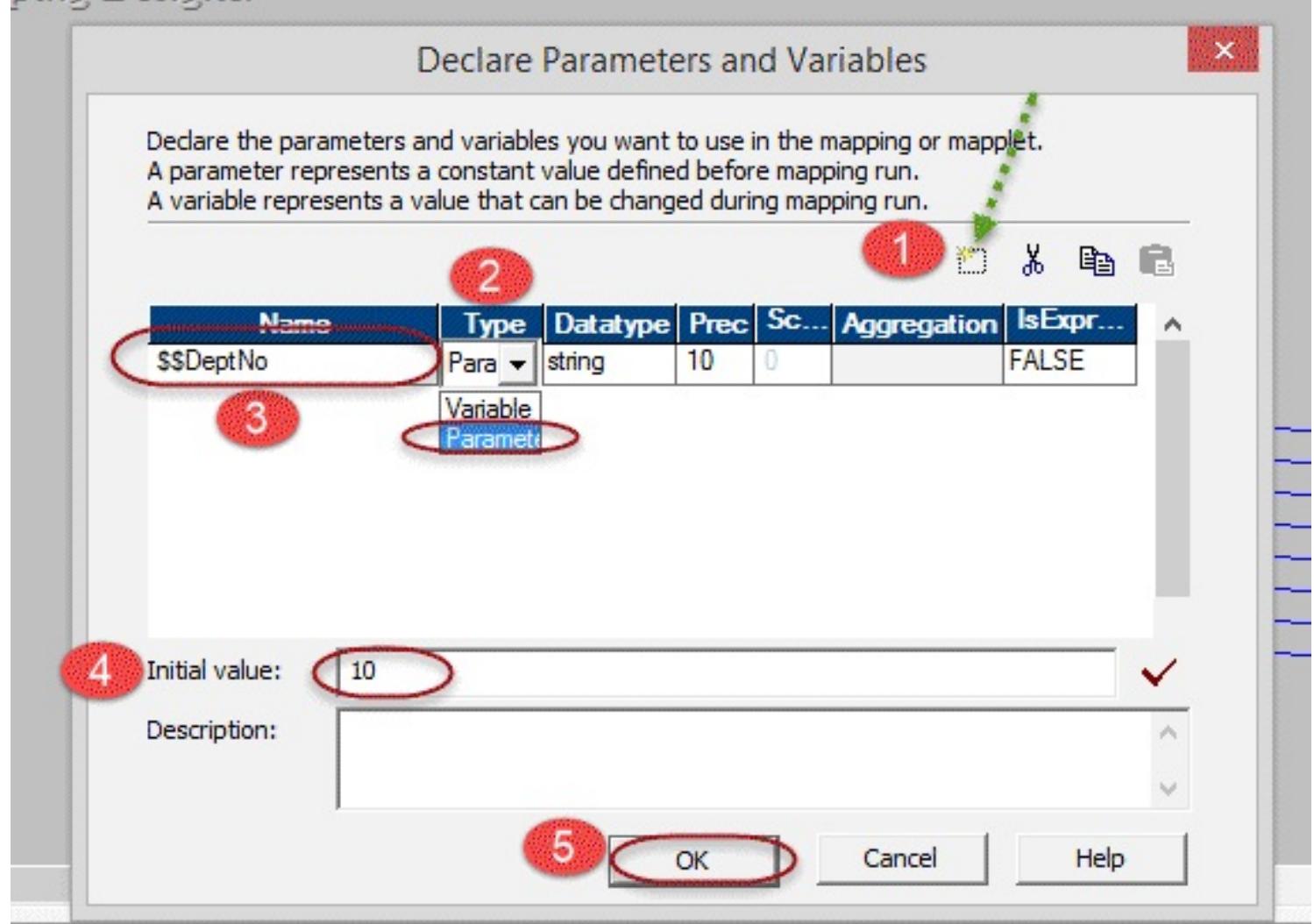
Step 1 – To Create mapping parameter - In mapping designer,

1. Select mappings menu
2. Select parameters and variables menu



Step 2- In next screen,

1. Click on add new variable menu
2. From drop down, select type as parameter
3. Enter parameter name as \$\$Deptno
4. Enter an initial value of 10
5. Select OK button

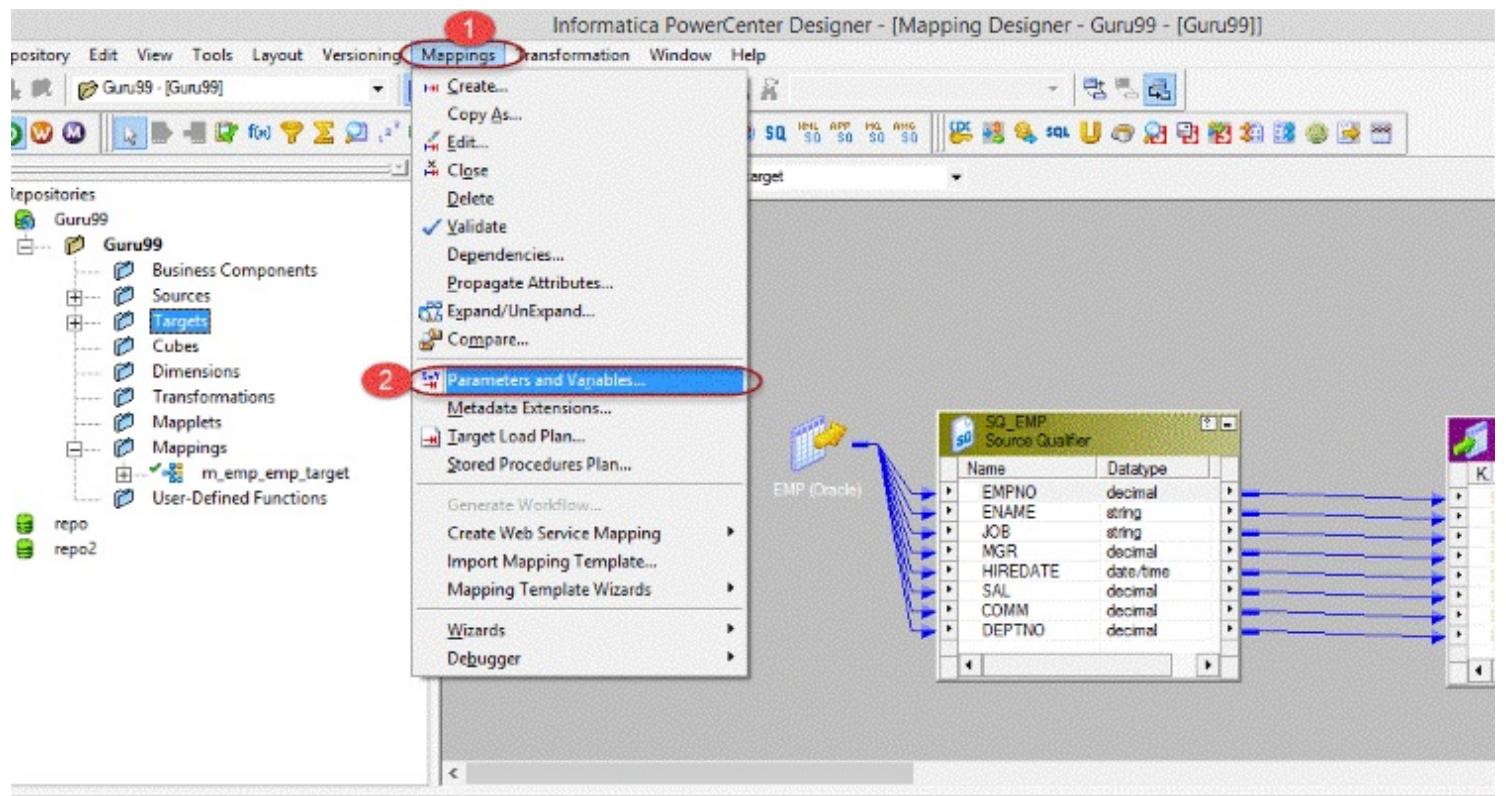


Now, you have created a mapping parameter deptno, with initially assigned value of 10, and this parameter can be referenced inside the mapping.

How to Create Mapping Variable

Step 1 – In mapping designer

1. Select mappings menu
2. Select parameters and variables menu



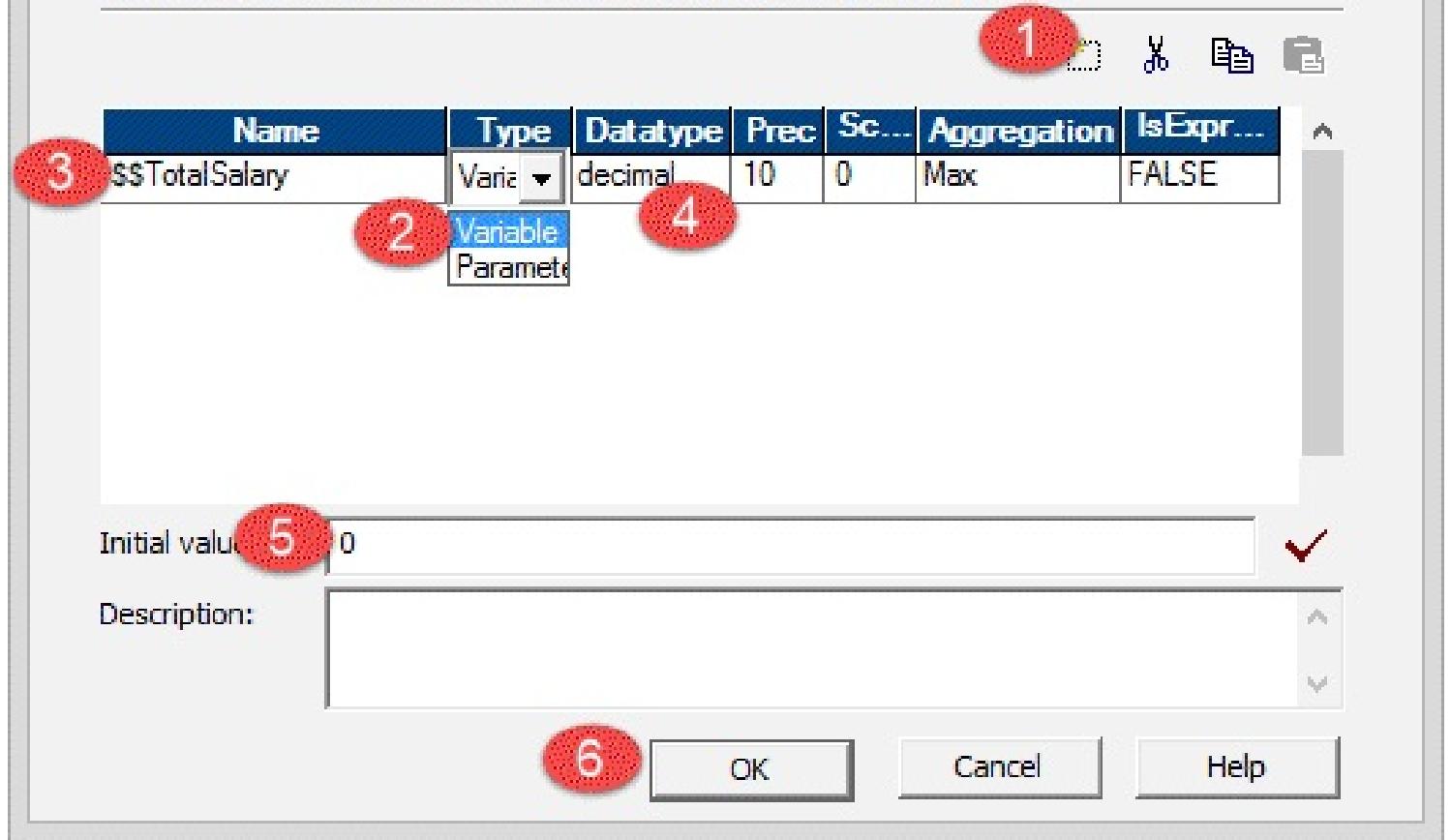
Step 2- On the next screen

1. Click on add new variable menu
2. From drop down, select type as variable
3. Enter variable name as \$\$TotalSalary
4. Select DataType as decimal
5. Enter an initial value of 0
6. Select OK button

Declare Parameters and Variables



Declare the parameters and variables you want to use in the mapping or mapplet.
A parameter represents a constant value defined before mapping run.
A variable represents a value that can be changed during mapping run.



This will create a mapping variable.

Note – mapping parameter and variable names always begin with \$\$.

Summary

Mappings are important in Informatica to match source data with target as per project requirements. We have discussed stage mappings and concept of Mapping Variables and Parameters.

Chapter 6: Workflows

A workflow is a group of instructions/commands to the integrations service. The integration service is an entity which reads workflow information from the repository, fetches data from sources and after performing transformation loads it into the target.

Workflow - It defines how to run tasks like **session task**, **command task**, **email task**, etc.

To create a workflow

1. You first need to create tasks
2. And then add those tasks to the workflow.

A Workflow is like an empty container, which has the capacity to store an object you want to execute. You add tasks to the workflow that you want to execute. In this tutorial, we are going to do following things in workflow.

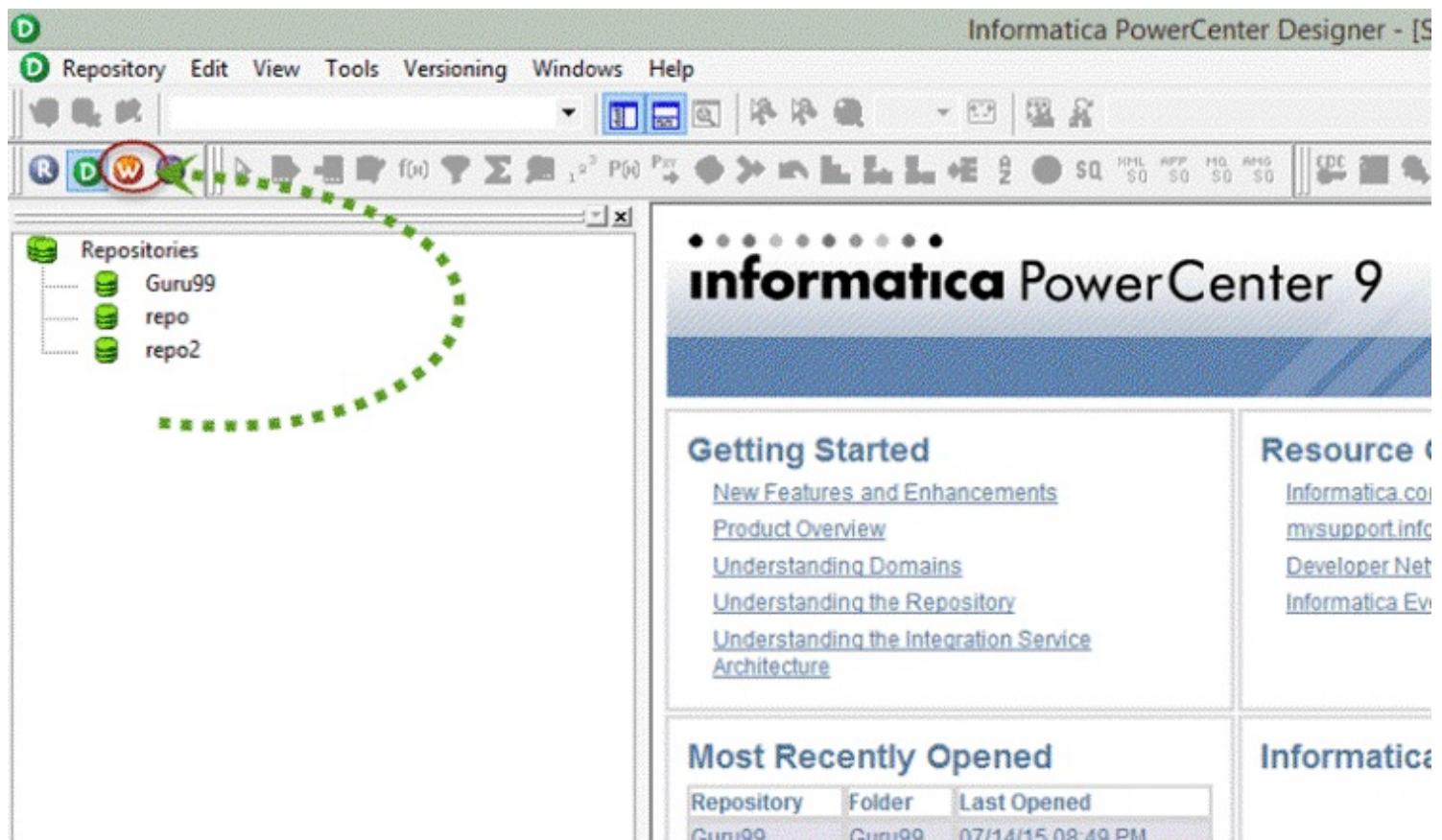


Workflow execution can be done in two ways

- **Sequence** : Tasks execute in the order in which they are defined
- **Event based** : Tasks gets executed based on the event conditions.

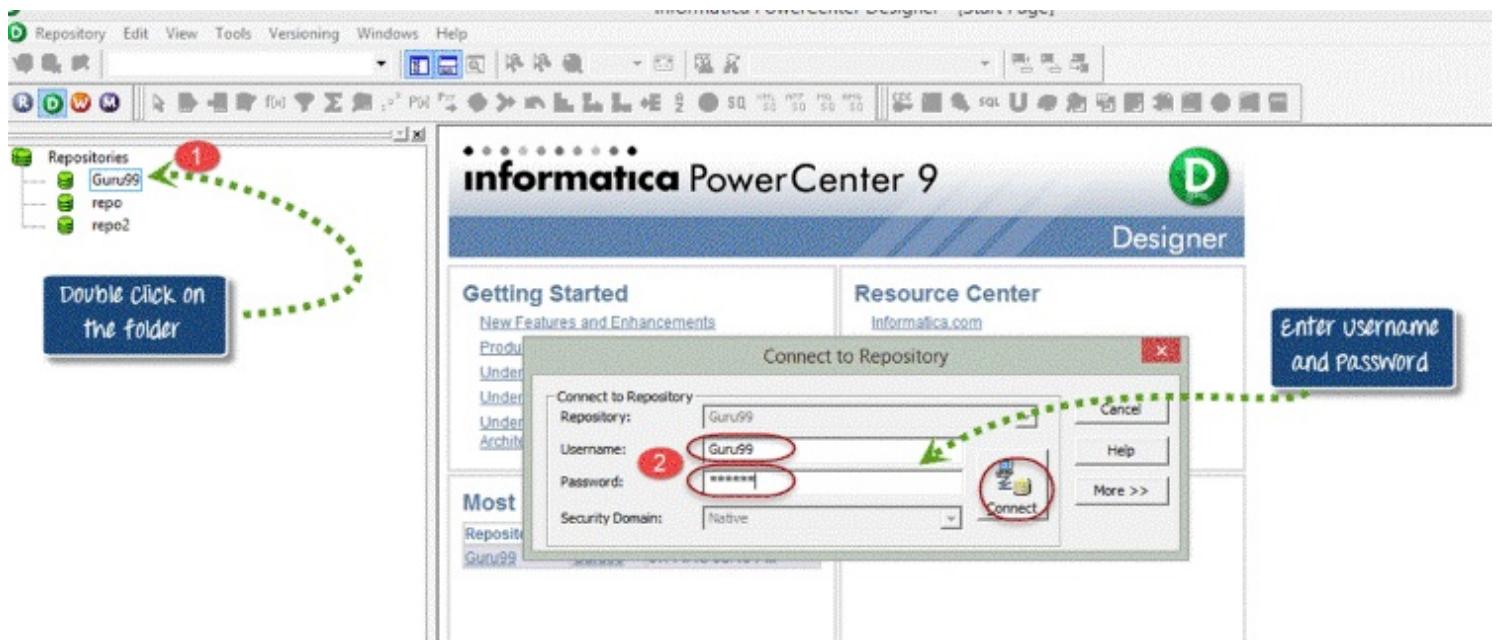
How to open Workflow Manager

Step1 – In the Informatica Designer, Click on the Workflow manager icon



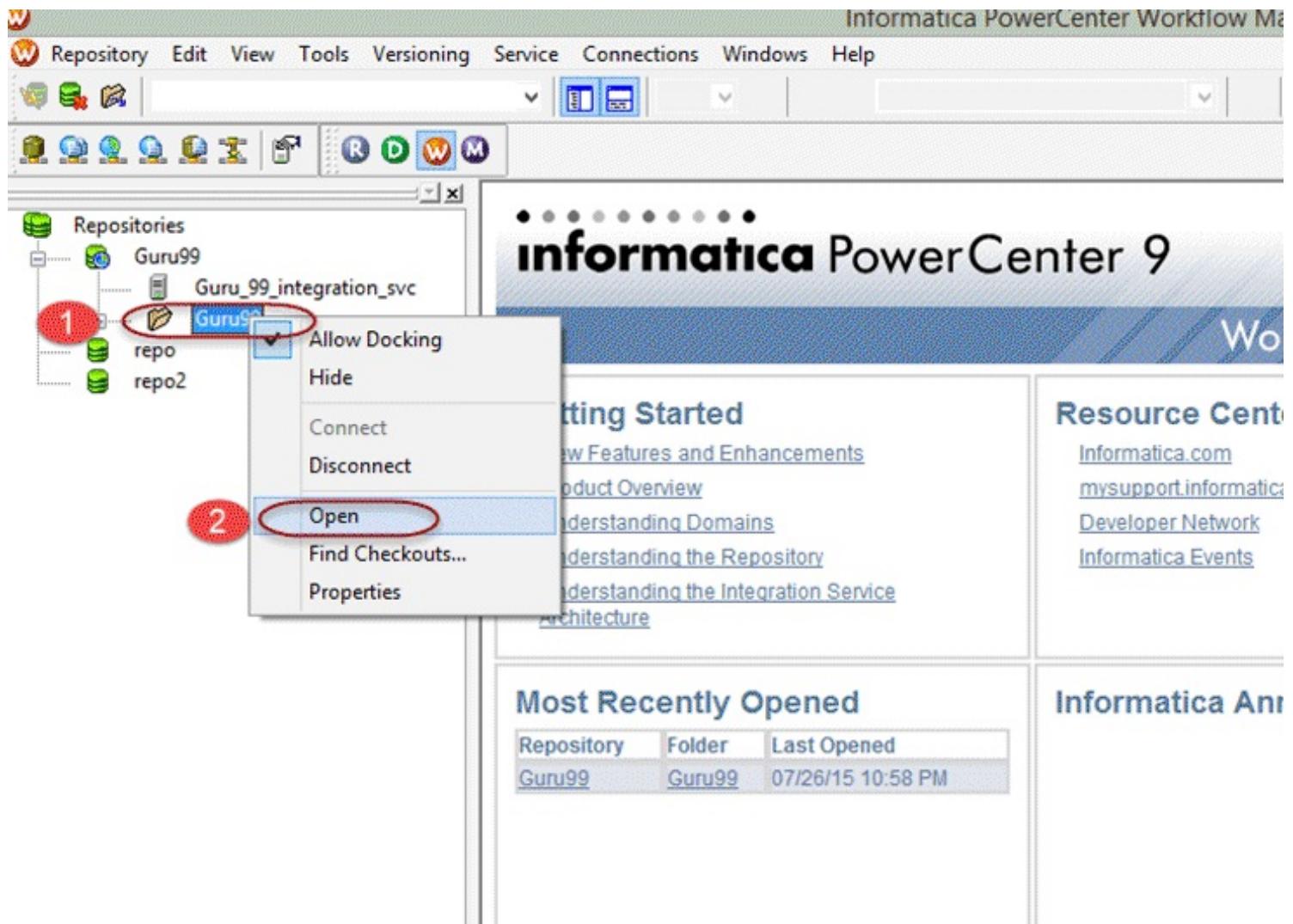
Step 2 – This will open a window of Workflow Manager. Then, in the workflow Manager.

1. We are going to connect to repository "guru99", so double click on the folder to connect.
2. Enter user name and password then select "Connect Button".

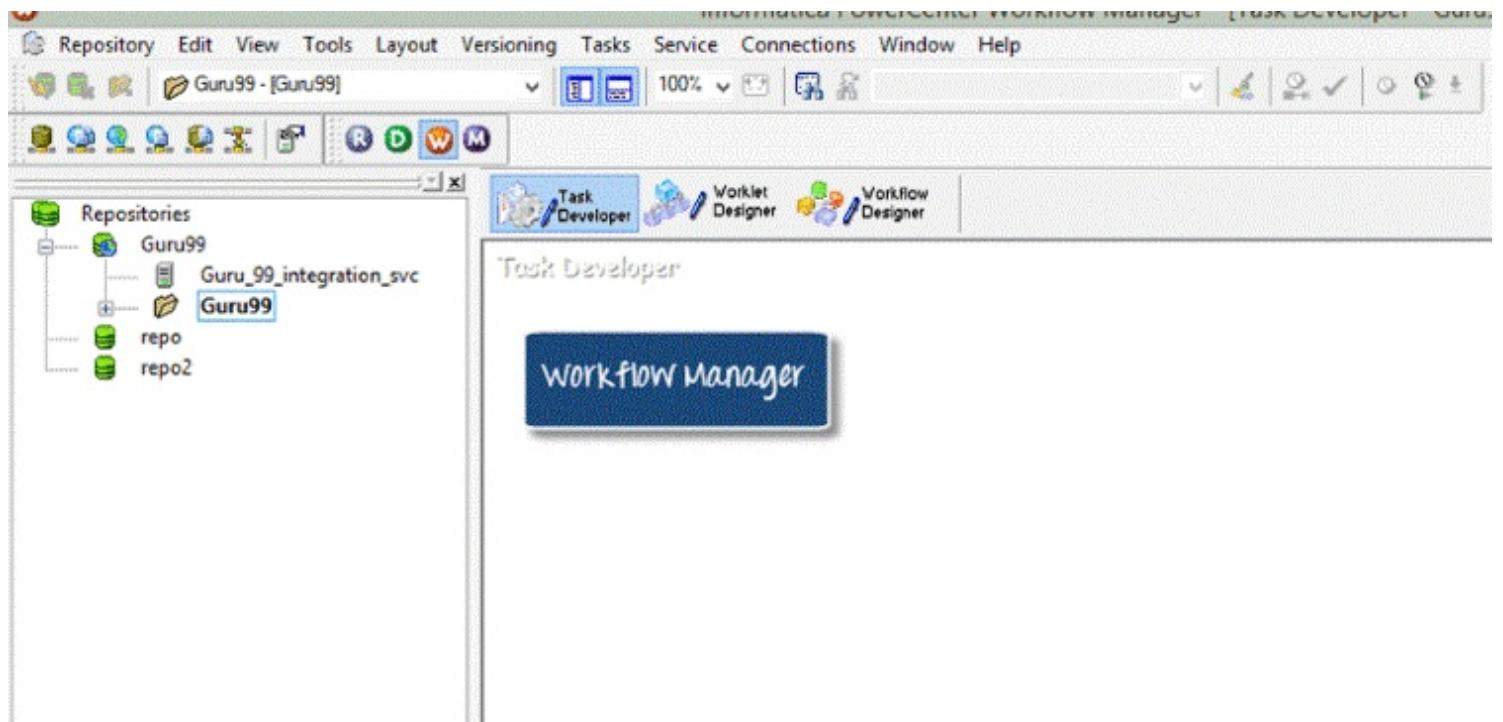


Step 3- In the workflow manager.

1. Right click on the folder
2. In the pop up menu, select open option



This will open up the workspace of Workflow manager.



How to Create Connections for Workflow Manager

To execute any task in workflow manager, you need to create **connections**. By using these connections, Integration Service connects to different objects.

For Example, in your mapping if you have source table in oracle database, then you will need oracle connection so that integration service can connect to the oracle database to fetch the source data.

Following type of connections can be created in workflow manager.

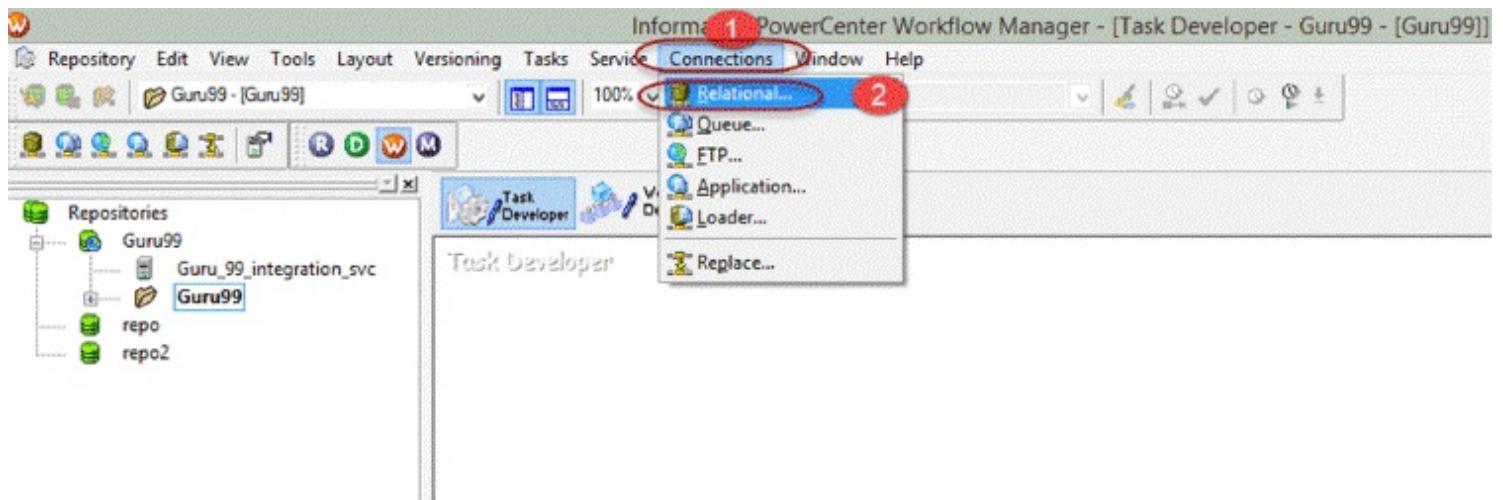
- Relational Connection
- Ftp Connection
- Queue
- Application

The choice of connection you will create, will depend on the type of source and target systems you want to connect. More often, you would be using **relational connections**.

To Create a Relational Connection

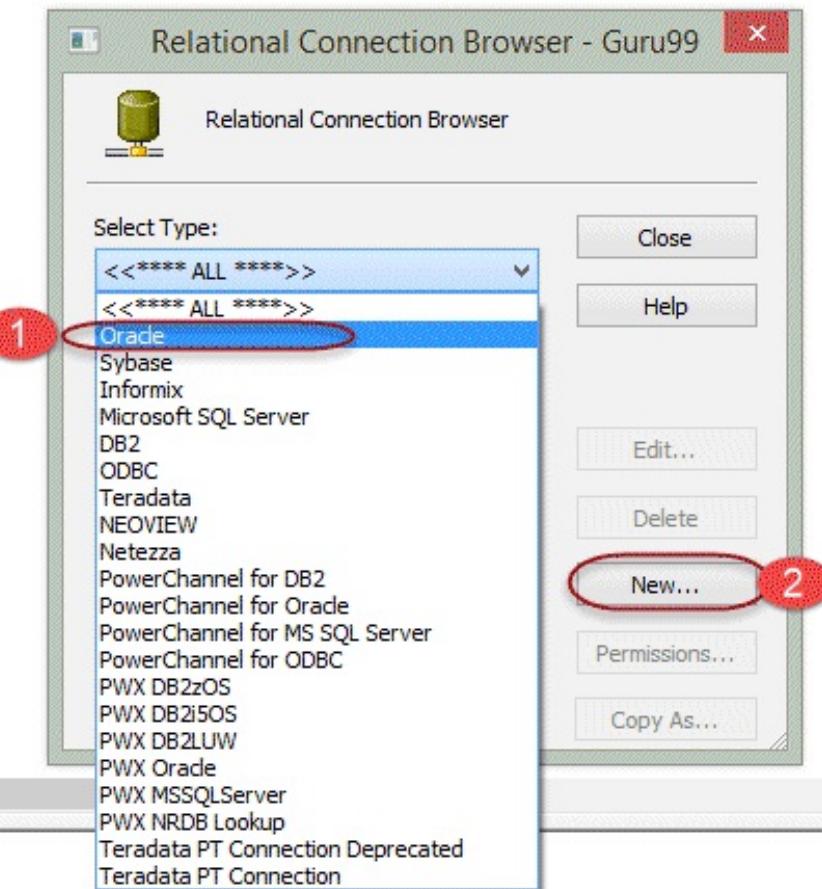
Step 1 – In Workflow Manager

1. Click on the Connection menu
2. Select Relational Option



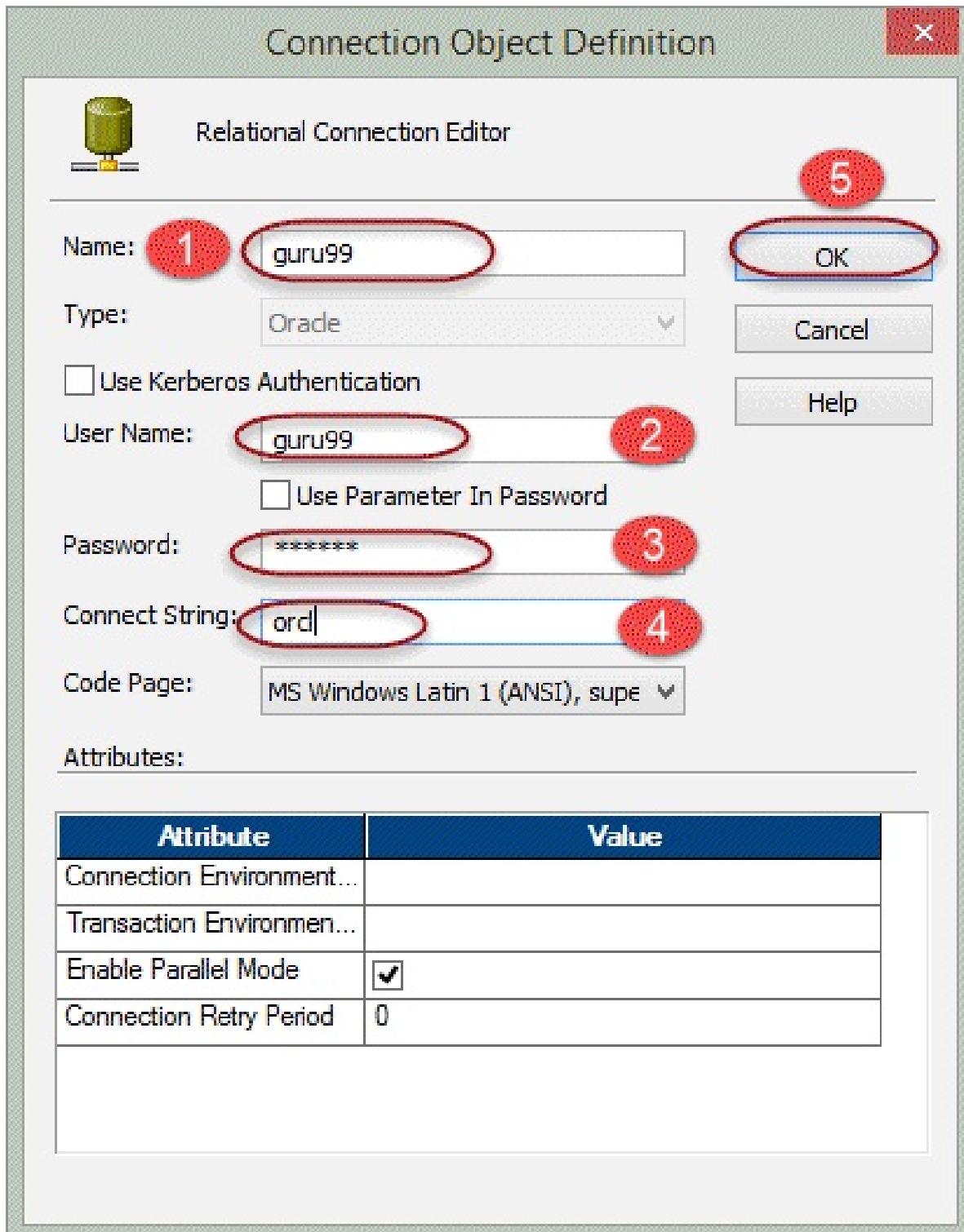
Step 2 – In the pop up window

1. Select Oracle in type
2. Click on the new button

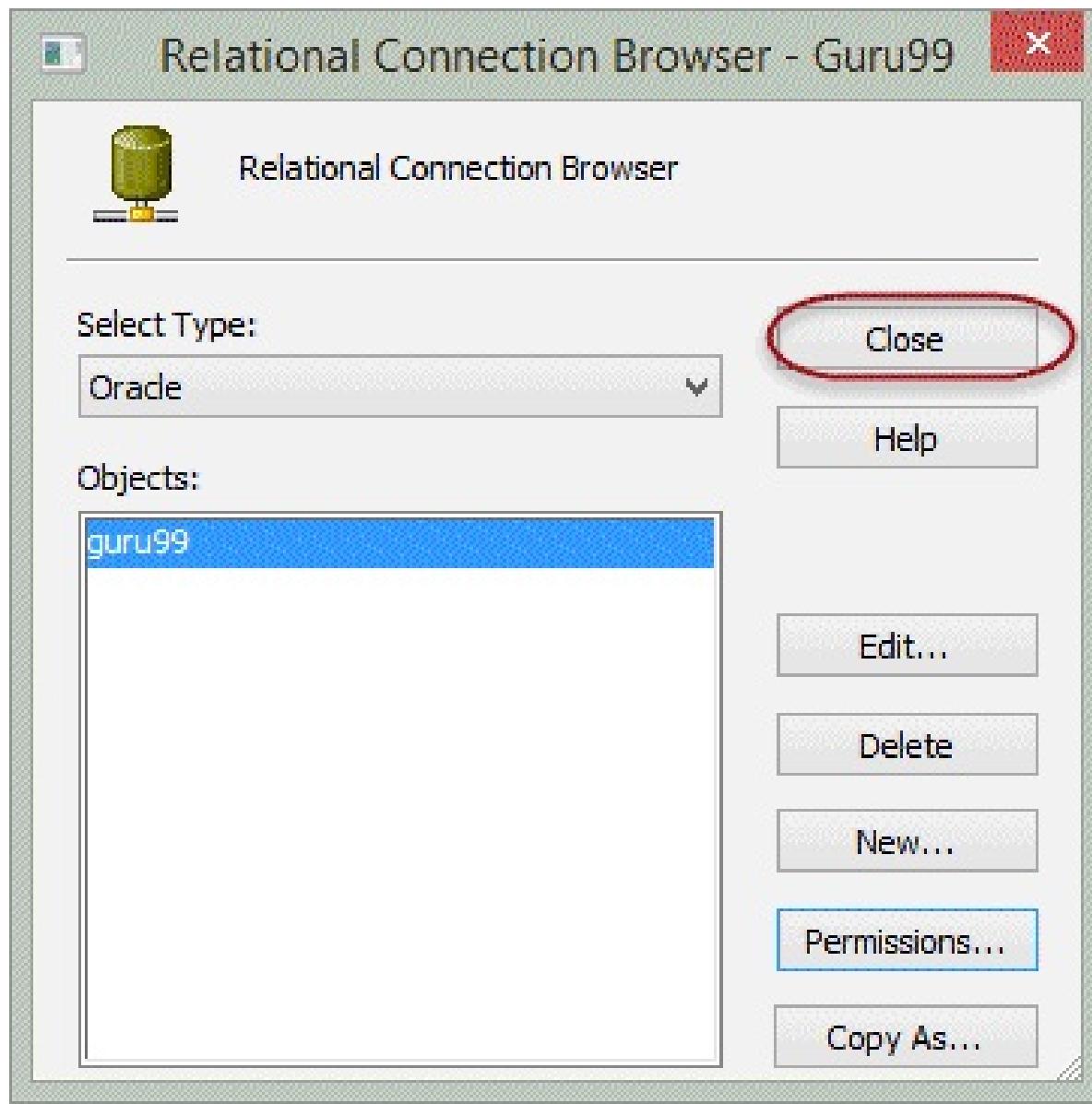


Step 3 – In the new window of connection object definition

1. Enter Connection Name (New Name-guru99)
2. Enter username
3. Enter password
4. Enter connection string
5. Leave other settings as default and Select OK button



Step 4 – You will return on the previous window. Click on the close button.

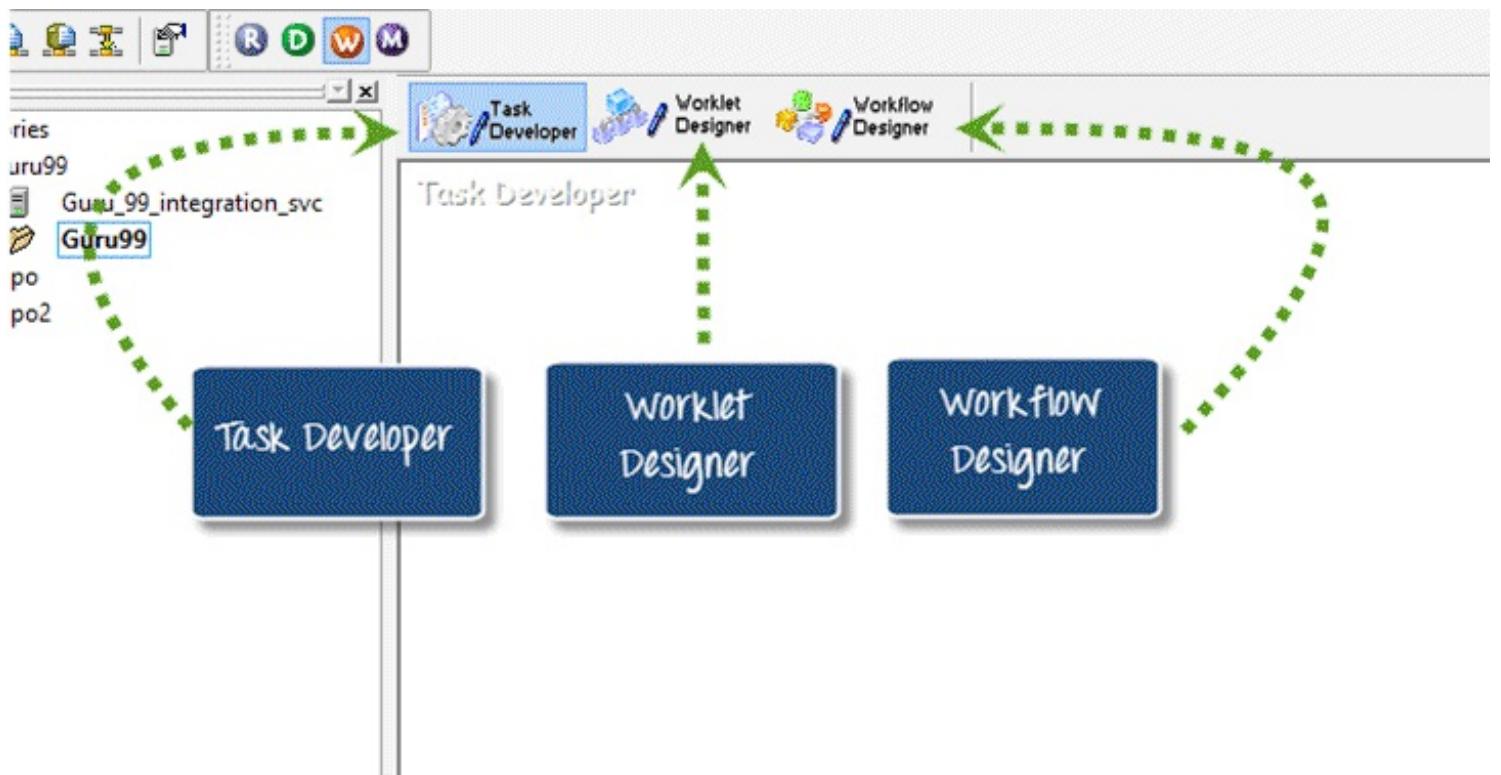


Now you are set with the relational connection in workflow manager.

Components of Workflow manager

There are three component tools of workflow manager that helps in creating various objects in workflow manager. These tools are

- Task Developer
- Worklet Designer
- Workflow Designer



Task Developer – Task developer is a tool with the help of which you can create reusable objects. Reusable object in workflow manager are objects which can be reused in multiple workflows. For Example, if you have created a command task in task developer, then you can reuse this task in any number of workflows.

The role of Workflow designer is to execute the tasks those are added in it. You can add any no of tasks in a workflow.

You can create three types of reusable tasks in task developer.

- Command task
- Session task
- Email task

Command task – A command task is used to execute different windows/unix commands during the execution of the workflow. You can create command task to execute various command based tasks. With help of this task you can execute commands **to create files/folders, to delete files/folders, to do ftp of files** etc.

Session Task - A session task in Informatica is required to run a mapping.

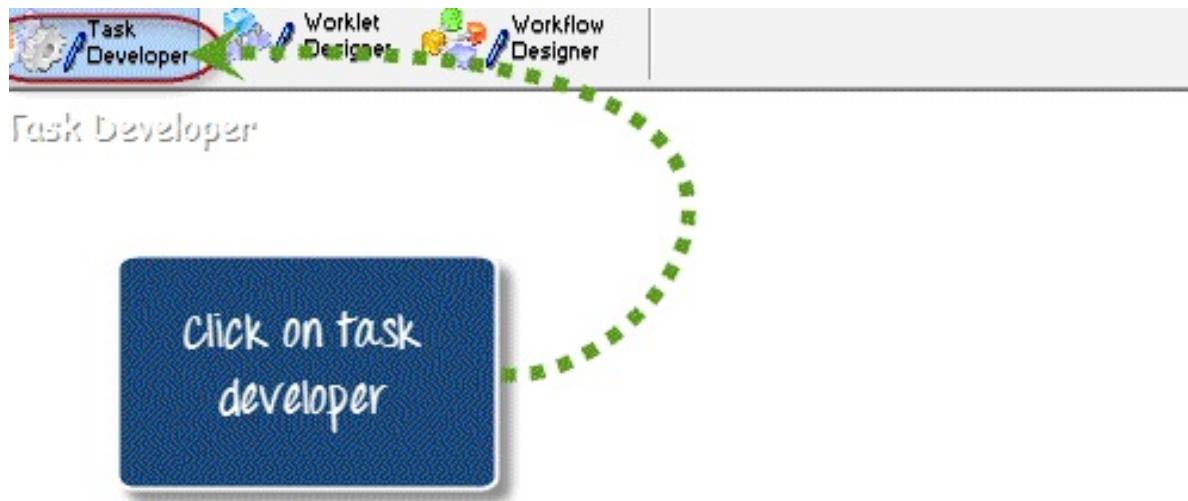
- Without a session task, you cannot execute or run a mapping
- A session task can execute only a single mapping. So, there is a one to one relationship between a mapping and a session

- A session task is an object with the help of which informatica gets to know how and where to execute a mapping and at which time
- Sessions cannot be executed independently, a session must be added to a workflow
- In session object cache properties can be configured and also advanced performance optimization configuration.

Email task - With the help of email task you can send email to defined recipients when the Integration Service runs a workflow. For example, if you want to monitor how long a session takes to complete, you can configure the session to send an email containing the details of session start and end time. Or, if you want the Integration Service to notify you when a workflow completes/fails, you can configure the email task for the same.

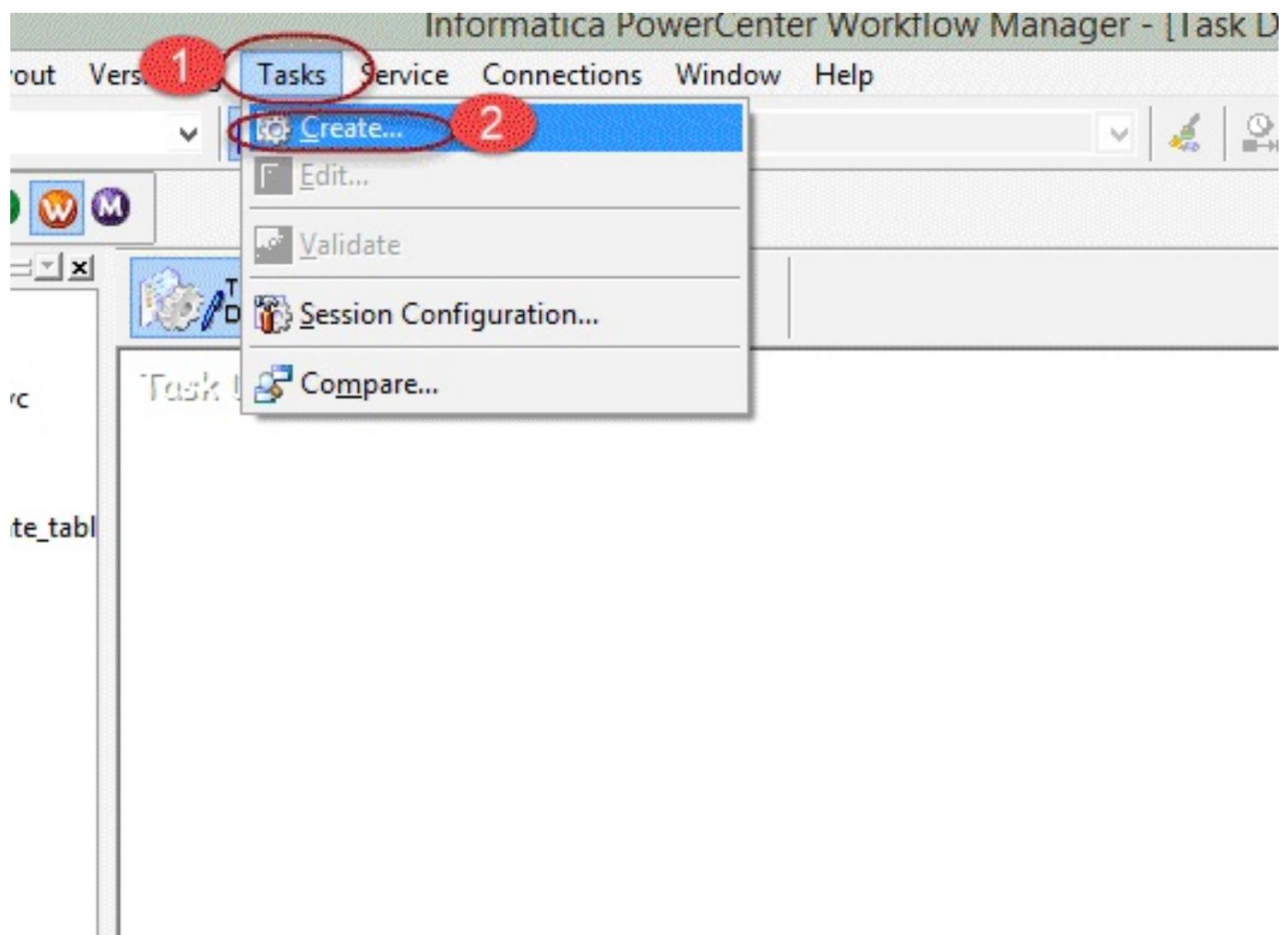
How to create command task

Step 1- To create a command task we are going to use Task Developer. In Workflow Manager, open the task developer by clicking on tab "task developer" from the menu.



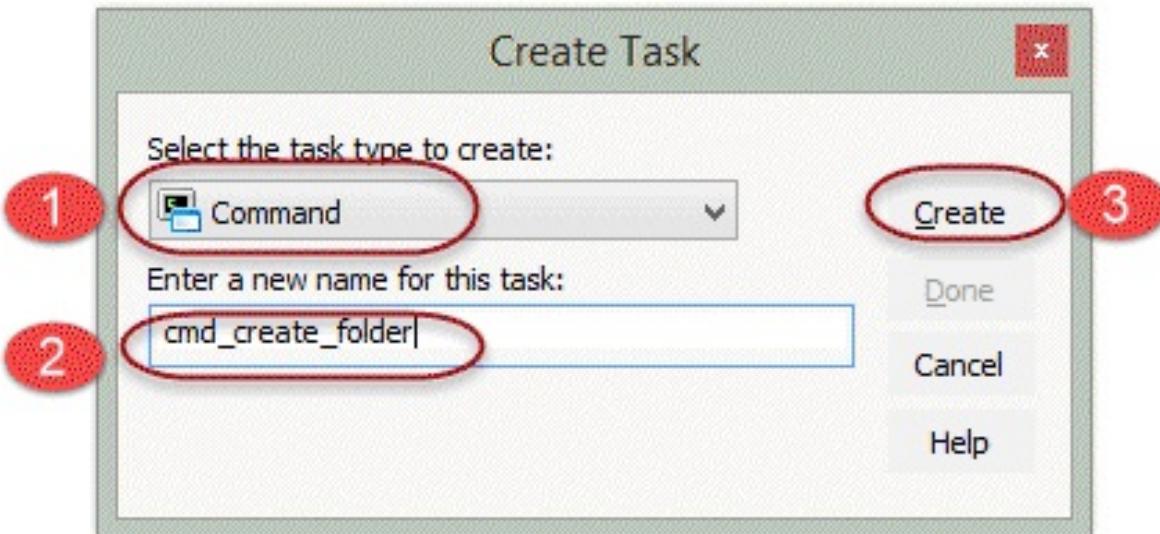
Step 2 – Once task developer is opened up, follow these steps

1. Select Tasks menu
2. Select Create option

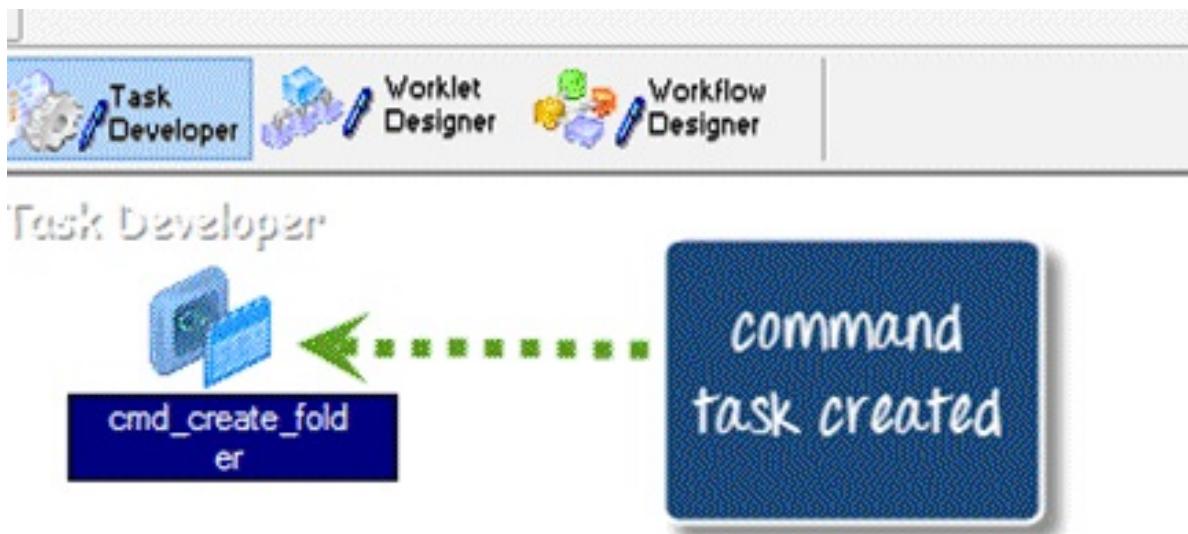


Step 3 – In the create task window

1. Select command as type of task to create
2. Enter task name
3. Select create button



This will create command task folder. Now you have to configure the task to add command in it, that we will see in next step.

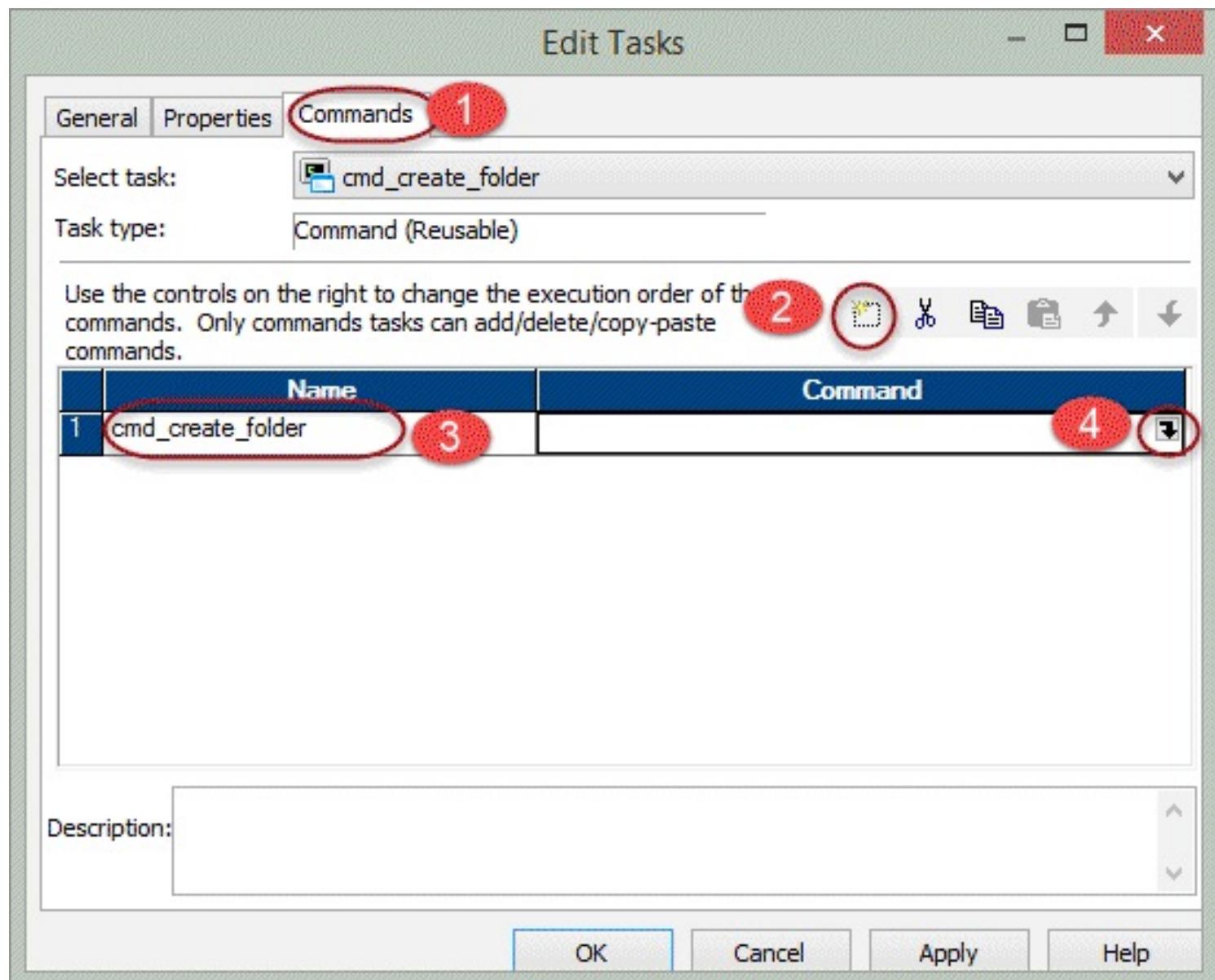


Step 4 – To configure the task, double click on the command task icon and it will open an "edit task window". On the new edit task window

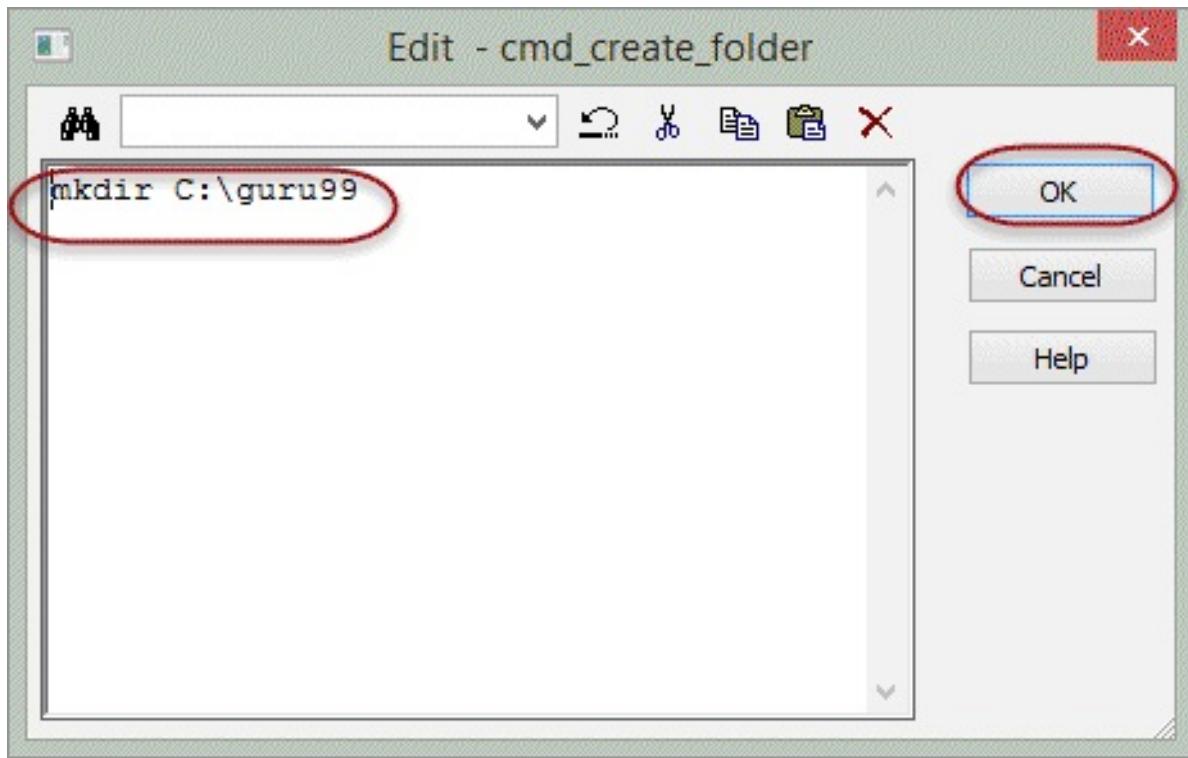
1. Select the commands menu
2. Click on the add new command icon
3. Enter command name

4. Click on the command icon to add command text

This will open a command editor box.



Step 5 – On the command editor box, enter the command "mkdir C:\guru99" (this is the windows command to create a folder named "guru99") and select OK.



Afther this step you will return to the edit tasks window and you will be able to see the command you added in to the command text box.

Step 6 – Click OK on the edit task window,

Edit Tasks

General Properties Commands

Select task: cmd_create_folder

Task type: Command (Reusable)

Use the controls on the right to change the execution order of the commands. Only command tasks can add/delete/copy-paste commands.



Name	Command
1 cmd_create_folder	mkdir C:\guru99

The entered command will appear here

Description:

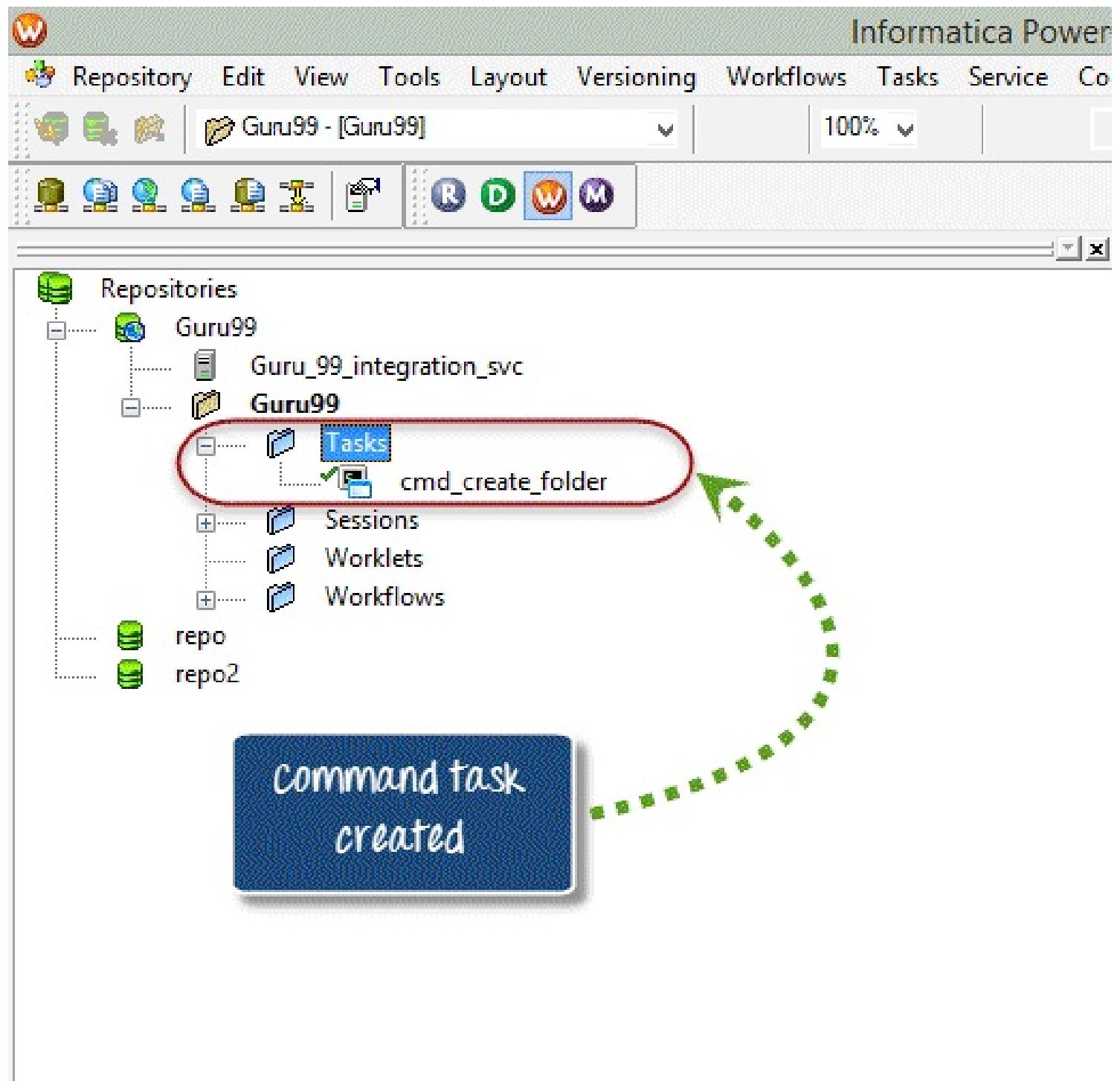
OK

Cancel

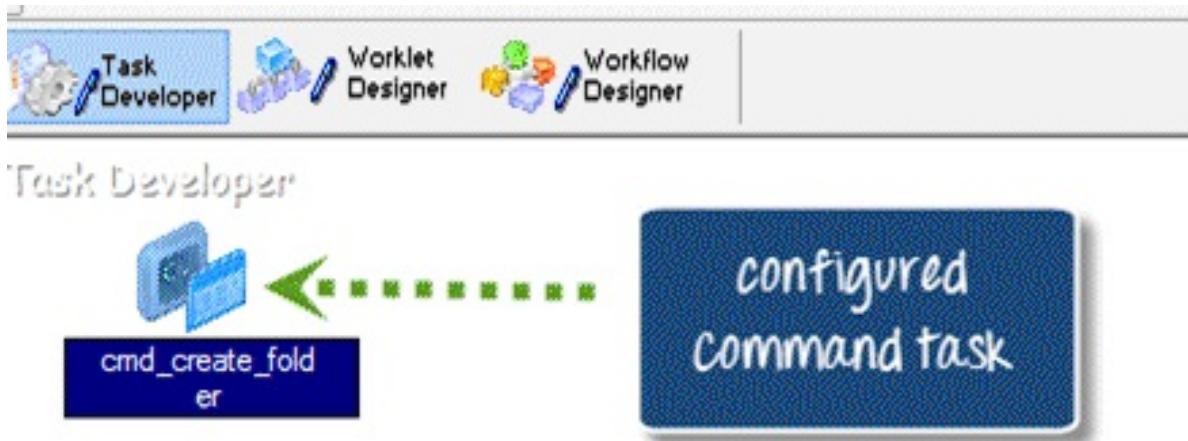
Apply

Help

The command task will be created in the task developer under "Guru99" repository.



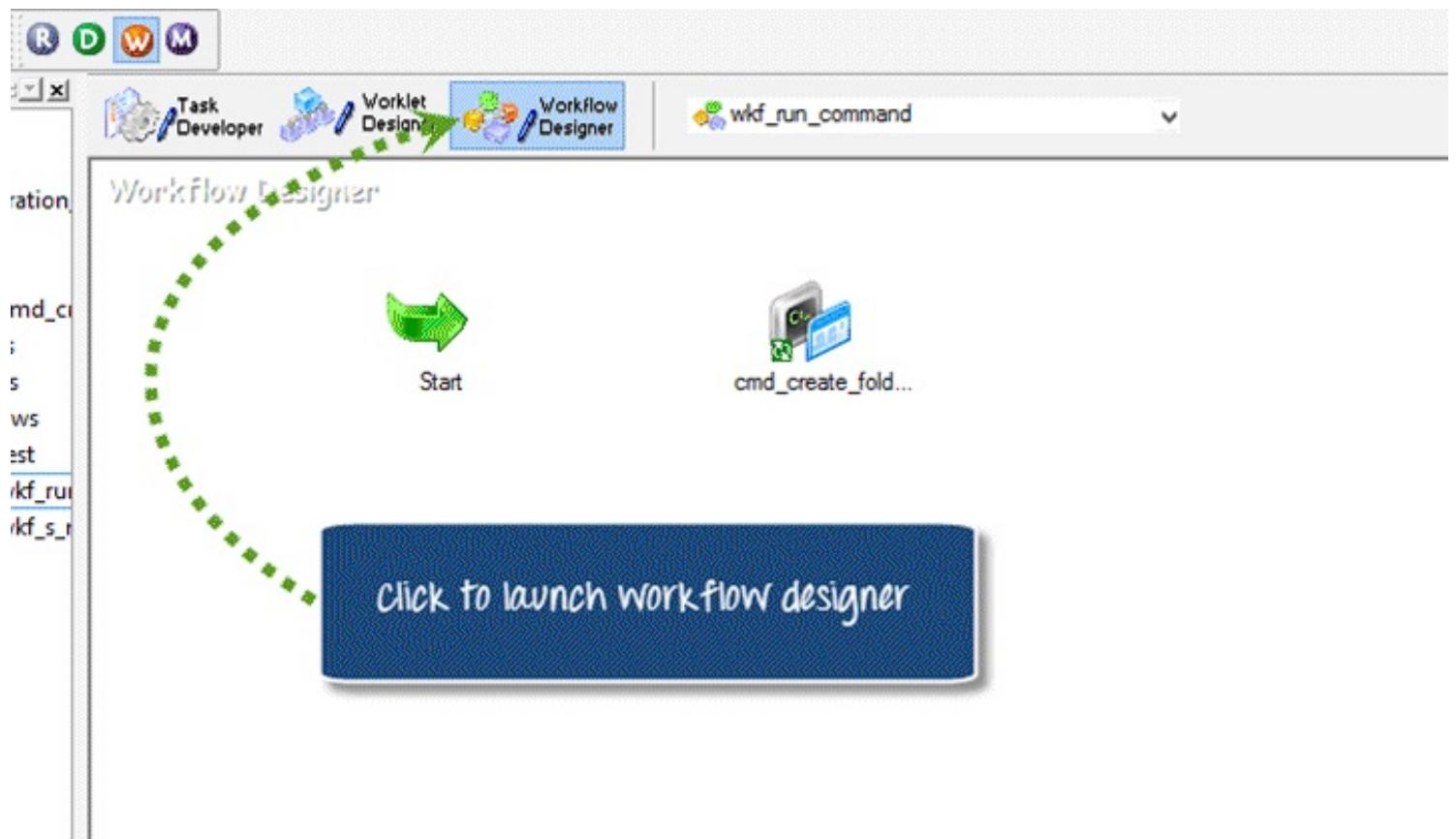
Note – use **ctrl+s** shortcut to save the changes in repository



How to create workflow to execute command task

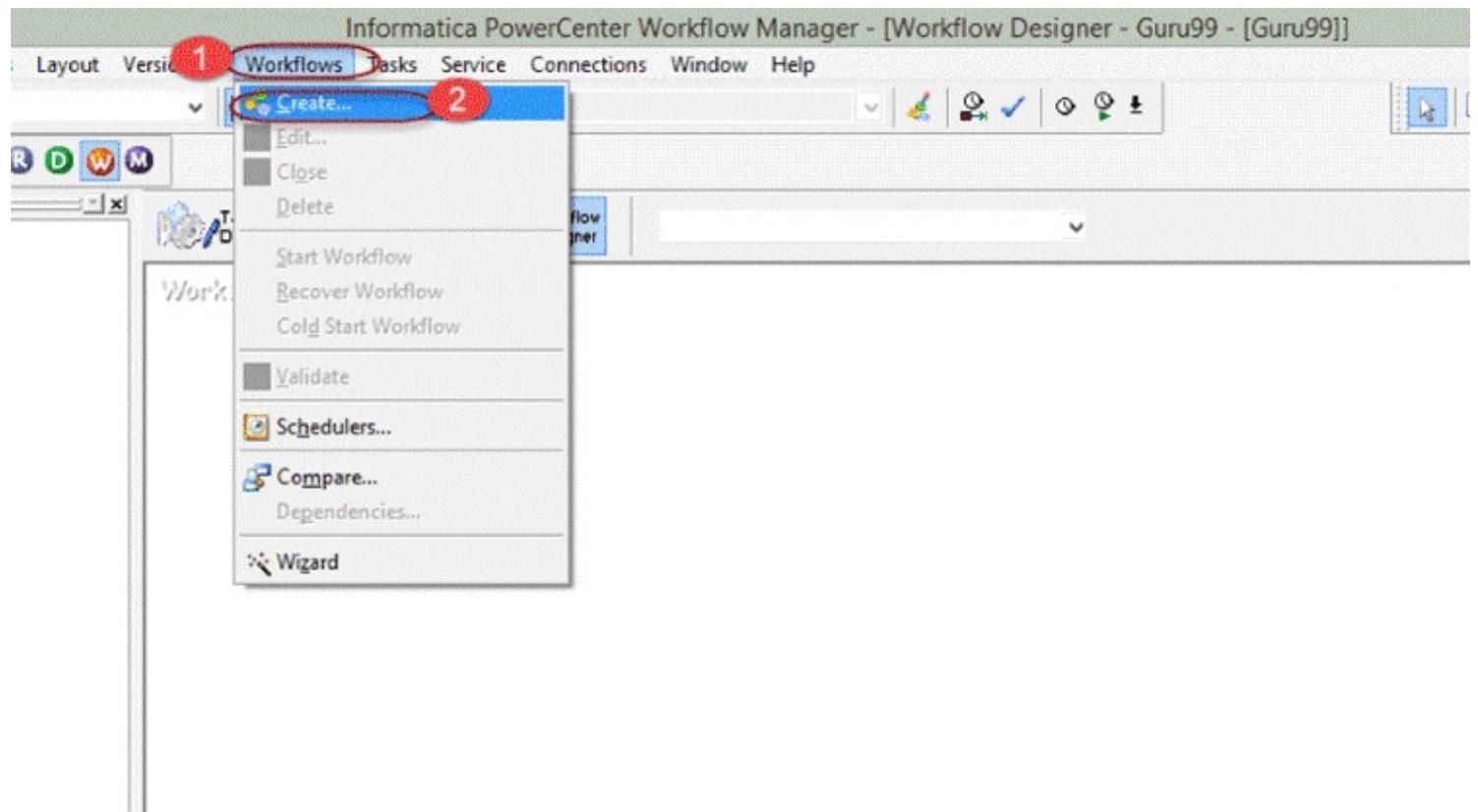
To execute command tasks you have to switch on to workflow designer. A workflow designer is a parent or container object in which you can add multiple tasks and when workflow is executed, all the added tasks will execute. To create a workflow

Step 1 – Open the workflow designer by clicking on workflow designer menu



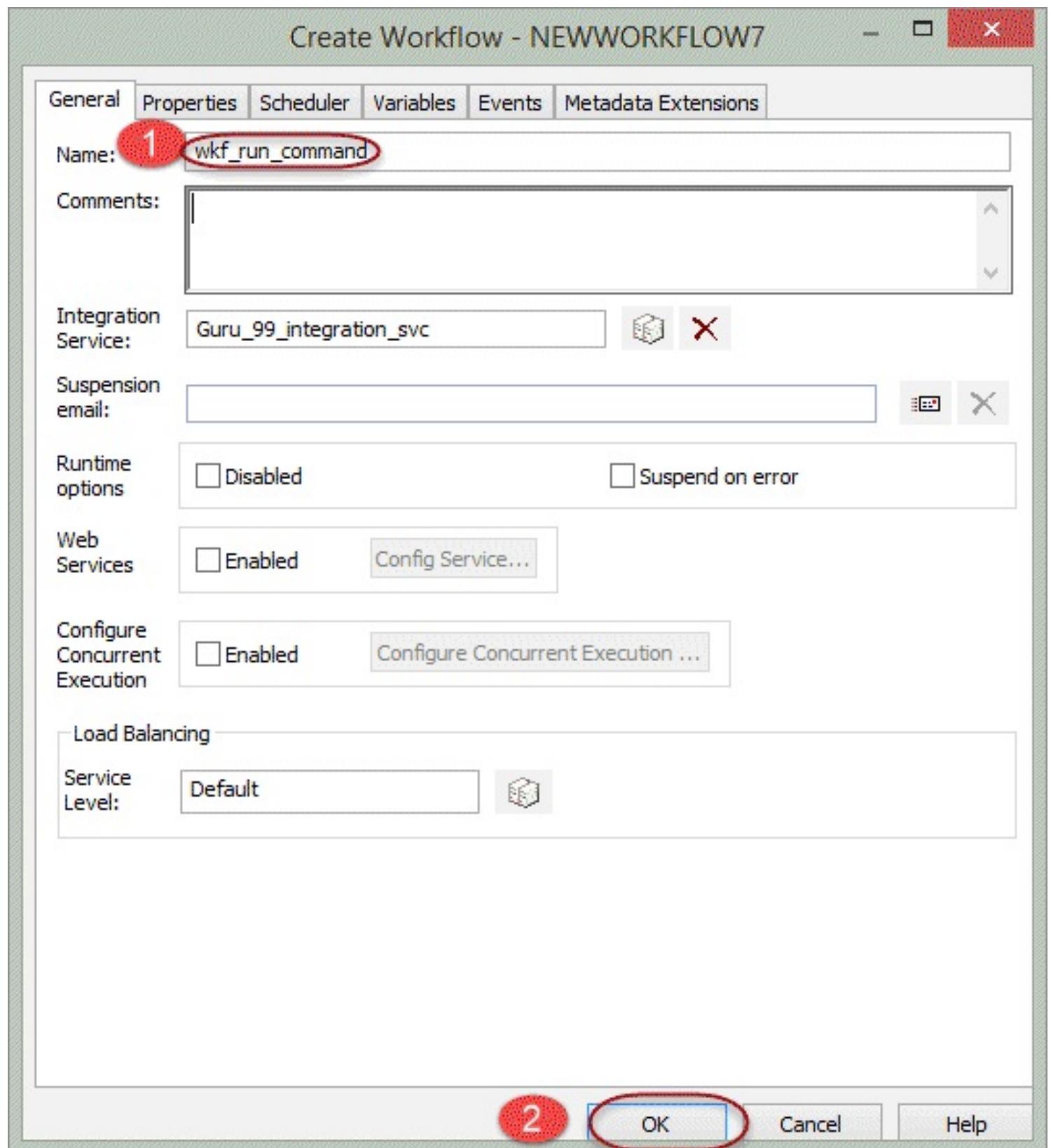
Step 2 – In workflow designer

1. Select workflows menu
2. Select create option



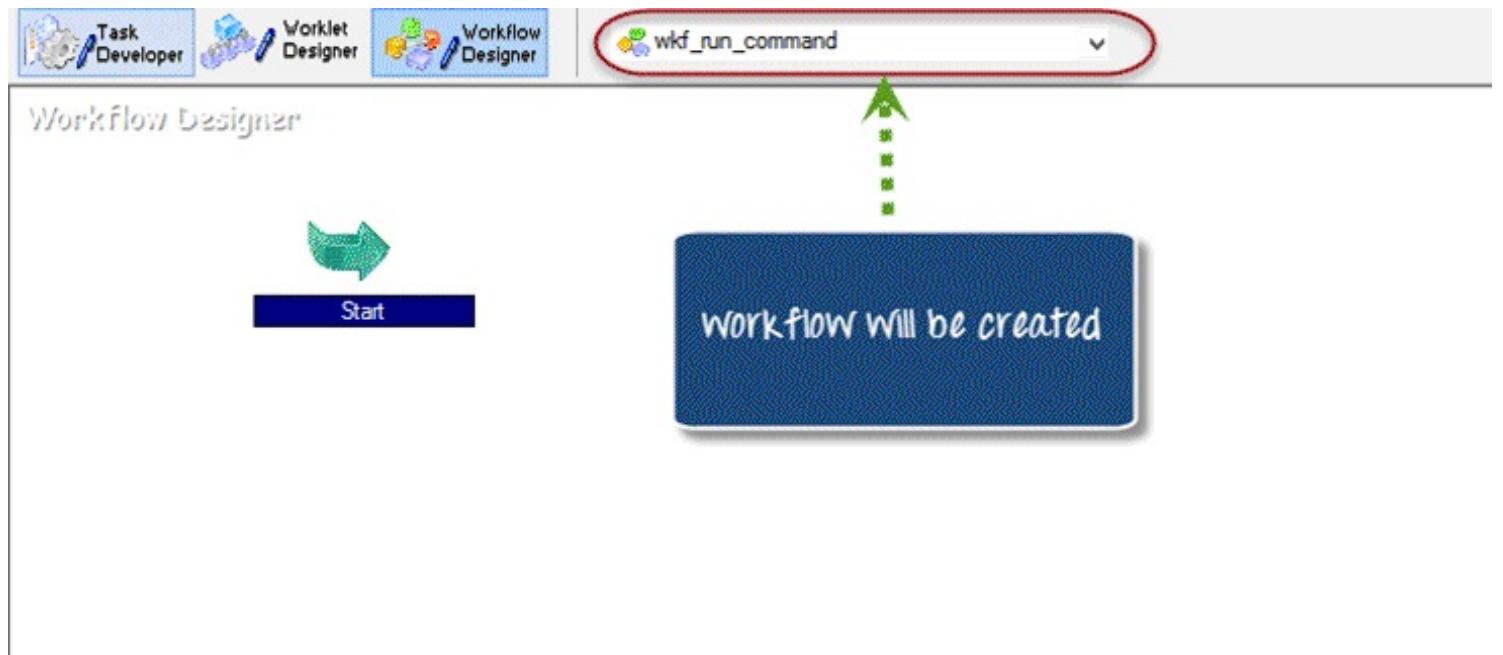
Step 3 – In create workflow window

1. Enter workflow name
2. Select OK Button (leave other options as default)



This will create the workflow.

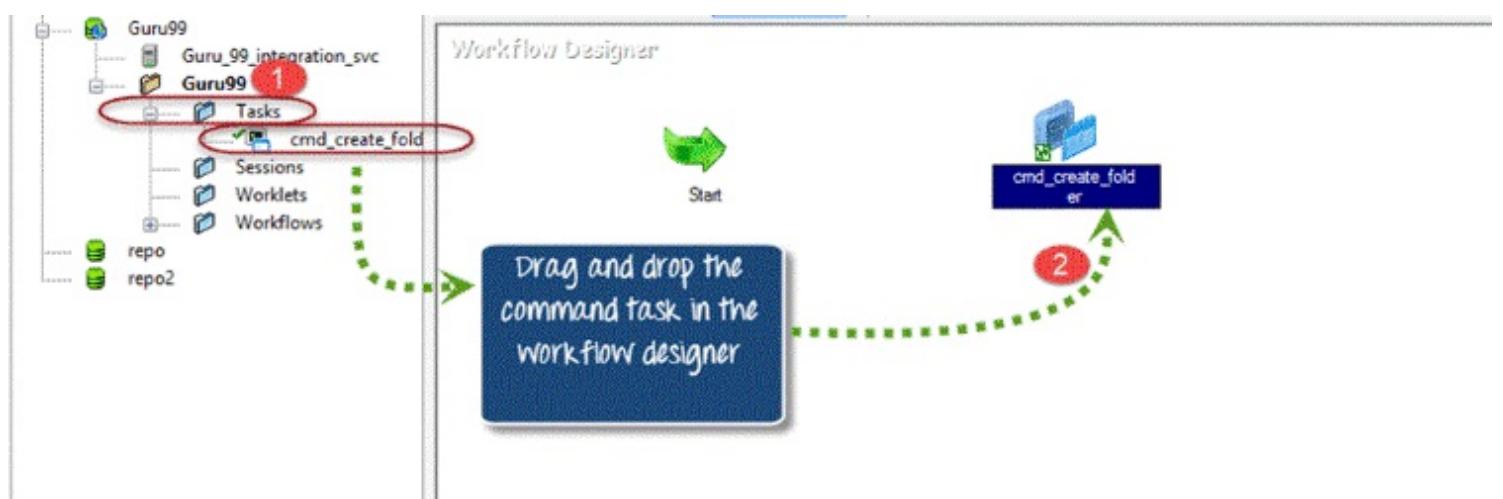
Naming Convention - Workflow names are prefixed with using 'wkf_', if you have a session named 's_m_employee_detail' then workflow for the same can be named as 'wkf_s_m_employee_detail'.



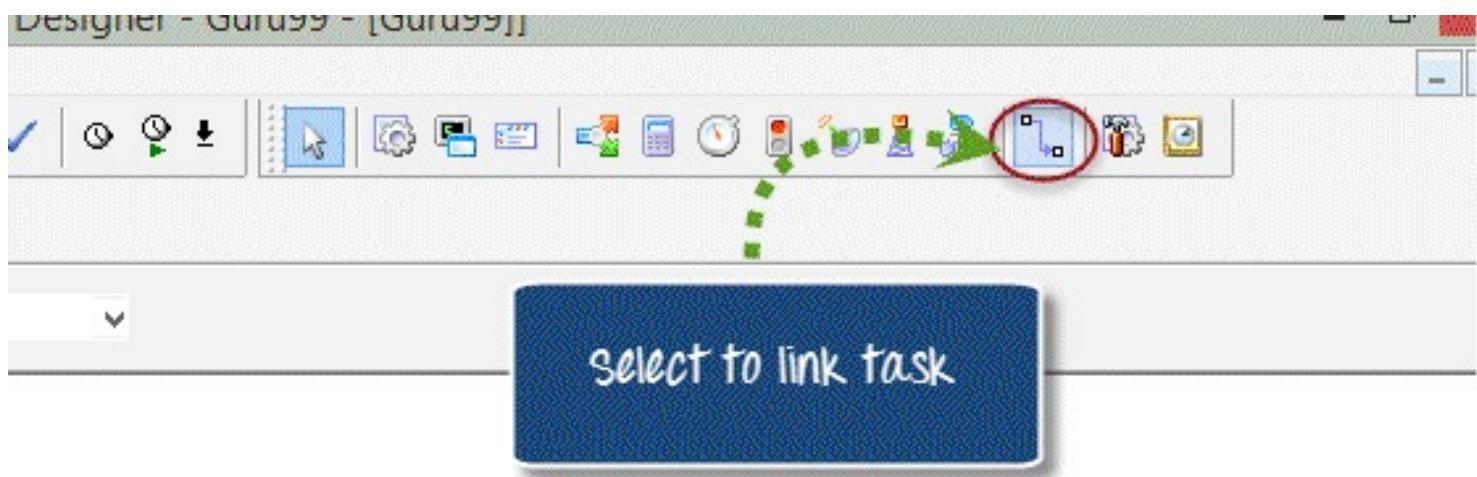
When you create a workflow, it does not consist of any tasks. So, to execute any task in a workflow you have to add task in it.

Step 4 - To add command task that we have created in Task developer to the workflow designer

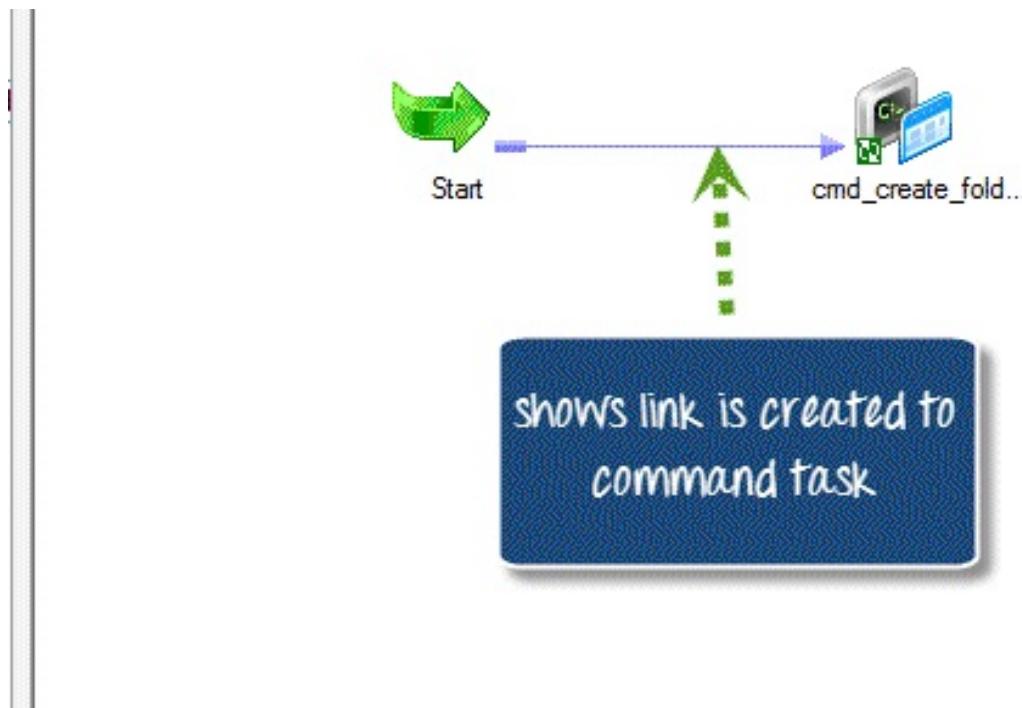
1. In the navigator tree, expand the tasks folder
2. Drag and drop the command task to workflow designer



Step 5 - Select the "link task option" from the toolbox from the top menu. (The link task option links various tasks in a workflow to the start task, so that the order of execution of tasks can be defined).



Step 6 – Once you select the link task icon, it will allow you to drag the link between start task and command task. Now select the start task and drag a link to the command task.

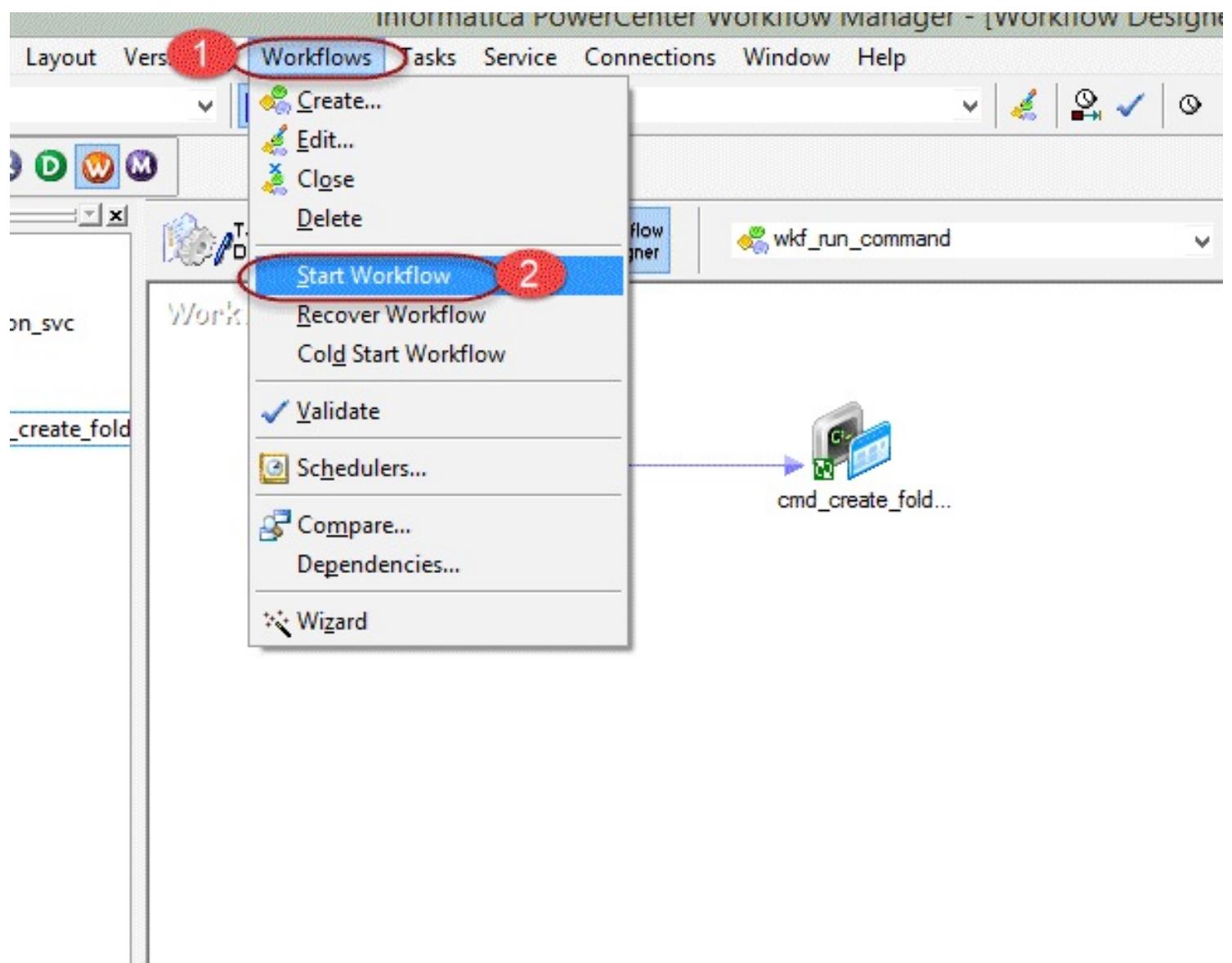


Now you are ready with the workflow having a command task to be executed.

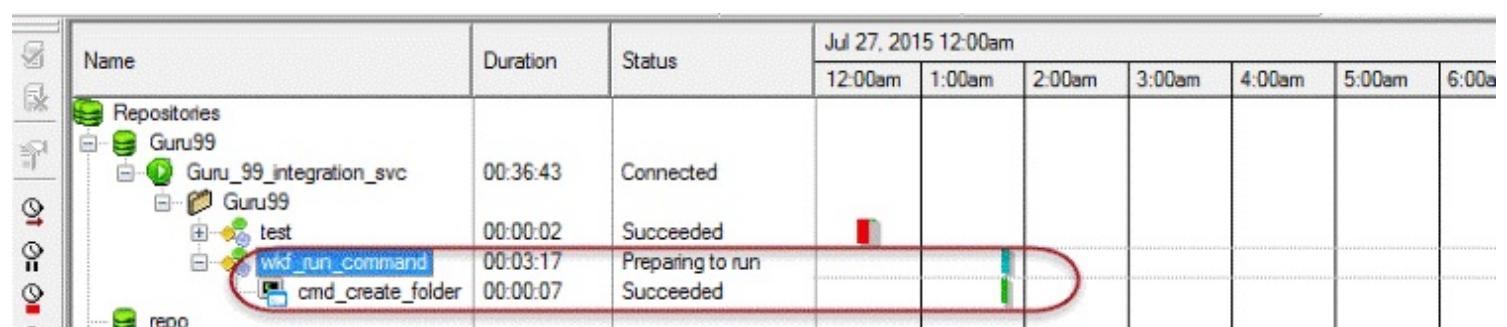
How to execute workflow

Step 1 – To execute the workflow

1. Select workflows option from the menu
2. Select start workflow option



This will open workflow monitor window and executes the workflow



Once the workflow is executed, it will execute the command task to create a folder (guru99 folder) in the defined directory.

Name	Date modified	Type	Size
\$Recycle.Bin	11-09-2015 01:55	File folder	
Boot	04-04-2014 00:44	File folder	
checkpoints	09-06-2015 18:53	File folder	
Documents and Settings	22-08-2013 20:15	File folder	
dotnetapps	12-07-2015 00:33	File folder	
drivers	25-06-2015 15:10	File folder	
guru99	27-07-2015 01:50	File folder	
inetpub	19-07-2015 01:35	File folder	
Informatica	09-06-2015 21:12	File folder	
Intel	20-05-2015 14:59	File folder	
miklass-ex1-008	08-02-2015 01:18	File folder	
MSOCache	02-06-2015 12:17	File folder	

folder is created after executing the command task

Session Task

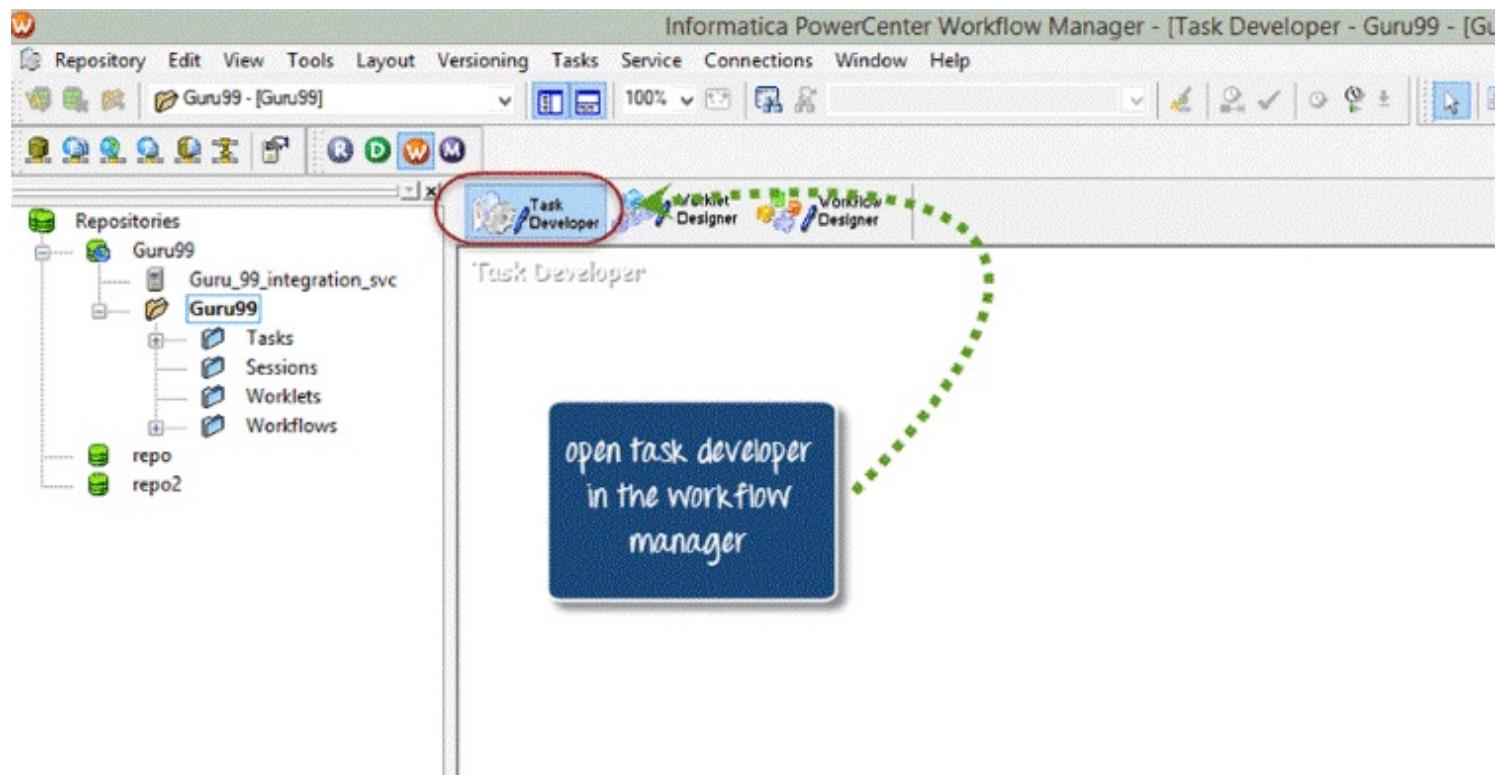
A session task in Informatica is required to run a mapping.

Without a session task, you cannot execute or run a mapping and a session task can execute only a single mapping. So, there is a one to one relationship between a mapping and a session. A session task is an object with the help of which Informatica gets to know how and where to execute a mapping and at which time. Sessions cannot be executed independently, a session must be added to a workflow. In session object cache properties can be configured and also advanced performance optimization configuration.

How to create a session task.

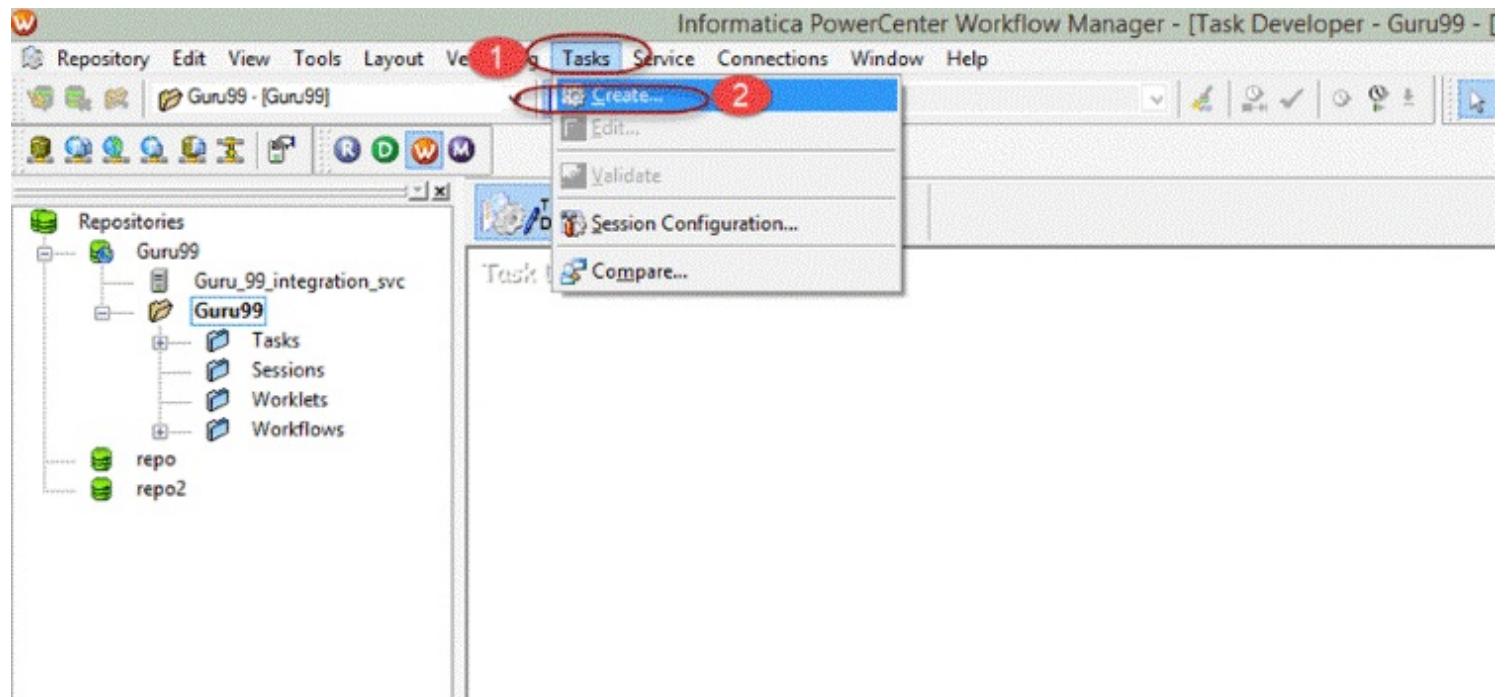
In this exercise you will create a session task for the mapping "m_emp_emp_target" which you created in the previous article.

Step1 – Open Workflow manager and open task developer



Step 2 – Now once the task developer opens, in the workflow manager go to main menu

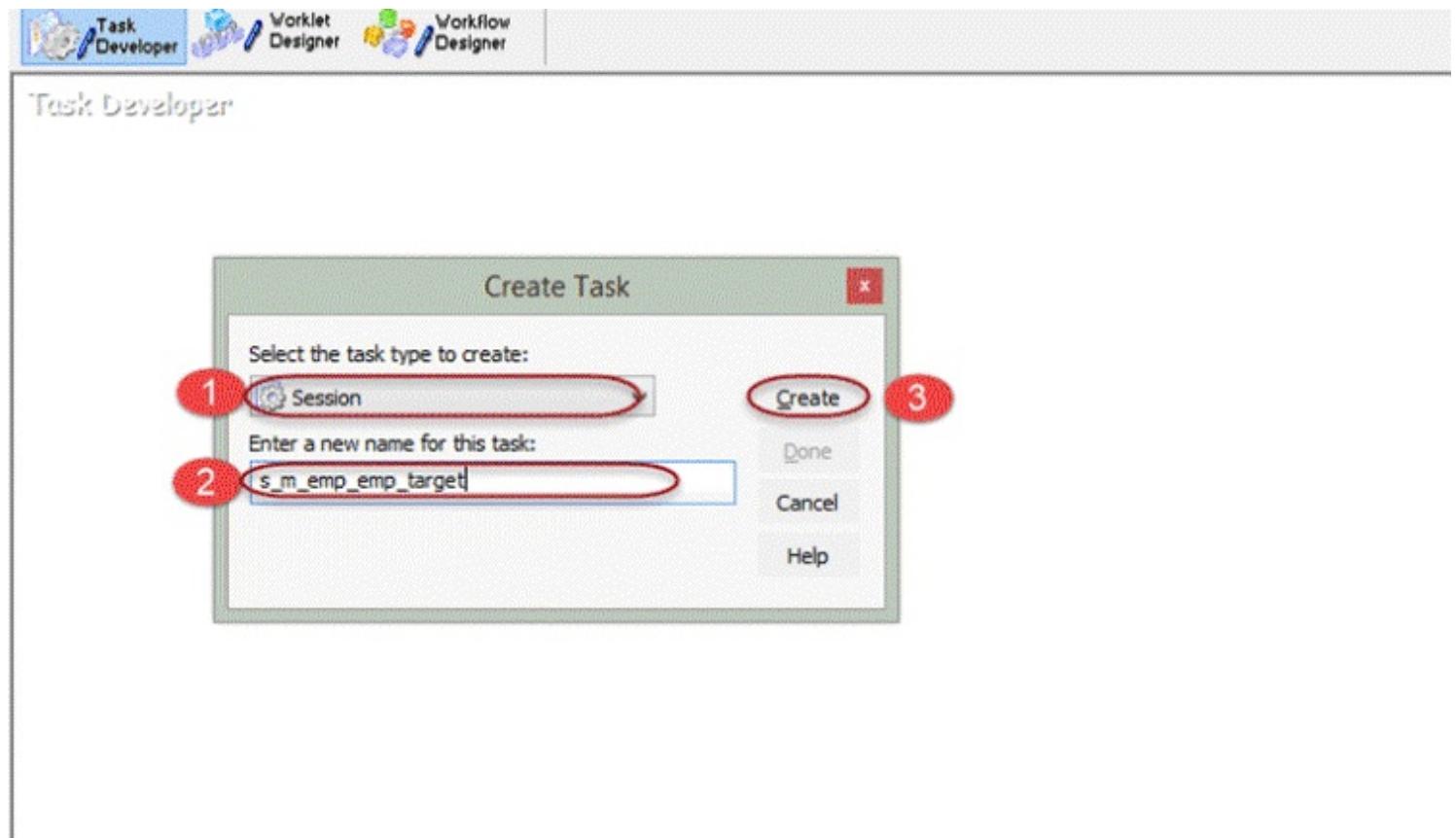
1. Click on task menu
2. Select create option



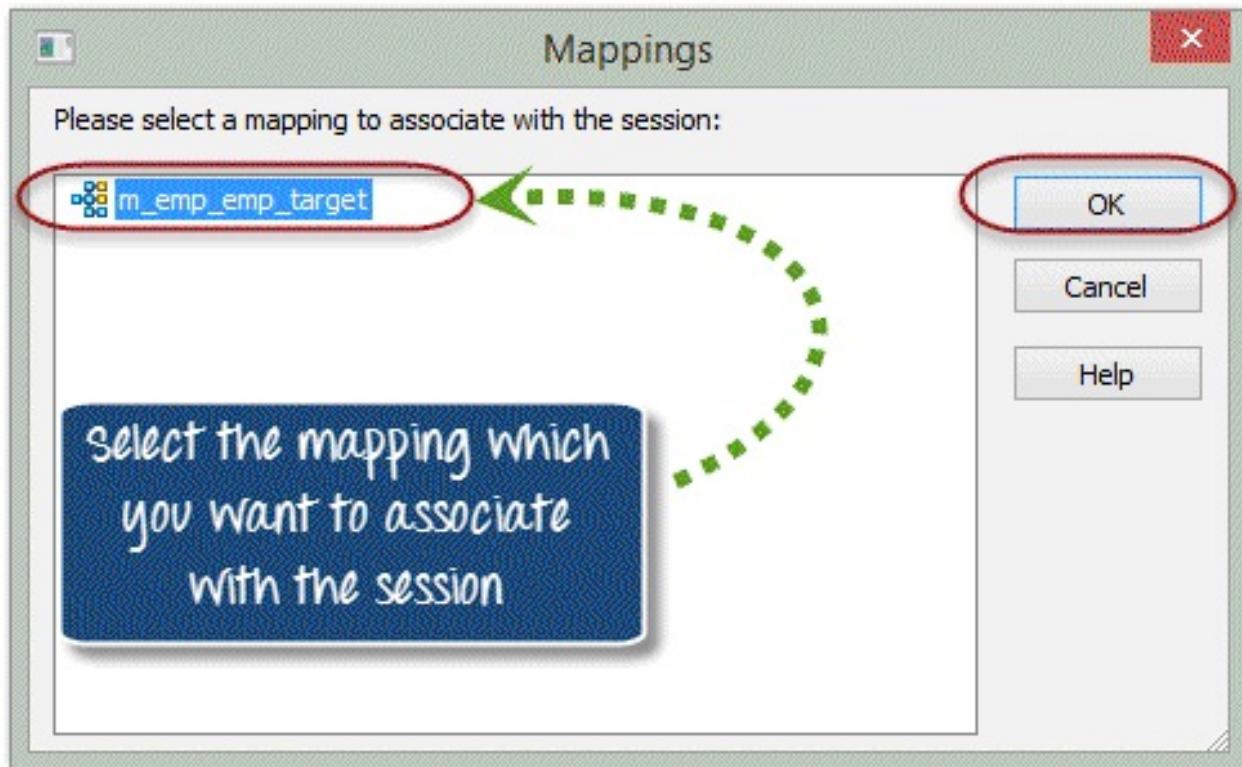
This will open a new window "Create Task"

Step 3 – In the create task window

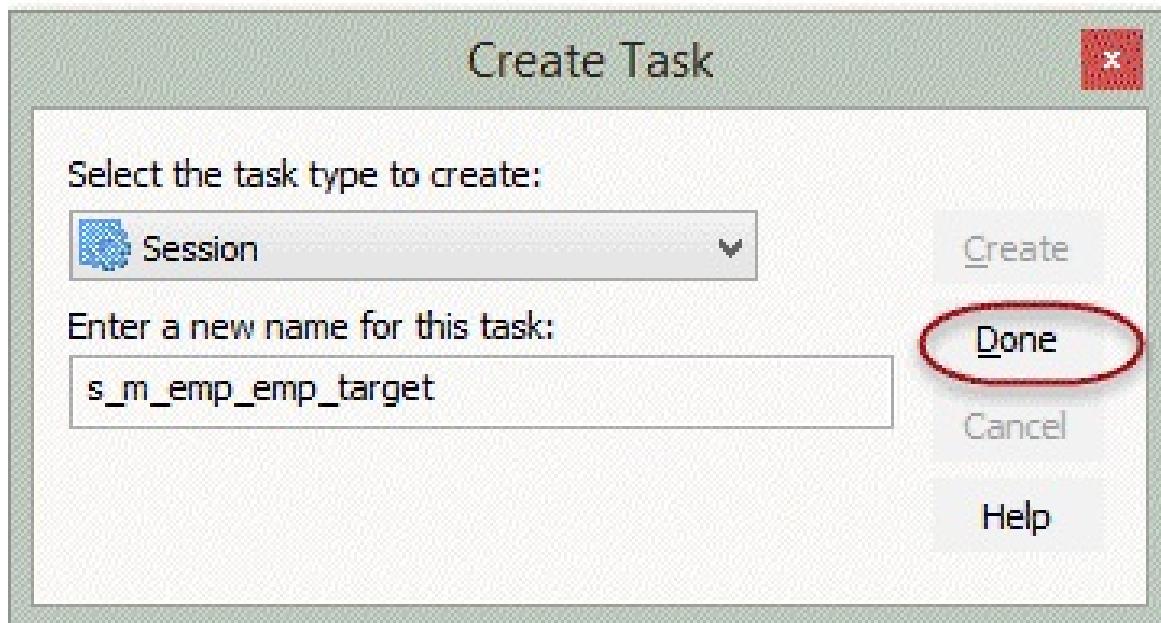
1. Select session task as type of task.
2. Enter name of task.
3. Click create button



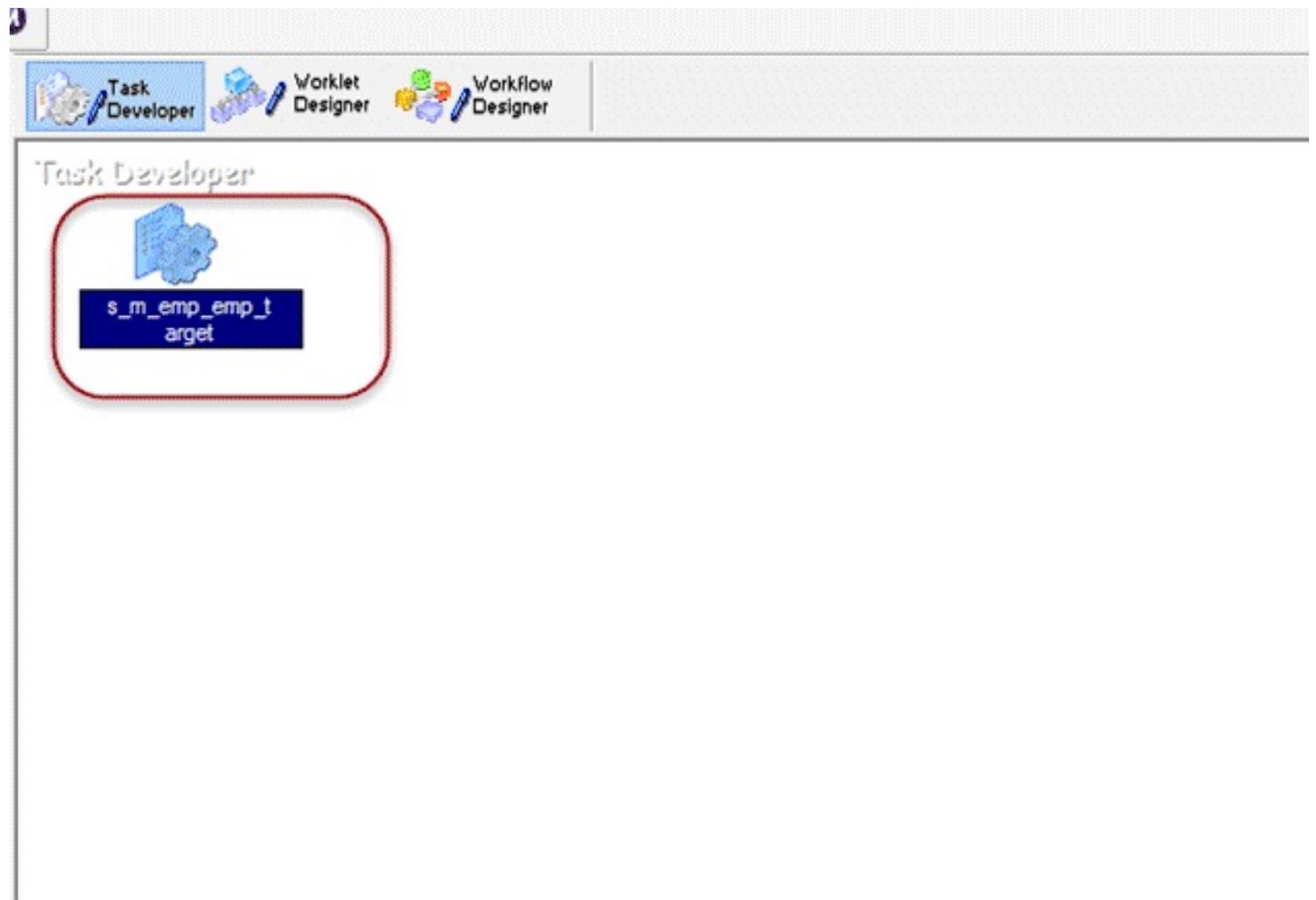
Step 4 – A window for selecting the mapping will appear. Select the mapping which you want to associate with this session, for this example select "m_emp_emp_target" mapping and click OK Button.



Step 5 – After that, click on "Done" button



Session object will appear in the task developer

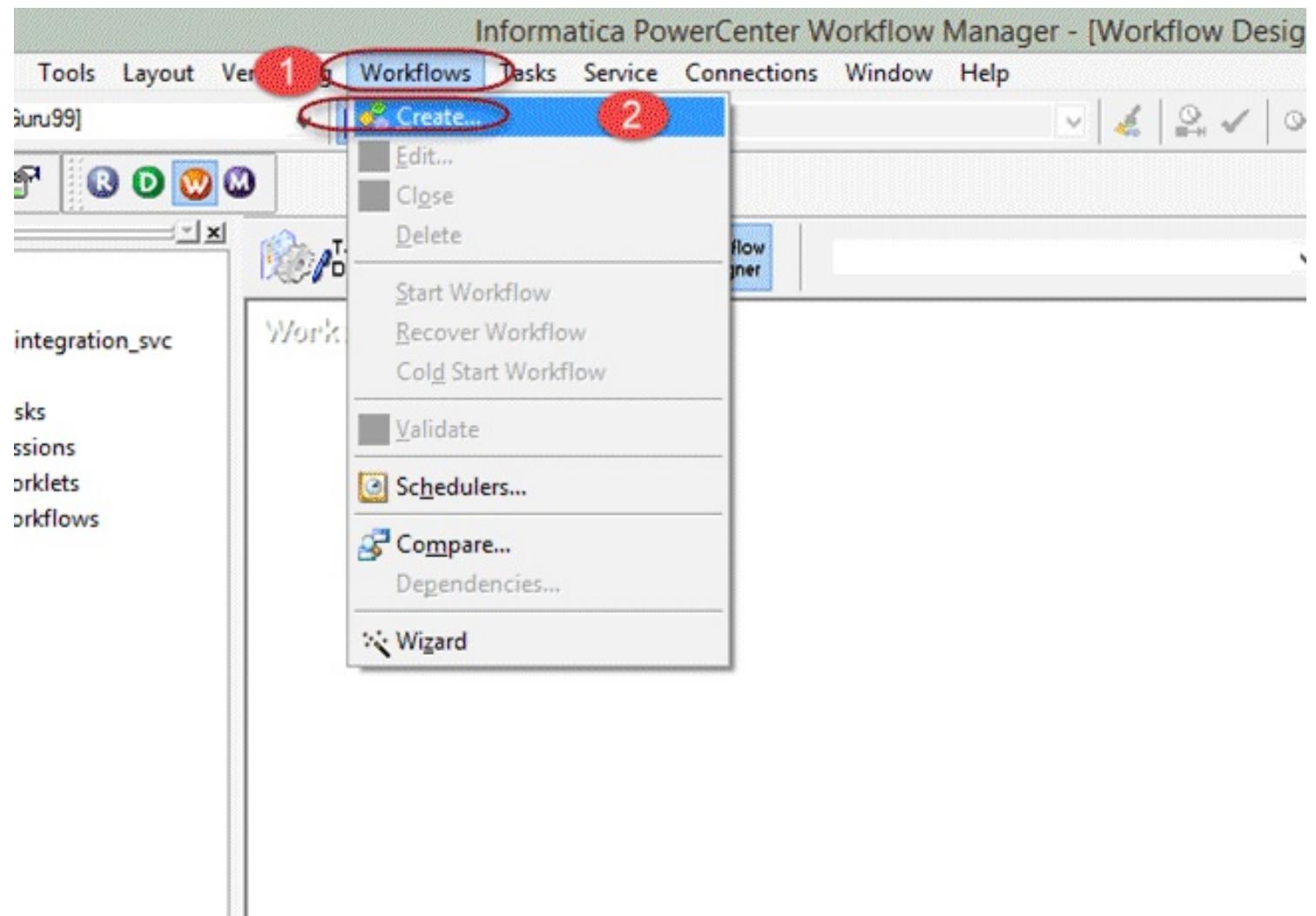


Step 6 – In this step you will create a workflow for the session task. Click on the workflow designer icon.



Step 7 – In the workflow designer tool

1. Click on workflow menu
2. Select create option



Step 8 – In the create workflow window

1. Enter workflow name
2. Select OK. (leave other properties as default, no need to change any properties)

Create Workflow - NEWWORKFLOW10

General Properties Scheduler Variables Events Metadata Extensions

Name: **wkf_s_m_emp_emp_target** 1

Comments:

Integration Service: Guru_99_integration_svc X

Suspension email: X

Runtime options Disabled Suspend on error

Web Services Enabled Config Service...

Configure Concurrent Execution Enabled Configure Concurrent Execution ...

Load Balancing

Service Level: Default

2 OK Cancel Help

In workflow manager a start task will appear, it's a starting point of execution of workflow.



Workflow Designer

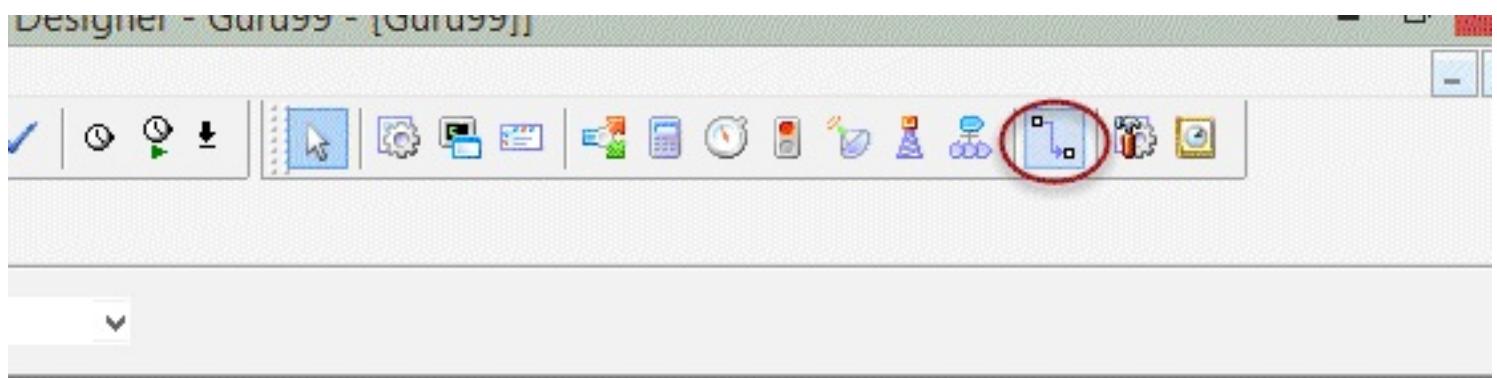


Step 9 – In workflow manager

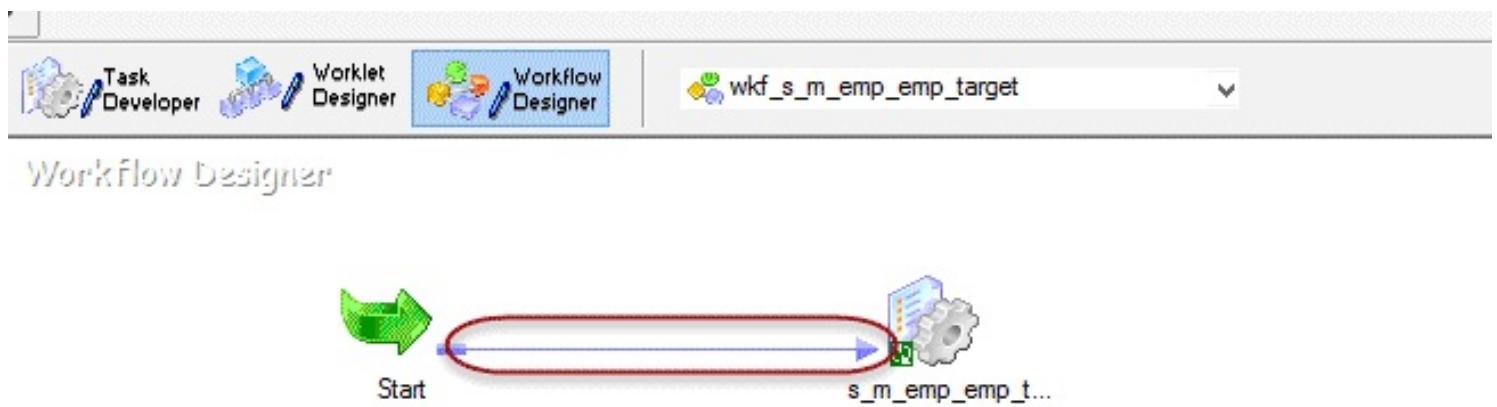
1. Expand the sessions folder under navigation tree.
2. Drag and drop the session you created in the workflow manager workspace.



Step 10 - Click on the link task option in the tool box.



Step 11 - Link the start task and session task using the link.

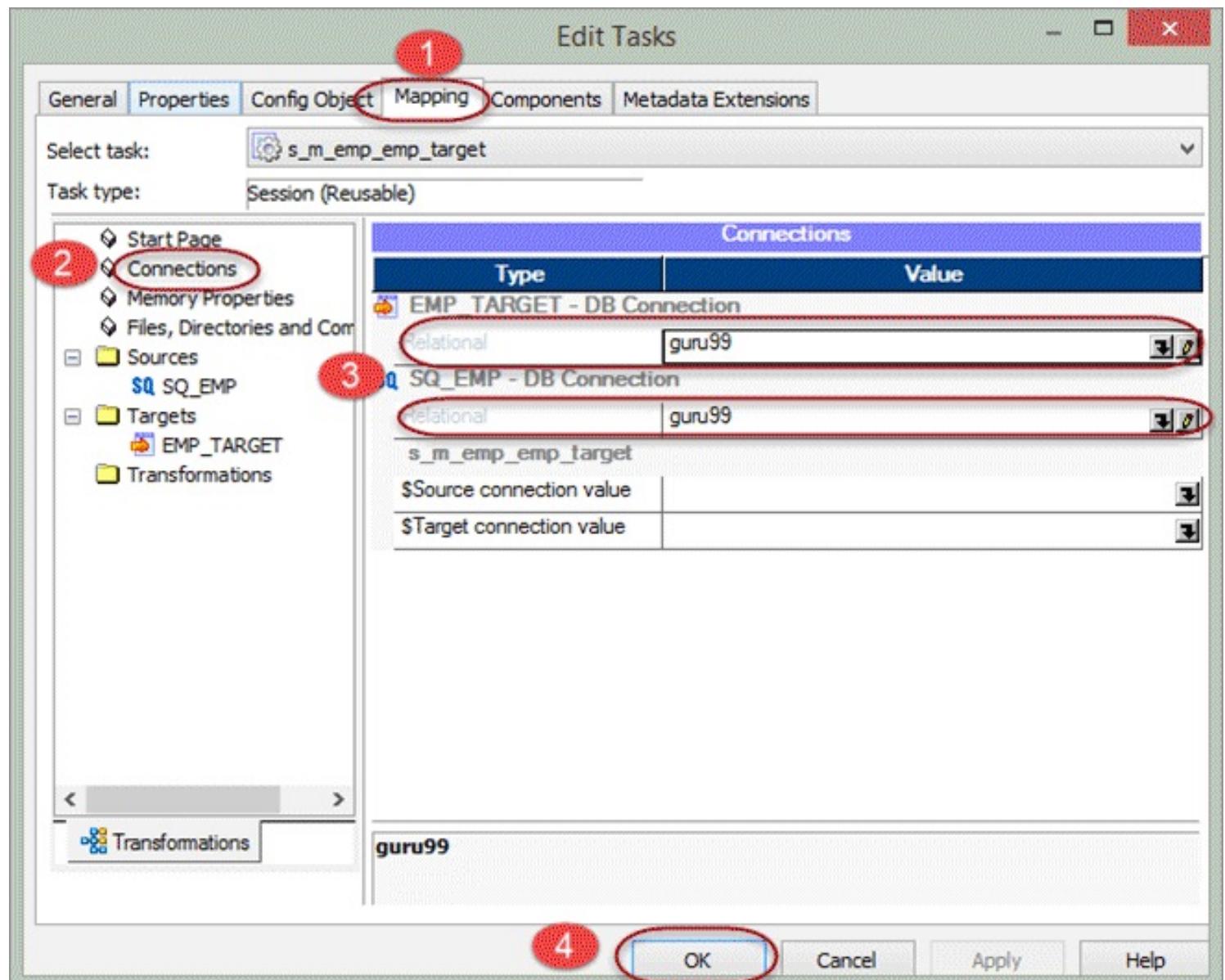


Step 12 – Double click on the session object in wokflow manager. It will open a task window to modify the task properties.

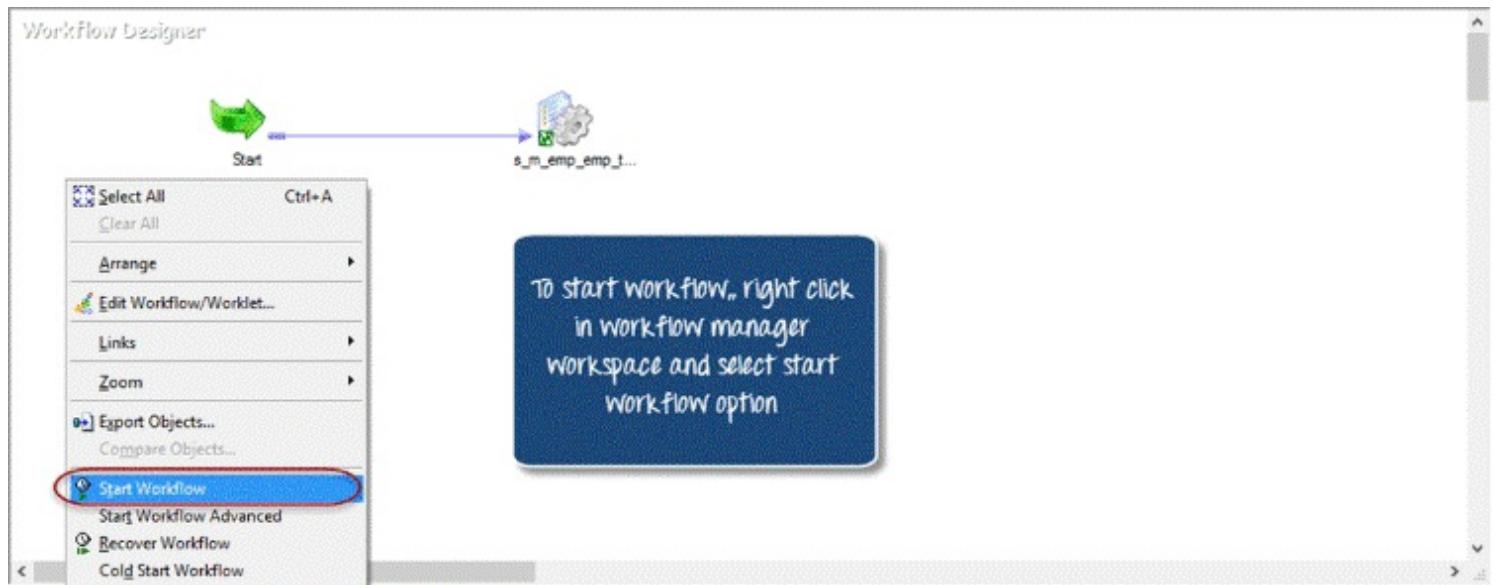
Step 13 – In the edit task window

1. Select mapping tab
2. Select connection property

3. Assign the connection to source and target, the connection which we created in early steps.
4. Select OK Button



Now your configuration of workflow is complete, and you can execute the workflow.



How to add multiple tasks to a start task

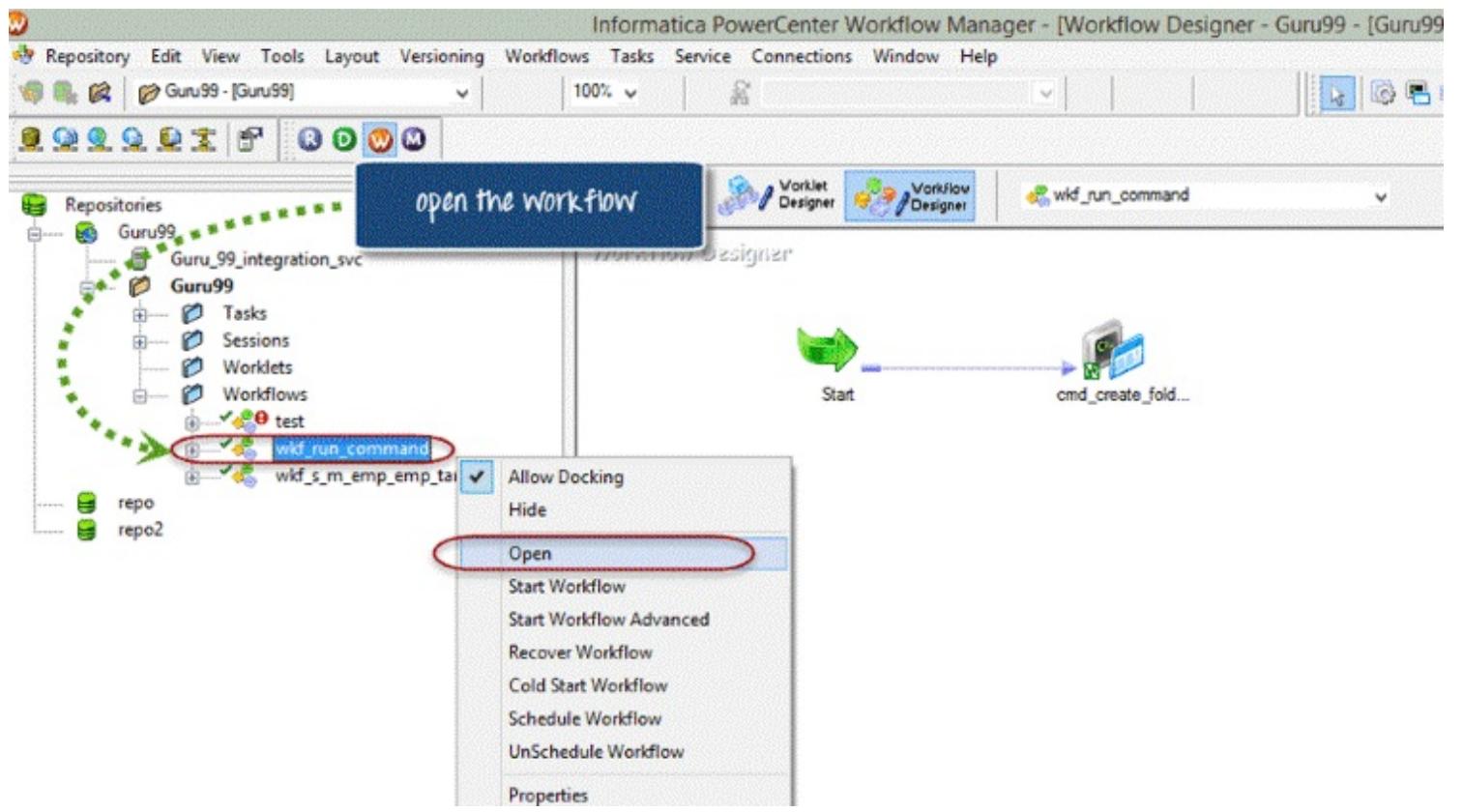
The start task is a starting point for the execution of workflow. There are two ways of linking multiple tasks to a start task.

1. Parallel
2. Serial

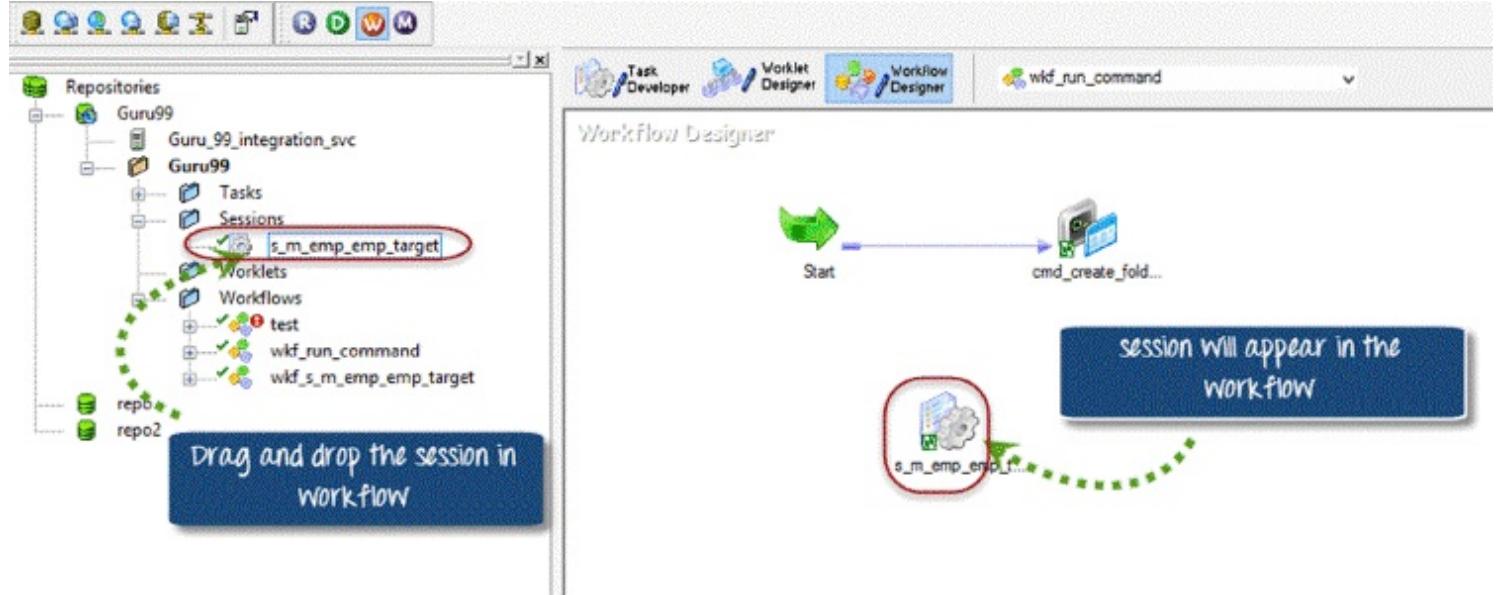
In parallel linking the tasks are linked directly to the start task and all tasks start executing in parallel at same time.

How to add tasks in parallel

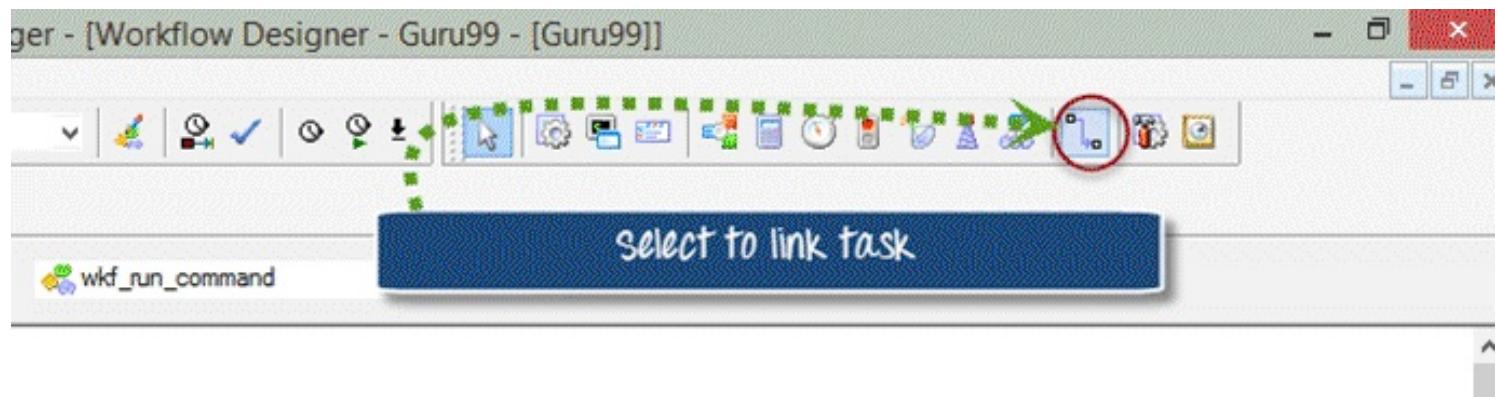
Step 1 – In the workflow manager, open the workflow "wkf_run_command"



Step 2 – In the workflow, add session task "s_m_emp_emp_target". (by selecting session and then drag and drop)

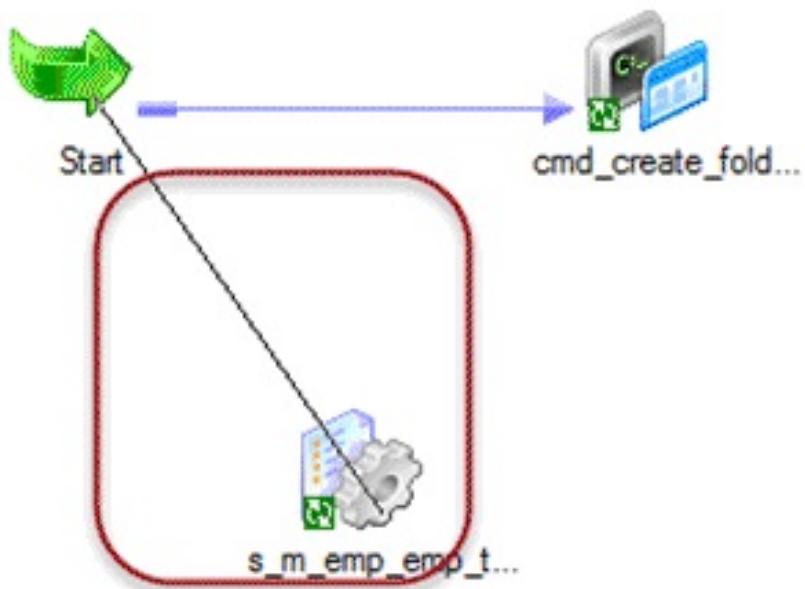


Step 3 – Select the link task option from the toolbox



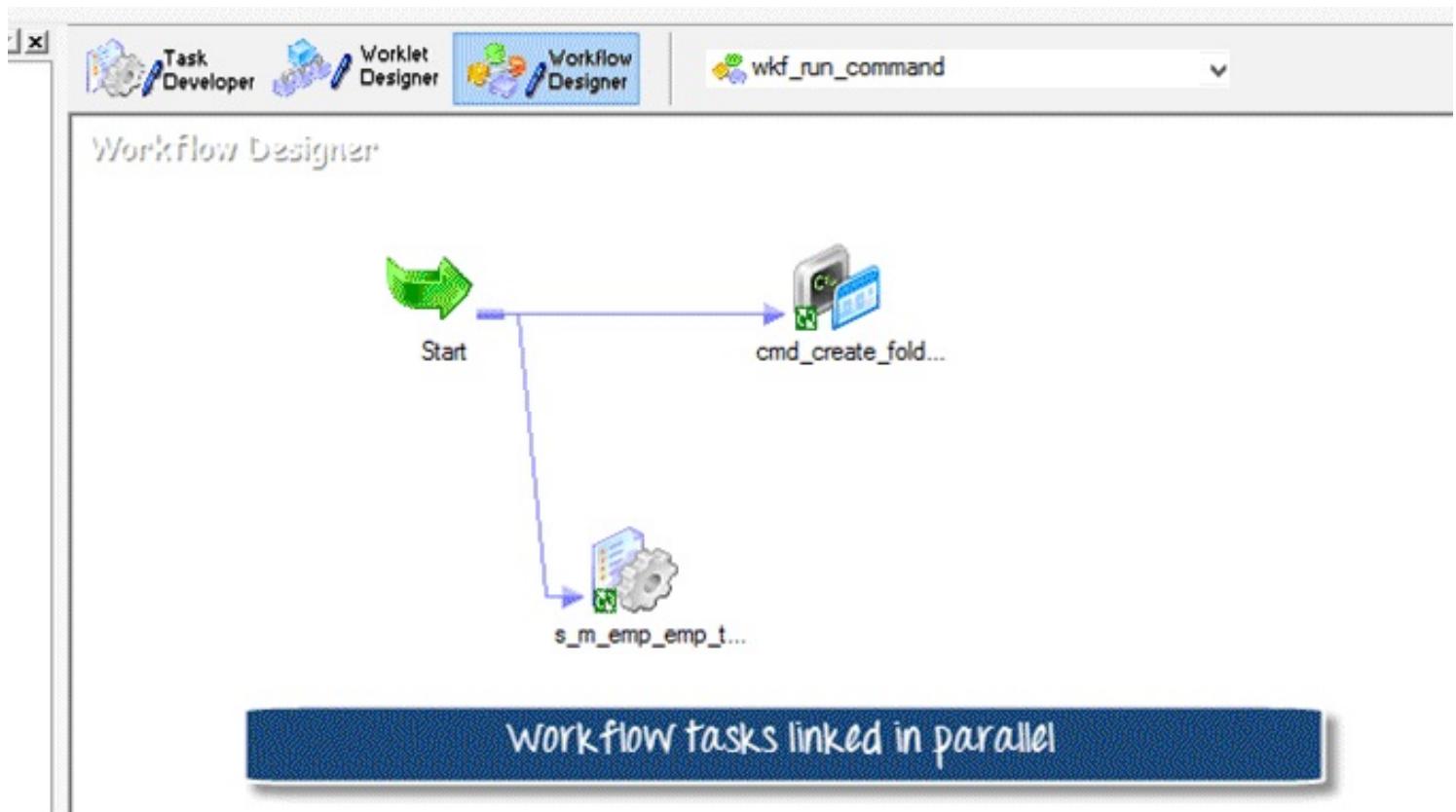
Step 4 - link the session task to the start task (by clicking on start tasks, holding the click and connecting to session task)

... designs



Link the session task to start task

After linking the session task, the workflow will look like this.



Step 5 – Start the workflow and monitor in the workflow monitor.

Workflow Designer

to start workflow, right click in workspace and select start workflow menu

Start Workflow

Workflow Run	Status Message	Node	Start Time	Completion Time	Status
wkf_run_command			01-08-2015 20:31:10	01-08-2015 20:31:14	Succeeded
wkf_run_command		node01_RAJESH	01-08-2015 20:31:10	01-08-2015 20:31:14	Succeeded
		node01_RAJESH	01-08-2015 20:31:11	01-08-2015 20:31:12	Succeeded

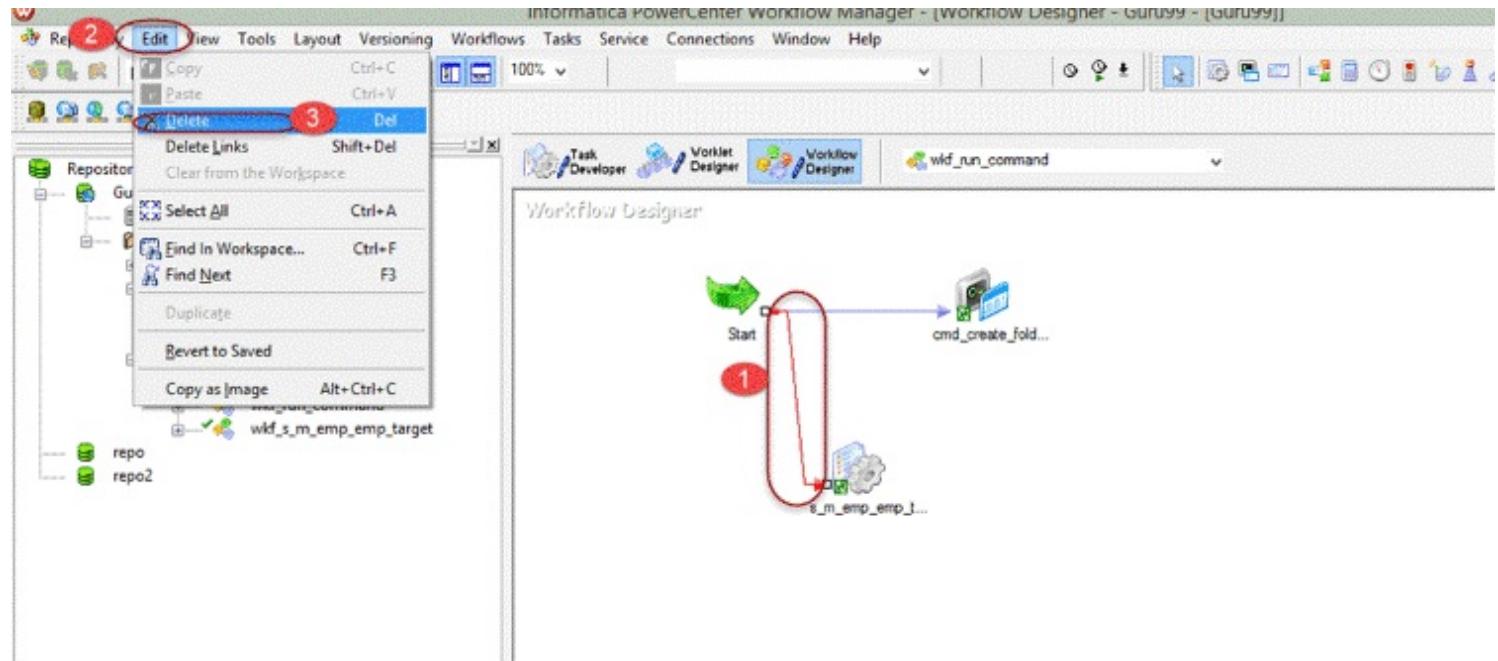
both tasks started together

How to add tasks in serial mode

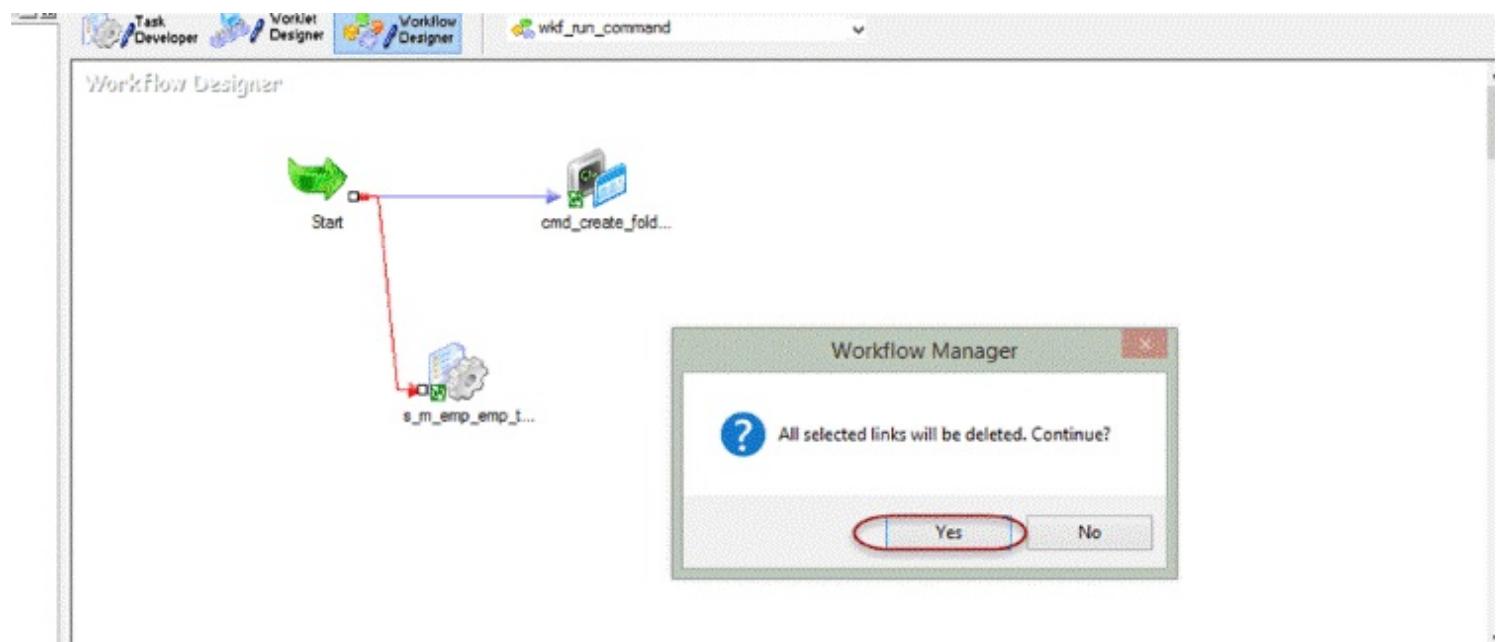
But before we add tasks in serial mode, we have to delete the task that we added to demonstrate parallel execution of task. For that

Step 1 – Open the workflow "w.kf_run_command"

1. Select the link to the session task.
2. Select edit option in the menu
3. Select delete option



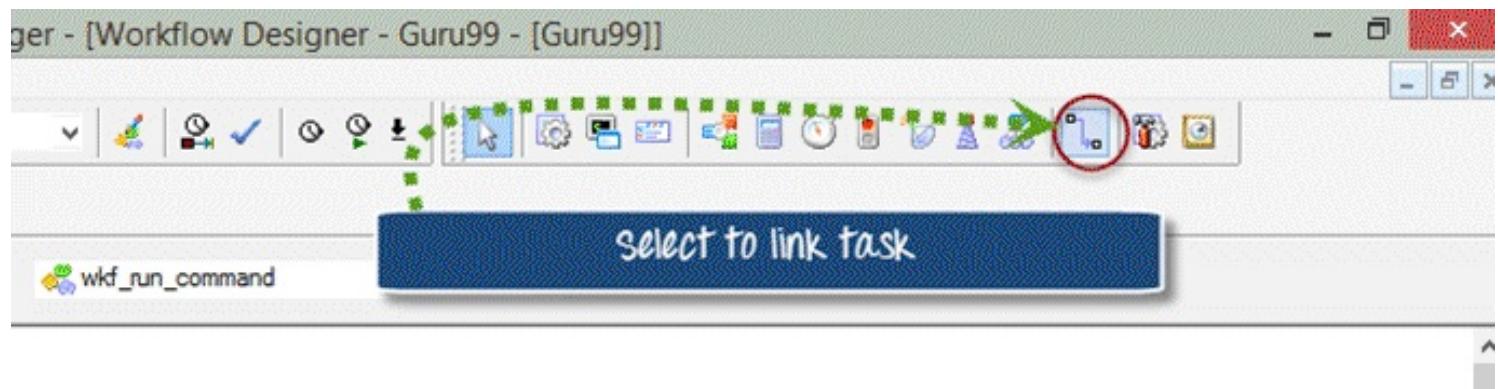
Step 2 – Confirmation dialogue box will appear in a window, select yes option



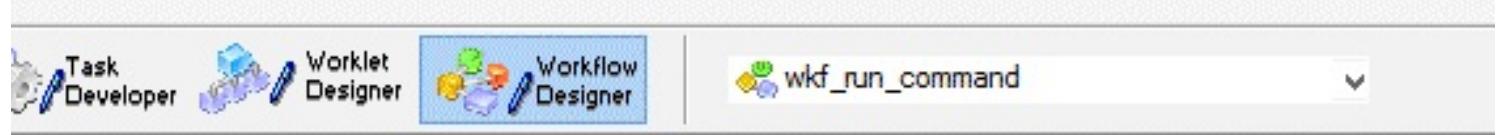
The link between the start task and session task will be removed.



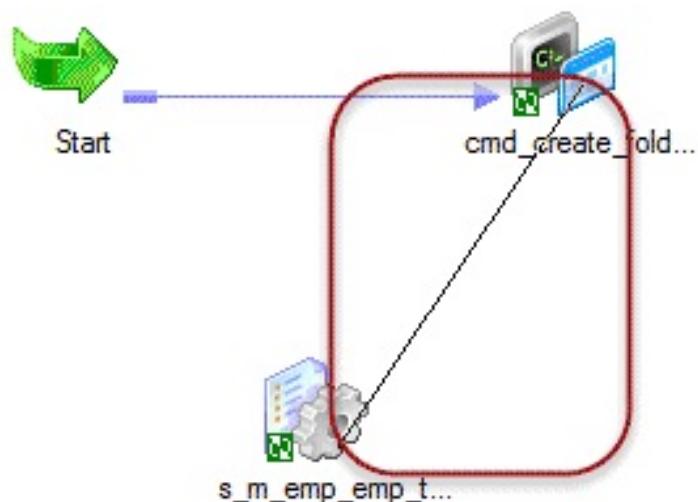
Step 3 – Now again go to top menu and select the link task option from the toolbox



Step 4 – link the session task to the command task

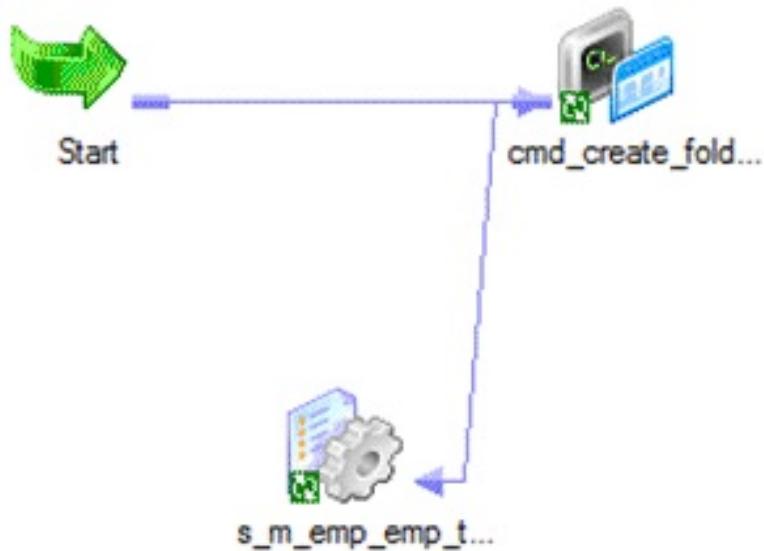


Workflow Designer



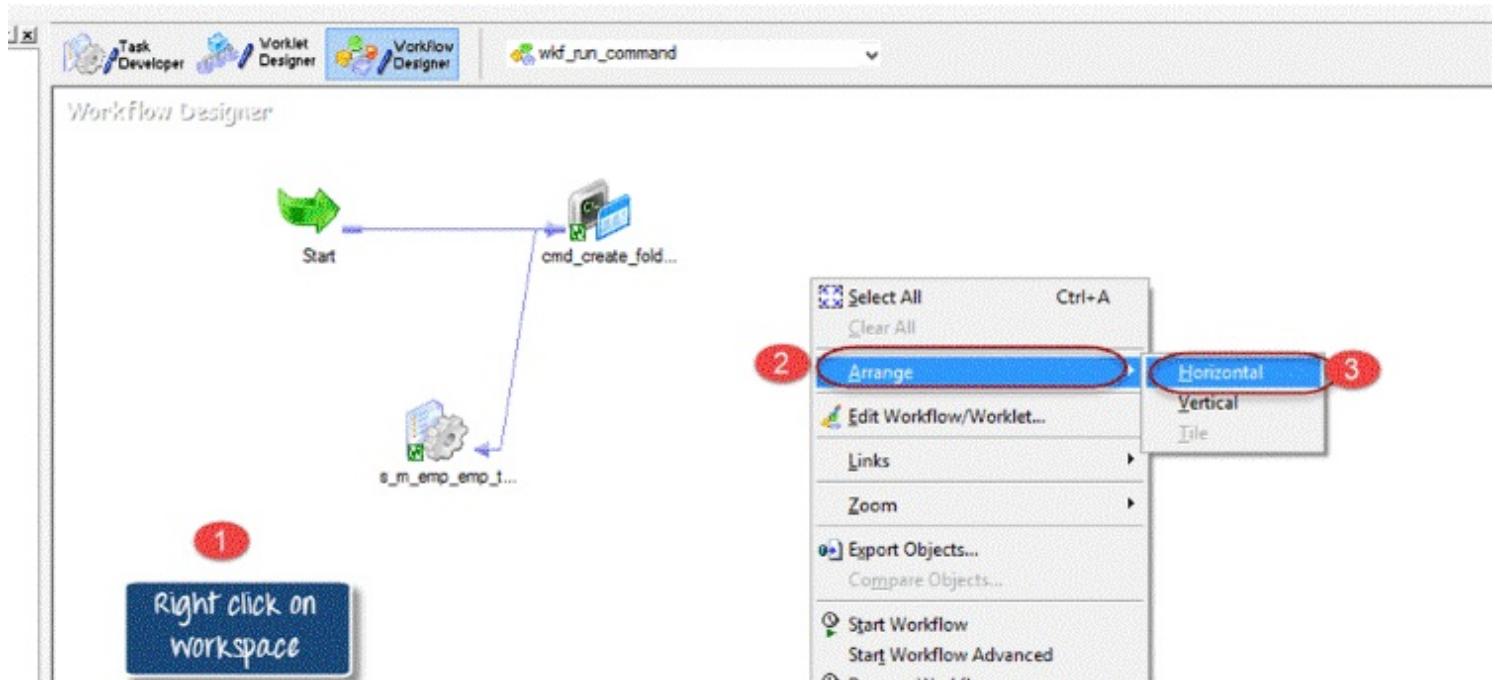
Link the command task to session task

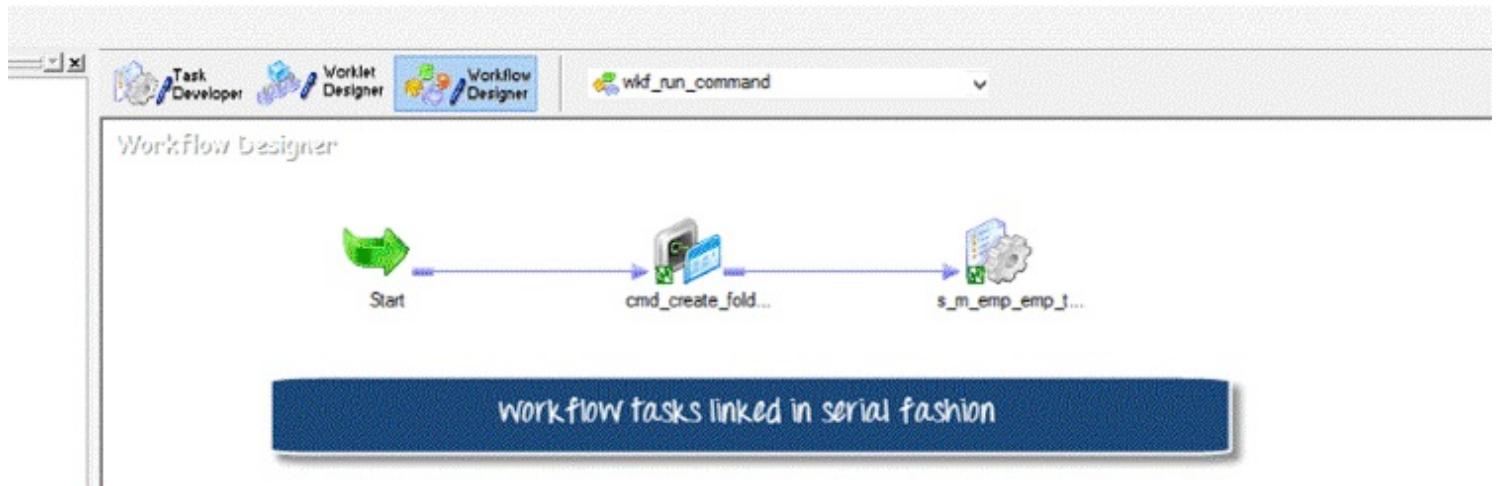
After linking the workflow will look like this



Step 5 - To make the visual appearance of workflow more clear

1. Right click on workspace of workflow
2. Select arrange menu
3. Select Horizontal option





If you start the workflow the command task will execute first and after its execution, session task will start.

Workflow Variable

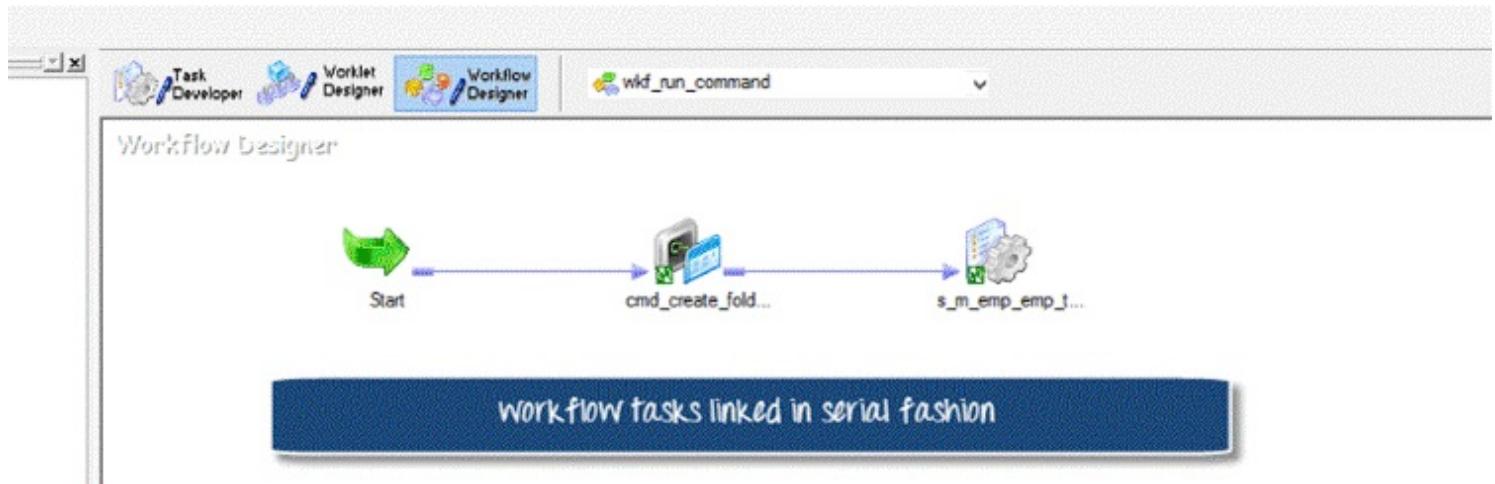
Workflow variables allows different tasks in a workflow to exchange information with each other and also allows tasks to access certain properties of other tasks in a workflow. For example, to get the current date you can use the inbuilt variable "sysdate".

Most common scenario is when you have multiple tasks in a workflow and in one task you access the variable of another task. For example, if you have two tasks in a workflow and the requirement is to execute the second task only when first task is executed successfully. You can implement such scenario using predefined variable in the workflow.

Implementing the scenario

We had a workflow "wkf_run_command" having tasks added in serial mode. Now we will add a condition to the link between session task and command task, so that, only after the success of command task the session task will be executed.

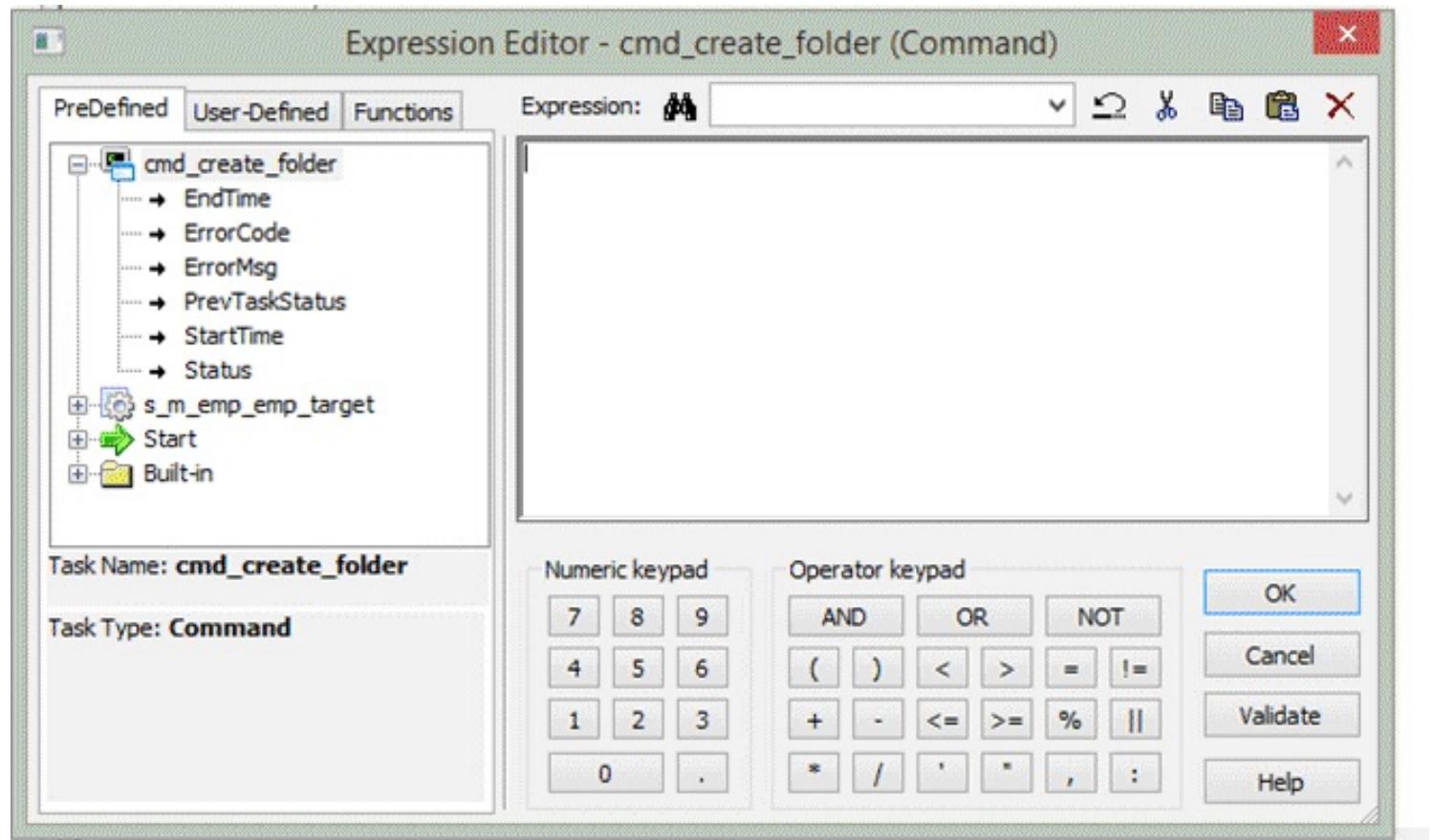
Step 1 - Open the workflow "wkf_run_command"



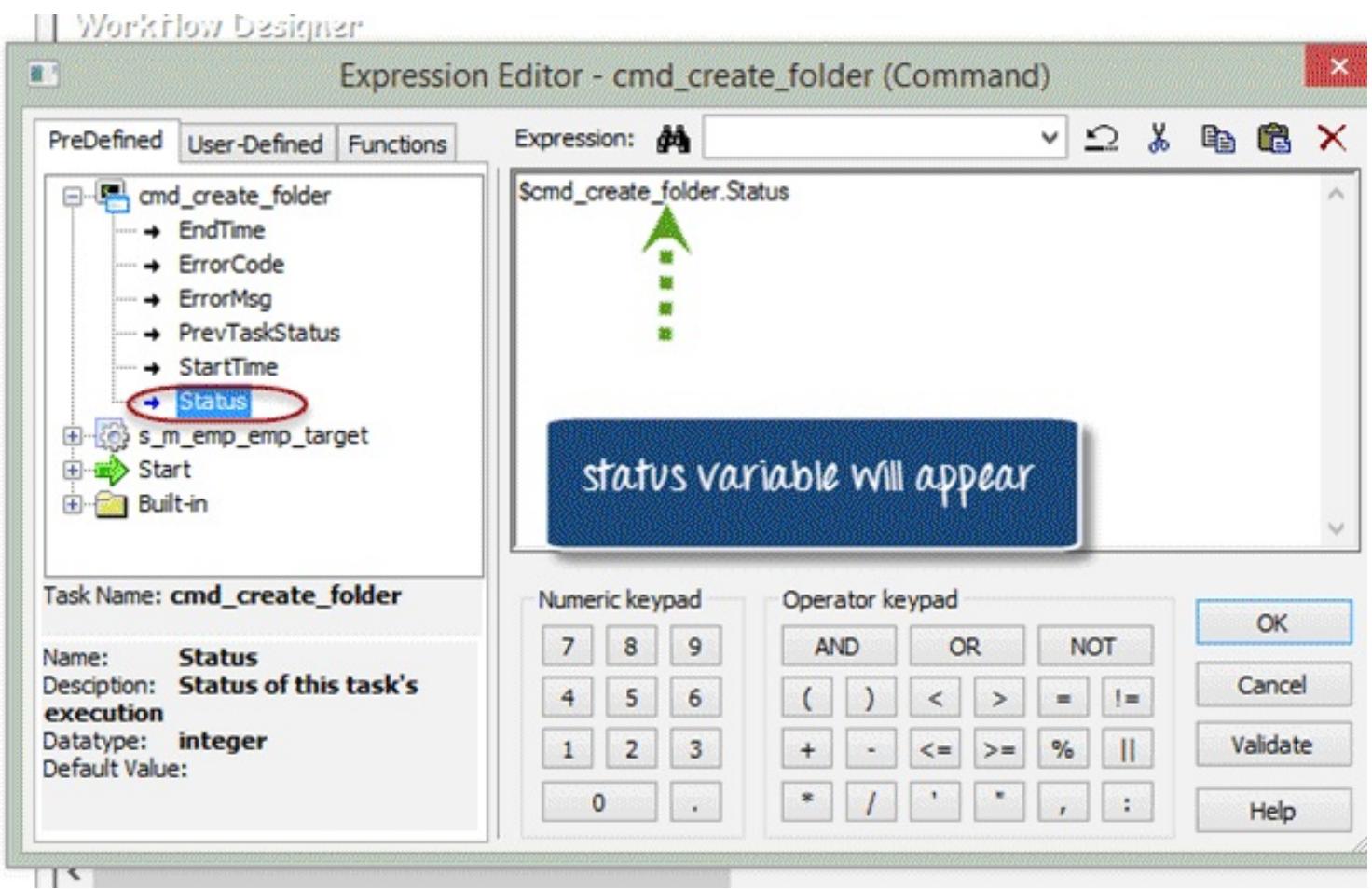
Step 2 - Double click on the link between session and command task



An Expression window will appear

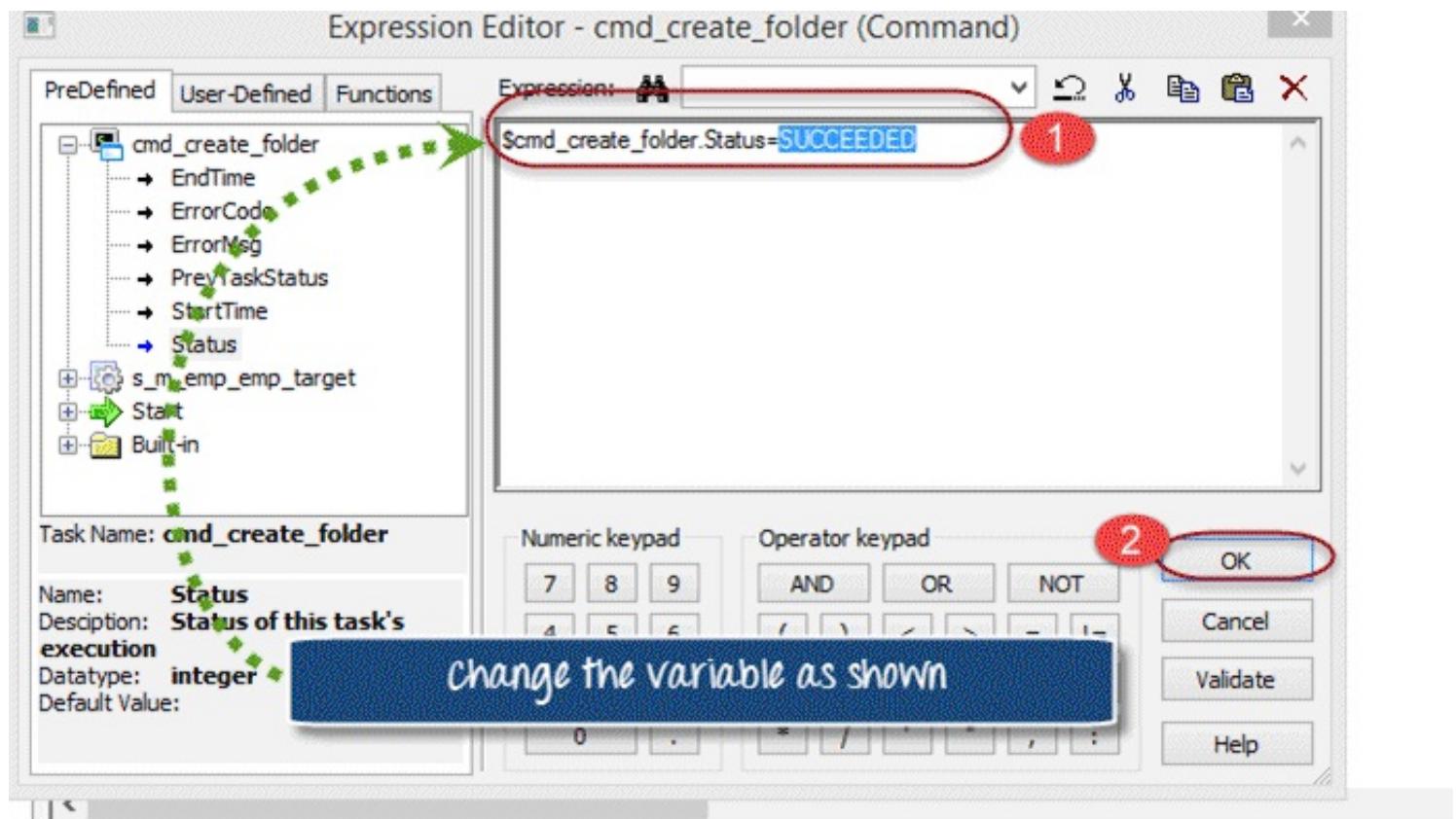


Step 3 – Double click the status variable under "cmd_create_folder" menu. A variable "\$cmd_create_folder.status" will appear in the editor window on right side.

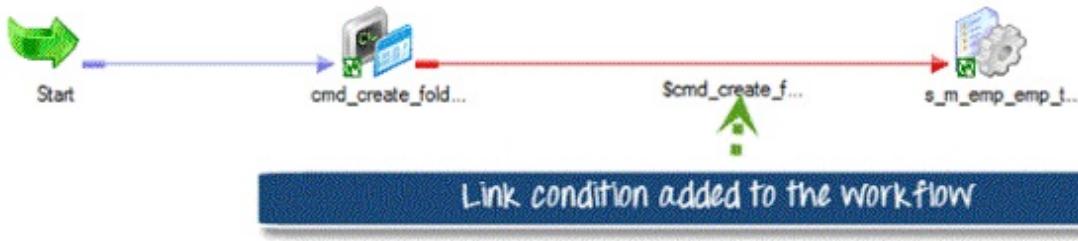


Step 4 - Now we will set the variable "\$cmd_create_folder.status" condition to succeeded status . which means when the previous tasks is executed and the execution was success, then only execute the next session task.

1. Change the variable to "\$cmd_create_folder.status=SUCCEEDED" value.
2. Click OK Button



The workflow will look like this



When you execute this workflow, the command task executes first and only when it succeeds then only the session task will get executed.

Workflow Parameter

Workflow parameters are those values which remain constant throughout the run. once their value is assigned it remains same. Parameters can be used in workflow properties and their values can be defined in parameter files. For example, instead of using hard coded connection value you can use a parameter/variable in the connection name and value can be defined in the parameter file.

Parameter files are the files in which we define the values of mapping/workflow variables or parameters. These files have the extension of ".par". As a general standard a parameter file is created for a workflow.

Advantages of Parameter file

- Helps in migration of code from one environment to other
- Allows easy debugging and testing
- Values can be modified with ease without change in code

Structure of parameter file

The structure of parameter file

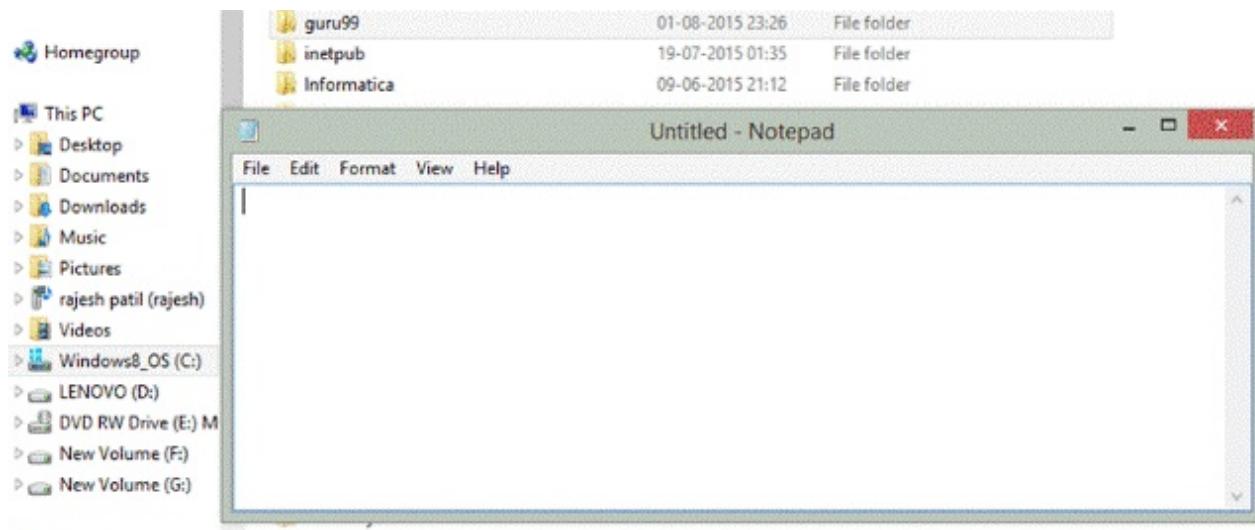
- [folder_name.WF:Workflow_name]
- \$Parameter_name=Parameter_value

Folder_name is the name of repository folder, workflow name is the name of workflow for which you are creating the parameter file.

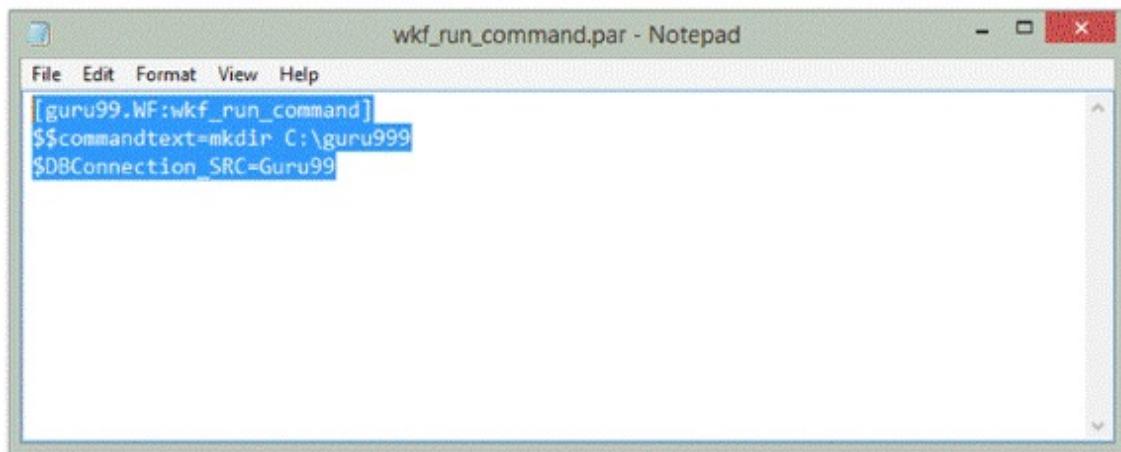
We will be creating a parameter file for the database connection "guru99" which we assigned in our early sessions for sources and targets.

How to create parameter file

Step 1 – Create a new empty file (notepad file)

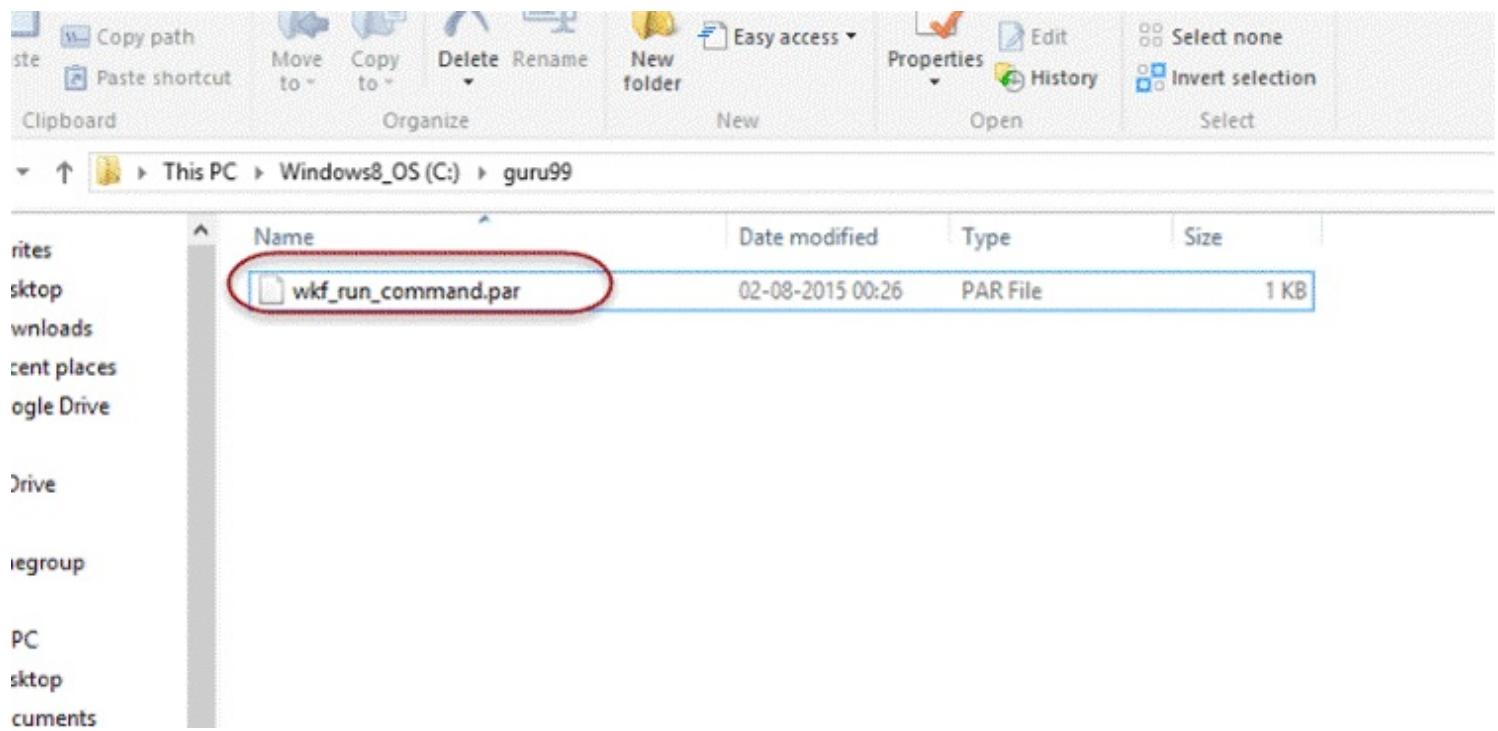


Step 2 – In the file enter text as shown in figure



```
[guru99.WF:wkf_run_command]
$$CommandText=mkdir C:\guru99
$DBConnection_SRC=Guru99
```

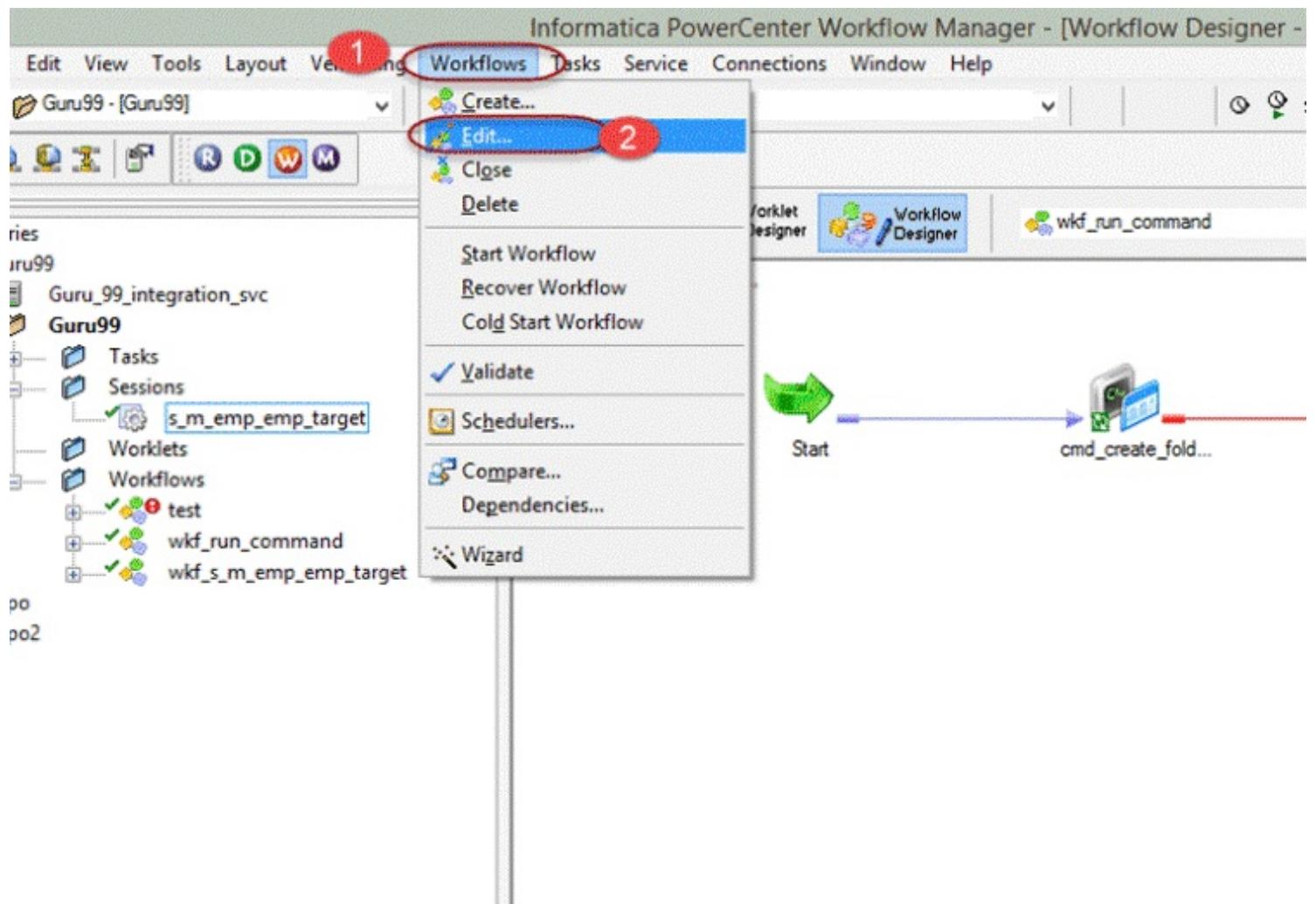
Step 3 – Save the file under a folder guru99 at the location "C:\guru99" as "wkf_run_command.par"



In the file we have created a parameter "\$DBConnection_SRC", we will assign the same to a connection in our workflow.

Step 4- Open the workflow "wkf_run_command"

1. Select workflows menu
2. Select edit option



Step 5 – This will open up edit workflow window, in this window

1. Go to properties tab menu
2. Enter the parameter file name as "c:\guru99\wkf_run_command.par"
3. Select OK Button

1

Edit Workflow - wkf_run_command

General Properties Scheduler Variables Events Metadata Extensions

Set the properties of the workflow/worklet in the grid below:

Attribute	Value
Parameter Filename	c:\guru99\wkf_run_command.par
Write Backward Compatible Workfl...	<input type="checkbox"/>
Workflow Log File Name	wkf_run_command.log
Workflow Log File Directory	\$PMWorkflowLogDir\
Save Workflow log by	By runs
Save workflow log for these runs	0
Enable HA recovery	<input type="checkbox"/>
Automatically recover terminated ta...	<input checked="" type="checkbox"/>
Maximum automatic recovery attem...	5

2

Parameter Filename

Parameter filename

3

OK Cancel Apply Help

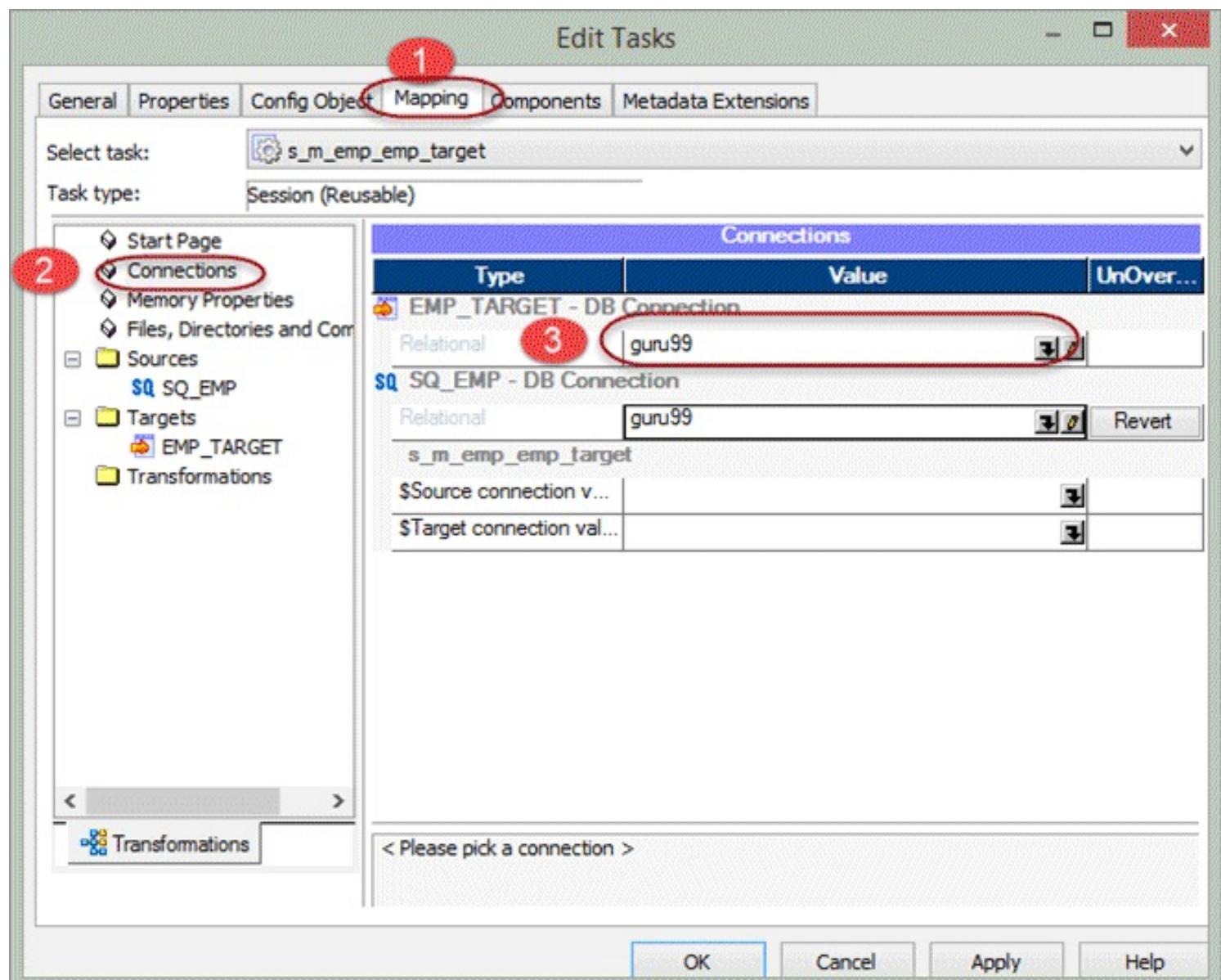
Now we are done with defining the parameter file content and point it to a workflow.

Next step is to use the parameter in session.

Step 6 - In workflow double click on the session "s_m_emp_emp_target", then

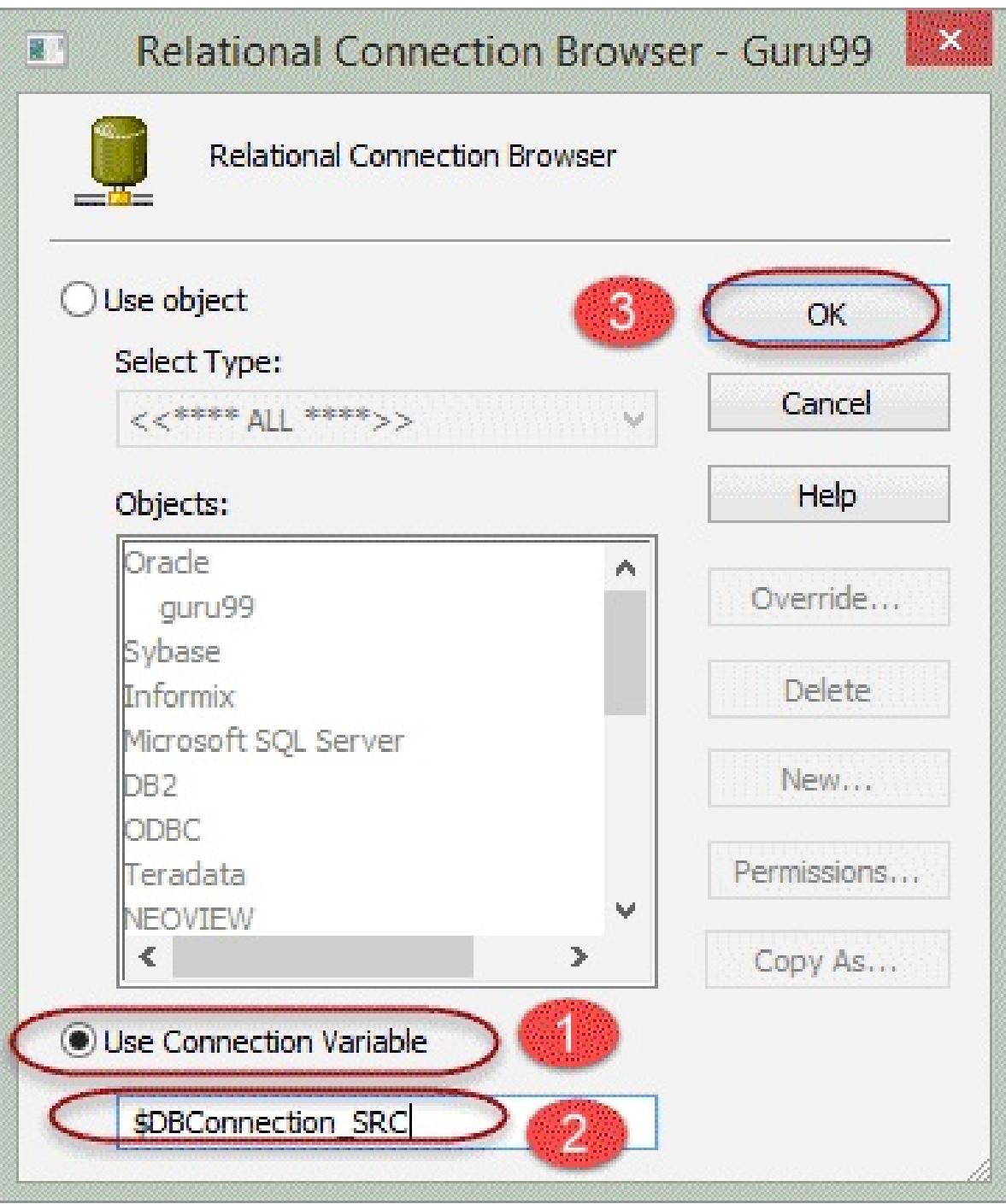
1. Select mappings tab menu

2. Select connection property in the left panel
3. Click on the target connection, which is hardcoded now as "guru99"

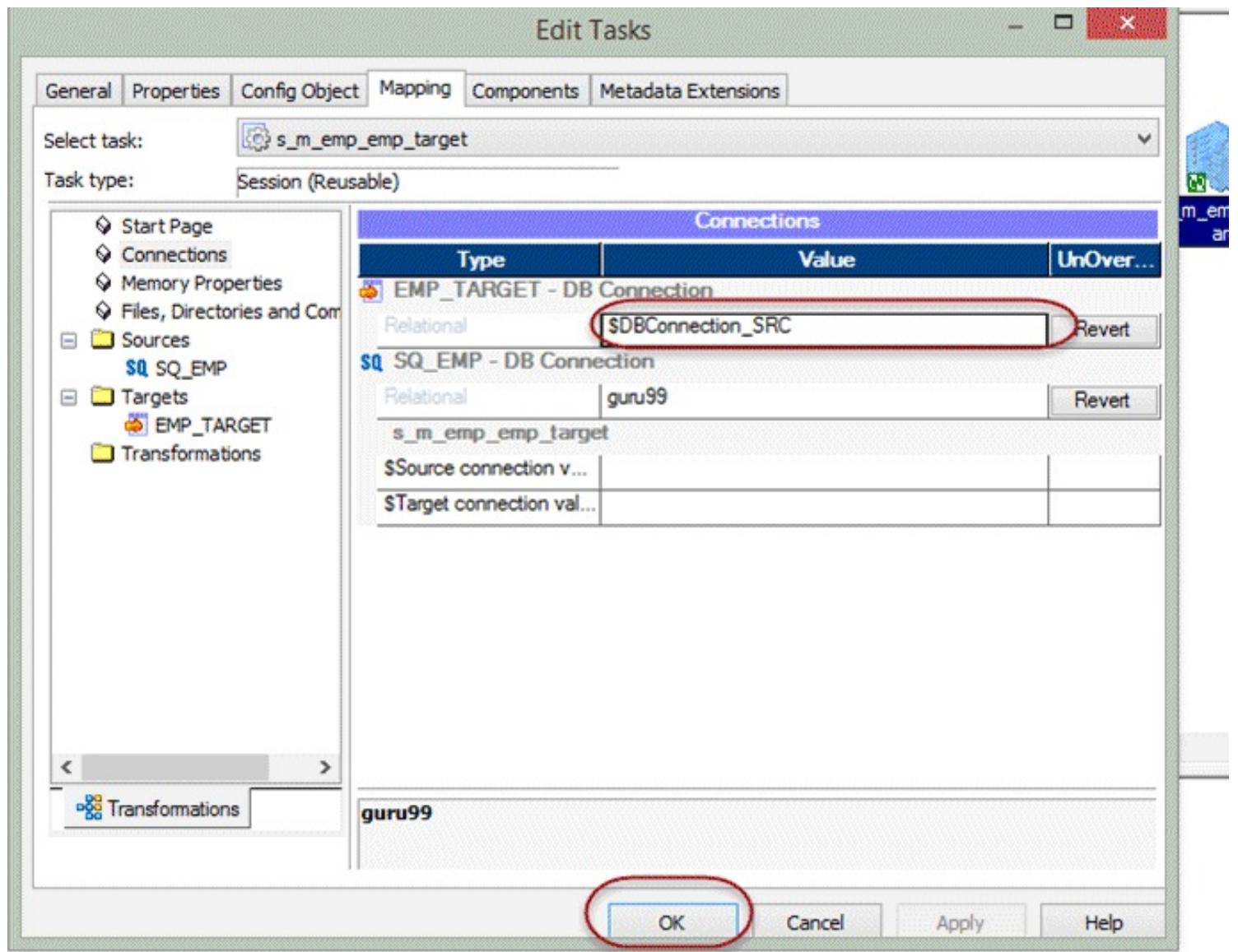


Step 7 - A connection browser window will appear, in that window

1. Select the option to use connection variable
2. Enter connection variable name as "\$DBConnection_SRC"
3. Select Ok Button



Step 8 – In the edit task window connection variable will appear for the target, Select OK button in the edit task window.



Now we are done with creating parameter for a connection and assigning its value to parameter file.

When we execute the workflow, the workflow picks the parameter file looks for the value of its parameters/variables in the parameter file and takes those values.

Workflow Monitor in Informatica

In our previous tutorial, we discussed on workflow -- which is nothing but a group of commands or instructions to the integration service. It defines how to run task like command task, session task, e-mail task, etc. To track everything is streamlined and executed in the desired order, we need a Workflow Monitor.

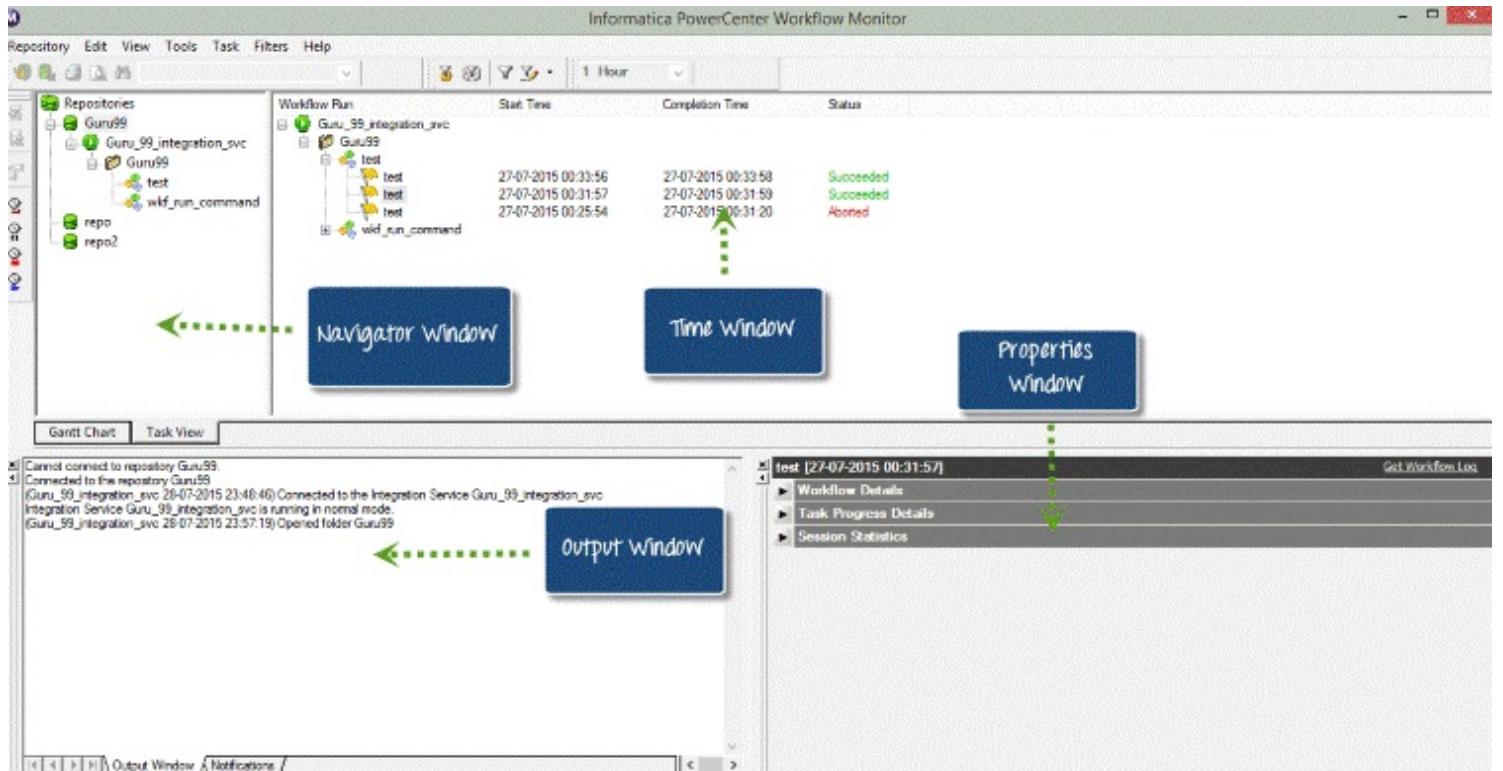
Workflow monitor is a tool with the help of which you can monitor the execution of workflows and tasks assigned to the workflow.

In workflow monitor you can,

- See the details of execution
- See the history of the workflow execution
- Stop, abort or restart workflows, and tasks
- Display the workflows those who are executed at least one time

Workflow monitor consists of following windows –

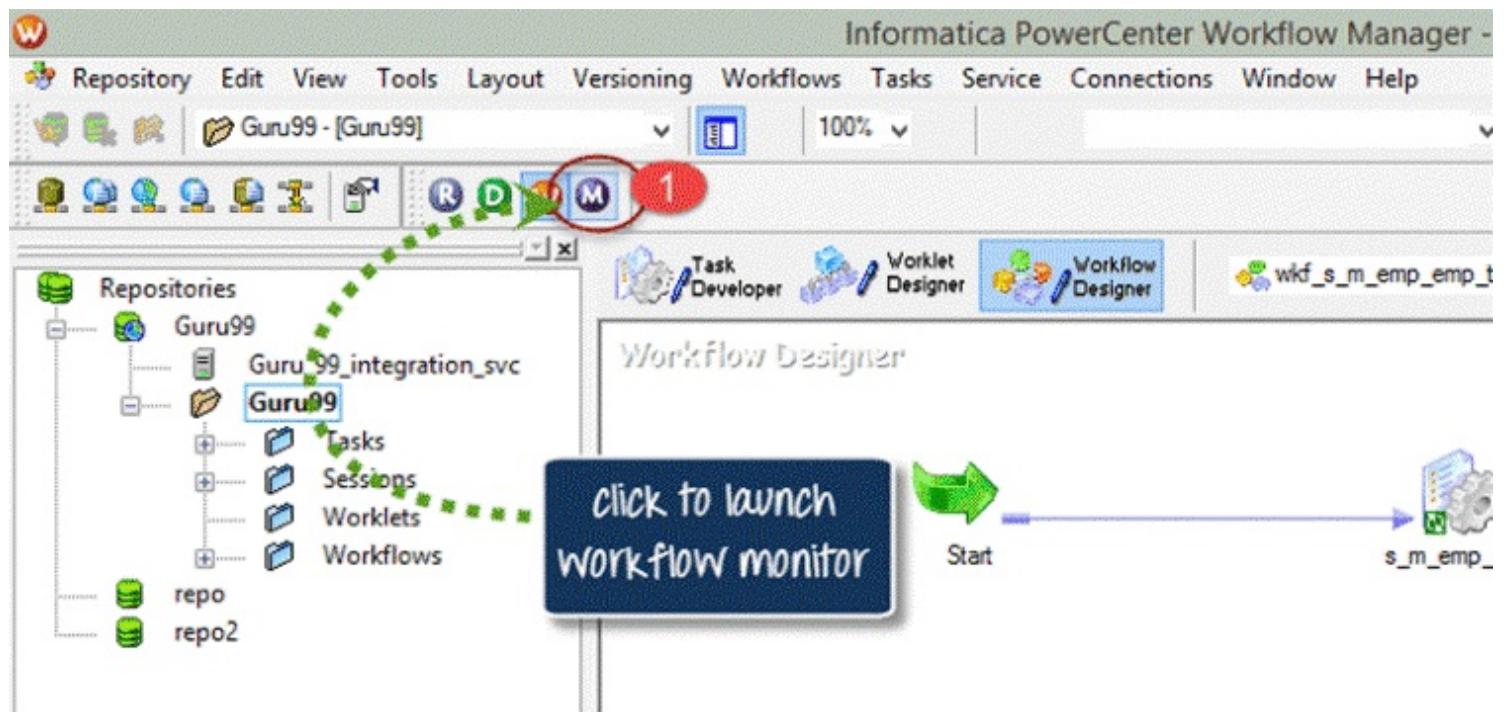
- Navigator window- shows the monitored repositories, folders & integration service
- Output window – displays the messages from integration services and repository
- Properties window – displays the details/properties about tasks and workflows
- Time window – displays the progress of the running tasks & workflows with timing details.



Now, let see what we can do in Workflow Monitor

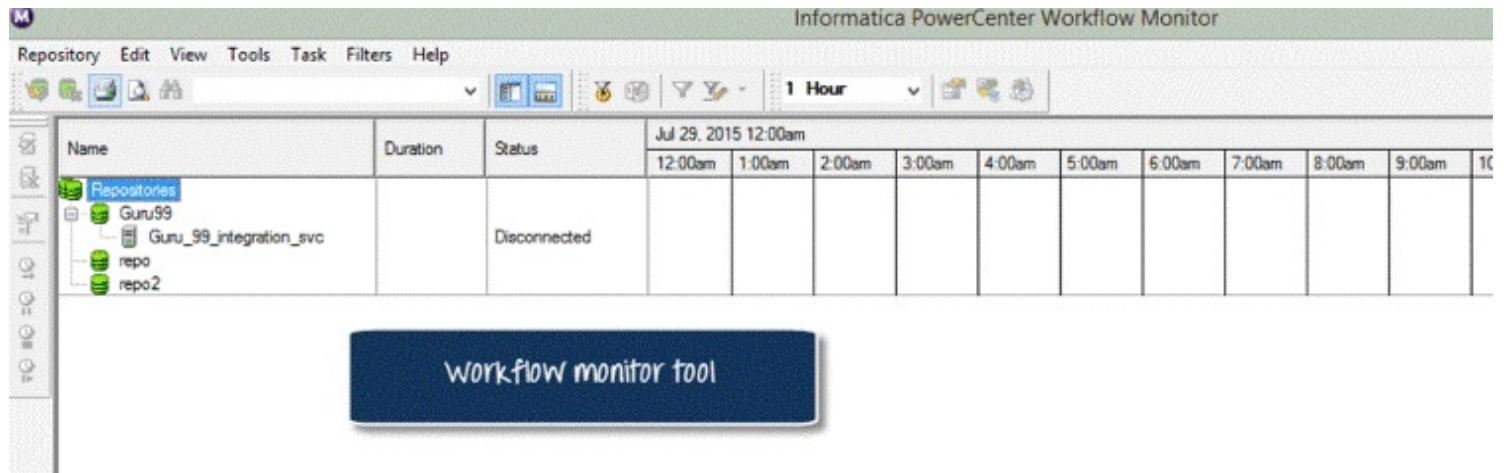
How to open Workflow Monitor

Step 1 – In Informatica Designer or Workflow manager toolbox, click on the workflow monitor icon



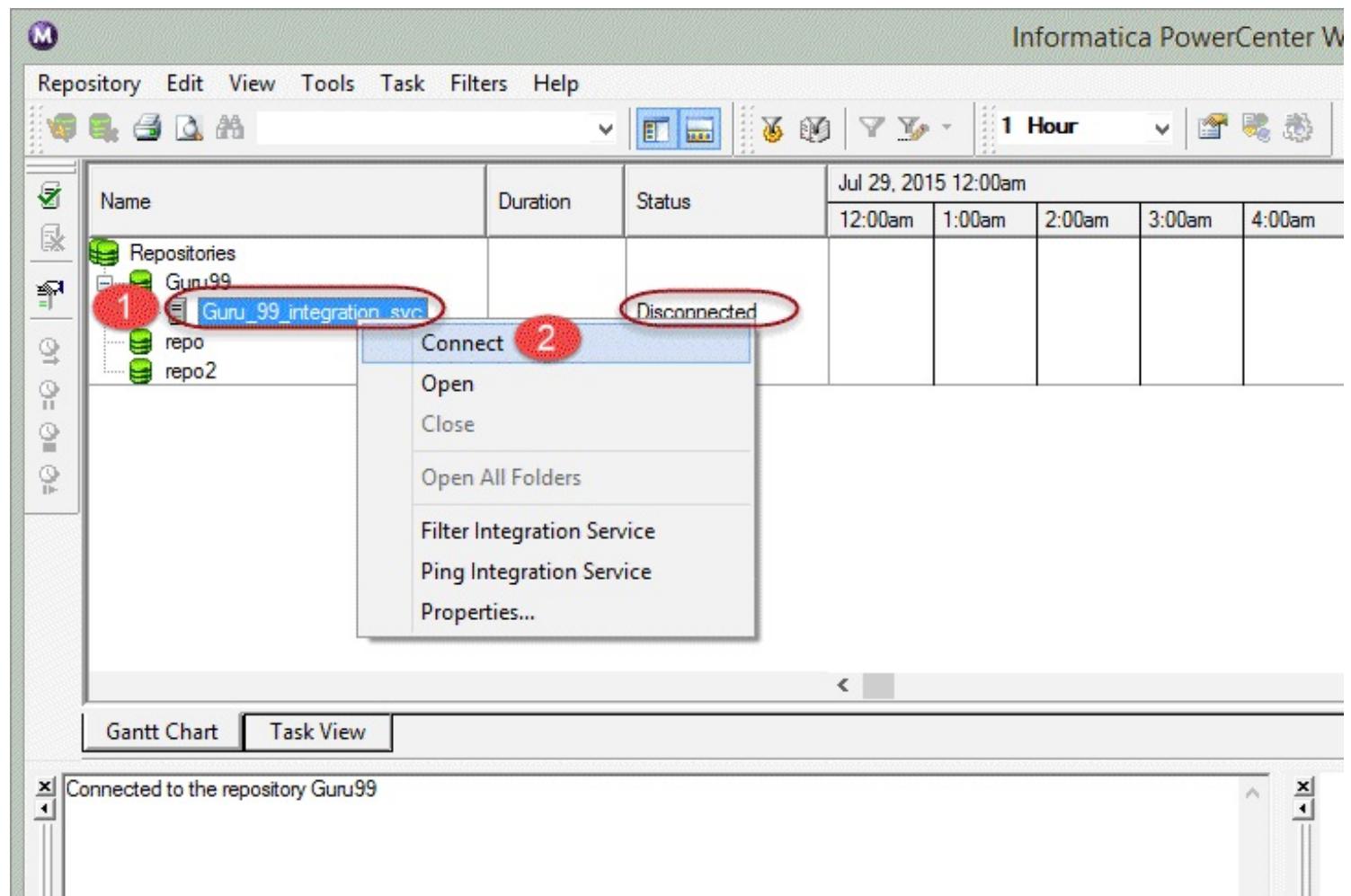
Step 2 – This will open workflow monitor window

In the workflow monitor tool, you will see the repositories and associated integration services on the left side. Under the status column, you will see whether you are connected or disconnected to integration service. If you are in the disconnected mode, you won't see any running workflows. There is a time bar which helps us to determine how long it took a task to execute.

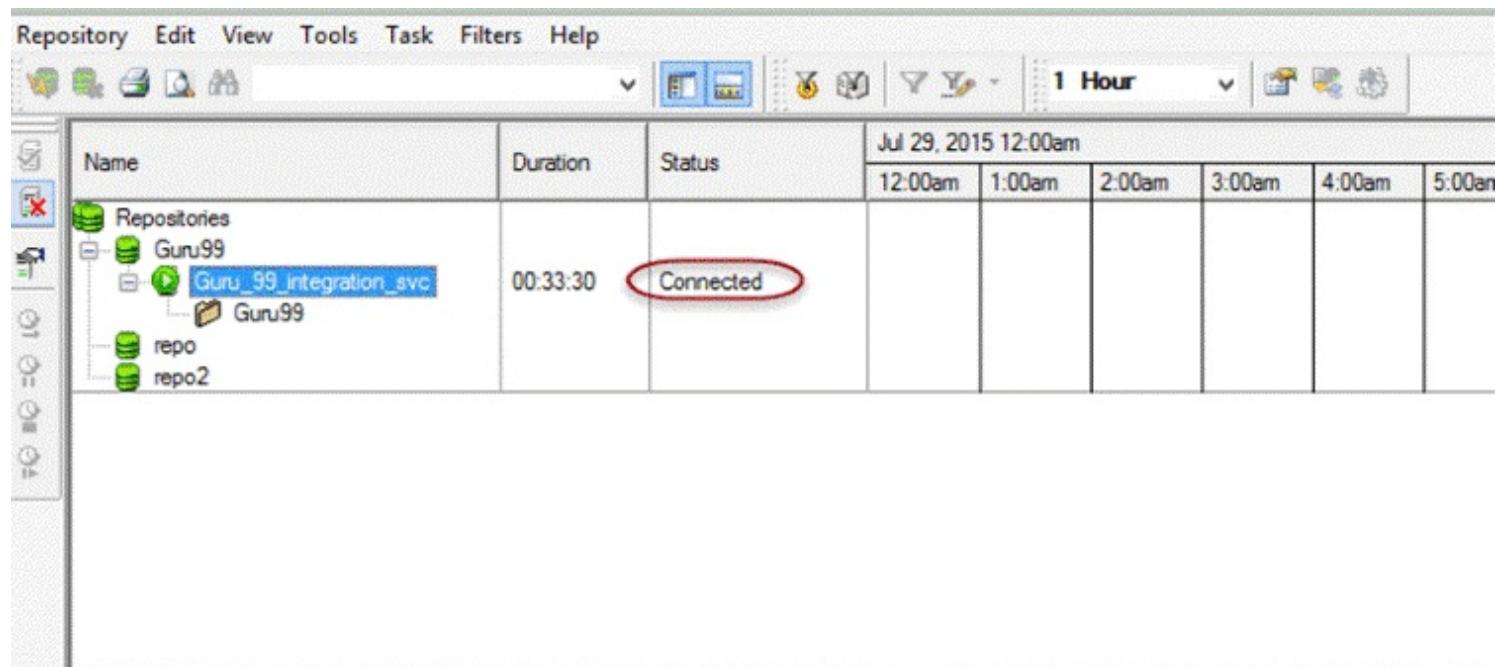


Step 3 – The workflow monitor is in a disconnected mode by default. To connect to integration service.

1. Right click on the integration service
2. Select connect option



After connecting, the monitor will show the status as connected.



Views in Workflow Monitor

There are two types of views available in Informatica workflow monitor

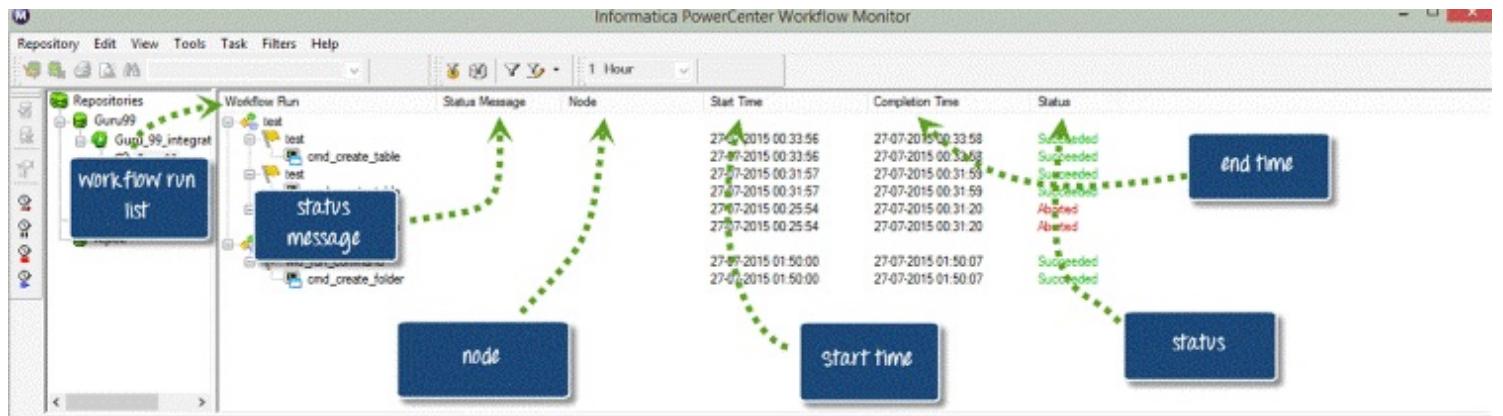
- Task view
- Gantt View

Task View

Task view displays the workflow runs in report format, and it is organized by workflow runs. It provides a convenient approach to compare workflow runs and filter details of workflow runs.

Task view shows the following details

- Workflow run list – Shows the list of workflow runs. It contains folder, workflow, worklet, and task names. It displays workflow runs in chronological order with the most recent run at the top. It displays folders and Integration Services alphabetically.
- Status message - Message from the Integration Service regarding the status of the task or workflow.
- Node - Node of the Integration Service executed the task.
- Start time - The time at which task or workflow started.
- Completion time – The time at which task or workflow completed the execution.
- Status - Shows status of the task or workflow, whether the workflow started, succeeded, failed or aborted.



Gantt Chart View

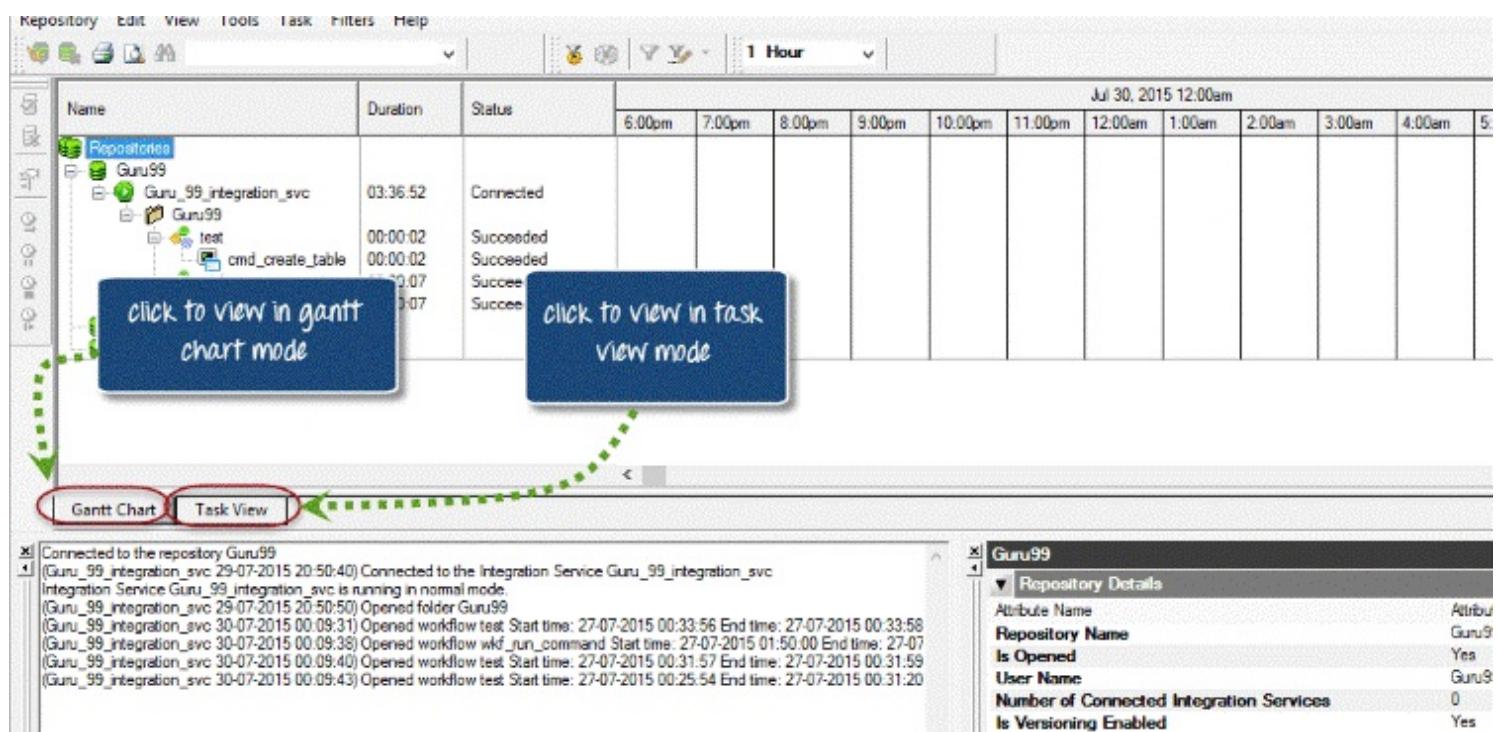
In Gantt chart view, you can view chronological view of the workflow runs. Gantt chart displays the following information.

- Task name – Name of the task in the workflow

- Duration – The time taken to execute the task
- Status – The most recent status of the task or workflow

To switch between Gantt chart and task views

To switch from Gantt chart to Task view or vice versa, click on the respective button as shown in the screenshot to change the mode.



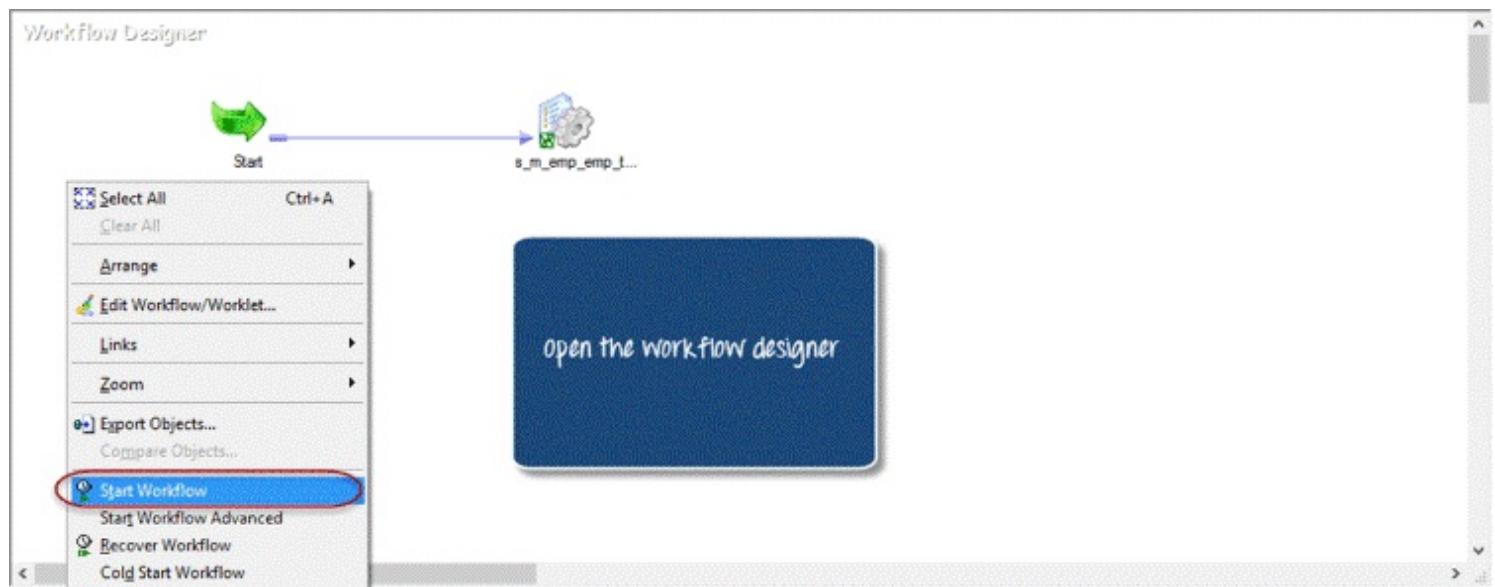
Example- How to monitor and view details

In previous examples, we have created a

- **Mapping "m_emp_emp_target"**: A mapping is a set of instructions on how to modify the data and processing of transformations that affects the record set.
- **Session "s_m_emp_emp_target"** : A session is a higher level object to a mapping which specifies the properties of execution. For example performance tuning options, connection details of sources/targets, etc.
- **Workflow "wkf_s_m_emp_emp_target"**: A workflow is a container for the session and other objects, and it defines the timing of the execution of tasks and the dependency or flow of execution.

Now, we will analyze the details of execution in this topic.

Step 1 – Restart the workflow designer, as described in previous topic



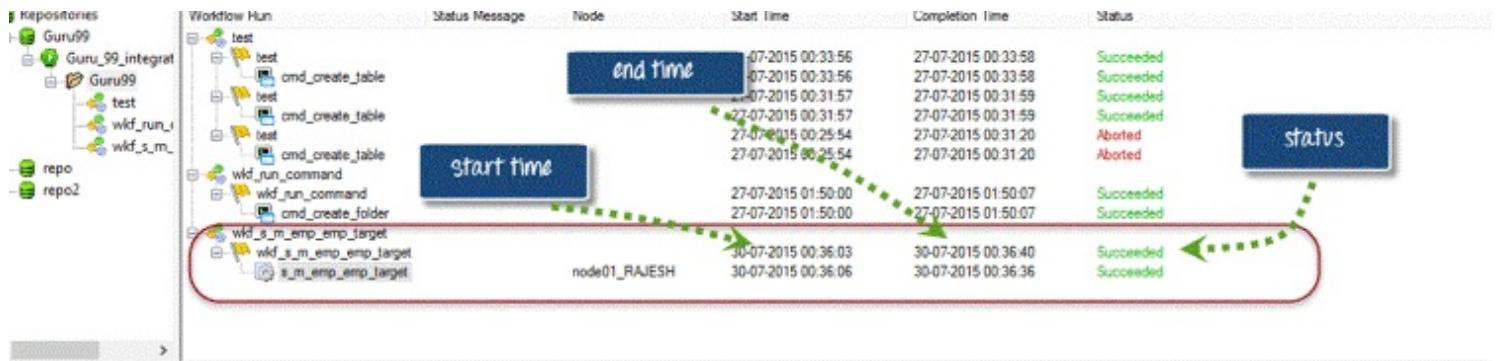
Step 2 – Go to workflow monitor and in the monitor window you will see details as shown in screen shot like repository, workflow run details, node details, workflow run start time, workflow run completion time and status.

Repository	Workflow Run	Status	Message	Node	Start Time	Completion Time	Status
Guru99	test	In progress	s_m_emp_emp_target	node01_RAJESH	30-07-2015 00:36:03	In progress	Running

Step 3 – Here you can view the currently running workflow, which is having status as "running".

Repository	Workflow Run	Status	Message	Node	Start Time	Completion Time	Status
Guru99	test	In progress	s_m_emp_emp_target	node01_RAJESH	30-07-2015 00:36:03	In progress	Running
		In progress	s_m_emp_emp_target	node01_RAJESH	30-07-2015 00:36:06	In progress	Running

Step 4 – Once the workflow execution completes, its status would change to succeeded/failed along with start and end time details.



Step 5 – To view the task details

1. Right click on task name
2. In the pop-up window select "get run properties"
3. A properties window would appear with the task details

1

2

3

Step 6 – Click on each of the menu of the properties window to view specific details.

Here we chose "Task Details" to view. It will display all the details like Instance Name, Task Type, Start Time, Integration Service Name, etc.

Task details -

The screenshot shows the 'Task Details' screen for task **s_m_emp_emp_target [30-07-2015 00:36:06]**. The 'Task Details' section is highlighted with a red box. A callout bubble labeled 'task details' points to this section. The 'Source/Target Statistics' section is also highlighted with a red box.

Attribute Name	Attribute Value
Instance Name	s_m_emp_emp_target
Task Type	Session
Integration Service Name	Guru_99_integration_svc
Node(s)	node01_RAJESH
Start Time	30-07-2015 00:36:06
End Time	30-07-2015 00:36:36
Recovery Time(s)	
Status	Succeeded
Status Message	
Deleted	No
Version Number	1

Source/Target Statistics

Statistic	Value
Applied Rows	14
Affected Rows	14
Rejected Rows	0
Total Throughput	14 rows/sec

Source and Target Statistics

Source and target statistics gives the details of source and target. For example, how many rows are fetched from the source and how many rows are populated in the target the current throughput, etc

In the following screen, 14 records are fetched from the source, and all 14 are populated in the target table.

- **Applied rows** signify how many records Informatica had tried to update or insert the target
- **Affected rows** signify how many numbers of applied rows were actually succeeded. Here all 14 rows are successfully loaded in the target, so the count is same for both.
- **Rejected rows** signify how many rows are dropped due to target constraint or other issues.

► Task Details**▼ Source/Target Statistics**

Transformation Name	Node	Applied Rows	Affected Rows	Rejected Rows	Throughput (Rows/Sec)	Throughput (MB/sec)
EMP_TARGET	node0...	14	14	0	2	154
SQ_EMP	node0...	14	14	0	14	1079

Source/target statistics

**► Partition Details****► Performance**

In this tutorial, you have learned how to open and monitor the workflows and tasks using workflow monitor.

Chapter 7: Debug Mappings & Session Objects

As we discussed in previous topics, a mapping is a collection of source and target objects linked together by a set of transformations. These transformations consist of a set of rules, which define the data flow and how the data is loaded into the targets.

The debugger is a very handy utility in Informatica. When we execute the mapping, Informatica executes the mapping based on the transformation logic defined.

Its execution is similar to batch mode execution where we cannot see the intermediate results, or how the data is getting modified from transformation to transformation.

Many a times we get the data in the target which is not as we expected. With the help of debugger, we **can analyze how the data is moving within the mapping. We can analyze the data row by row in a debugger.**

The use of debugger requires a session instance, we can either reuse an existing session which was created earlier to run the mapping, or we can create a new debug session instance. Since during debugging the data will be fetched from the source, so we have the option of what to do with that data.

- Either we can discard the data or
- The data can be loaded into the target

These options can be set during the configuration of the debugger.

To monitor the debugger, there are two windows in Informatica designer

- Target window
- Instance window

In target window, the target record which is going to be inserted into the target table is shown. In instance window, you can select any of the transformation instances of the mapping. After that, all the ports of that transformation are visible in this window, and you can analyze the data, how it is getting transformed.

Steps to use debugger in a mapping

Informatica designer provides the feature to use the debugger to debug mappings. It means we can debug and analyze any existing mapping. The requirement for the debugger is that the mapping should be a valid mapping.

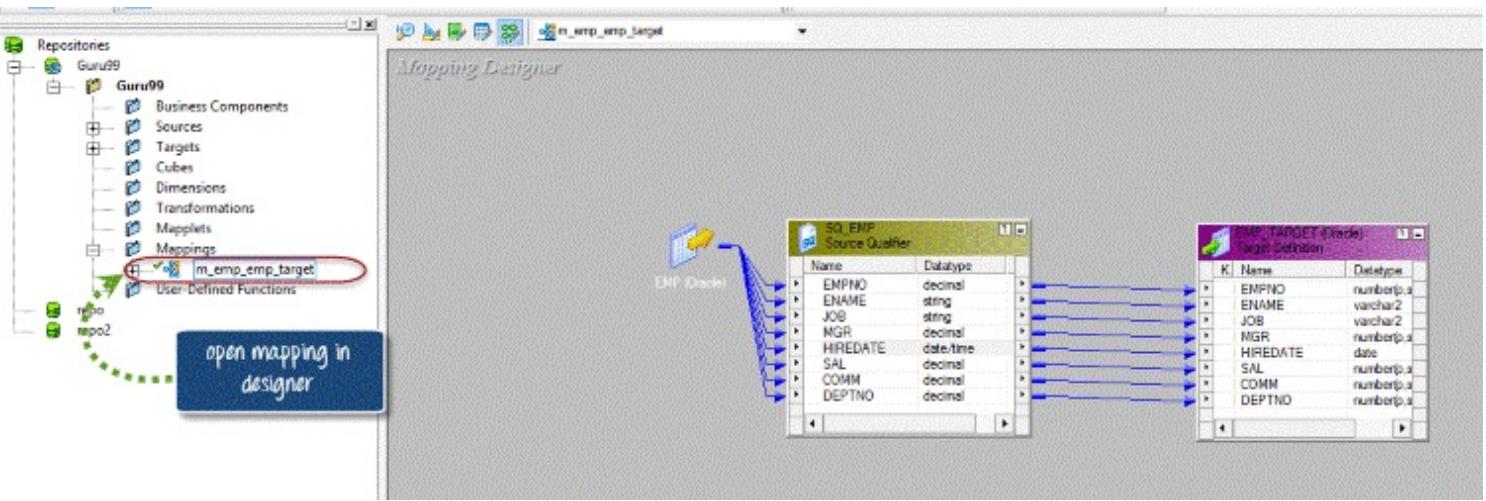
In the exercise, we will debug our existing mapping "m_emp_emp_target" which we created earlier. In our source data, we will check for a particular record of employee_name ='SCOTT'. For Scott, we will verify the values are fetched in Informatica and whether the record is pushed to the target correctly or not.

Our source data for the "emp" table looks like this.

EMPNO	ENAME	JOB	MGR	SAL	HIREDATE	DEPTNO
7369	SMITH	CLERK	7902	800	17-DEC-80	20
7499	ALLEN	SALESMAN	7698	1600	20-FEB-81	30
7521	WARD	SALESMAN	7698	1250	22-FEB-81	30
7566	JONES	MANAGER	7839	2975	02-APR-81	20
7654	MARTIN	SALESMAN	7698	1250	28-SEP-81	30
7698	BLAKE	MANAGER	7839	2850	01-MAY-81	30
7782	CLARK	MANAGER	7839	2450	09-JUN-81	10
7788	SCOTT	ANALYST	7566	3000	09-DEC-82	20
7839	KING	PRESIDENT		5000	17-NOV-81	10
7844	TURNER	SALESMAN	7698	1500	08-SEP-81	30
7876	ADAMS	CLERK	7788	1100	12-JAN-83	20
EMPNO	ENAME	JOB	MGR	SAL	HIREDATE	DEPTNO
7900	JAMES	CLERK	7698	950	03-DEC-81	30
7902	FORD	ANALYST	7566	3000	03-DEC-81	20
7934	MILLER	CLERK	7782	1300	23-JAN-82	10

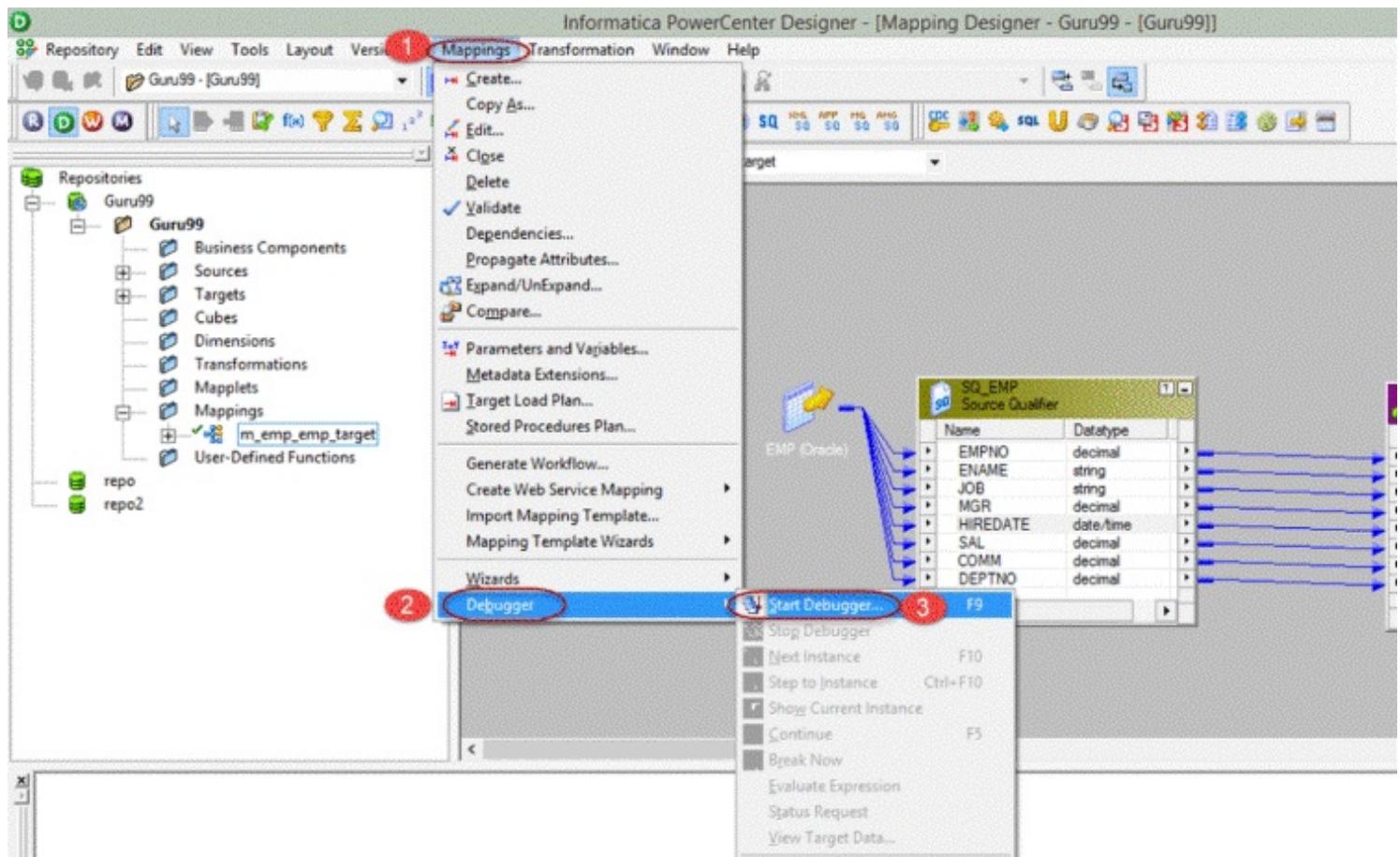
For the ename "SCOTT", using debugger we will check that what value for the "sal" (salary) column is fetched in the Informatica.

Step 1 – Open Informatica designer and in designer open the mapping "m_emp_emp_target" by double clicking on it. It will open source qualifier and target definition tables showing mapping inside Mapping Designer.

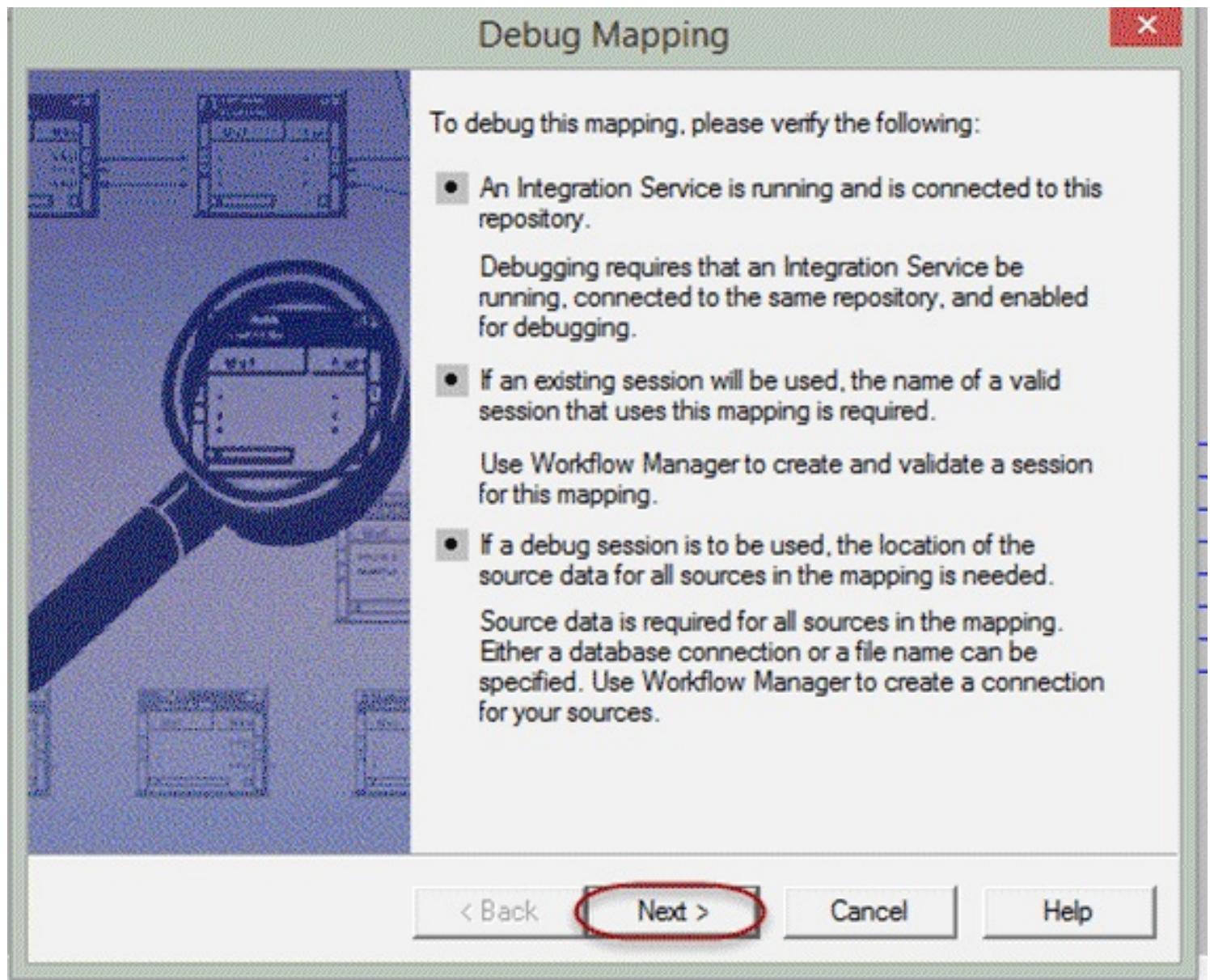


Step 2 – Now start the debugger

1. Click on the mappings menu option
2. Select Debugger from the options
3. Select "start debugger" option



Step 3 – Select next on debug mapping window



Step 4 – In the next window, you will get an option to select session type. You can use existing session, existing reusable session, or you can create a debug session instance.

When you use the existing session, Informatica will use the earlier session which you created for this mapping. When you select the create debug session instance, a new session instance will be created for debugging.

In this example, we will select existing reusable session instance. The integration service will be the default.

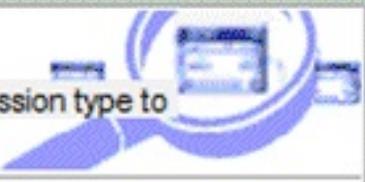
1. Select "use an existing session instance" option
2. Click next

Debug Mapping



Select Integration Service and Session Type

This allows you to specify which registered Integration Service and session type to use.



To debug this mapping, an Integration Service will be used to run the mapping in debug mode.

Please select an Integration Service to use:

Guru_99_integration_svc



Session

1

- Use an existing session instance
- Use an existing reusable session
- Create a debug session instance

A session is required to run the mapping on the Integration Service. You can either use an existing session or you can create a debug session.

If an existing session is used, all of its settings will be used by the Integration Service.

If a debug session is created, many of the session parameters cannot be specified and default values will be used.

2

< Back

Next >

Cancel

Help

Step 5 – In the new window, Informatica will prompt you to select the existing session for the debugging.

Select the session "s_m_emp_emp_target" and click next button

Debug Mapping



Select Session

This allows you to specify which session to use.

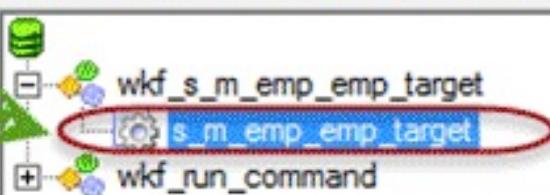


To debug this mapping, an existing session instance will be used.

The following valid sessions are available:

Select the session for debugger

Please select which session instance you would like to use to debug this mapping:



< Back

Next >

Cancel

Help

Step 6 - In the next window there will be an option "Discard target data", we select this option as during debugging we don't want to load any data in our target table.

1. Select option "Discard target data."
2. Select target table "emp_target"
3. Click finish button.

Target Options



Target Table Options

This allows you to select some of the target table options.



The data for the target tables can either be discarded or written to the target location. If the data is discarded, you can still view the data, row by row, while the mapping is running.

1

Discard target data

To view the target table data in real time, select the target tables below. Those target tables which are checked will be shown during the mapping run. This can be changed later when the mapping is being debugged.

2

EMP_TARGET

3

< Back

Finish

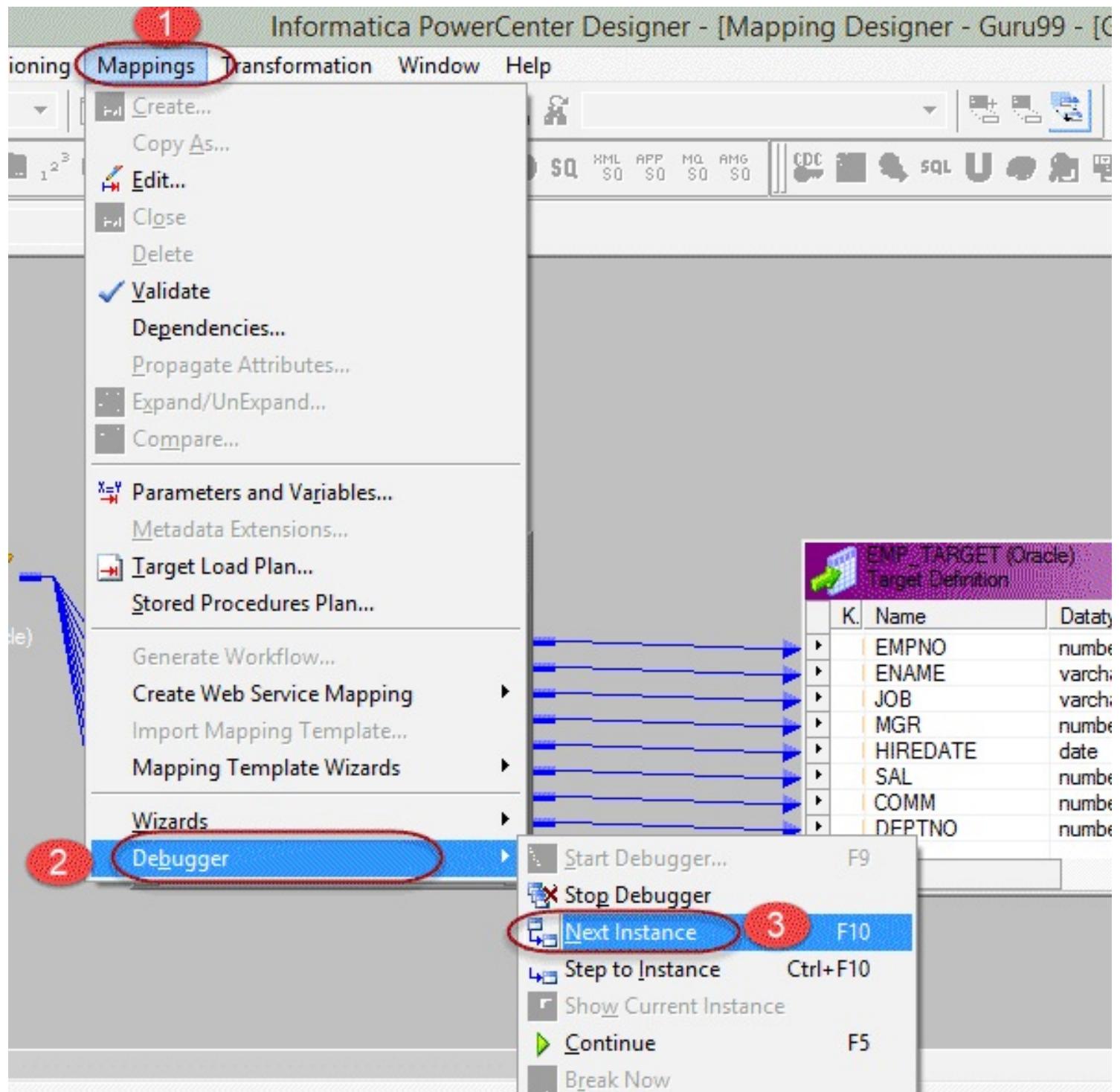
Cancel

Help

With the above steps, we have configured and started the debugger. Now we can debug the mapping.

Step 7 - The debugger has started, but it hasn't fetched any rows from the source. During debugging it fetches data from source row by row. To start fetch first row from the source

1. Select mappings option from the menu
2. Select debugger option
3. Select next instance option



Step 8 - In the instance window you will see the first fetched record from the source qualifier.

The screenshot shows the Oracle Data Integrator interface with two windows:

- target window**: Shows the source classifier for the EMP table (Oracle) with columns: Name, Datatype. The columns are: EMPNO (decimal), ENAME (string), JOB (string), MGR (decimal), HIREDATE (date/time), SAL (decimal), COMM (decimal), DEPTNO (decimal).
- instance window**: Shows the instance details for the SQ_EMP table. The columns are: K, Name, Datatype. The rows are: EMPNO (numberp,s), ENAME (varchar2), JOB (varchar2), MGR (numberp,s), HIREDATE (date), SAL (numberp,s), COMM (numberp,s), DEPTNO (numberp,s).

Below these windows is the mapping editor:

- Target Instance:** EMP_TARGET
- Instance:** SQ_SQL_EMP

Port	Value	Null	Data Type	RowID
EMPNO	7369.000000000000...	<input type="checkbox"/>	decimal	1
ENAME	SMITH	<input type="checkbox"/>	string	1
JOB	CLERK	<input type="checkbox"/>	string	1
MGR	7902.000000000000...	<input type="checkbox"/>	decimal	1
HIREDATE	12/17/1980 00:00:00	<input type="checkbox"/>	date/time	1
SAL	800.00	<input type="checkbox"/>	decimal	1
COMM	<No data available>	<input checked="" type="checkbox"/>	decimal	1
DEPTNO	20.00000000000000...	<input type="checkbox"/>	decimal	1
<<ROW TYPE>>	Insert			

If you analyze the instance window, it will be showing the first record – ename "SMITH" row. This isn't the record which we are looking for. So we can move to the next instance.

Step 9 - To go to next instance (to move to next row)

- Press F10 button (shortcut for going to next instance, instead of going to mapping option then debugger option)

After moving to next instance, the data in instance window will change to next record.

The screenshot shows the instance window for the SQ_SQL_EMP table after pressing F10. The data has changed to the second record:

Port	Value	Null	Data Type	RowID
EMPNO	7499.000000000000...	<input type="checkbox"/>	decimal	2
ENAME	ALLEN	<input type="checkbox"/>	string	2
JOB	SALESMAN	<input type="checkbox"/>	string	2
MGR	7698.000000000000...	<input type="checkbox"/>	decimal	2
HIREDATE	02/20/1981 00:00:00	<input type="checkbox"/>	date/time	2
SAL	1600.00	<input type="checkbox"/>	decimal	2
COMM	300.00	<input type="checkbox"/>	decimal	2
DEPTNO	30.00000000000000...	<input type="checkbox"/>	decimal	2
<<ROW TYPE>>	Insert			

Step 10 - Keep pressing F10 button, till you reach to the required record of "SCOTT" employee. Once you reach that point, the instance window will be like as shown below

Instance: SQ SQ_EMP

Port	Value	Null	Data Type	RowID
EMPNO	7788.0000000000...	<input type="checkbox"/>	decimal	8
ENAME	SCOTT	<input checked="" type="checkbox"/>	string	8
JOB	ANALYST	<input type="checkbox"/>	string	8
MGR	7566.0000000000...	<input type="checkbox"/>	decimal	8
HIREDATE	12/09/1982 00:0:0...	<input type="checkbox"/>	date/time	8
SAL	3000.00	<input type="checkbox"/>	decimal	8
COMM	<No data availa...	<input checked="" type="checkbox"/>	decimal	8
DEPTNO	20.000000000000...	<input type="checkbox"/>	decimal	8
<<ROW TYPE>>	Insert			

This is the instance, which we were looking for. So in the instance window, we can analyze the data. Whether it is coming correctly, or any value is missing.

In the source data, the details of the "SCOTT" employee were like job = "ANALYST", manager id = "7566", sal "3000."

And in the debugger, for the employee "SCOTT" we have analyzed all these details, and they are matching with the source data, so we are assured that the records were fetched correctly by source qualifier.

Using breakpoint in debugger

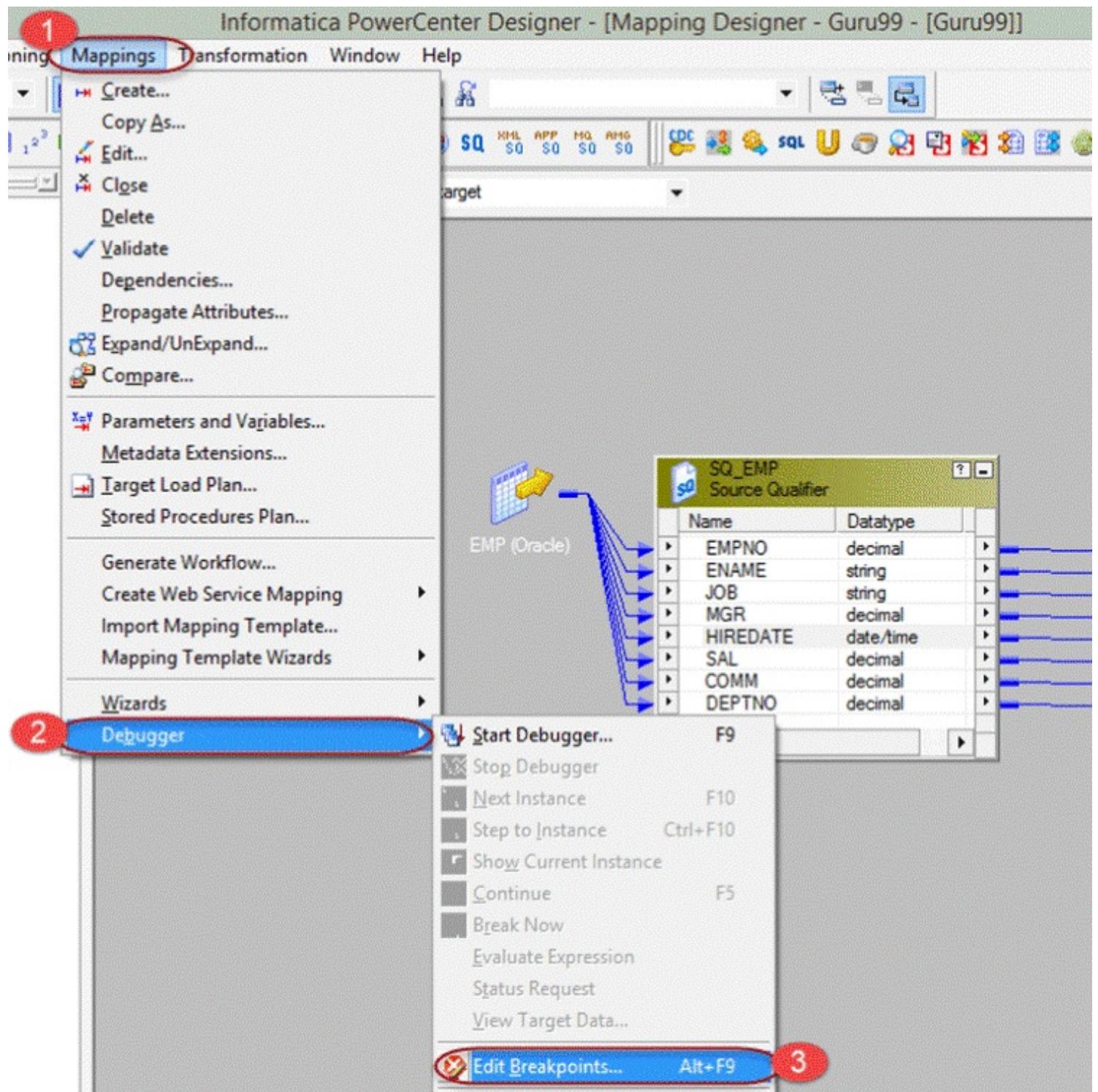
When we are looking for a particular row of data like in the previous example, then instead of going to every instance of source data we can define breakpoints.

A breakpoint is a condition or a set of a condition defined during configuration of debugging such that when that condition becomes true the debugger halts to that particular instance.

For example if we define deptno=20 in a breakpoint, then whenever for a record if this condition is true, then the debugger will stop to that instance of data.

Step 1- To configure the breakpoint

1. Go to mappings menu
2. Select debugger option
3. Select "edit breakpoints" option



Step 2 – You will see an "edit breakpoint" window as shown below

Edit Breakpoints

Instance name:

<<***** ALL INSTANCES *****>>

OK

Cancel

Help

Breakpoint type

Data

Number of rows to skip before stopping:

Error

Number of errors to skip before stopping:

Condition



Port

Operator

Type

Value

Breakpoints:

Add

Remove

Remove All

Step 3 – In the breakpoint window, click on the add button

Edit Breakpoints

Instance name: <<***** ALL INSTANCES *****>> ▾

Breakpoint type:

Data Number of rows to skip before stopping: []

Error Number of errors to skip before stopping: []

Condition

Port	Operator	Type	Value
------	----------	------	-------

Breakpoints:

Add
Remove
Remove All

Step 4 – In the new window

1. Select the SQ_EMP as instance
2. Select OK button

Edit Breakpoints

Instance name:

<<***** ALL INSTANCES *****>>

OK

Breakpoint type

Cancel

Select Instance

Select the Instance to add a breakpoint to:

<<***** GLOBAL BREAKPOINTS *****>>

\$[SQ_EMP]

1

2

OK

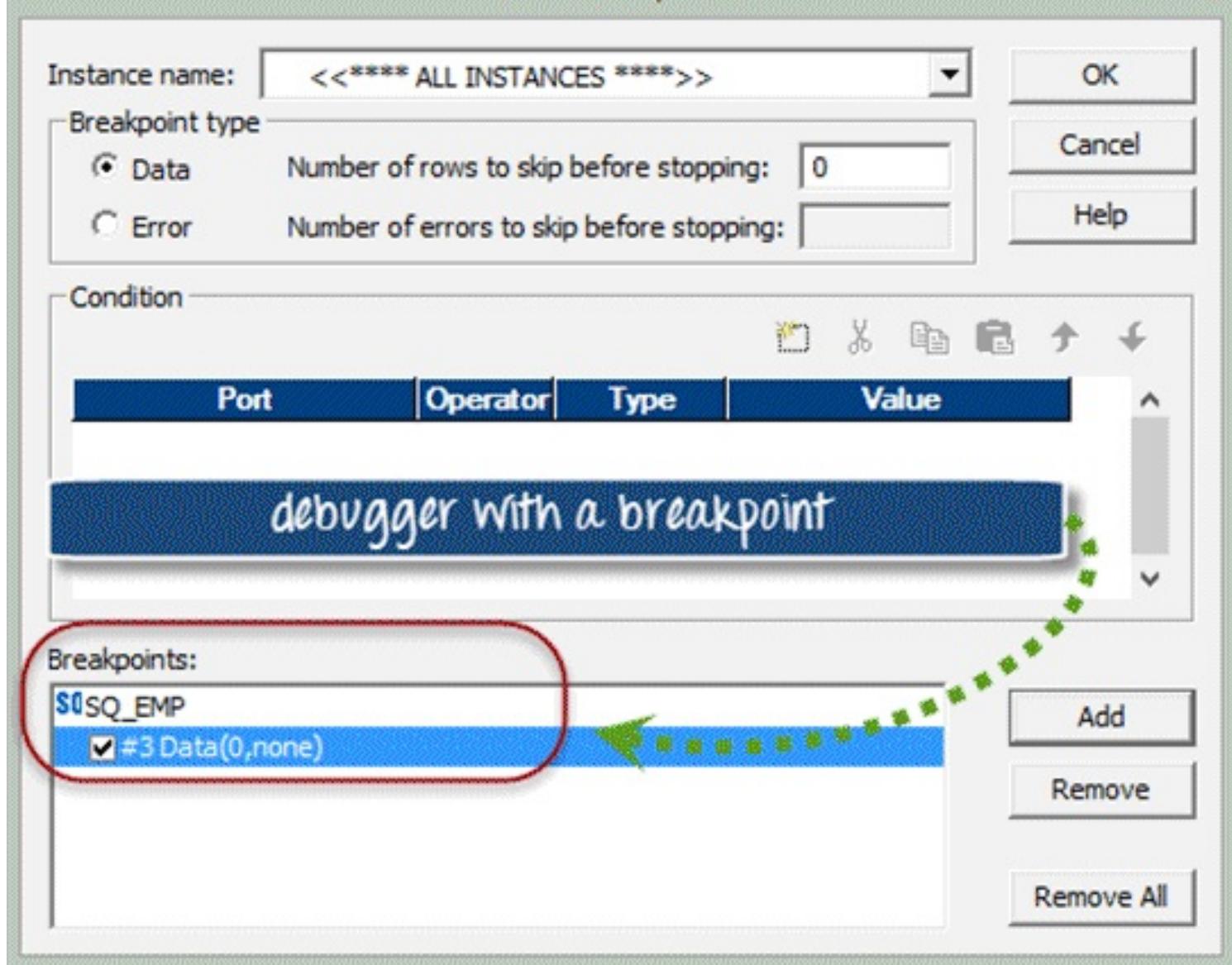
Cancel

Remove

Remove All

Step 5 – The previous window will look as shown below

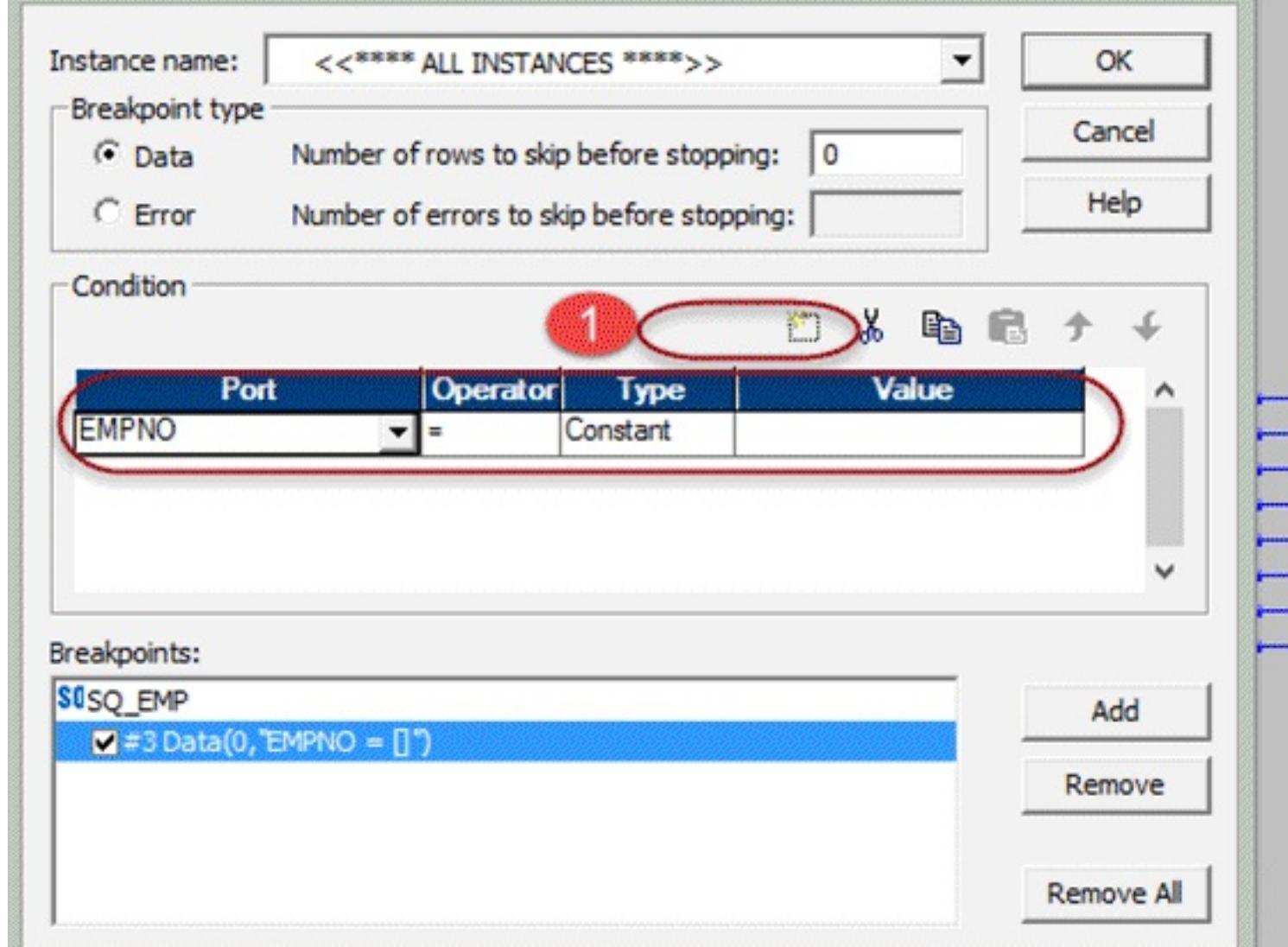
Edit Breakpoints



Step 6 – You have a breakpoint, now you have to define the breakpoint condition. To define breakpoint condition

1. Click on "add new condition" icon

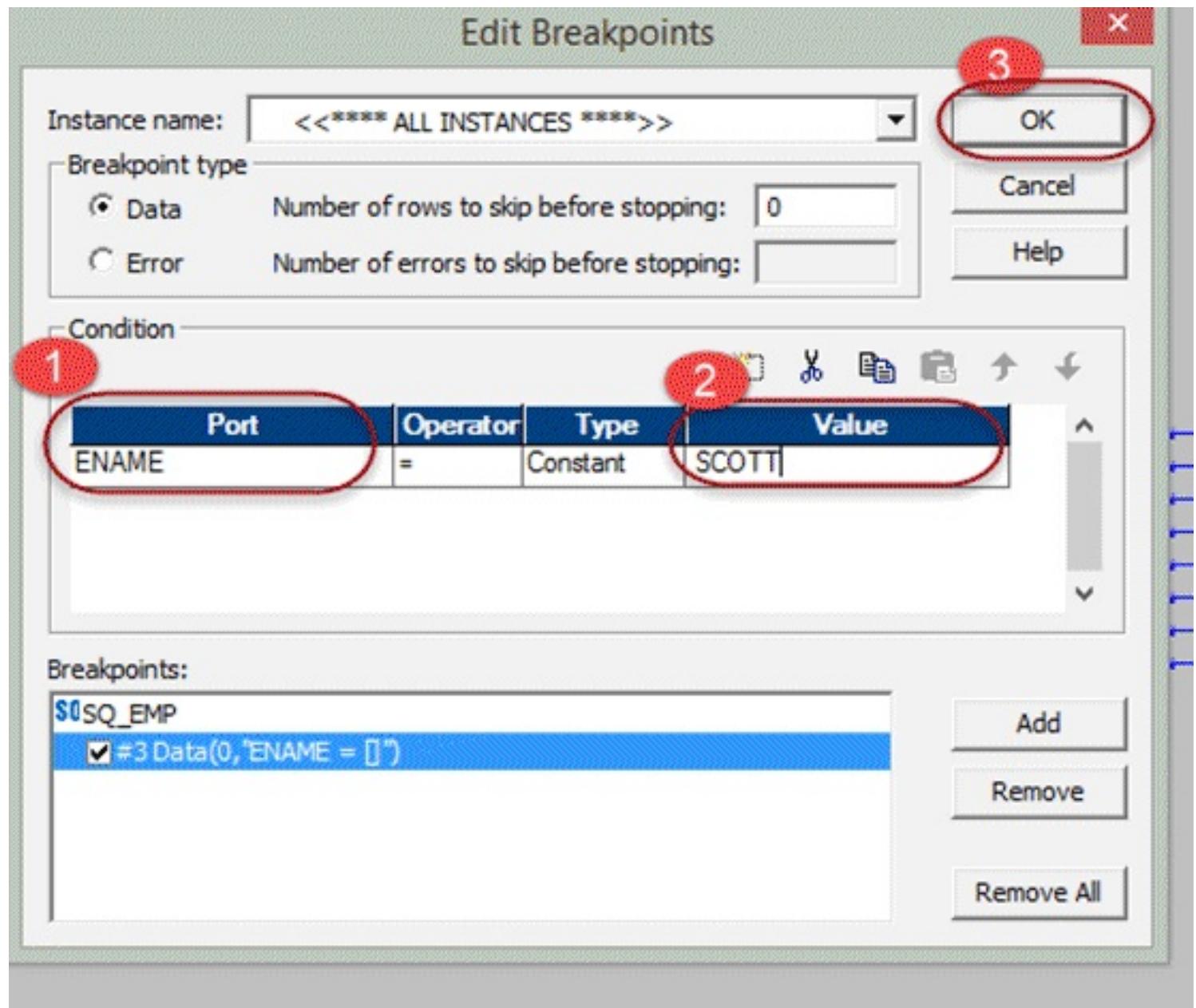
Edit Breakpoints



A new row under condition tab will appear. Here we can configure the condition.

Step 7 - In the condition table

1. In port drop down, select "ENAME" option (as we are debugging for record ename='SCOTT')
2. In the value box, enter value "SCOTT."
3. Select OK button

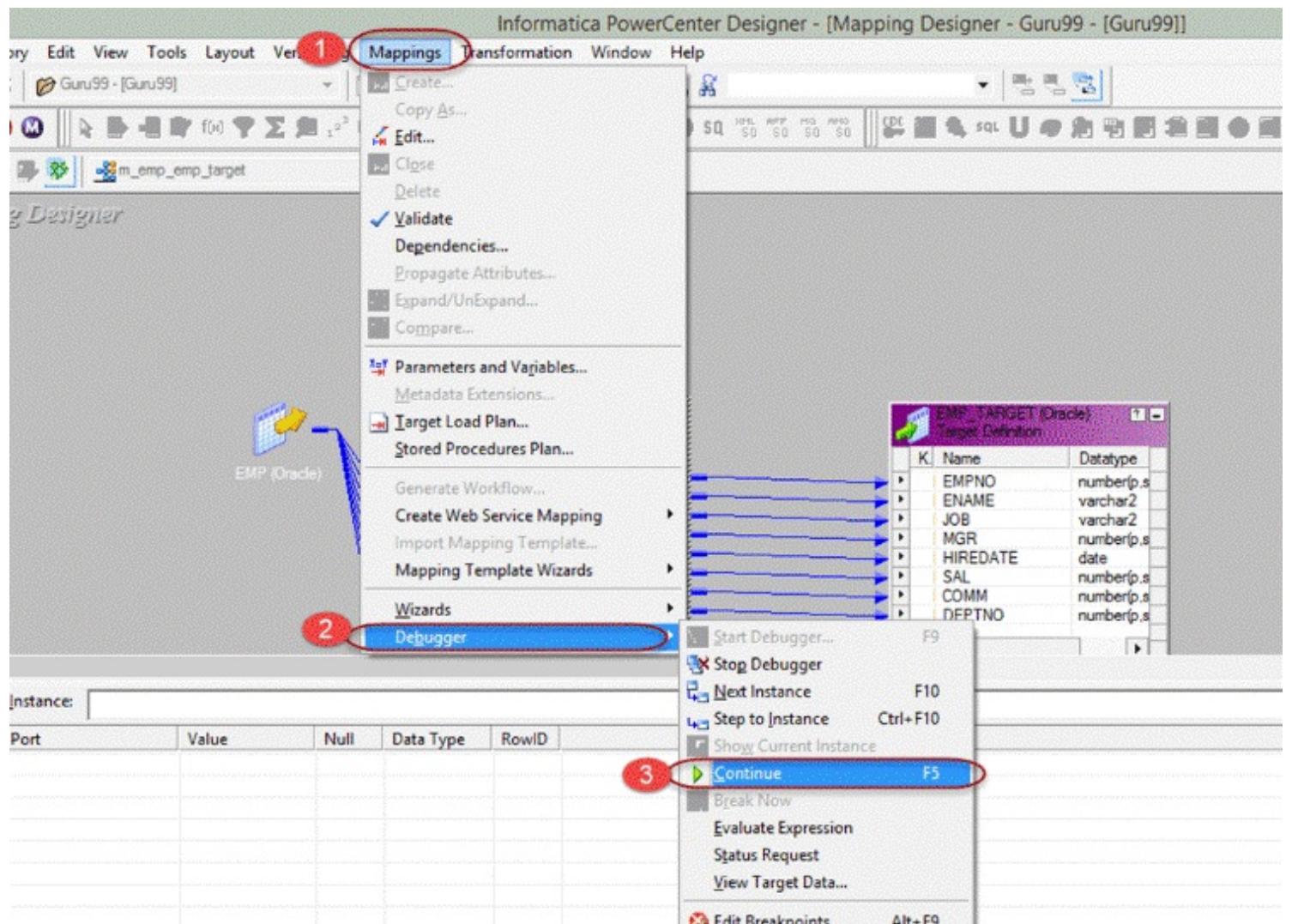


Step 8 – Now start the debugger as mentioned in the previous topic

Go to mapping menu -> select debugger option -> select start debugger -> select session instance

Step 9 – Once debugger is started

1. Select mappings menu
2. Select debugger option
3. Select continue option



The debugger will go to each instance of row itself and will check for the condition "ename=SCOTT". The record for which this condition is true, the debugger will halt on that instance. So when the data volume is high, then instead of going for each instance of source record you can configure a breakpoint, and you can directly jump to that record.

Port	Value	Null	Data Type	RowID
EMPNO	7788.0000000000...		decimal	8
ENAME	SCOTT		string	8
JOB	ANALYST		string	8
MGR	7566.0000000000...		decimal	8
HIREDATE	12/09/1982 00:00:00		date/time	8
SAL	3000.00		decimal	8
COMM	<No data availa...		decimal	8
DEPTNO	20.000000000000...		decimal	8
<<ROW TYPE>>	Insert			

Once the Informatica is done with going through all the instances, the debugger is stopped automatically. If you want to analyze the data again, you have to restart the

debugger again.

Tracing in Informatica

Whenever you execute a session for a mapping, a session log is generated. It consists of the details of the execution like

- Timing of the events
- Error details
- Individual transformation statistics etc.

Inside the transformations of a mapping, you can configure and set how much details Informatica writes inside the session log.

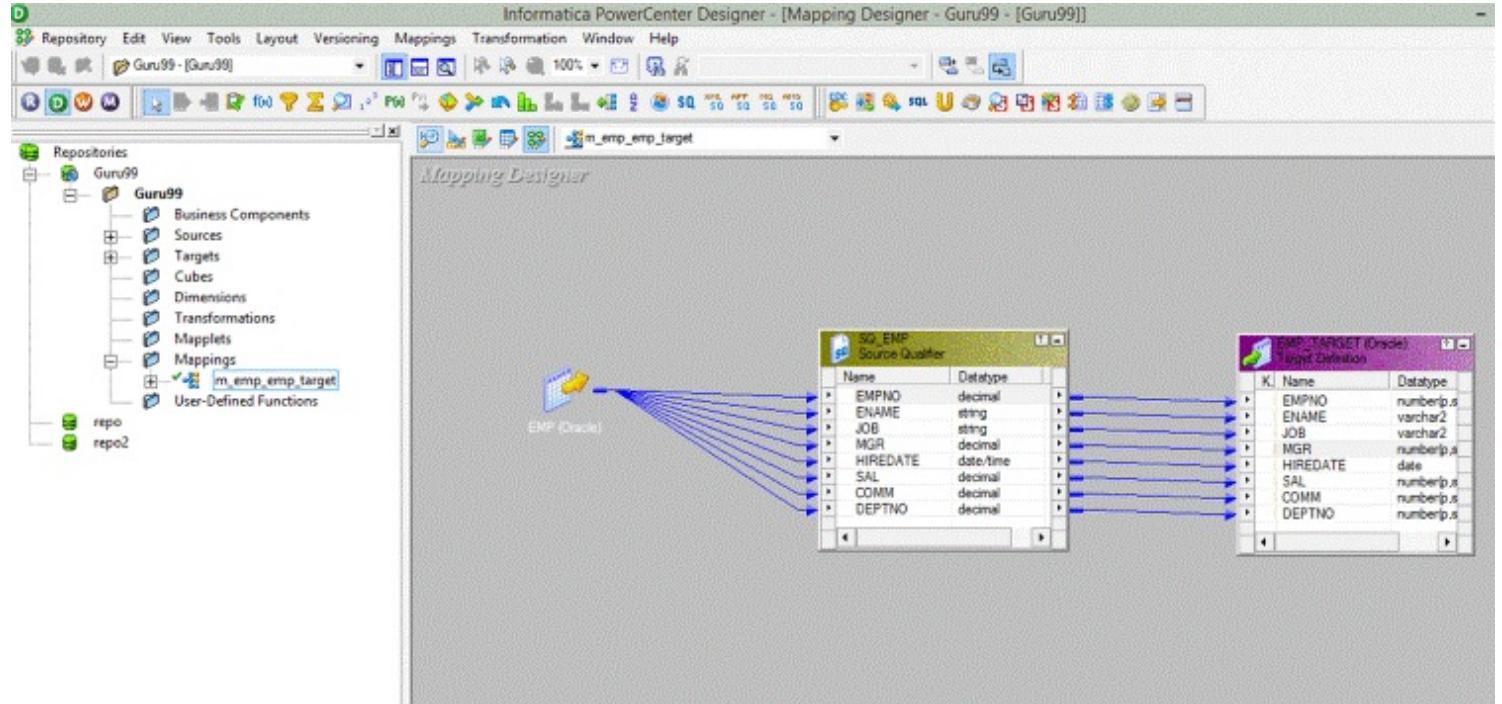
There are four levels which can be set for tracing inside the transformations.

Different Log Modes	Activities in different Log Modes
• Normal	<ul style="list-style-type: none">• In normal mode, Informatica writes about the transformation initialization and status information, errors encountered and summarizes sessions result.
• Terse	<ul style="list-style-type: none">• In Terse Mode, Informatica writes only about initialization information and error details. Terse mode writes minimum details into the session log.
• Verbose initialization	<ul style="list-style-type: none">• In Verbose Initialization mode, Informatica writes the log details similar to normal mode and in addition also writes about the data and index files used and transformation statistics.
• Verbose data	<ul style="list-style-type: none">• In Verbose data, Informatica writes log similar to verbose initialization but in addition also writes every data which passes through the transformation.

Note: By default the tracing level is set to normal, it is recommended to use verbose data mode only when you are debugging the mapping. The terse mode can be used to improve the performance, as it writes the minimum details. So the overhead of writing log will be minimum.

How to set tracing level in a transformation

Step 1 – Open the mapping in Informatica designer, for which you want to set the tracing level



Step 2 – Double click on the transformation (Source Qualifier transformation "SQ_EMP")

It will open edit transformation window.

Edit Transformations

Transformation Ports Properties Sources Metadata Extensions

Select transformation: SQ SQ_EMP

Transformation type: Source Qualifier

	Port Name	Datatype	Prec	Scale	I	O
1	EMPNO	decimal	4	0	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
2	ENAME	string	10	0	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
3	JOB	string	9	0	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
4	MGR	decimal	4	0	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
5	HIREDATE	date/time	29	9	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
6	SAL	decimal	7	2	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
7	COMM	decimal	7	2	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
8	DEPTNO	decimal	2	0	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>

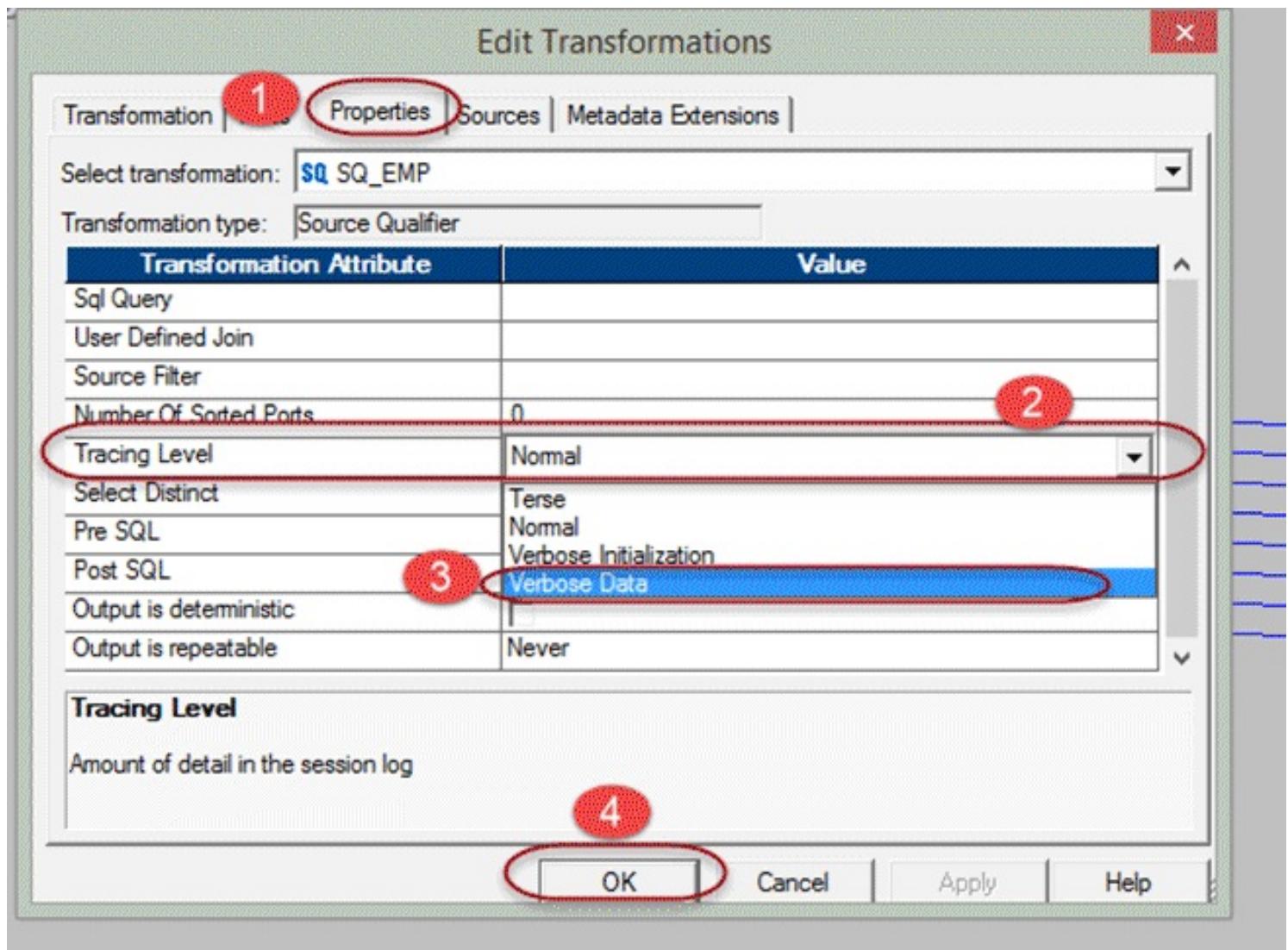
Default value:

Description:

OK Cancel Apply Help

Step 3 – In the edit transformation window

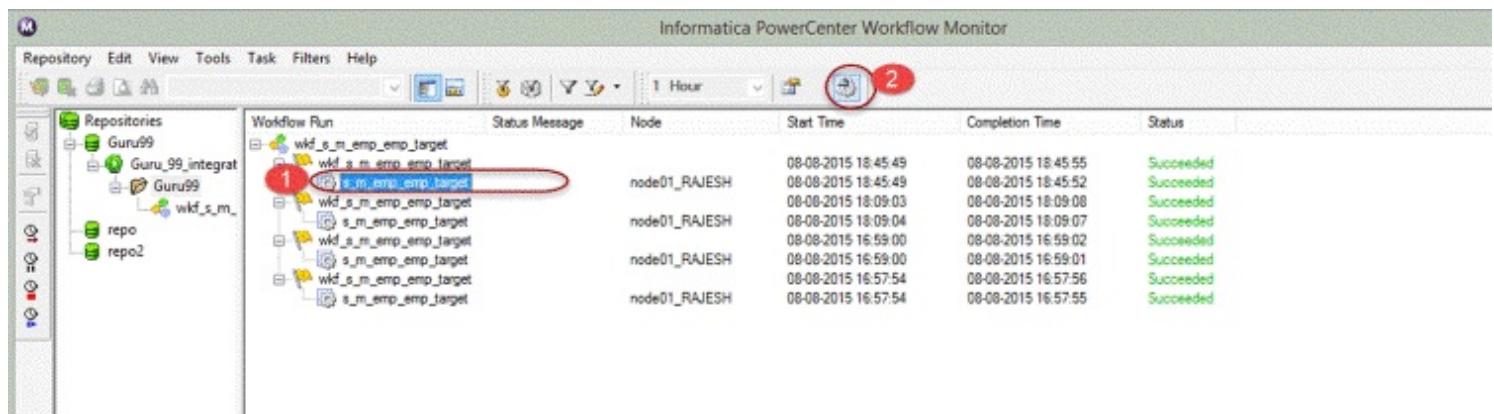
1. Click the properties tab
2. Select the Tracing level option
3. From the drop down select Verbose data
4. Select OK button



Step 4 – Save the mapping and execute the mapping (use **ctrl+s** keyboard shortcut to save the mapping)

Step 5- To view the log

1. Open workflow monitor and select the session which was executed in the last step
2. Click on "session log" option to open the session log for the mapping



This will open session log window for the mapping

The session log provides the detail about how your session got executed. It provides the timing details when execution started and stopped. It provides the basic idea about the performance. It mentions which database connection you are using, what parameter file you are using. It also summarizes about the source and targets by mentioning how many source records were fetched, how many records were loaded into the target, etc.

Log Events for Session: s_m_emp.emp_target					
Save As...	Copy	Find...	Refresh		
Severity	Timestamp	Node	Thread	Message Co.	Message
INFO	08-08-2015 18:45:49	node01_RAJEE	DIRECTOR	VAR_27028	Use persisted repository value [0.00000000000000] for mapping variable[\$\$TotalSalary].
INFO	08-08-2015 18:45:49	node01_RAJEE	DIRECTOR	TM_6014	Initializing session [s_m_emp.emp_target] at [Sat Aug 08 18:45:49 2015].
INFO	08-08-2015 18:45:49	node01_RAJEE	DIRECTOR	TM_6633	Repository Name: [Guru99]
INFO	08-08-2015 18:45:49	node01_RAJEE	DIRECTOR	TM_6694	Server Name: [Guru_99_integration_svc]
INFO	08-08-2015 18:45:49	node01_RAJEE	DIRECTOR	TM_6696	Folder: [Guru99]
INFO	08-08-2015 18:45:49	node01_RAJEE	DIRECTOR	TM_6698	Workflow: [obj_s_m_emp.emp_target] Run Instance Name: [] Run Id: [16]
INFO	08-08-2015 18:45:49	node01_RAJEE	DIRECTOR	TM_5101	Mapping name: m_emp.emp_target [version CheckedOut].
INFO	08-08-2015 18:45:49	node01_RAJEE	DIRECTOR	TM_5964	Date format for the Session is [MM/DD/YYYY HH24.MI.SS.US]
INFO	08-08-2015 18:45:49	node01_RAJEE	DIRECTOR	TM_6708	Using configuration property [EnableDataEncryption=no]
INFO	08-08-2015 18:45:49	node01_RAJEE	DIRECTOR	TM_6708	Using configuration property [StoreHPersistenceInDB=no]
INFO	08-08-2015 18:45:49	node01_RAJEE	DIRECTOR	TM_5703	Session [s_m_emp.emp_target] is run by 64-bit Integration Service [node01_RAJEE], version [9.6.1], build [607].
INFO	08-08-2015 18:45:49	node01_RAJEE	MANAGER	PETL_24058	Running Partition Group [1].
INFO	08-08-2015 18:45:49	node01_RAJEE	MANAGER		
INFO	08-08-2015 18:45:49	node01_RAJEE	MANAGER		
INFO	08-08-2015 18:45:49	node01_RAJEE	MAPPING	TM_5105	Using low precision processing.
INFO	08-08-2015 18:45:49	node01_RAJEE	MAPPING	TM_6180	Deadlock retry logic will not be implemented.
INFO	08-08-2015 18:45:49	node01_RAJEE	MAPPING	TM_5187	Session target-based commit interval is [10000].
INFO	08-08-2015 18:45:49	node01_RAJEE	MAPPING	TM_5307	DTM error log disabled.
INFO	08-08-2015 18:45:49	node01_RAJEE	MAPPING	TE_7022	TShimWriter initialized.
INFO	08-08-2015 18:45:49	node01_RAJEE	MAPPING	CMM_1053	SQ_EMP Exchange Row Definition Info ([0] Names(EMPNO->EMPNO) Native-type: 3005 (number(p,s)) C-type: 1005 (Double) Precision:4 Scale: 0 [1] Names(ENAME->ENAME) Native-type: 3005 (varchar2) C-type: 1003 (Char) Precision: 10 Scale: 0 [2] Names(JOB->JOB) Native-type: 3009 (varchar2) C-type: 1003 (Char) Precision: 9 Scale: 0 [3] Names(MGR->MGR) Native-type: 3005 (number(p,s)) C-type: 1005 (Double) Precision:4 Scale: 0 [4] Names(HREDATE->HREDATE) Native-type: 3002 (date) C-type: 1007 (Date/Time) Precision: 29 Scale: 9 [5] Names(SAL->SAL) Native-type: 3005 (number(p,s)) C-type: 1005 (Double) Precision:4 Scale: 0)

Step 6 – Scroll down in the log, and you can see additional log details captured including the data records.

Log Events for Session: s_m_emp.emp_target

Event	Timestamp	Node	Thread	Message Content
INFO	08-08-2015 18:45:49	node01_RAJE	READER_1_1_1	RR_4050 First row returned from database to reader : (Sat Aug 08 18:45:49 2015) Read [14] rows, read [0] error rows for source table [EMP] instance name [EMP]
INFO	08-08-2015 18:45:49	node01_RAJE	READER_1_1_1	SQ_EMP Exchange: Control-[EOF]
INFO	08-08-2015 18:45:49	node01_RAJE	READER_1_1_1	CMN_1053 SQ_EMP Exchange: Rowdata: (RowType=0[insert]) Src Rowid=1 Targ Rowid=1 EMPNO (EMPNO Double) : "7369.000000000000" ENAME (ENAME Char 10) : "SMITH" JOB (JOB Char 9) : "CLERK" MGR (MGR Double) : "7802.000000000000" HIREDATE (HIREDATE Date) : "12/17/1980 00:00:00.000000000" SAL (SAL Double) : "300.000000000000" COMM (COMM Double) : "NULL" DEPTNO (DEPTNO Double) : "20.00000000000000"
				session log in verbose data mode
				SQ_EMP Exchange: Rowdata: (RowType=0[insert]) Src Rowid=2 Targ Rowid=2 EMPNO (EMPNO Double) : "4999.000000000000" ENAME (ENAME Char 10) : "ALLEN" JOB (JOB Char 9) : "SALESMAN" MGR (MGR Double) : "7698.000000000000" HIREDATE (HIREDATE Date) : "02/20/1981 00:00:00.000000000" SAL (SAL Double) : "1600.000000000000" COMM (COMM Double) : "300.000000000000" DEPTNO (DEPTNO Double) : "30.00000000000000"
				SQ_EMP Exchange: Rowdata: (RowType=0[insert]) Src Rowid=3 Targ Rowid=3 EMPNO (EMPNO Double) : "7521.000000000000" ENAME (ENAME Char 10) : "WARD" JOB (JOB Char 9) : "SALESMAN" MGR (MGR Double) : "7698.000000000000" HIREDATE (HIREDATE Date) : "02/22/1981 00:00:00.000000000" SAL (SAL Double) : "1250.000000000000" COMM (COMM Double) : "500.000000000000" DEPTNO (DEPTNO Double) : "30.00000000000000"
				SQ_EMP Exchange: Rowdata: (RowType=0[insert]) Src Rowid=4 Targ Rowid=4 EMPNO (EMPNO Double) : "7566.000000000000" ENAME (ENAME Char 10) : "JONES" JOB (JOB Char 9) : "MANAGER"

In this way, you can set the tracing level in mappings to capture the additional details for debugging.

In this tutorial, you have learned how to set, configure and execute the debugger and how to set tracing level in mappings. These options provide you the ability to debug the mappings.

Session Objects in Informatica

The session object is a set of instructions that instructs Informatica how and when to move the data from source to targets.

A session object is a task, just like other tasks that we create in workflow manager. Any session you create must have a mapping associated with it.

A session can have a single mapping at a time and once assigned, it cannot be changed. To execute a session task, it must be added to a workflow.

A session can be a reusable object or non-reusable. When you create a session in task developer, then it can be reused, but when you create a session in workflow designer, then it is non-reusable.

A reusable session can be added to multiple workflows.

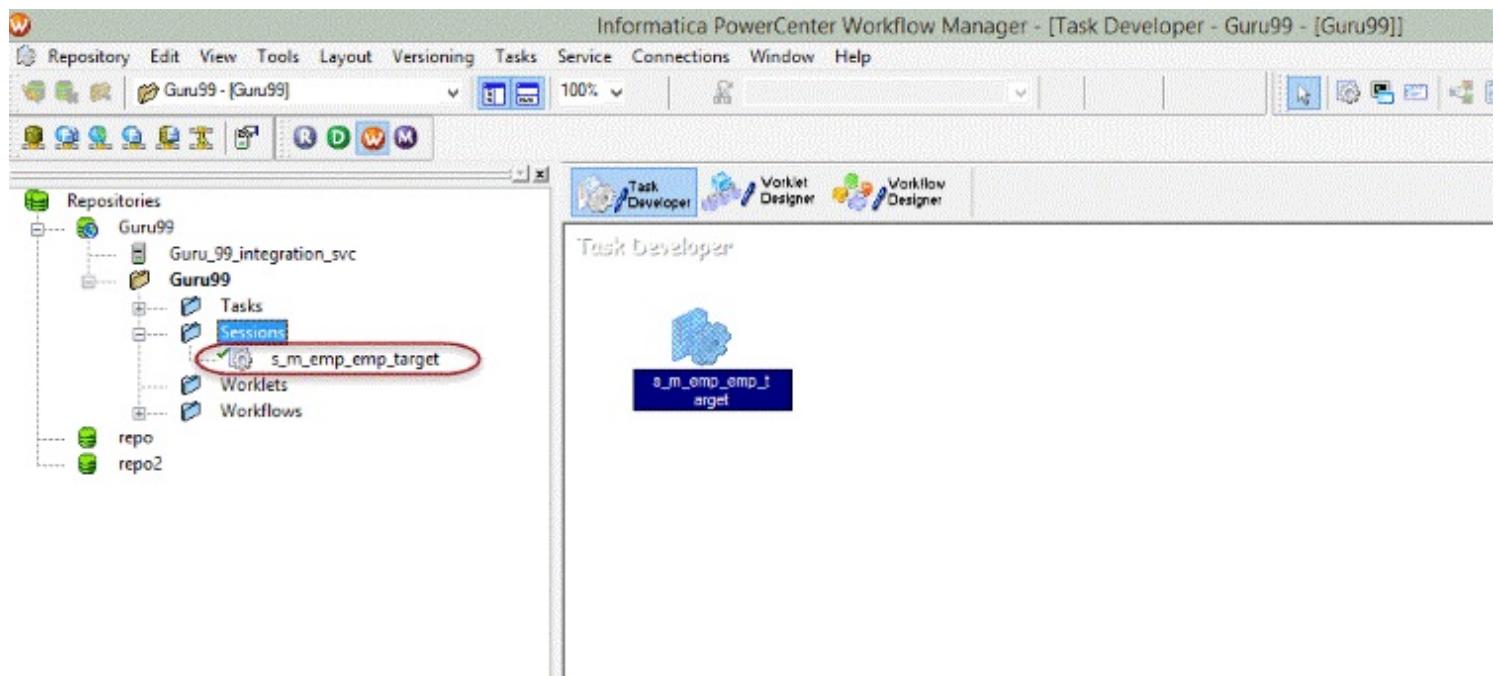
Properties Of Session

Using the properties of the session you can configure various characteristics of the session like pre and post SQL scripts, log file name and path, memory properties, etc.

You can also override mapping properties in the session properties. In this section, we will discuss the following important properties of the session.

- Treat source rows as
- Commit Interval
- Session log file Property
- Test Load Property
- Log options
- Error Handling
- Source/Target Properties
- Connections

Step 1) Open the session "s_m_emp_emp_target" in task developer, which we created in the earlier tutorial.



Step 2) Double click on the session icon inside Task Developer to open edit task window.

Edit Tasks

General Properties Config Object Mapping Components Metadata Extensions

Select task:

Task type: Session (Reusable)

Description:

Mapping name:

Step 3) Inside the "Edit Task" window clicks on the properties tab.

Edit Tasks

General **Properties** Config Object Mapping Components Metadata Extensions

Select task:

Task type: Session (Reusable)

Description:

Mapping name:

OK Cancel Apply Help

Step 4) In properties tab, it will show the properties of the session

Edit Tasks

General Properties Config Object Mapping Components Metadata Extensions

Select task: **s_m_emp_emp_target**

Task type: Session (Reusable)

Attribute	Value
General Options	
Write Backward Compatible Session Log File	<input type="checkbox"/>
Session Log File Name	s_m_emp_emp_target.log
Session Log File directory	\$PMSSessionLogDir\
Parameter Filename	
Enable Test Load	<input type="checkbox"/>
Number of rows to test	1
\$Source connection value	
\$Target connection value	
Treat source rows as	Delete
Commit Type	Target
Commit Interval	10000
Commit On End Of File	<input checked="" type="checkbox"/>
Rollback Transactions on Errors	<input type="checkbox"/>
Recovery Strategy	Fail task and continue workflow

General Options

General

OK Cancel Apply Help

Treat Source Rows As Property

This property allows you to define how the source data affects the target table. For example, you can define that the source record should be inserted or deleted from the target.

This property has four options –

- Insert
- Update
- Delete
- Data-driven

Edit Tasks

General Properties Config Object Mapping Components Metadata Extensions

Select task: **s_m_emp_emp_target**

Task type: Session (Reusable)

Attribute	Value
General Options	
Write Backward Compatible Session Log File	<input type="checkbox"/>
Session Log File Name	s_m_emp_emp_target.log
Session Log File directory	\$PMSSessionLogDir\
Parameter Filename	
Enable Test Load	<input type="checkbox"/>
Number of rows to test	1
\$Source connection value	
\$Target connection value	
Treat source rows as	Delete
Commit Type	Insert
Commit Interval	Delete
Commit On End Of File	Update
Rollback Transactions on Errors	<input type="checkbox"/>
Recovery Strategy	Fail task and continue workflow
Treat source rows as	
Specify the source row type.	

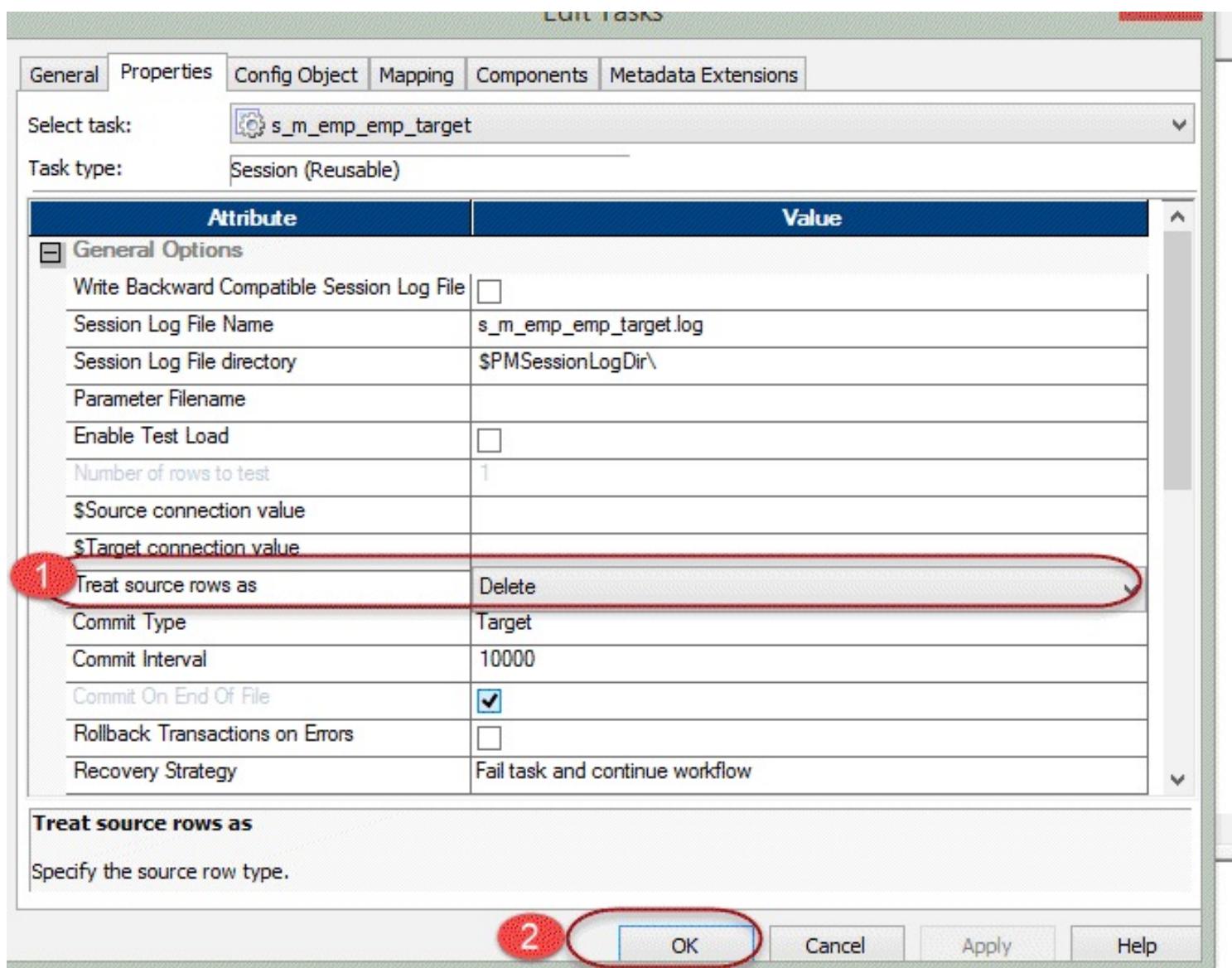
OK Cancel Apply Help

- When this property is set to **insert**, the source data will be marked to be inserted. It means the data will only be inserted.
- When the property is set to **update**, the target data will be updated by the source data. For updating of data primary key needs to be defined in the target table.
- When property is set to **delete** the source data which is already present in the target will be deleted from the target table. For this property to execute and apply the changes, the primary key should be defined in the target table.
- With the property set to **data driven**, the Informatica checks what source records are marked. If in a mapping the source records are marked as insert then records will be inserted into the target. If records are marked as an update in the mapping, then the records will be updated in the target. So what operation will be performed at the target depends on how records are handled inside the mapping.

How To Make Treat Source Rows – Delete

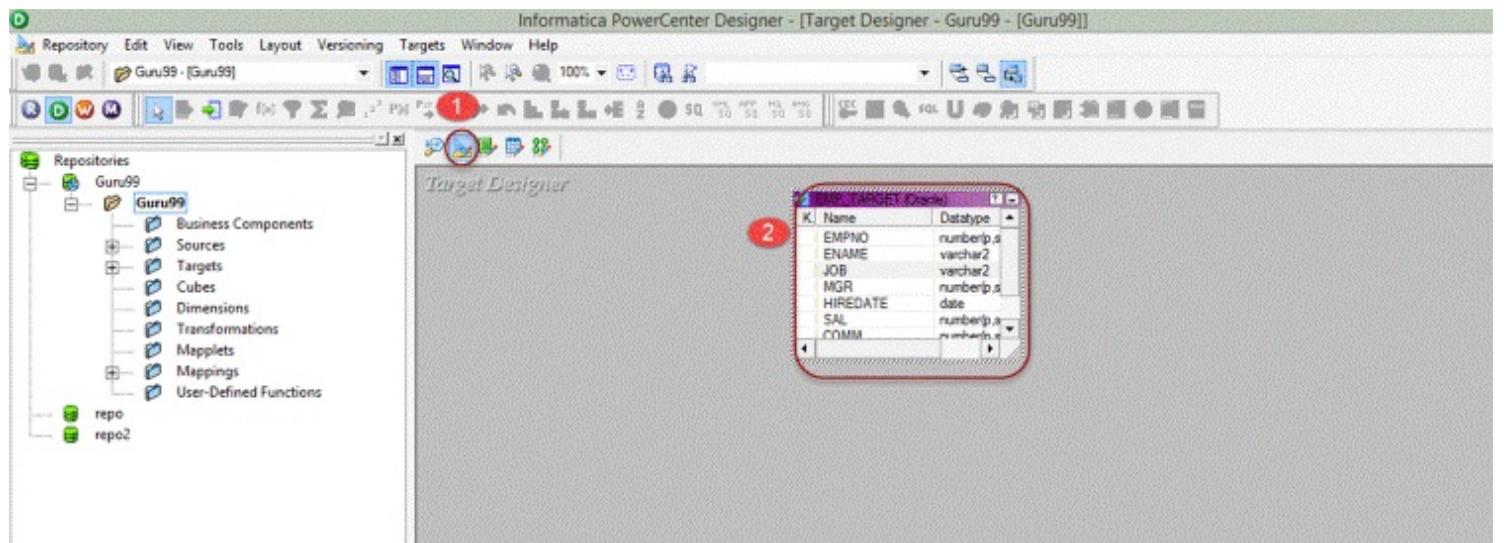
Step 1 -

1. In the property tab of the session task, select "Delete" option in "Treat Source Rows as"
2. Select OK Button



Step 2 – To define primary key in target table, open Informatica designer

1. Open target designer
2. Double click on the "emp_target" table



This will open an "Edit Table" for our target table.

Step 3 – In the edit window of target table

1. For the EmpNo column, select key type as "primary key" from the scroll down menu and
2. Select OK button.

Edit Tables

Table	Columns	Indexes	Metadata Extensions		
Select table: EMP_TARGET					
Column Name	Datatype	Prec	Scale	Not...	Key type
1 EMPNO	number(p,s)	4	0	<input checked="" type="checkbox"/>	NOT A KEY
2 ENAME	varchar2	10	0	<input type="checkbox"/>	NOT A KEY
3 JOB	varchar2	9	0	<input type="checkbox"/>	PRIMARY KEY
4 MGR	number(p,s)	4	0	<input type="checkbox"/>	FOREIGN KEY
5 HIREDATE	date	19	0	<input type="checkbox"/>	PRIMARY/FOREIGN
6 SAL	number(p,s)	7	2	<input type="checkbox"/>	NOT A KEY
7 COMM	number(p,s)	7	2	<input type="checkbox"/>	NOT A KEY
8 DEPTNO	.	.	.	<input type="checkbox"/>	NOT A KEY

Primary table:

Primary column:

Description:

2 OK Cancel Apply Help

Step 4 – Save the changes in Informatica and execute the workflow for this mapping.

When you execute this mapping, **the source records which are already present in the target will get deleted.**

Commit Interval - Property

This property defines the interval after which the Informatica makes commit operation to the target table.

For example, if you are inserting 20,000 records in a target table, and you define commit interval as 5,000, then after every 5,000 insertions of records in the target, a commit operation will be performed.

Edit Tasks

General Properties Config Object Mapping Components Metadata Extensions

Select task: s_m_emp_emp_target

Task type: Session (Reusable)

Attribute	Value
Session Log File Name	s_m_emp_emp_target.log
Session Log File directory	\$PMSessionLogDir\
Parameter Filename	
Enable Test Load	<input type="checkbox"/>
Number of rows to test	1
\$Source connection value	
\$Target connection value	
Treat source rows as	Insert
Commit Type	Target
Commit Interval	10000
Commit On End Of File	<input checked="" type="checkbox"/>
Rollback Transactions on Errors	<input type="checkbox"/>
Recovery Strategy	Fail task and continue workflow
Java Classpath	

Performance

Commit Interval

Default value for commit interval

OK Cancel Apply Help

Session Log File Name & Session Log File directory

Edit Tasks

General Properties Config Object Mapping Components Metadata Extensions

Select task: **s_m_emp_emp_target**

Task type: Session (Reusable)

Attribute	Value
General Options	
Write Backward Compatible Session Log File	<input type="checkbox"/>
Session Log File Name	s_m_emp_emp_target.log
Session Log File directory	\$PMSessionLogDir\
Parameter Filename	
Enable Test Load	<input type="checkbox"/>
Number of rows to test	1
\$Source connection value	
\$Target connection value	
Treat source rows as	Insert
Commit Type	Target
Commit Interval	10000
Commit On End Of File	<input checked="" type="checkbox"/>
Rollback Transactions on Errors	<input type="checkbox"/>
Recovery Strategy	Fail task and continue workflow

Session Log File directory

Log file directory

OK Cancel Apply Help

Configure this property to modify

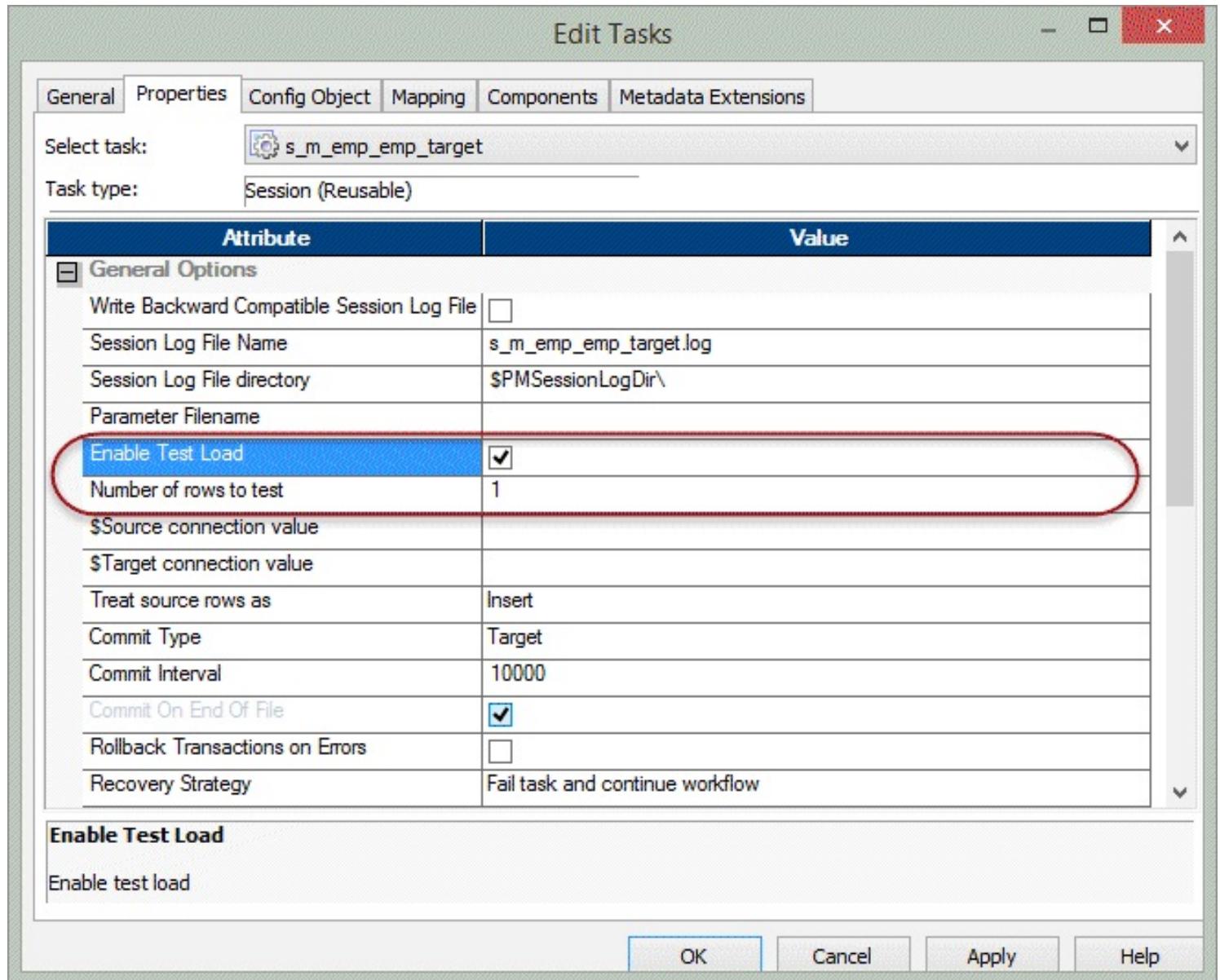
- Default session log file name and
- Path of the log file

The \$PMSessionLogDir\ is an Informatica variable and in windows it points to the following default location "C:\Informatica\9.6.1\server\infa_shared\SessLogs".

Enable Test Load

Using this property, you can test your session and mappings. When you use this feature and execute the sessions, records are fetched from the sources but they don't get loaded into the target. So this feature helps in testing the correctness of mappings, parameter files, functioning of various transformations inside the mapping.

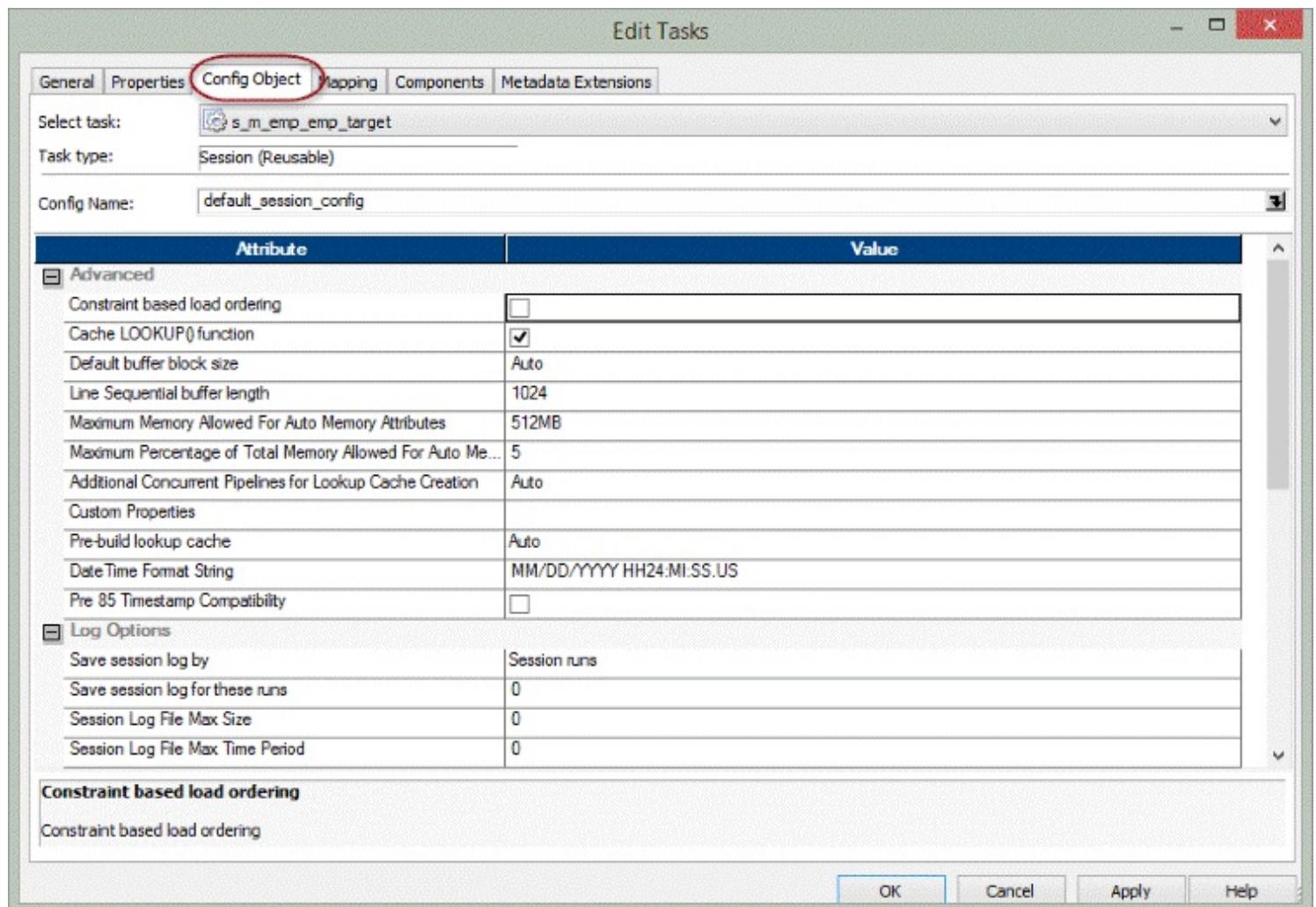
If you enable this feature, then there is another property – No of Rows to Test, this property should be configured for the no of records which you want to be fetched from the source for the test load.



Memory Properties

Memory properties give us the flexibility to fine tune the memory allocated to the Informatica for performance optimizations. When there are high bottleneck and performance is poor then you can try to improve the performance using the memory properties.

To configure memory properties click on the "config object" tab of the edit task window. It will open another window where you can configure the changes.



In this section, you can configure the memory properties. For example, default buffer block size, sequential buffer length, etc. Changes to this properties will determine how much memory should be allocated to Informatica services for their operation.

Edit Tasks

General	Properties	Config Object	Mapping	Components	Metadata Extensions
Select task:	s_m_emp_emp_target				
Task type:	Session (Reusable)				
Config Name:	default_session_config				
Attribute	Value				
Advanced					
Constraint based load ordering	<input type="checkbox"/>				
Cache LOOKUP() function	<input checked="" type="checkbox"/>				
Default buffer block size	Auto				
Line Sequential buffer length	1024				
Maximum Memory Allowed For Auto Memory Attributes	512MB				
Maximum Percentage of Total Memory Allowed For Auto Me...	5				
Additional Concurrent Pipelines for Lookup Cache Creation	Auto				
Custom Properties					
Pre-build lookup cache	Auto				
Date Time Format String	MM/DD/YYYY HH24.MI.SS.US				
Pre 85 Timestamp Compatibility	<input type="checkbox"/>				
Log Options					
Save session log by	Session runs				
Save session log for these runs	0				
Session Log File Max Size	0				
Session Log File Max Time Period	0				
Default buffer block size					
Default buffer block size					
OK Cancel Apply Help					

Log options

Edit Tasks

General Properties Config Object Mapping Components Metadata Extensions

Select task: **s_m_emp_emp_target**

Task type: Session (Reusable)

Config Name: default_session_config

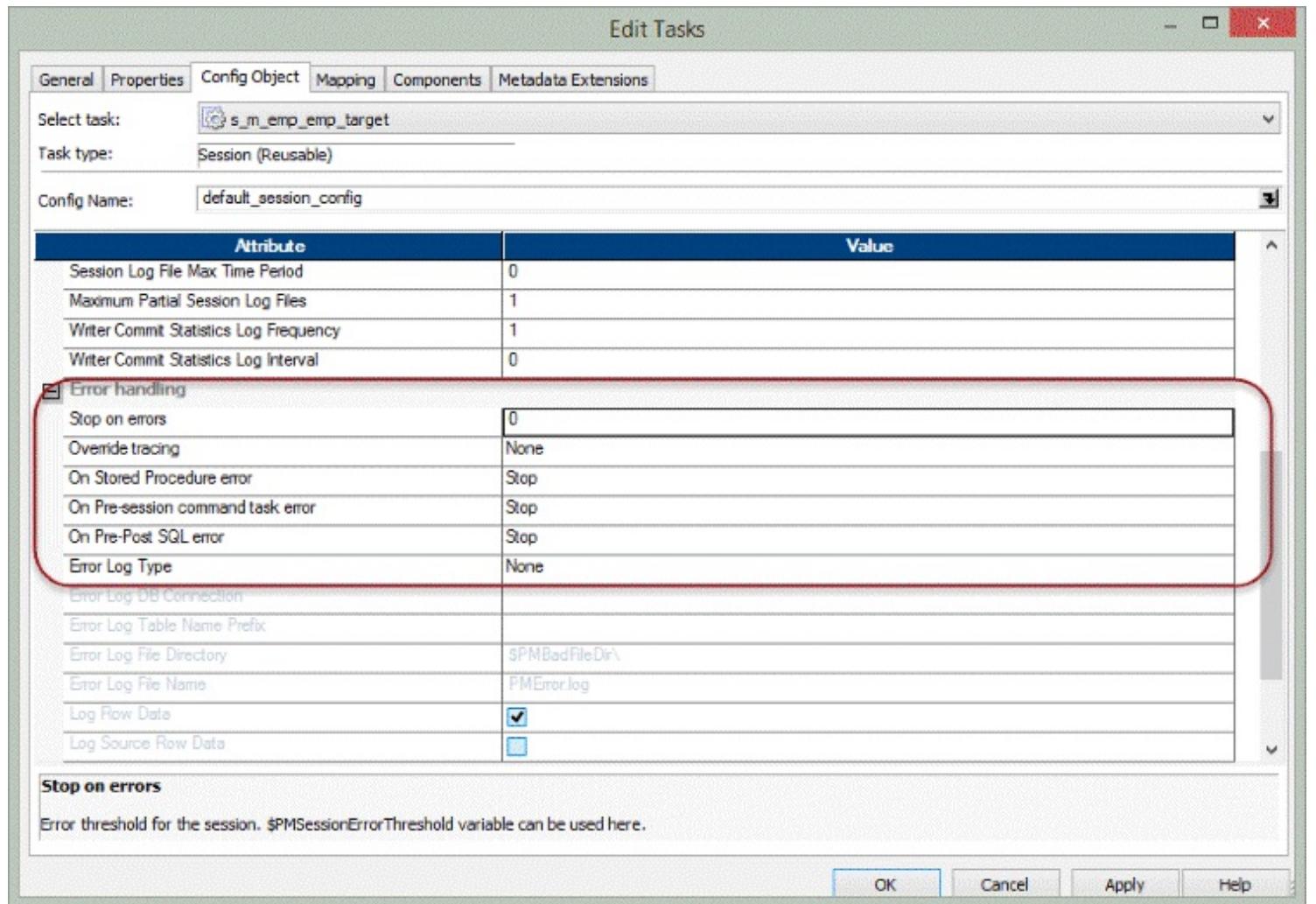
Attribute	Value
Date Time Format String	MM/DD/YYYY HH24:MI:SS.US
Pre 85 Timestamp Compatibility	<input type="checkbox"/>
Log Options	
Save session log by	Session runs
Save session log for these runs	0
Session Log File Max Size	0
Session Log File Max Time Period	0
Maximum Partial Session Log Files	1
Writer Commit Statistics Log Frequency	1
Writer Commit Statistics Log Interval	0
Error handling	
Stop on errors	0
Override tracing	None
On Stored Procedure error	Stop
On Pre-session command task error	Stop
On Pre-Post SQL error	Stop
Error Log Type	None
Default buffer block size	
Default buffer block size	

OK Cancel Apply Help

s_m_emp_emp_target in folder [Guru99] last saved by user [Guru99].

In this property section, you can configure the log properties of the session. You can set the no for how many no of logs you want to save for a session, session log file max size.

Error Handling



In this section, you can configure the error properties for the session.

Using **Stop on errors** you can configure after how many errors the session has to be stopped.

Using **override tracing** you can override the mapping tracing levels.

You can also configure the behaviour of the session for various errors encountered for example stored procedure error, pre-post SQL error, etc.

Mapping and Source/Target Properties

In the mapping tab of the edit task window of the session, you can configure the properties related to the mapping and its sources/targets and transformations. With this section of properties, you can override the properties of the source and targets. For the sources and targets we can override the table names, we can apply table name prefixes. Inside the mappings we can configure the properties of various transformations, sources and targets, in addition to that in this section we can also

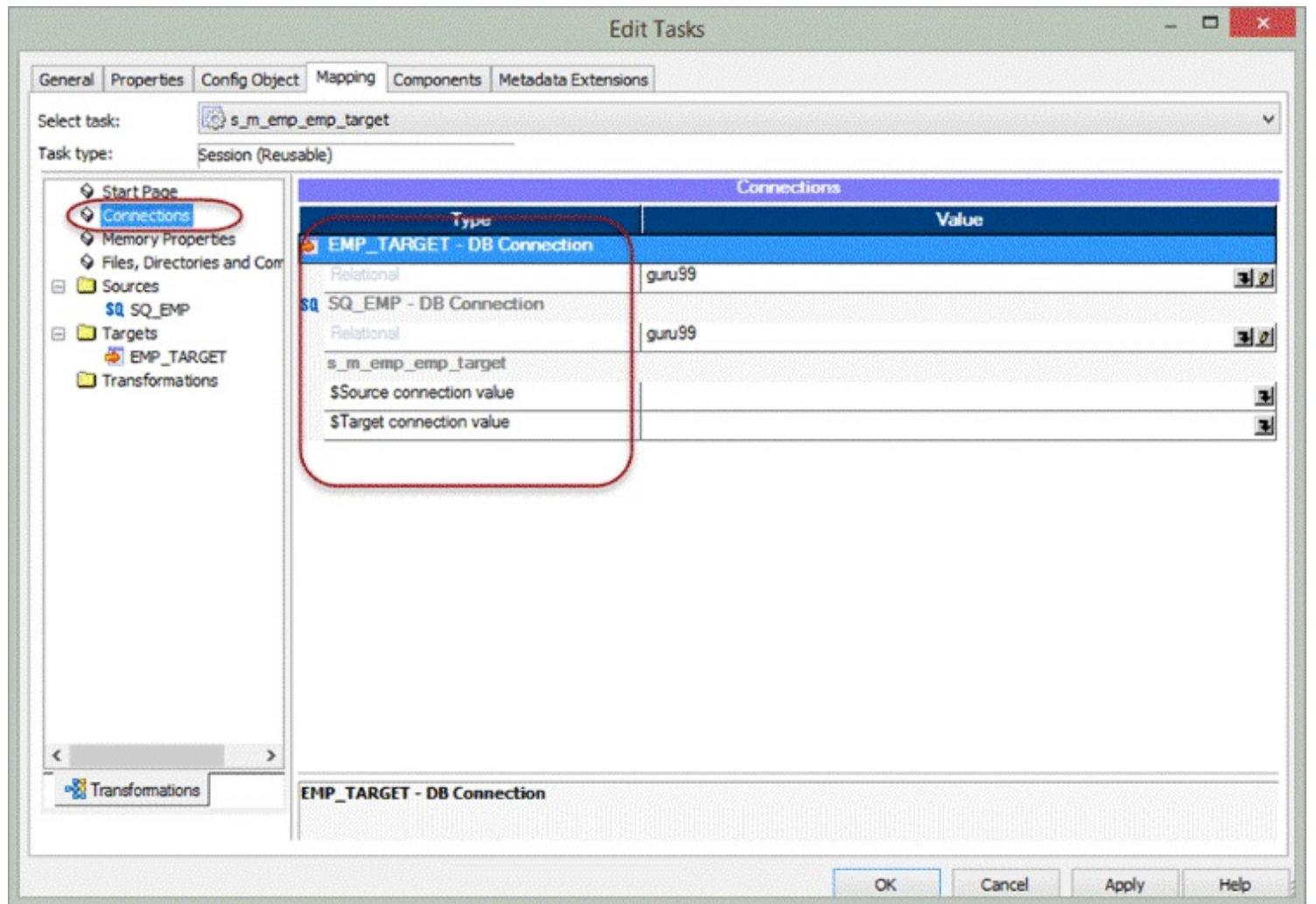
review and override those properties. It's like a single place to review all those properties.

The screenshot shows the 'Edit Tasks' window for a mapping named 's_m_emp_emp_target'. The 'Mapping' tab is selected. On the left, a tree view shows 'Connections' (selected), 'Memory Properties', 'Files, Directories and Com', 'Sources' (with 'SQ_SQL_EMP' selected), 'Targets' (with 'EMP_TARGET' selected), and 'Transformations'. The main area is titled 'Connections' and lists two connections: 'EMP_TARGET - DB Connection' and 'SQ_EMP - DB Connection'. Both are of type 'Relational'. The 'EMP_TARGET' connection has a value 'gun99' and is associated with source connection '\$Source connection value' and target connection '\$Target connection value'.

Type	Value
EMP_TARGET - DB Connection	gun99
SQ_EMP - DB Connection	gun99

Connection Properties in Mapping

Using this property, you can define database connections for the source and targets.



Source Properties

In this section, you can configure the properties related to the source of the mapping. You can configure pre and post SQL scripts for the source.

Using SQL query property, you can override the SQL for the source. You can also override the source table name in this section.

The screenshot shows the Informatica PowerCenter interface with the following details:

- Task:** s_m_emp_emp_target
- Task Type:** Session (Reusable)
- Sources:** SQ_SQL_EMP (highlighted with a red circle)
- Targets:** EMP_TARGET
- Transformations:** None

Properties Window (SQ_SQL_EMP - Source Qualifier):

Attribute	Value
User Defined Join	
Number Of Sorted Ports	0
Tracing Level	Verbose Data
Select Distinct	<input type="checkbox"/>
Pre SQL	
Post SQL	
Sql Query	
Source Filter	
EMP - Source	
Owner Name	
Source Table Name	

Target Properties

In this section, you can configure the details of the target. You can define whether target load has to be a bulk load or a normal mode.

In bulk load, the performance gain is achieved as during the load there are no redo log buffers managed by the database.

On the other hand, normal load is slower as compared to bulk load, but in case of failure database recovery is possible.

You can also define the property to truncate the target table before populating it. It means before loading any records in the target, the target table will be truncated, and then the load will be performed. This property is useful when we create mappings for stage load.

We can also define target table pre SQL and post SQL. Pre SQL is the piece of SQL code which will be executed before performing insert in the target table, and post SQL code will be executed after the load of target table is completed.

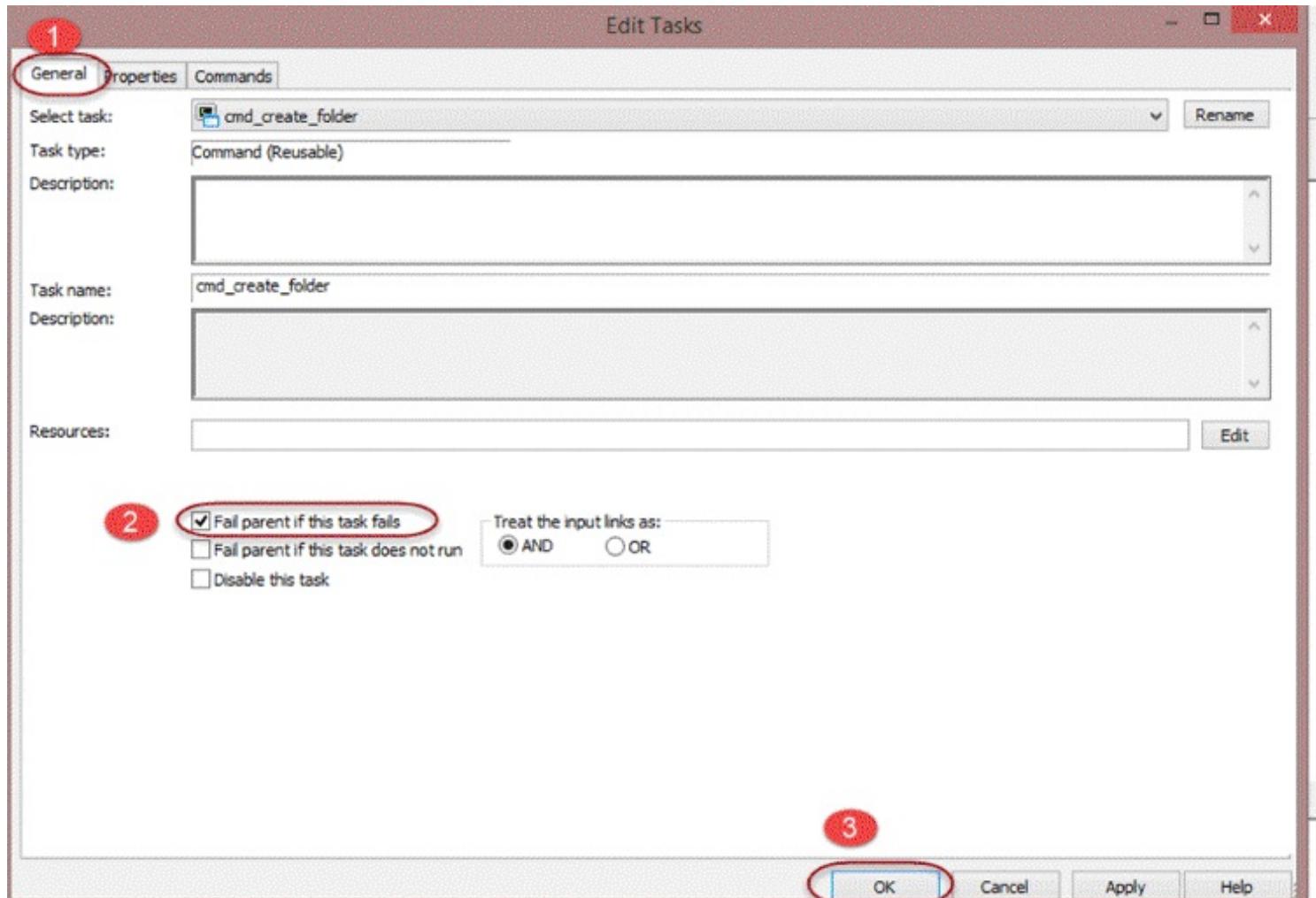
Success or Failure Of Session Task

When you have multiple sessions inside a workflow, then there can be a scenario where one or more session fails. In such condition, there comes a question of what would be the status of the workflow because you are having a workflow in which few tasks have failed, and few task got succeeded. To handle such conditions, Informatica provides the option to set this failure specific property inside the workflow. To configure such behaviour –

Step 1 – Open the workflow "wkf_run_command", which we created earlier

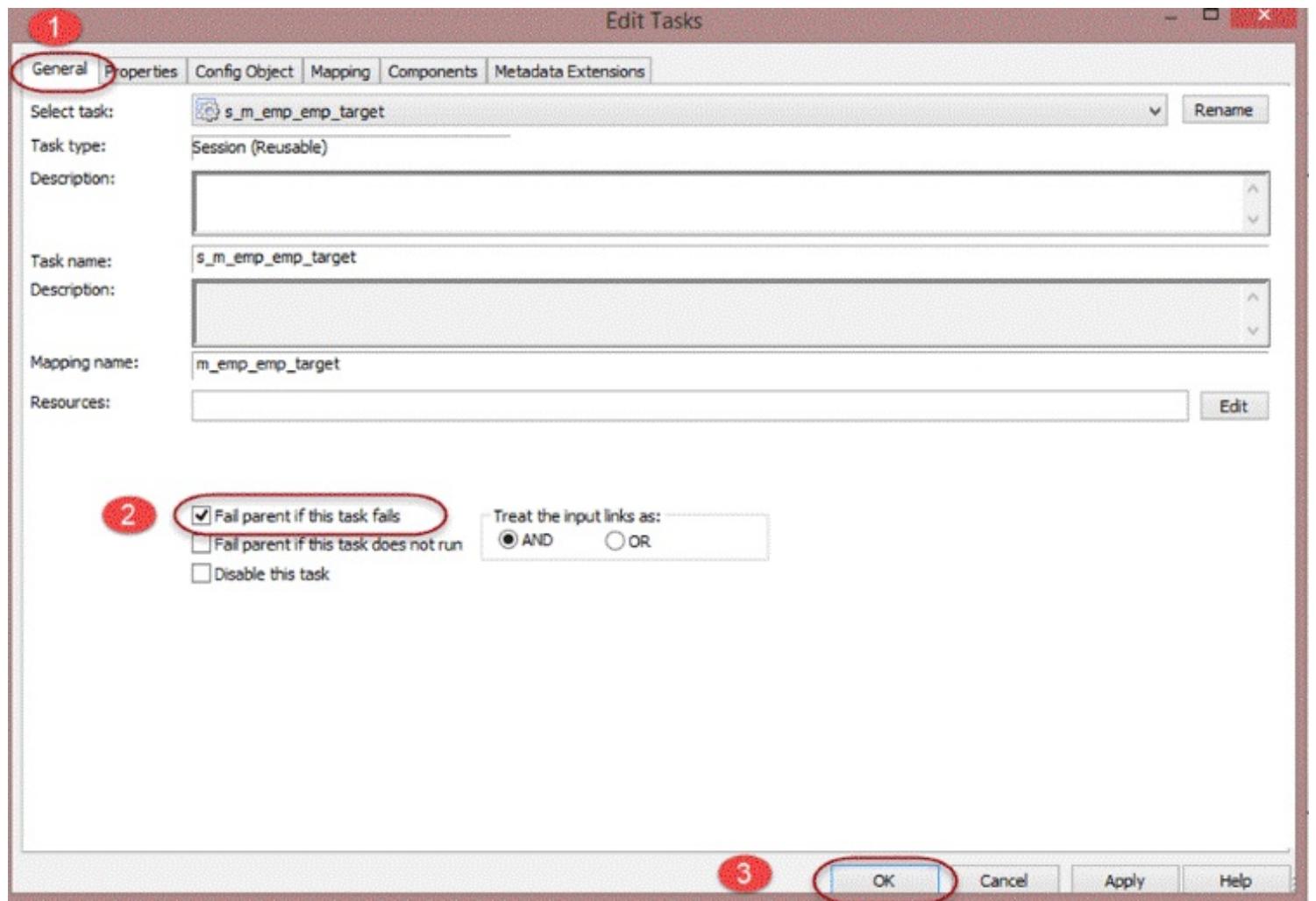
Step 2 – Double click on the command task, this will open edit task window

1. Select the general tab
2. Select the check box against property "fail parent if this task fails"
3. Select Ok Button



Step 3 – Double click on the session task, this will open edit task window

1. Select the general tab
2. Select the check box against property "fail parent if this task fails"
3. Select Ok Button



Step 4 – Save the changes using a ctrl+s shortcut.

When you execute this workflow after making the above changes if any of the tasks fails the workflow status will be made as failed. so you can identify that during the execution of your workflow some of its task has failed.

Chapter 8: Transformations

Transformations are the objects in Informatica which creates, modifies or passes data to the defined target structures (tables, files or any other target).

The purpose of the transformation in Informatica is to modify the source data as per the requirement of target system. It also ensures the quality of the data being loaded into the target.

Informatica provides various transformations to perform specific functionalities.

For example, performing tax calculation based upon source data, data cleansing operation, etc. In transformations, we connect the ports to pass data to it, and transformation returns the output through output ports.

Classification of Transformation

Transformation is classified into two categories, one based on connectivity, and other based on the change in no of rows. First we will look the transformation based on connectivity.

Types of transformation based on connectivity

- Connected Transformations
- Unconnected Transformations

In Informatica, during mappings the transformations which are connected to other transformations are called connected transformations.

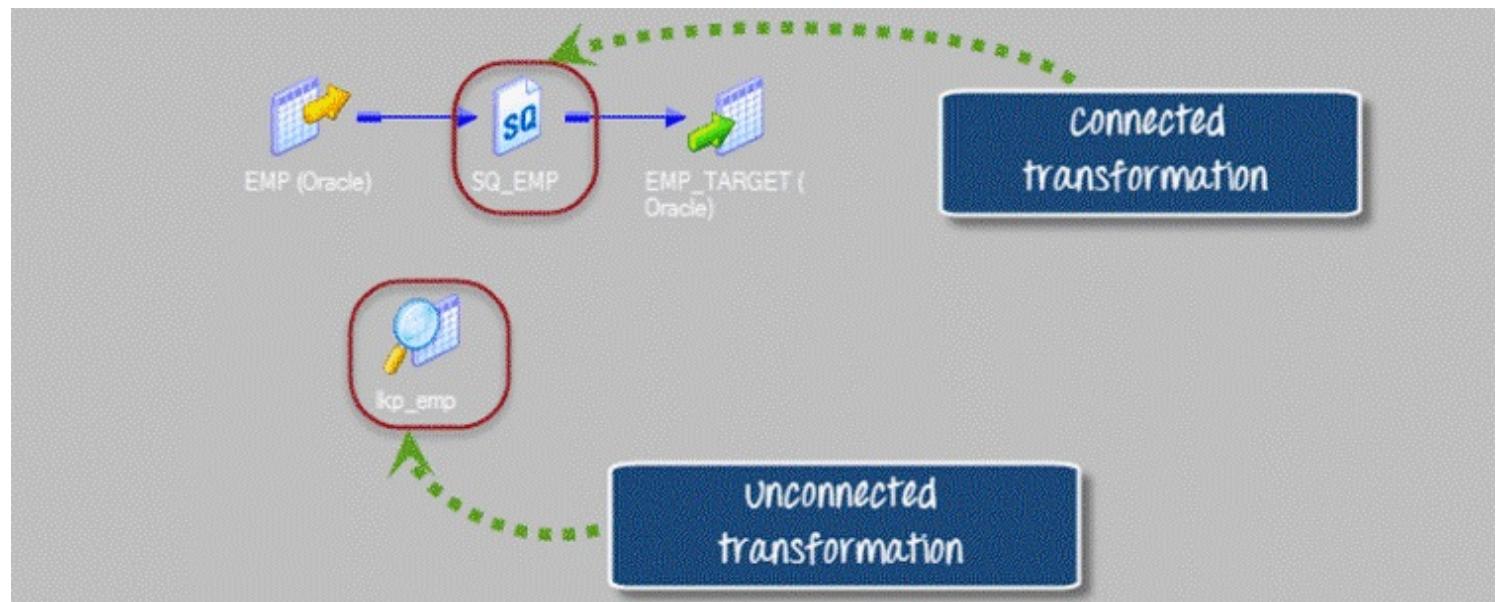
For example, Source qualifier transformation of Source table EMP is connected to filter transformation to filter employees of a dept.

Those transformations that are not connected to any other transformations are called unconnected transformations.

Their functionality is used by calling them inside other transformations like Expression transformation. These transformations are not part of the pipeline.

The connected transformations are preferred when for every input row, transformation is called or is expected to return a value. For example, for the zip codes in every row, the transformation returning city name.

The unconnected transformations are useful when their functionality is only required periodically or based upon certain conditions. For example, calculation the tax details if tax value is not available.



Types of transformations based on the change in no of rows

- Active Transformations
- Passive Transformations

Active Transformations are those who modifies the data rows and the number of input rows passed to them. For example, if a transformation receives ten number of rows as input, and it returns fifteen number of rows as an output then it is an active transformation. The data in the row is also modified in the active transformation.

Passive transformations are those who does not change the number of input rows. In passive transformations the number of input and output rows remain the same, only data is modified at row level.

In the passive transformation, no new rows are created, or existing rows are dropped.

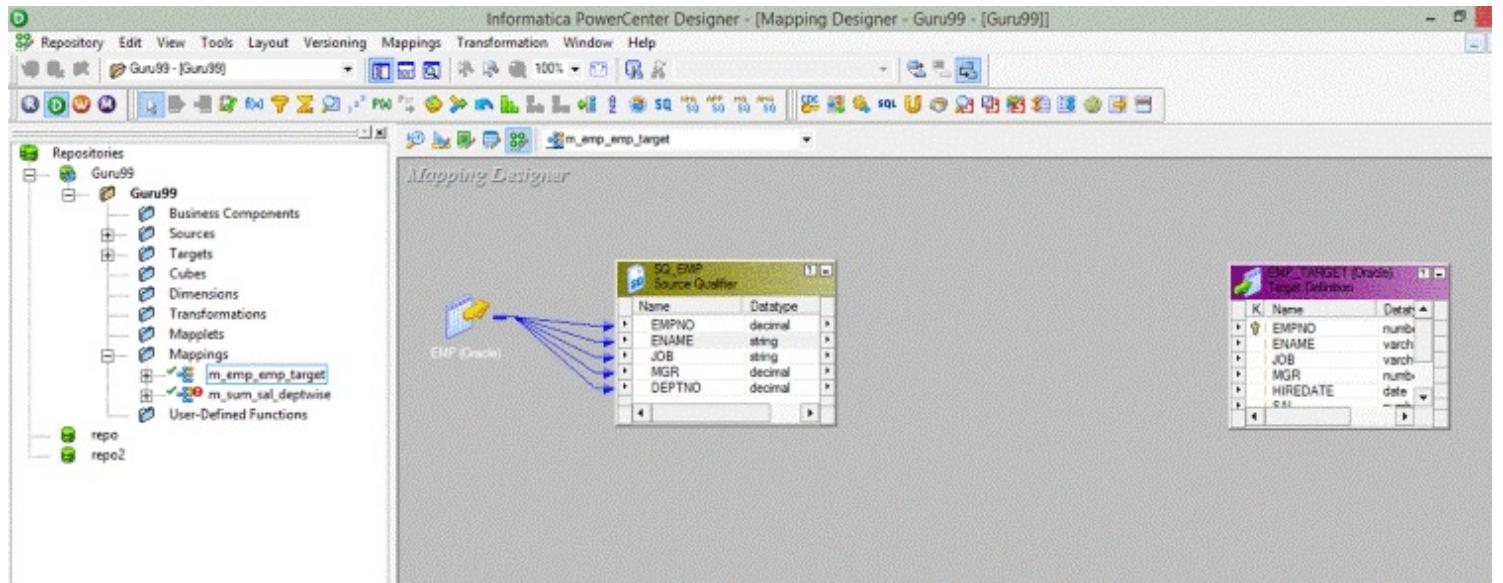
Filter Transformation

Using the filter transformation, we can filter the records based on the filter condition. Filter transformation is an active transformation as it changes the no of records.

For example, for loading the employee records having deptno equal to 10 only, we can put filter transformation in the mapping with the filter condition deptno=10. So only those records which have deptno =10 will be passed by filter transformation, rest other records will be dropped.

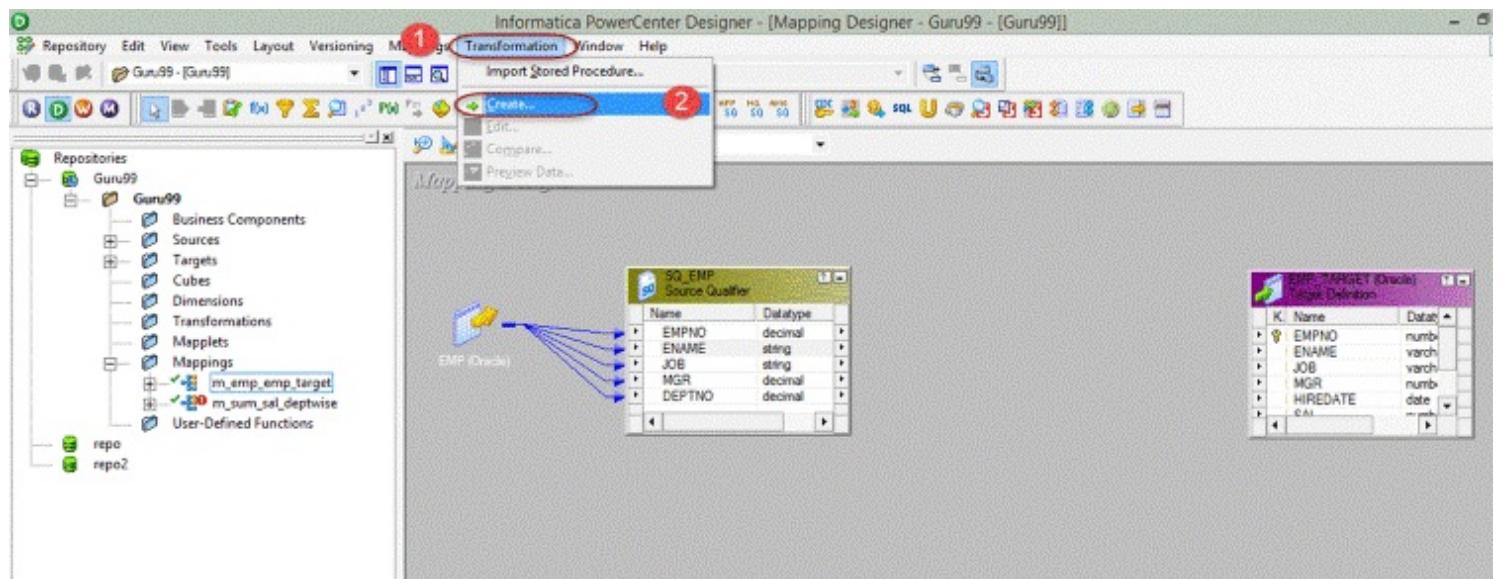
How to use filter transformation-

Step 1 – Create a mapping having source "EMP" and target "EMP_TARGET"



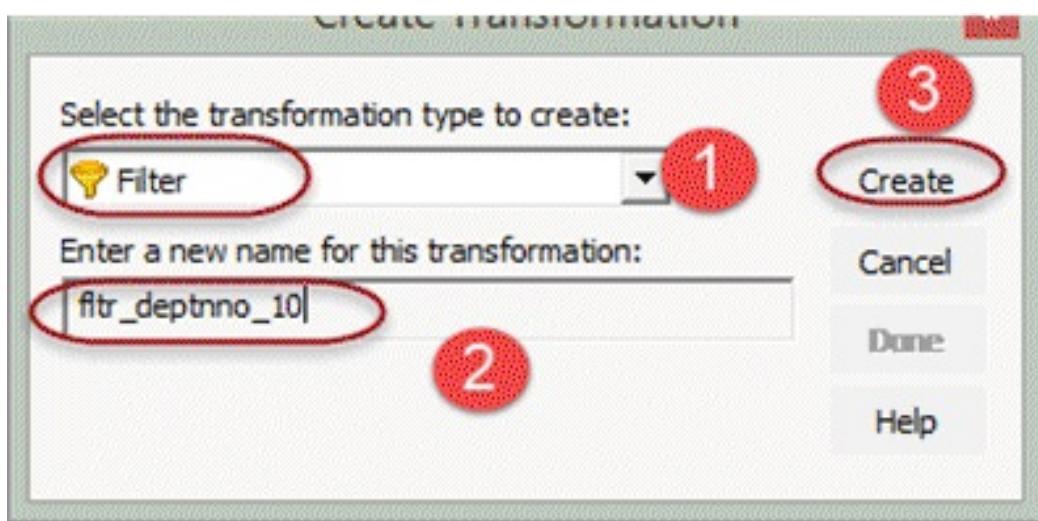
Step 2 – Then in the mapping

1. Select Transformation menu
2. Select create option

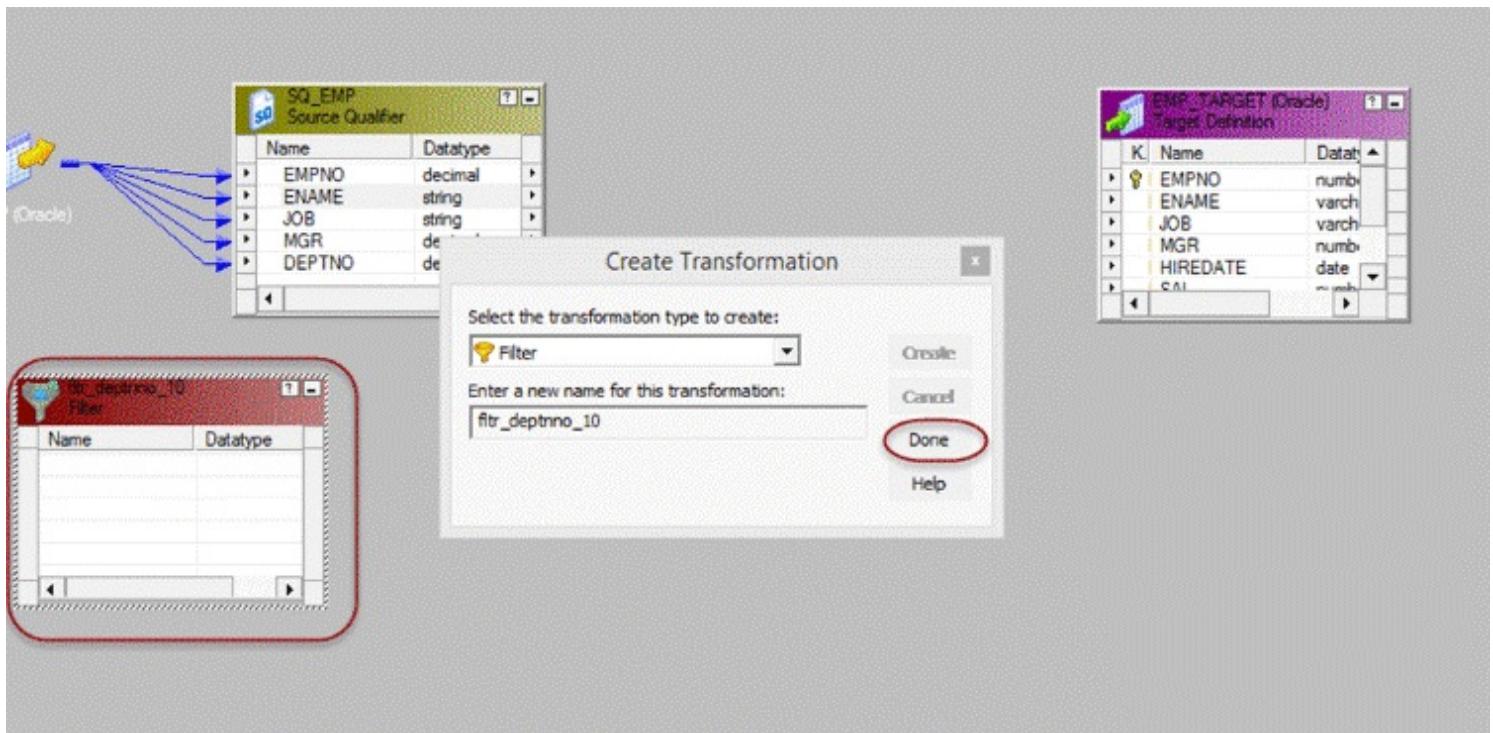


Step 3 - Then in the create transformation window

1. Select Filter Transformation from the list
2. Enter Transformation name "fltr_deptno_10"
3. Select create option

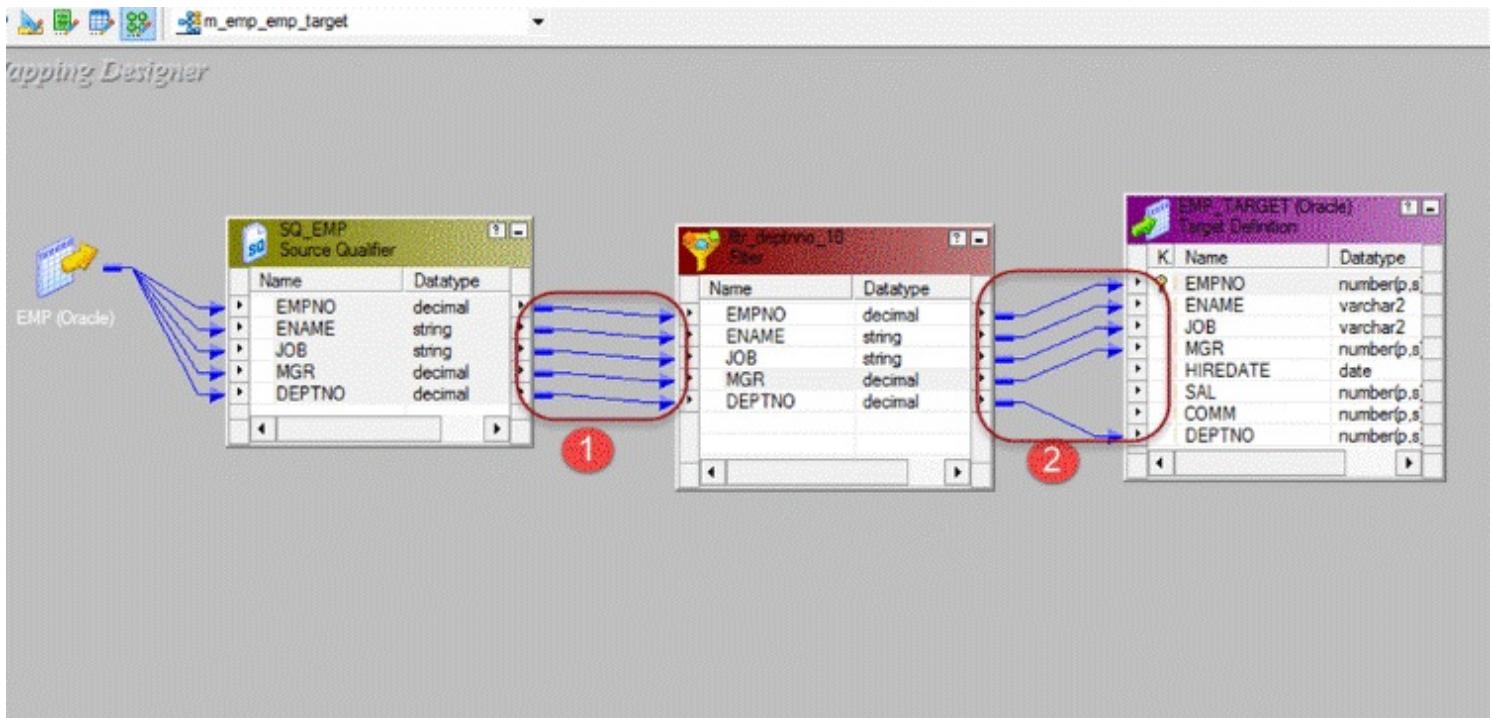


Step 4 – The filter transformation will be created, Select "Done" button in the create transformation window



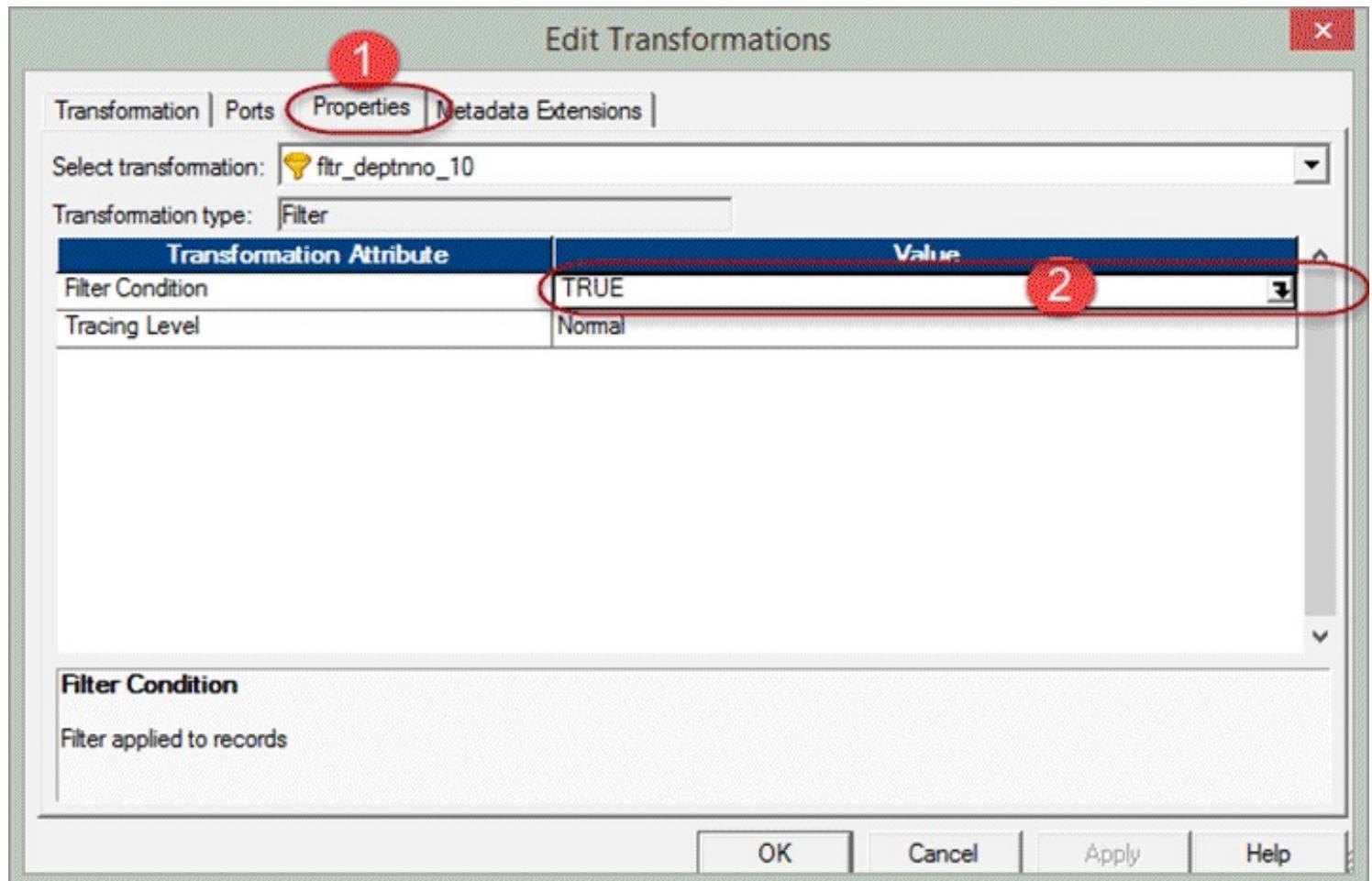
Step 5 – In the mapping

1. Drag and drop all the Source qualifier columns to the filter transformation
2. Link the columns from filter transformation to the target table



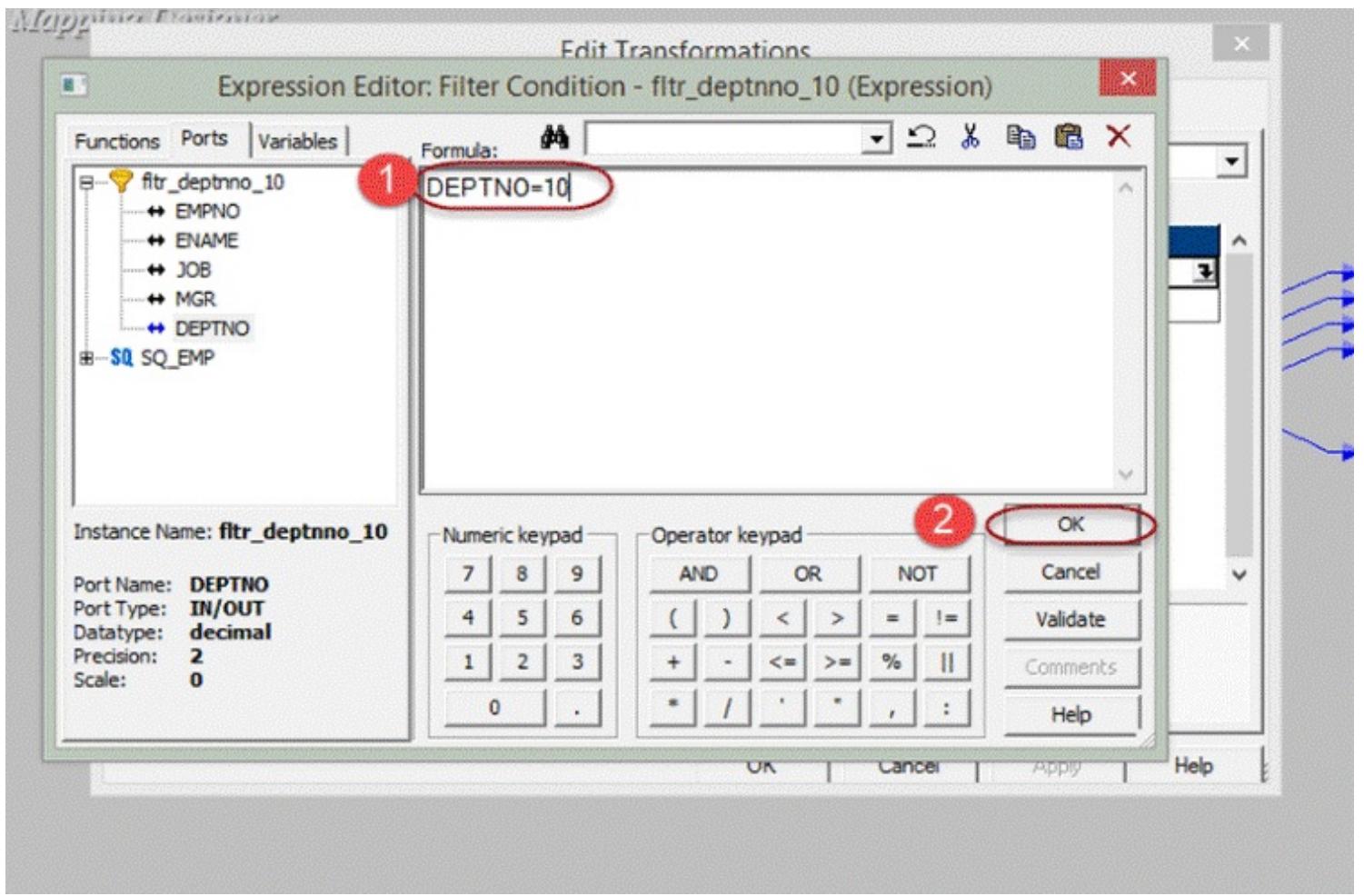
Step 6 – Double click on the filter transformation to open its properties, and then

1. Select the properties menu
2. Click on the Filter condition editor

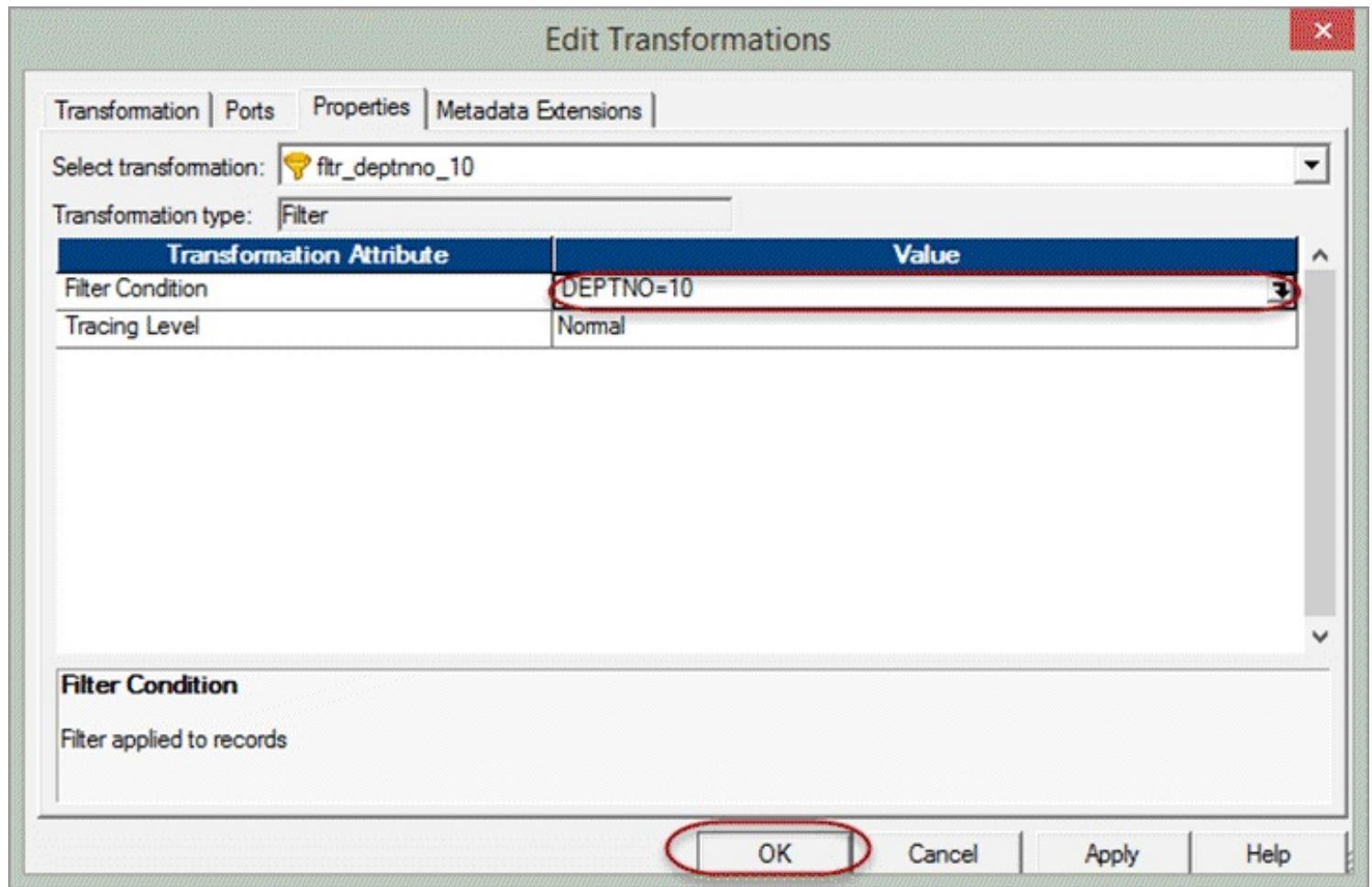


Step 7 – Then in the filter condition expression editor

1. Enter filter condition – deptno=10
2. Select OK button



Step 8 – Now again in the edit transformation window in Properties tab you will see the filter condition, select OK button



Now save the mapping and execute it after creating session and workflow. In the target table, the records having deptno=10 only will be loaded.

In this way, you can filter the source records using filter transformation.

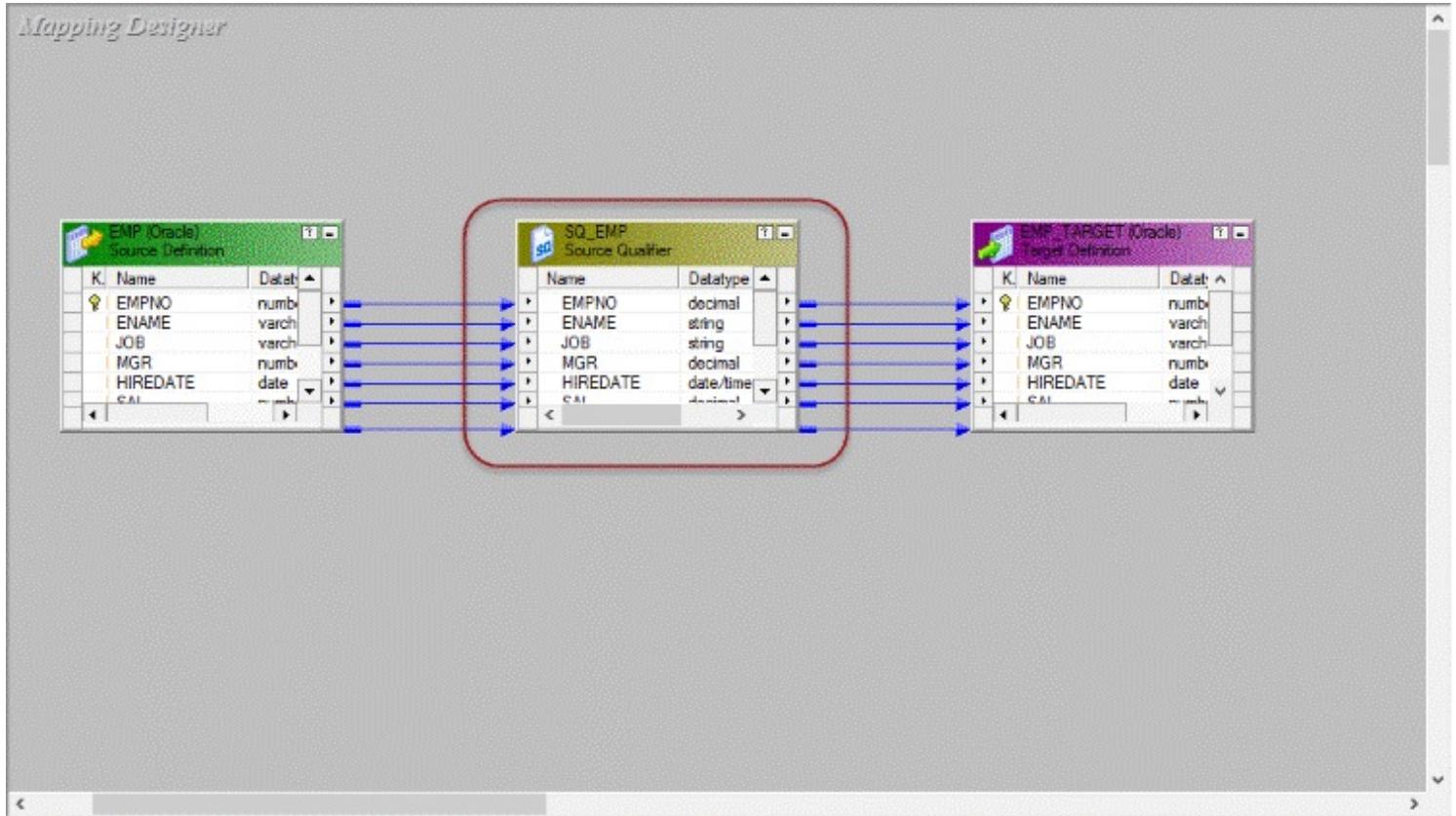
Source Qualifier Transformation

Whenever we add a relational source or a flat file to a mapping, a source qualifier transformation is required. When we add a source to a mapping, source qualifier transformation is added automatically. With source qualifier, we can define and override how the data is fetched from the source.

In the following example we would be modifying source qualifier of our mapping "m_emp_emp_target", so instead of returning all the columns it will return only selected columns.

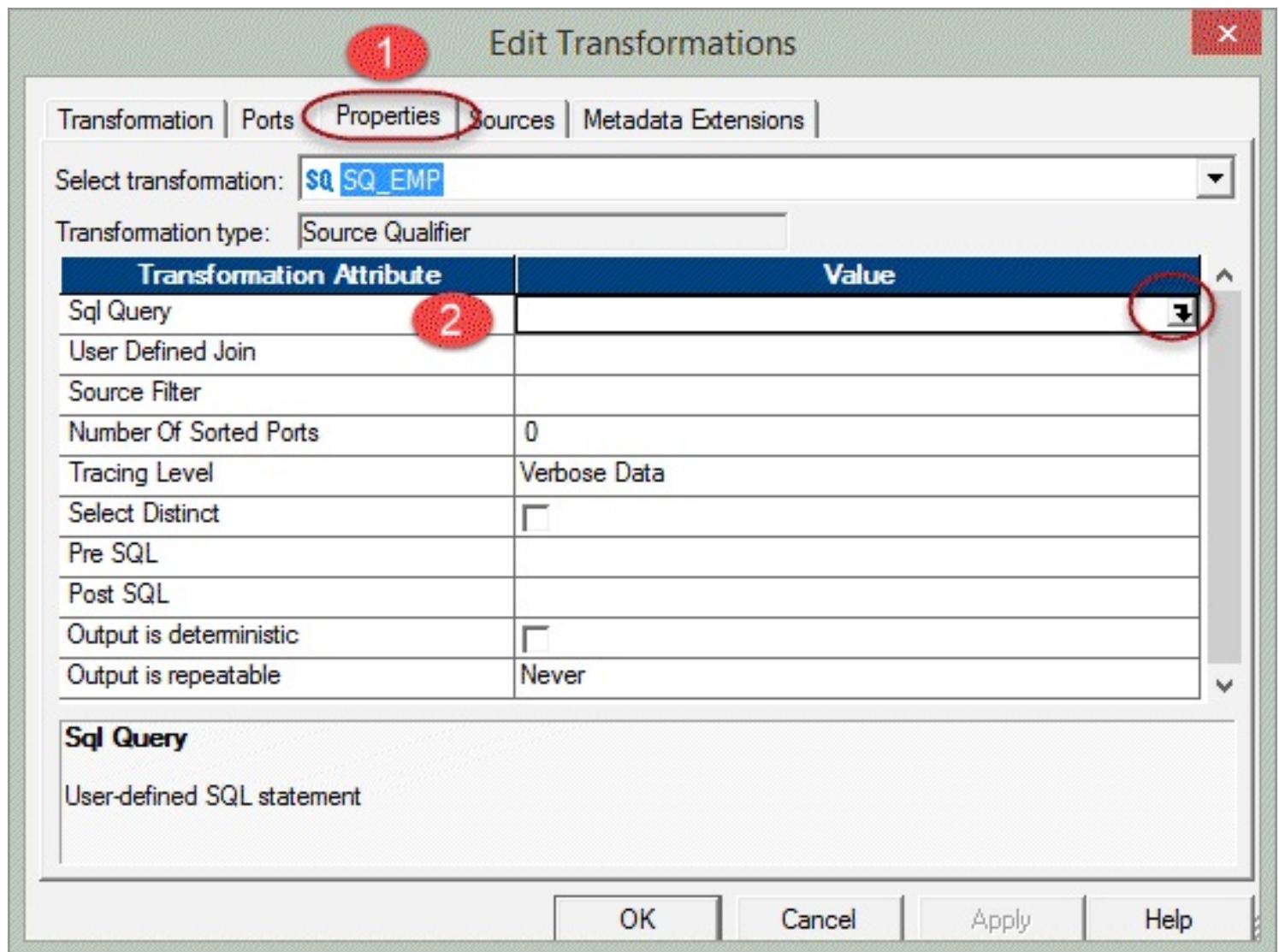
Step 1 – Open mapping "m_emp_emp_target" in mapping designer.

Mapping Designer



Step 2 – Double click on the source Qualifier transformation "SQ_EMP". It will open edit transformation property window for it. Then

1. Click on the properties tab
2. Click on the SQL Query Modify option, this will open an SQL editor window



Step 3 – In the SQL editor window

1. Enter the following query

```
SELECT EMPNO, ENAME, JOB, MGR FROM EMP
```

Note – we are selecting the columns EMPNO, ENAME, JOB & MANAGER from the source, so we have kept only those in the select query

2. Select OK Button

Edit Transformations

Transformation | Ports Properties | Sources | Metadata Extensions |

Select transformation: **SQ SQ_EMP**

Transformation type: Source Qualifier

Transformation Attribute	Value
Sql Query	SELECT EMPNO, ENAME, JOB, MGR...
User Defined Join	
Source Filter	
Number Of Sorted Ports	0
Tracing Level	Verbose Data
Select Distinct	<input type="checkbox"/>
Pre SQL	
Post SQL	
Output is deterministic	<input type="checkbox"/>
Output is repeatable	Never

Sql Query

User-defined SQL statement

OK

Cancel

Apply

Help

Step 4 – In the "edit transformations" window,

1. Select Ports tab from the menu
2. Under ports tab, you will see all the ports. Keep only the ports EMPNO, ENAME, JOB, MGR and delete other ports

Edit Transformations

Transformation Ports Properties Sources

Select transformation: SQ SQ_EMP

Transformation type: Source Qualifier

Delete ports using the delete option

Port Name Datatype Prec Scale I O

1	EMPNO	decimal	4	0	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
2	ENAME	string	10	0	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
3	JOB	string	9	0	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
4	MGR	decimal	4	0	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
5	HIREDATE	date/time	29	9	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
6	SAL	decimal	7	2	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
7	COMM	decimal	7	2	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
8	DEPTNO	decimal	2	0	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>

Delete these ports

Default value:

Description:

OK Cancel Apply Help

A red circle highlights the 'Delete' icon in the toolbar above the table, and another red circle highlights the 'Delete' icon in the bottom right corner of the dialog.

Step 5 – After deletion of ports, Select OK Button

Edit Transformations

Transformation Ports Properties Sources Metadata Extensions

Select transformation: **SQ SQ_EMP**

Transformation type: Source Qualifier

	Port Name	Datatype	Prec	Scale	I	O
1	EMPNO	decimal	4	0	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
2	ENAME	string	10	0	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
3	JOB	string	9	0	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
4	MGR	decimal	4	0	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>

Default value:

Description:

OK Cancel Apply Help

Now, again click on properties tab in Edit Transformations window, and you will see only those data that you have selected.

Edit Transformations

Transformation | Ports | Properties | Sources | Metadata Extensions |

Select transformation: **SQ SQ_EMP**

Transformation type: Source Qualifier

Transformation Attribute	Value
Sql Query	SELECT EMPNO, ENAME, JOB, MGR...
User Defined Join	
Source Filter	
Number Of Sorted Ports	0
Tracing Level	Verbose Data
Select Distinct	<input type="checkbox"/>
Pre SQL	
Post SQL	
Output is deterministic	<input type="checkbox"/>
Output is repeatable	Never

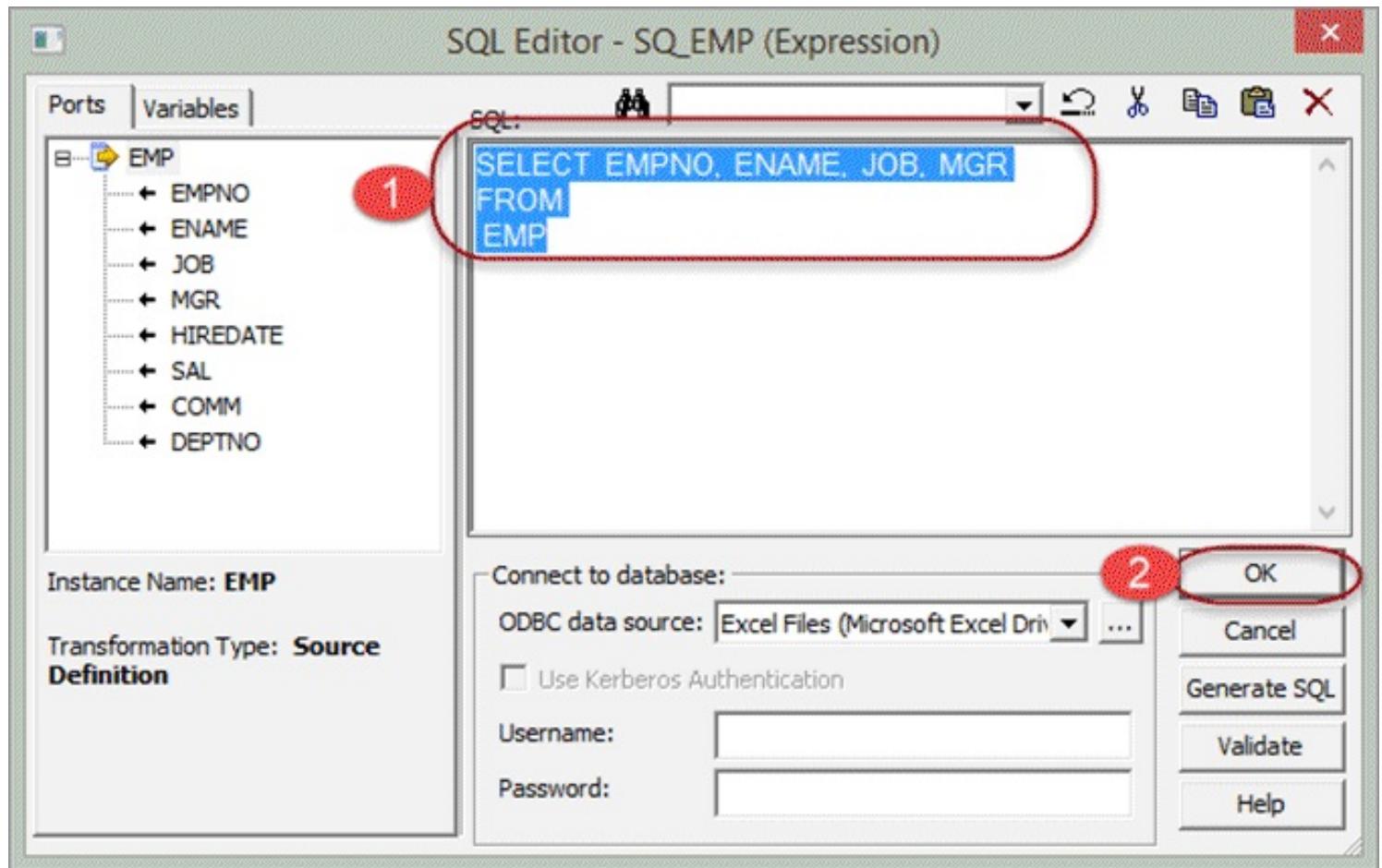
Sql Query

User-defined SQL statement

OK Cancel Apply Help

When you click on "OK" button it will open SQL Editor Window, and

1. It will confirm the data you have selected are correct and ready for loading into the target table
2. Click on OK button to process further



Save the mapping (using **ctrl+s** Shortcut) and execute the workflow, after execution only the selected columns will be loaded into the target.

In this way, you can override in source qualifier what columns needs to be fetched from the source & this is the only way to override what specific columns will be brought inside the mapping.

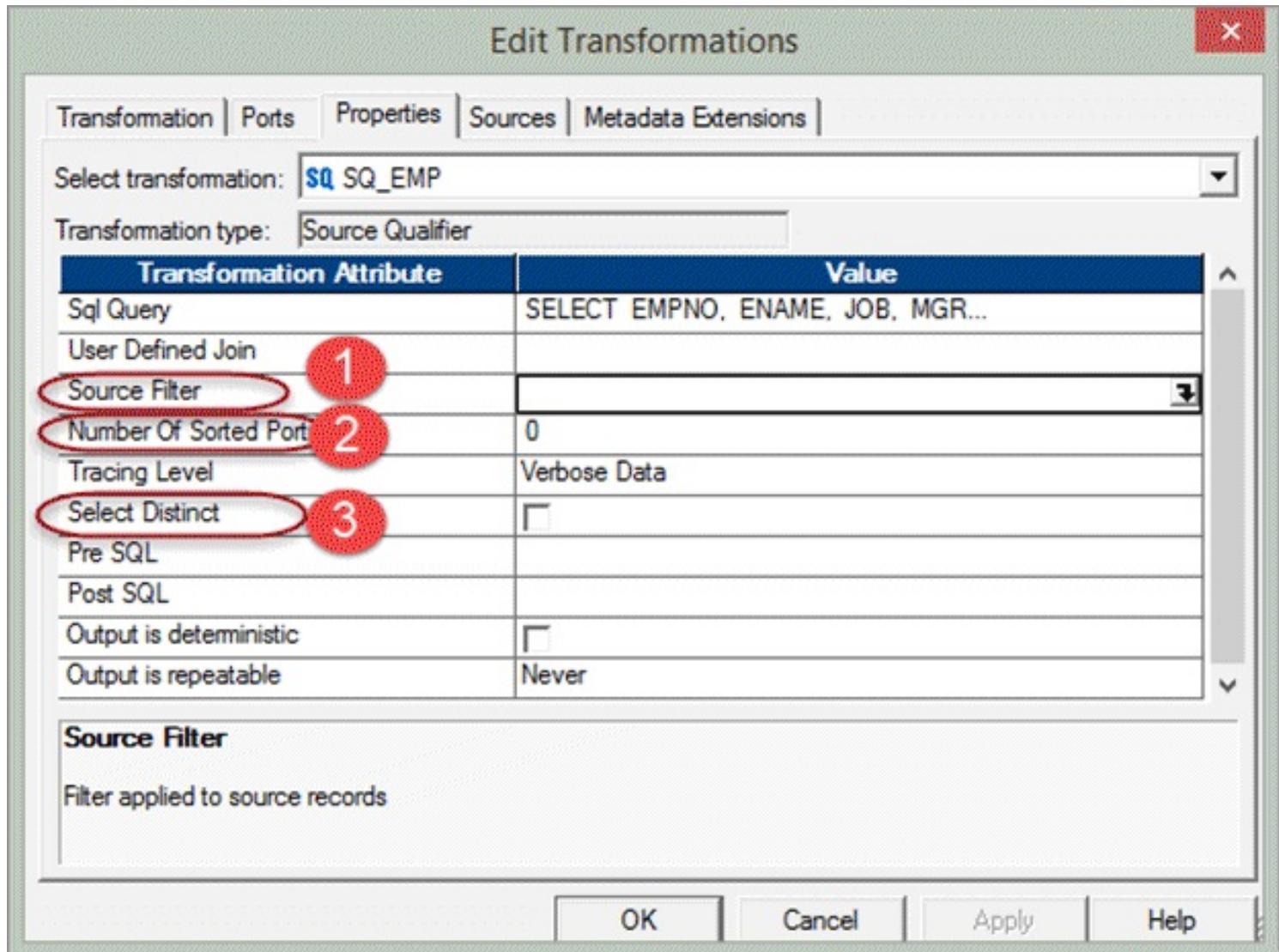
Properties of Source Qualifier

You can use various properties of Source Qualifier, to determine what type of source data needs to transform to target table.

- 1. Source Filter** – Using the source filter property you can filter the number of source records. For example, you want to fetch only the employees of deptno 10, then you can enter the filter condition `deptno=10` in source filter property and execute the data.
- 2. Number for sorted ports** – In source qualifier transformation, you can also sort the input records based on the ports number. So when the data is passed on to the transformations inside the mapping, it will read the port number and sort the data accordingly.

As data can be sorted based on a single or multiple ports, you have to give the number of ports which will be used in sorting. If you give value as 1, then only empno data will be sorted. If you give value as 2 then on empno and ename on both columns data will be sorted.

3. **Select Distinct** – you can fetch only distinct records from the source using this property. When you select the select distinct option, only distinct combination of source data will be fetched by source qualifier.

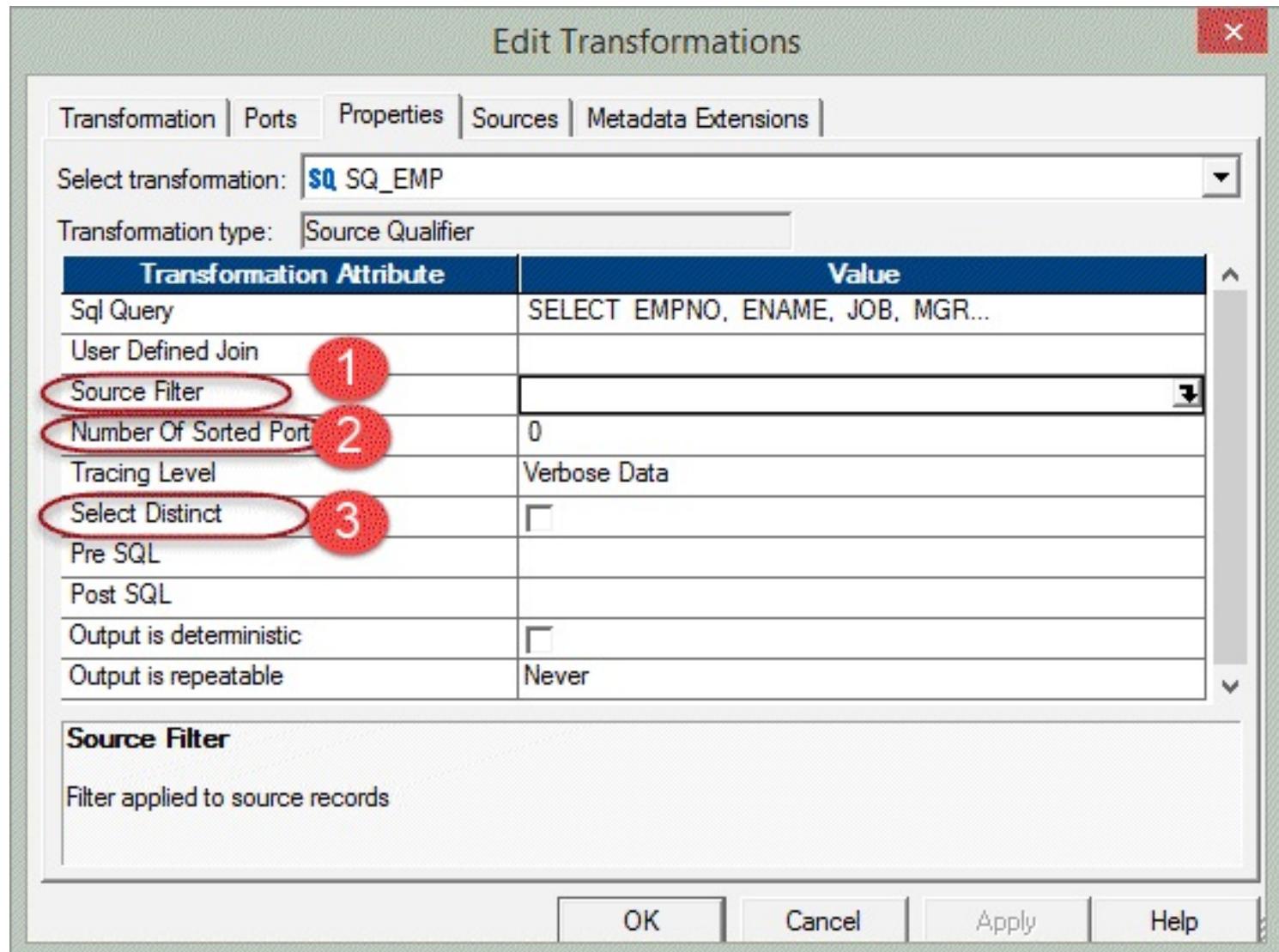


Aggregator Transformation

The aggregator transformation performs aggregate calculations like sum, average, etc.

For example, if you want to calculate the sum of salaries of all employees department wise, we can use the Aggregator Transformation.

The aggregate operations are performed over a group of rows, so a temporary placeholder is required to store all these records and perform the calculations.



For this, aggregator cache memory is used. This is a temporary main memory which is allocated to the aggregator transformation to perform such operations.

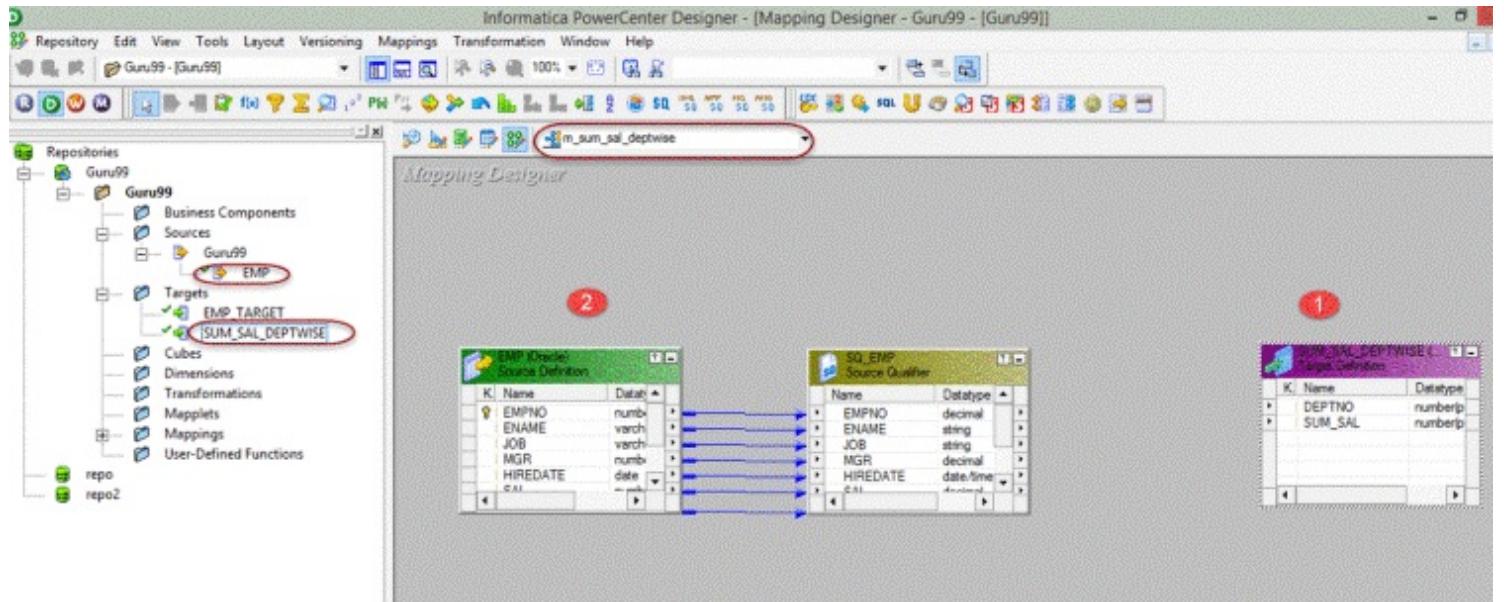
In this example, we will calculate the sum of salaries department wise. For this, we require a new column to store this sum. So, first of all, we will prepare a new column.

Step 1 – Create a new database target table, for example, say "sum_sal_deptwise", using the below script. You will see the new database target table is created under Target folder in next step.

Step 2 – Create a New mapping "m_sum_sal_deptwise".

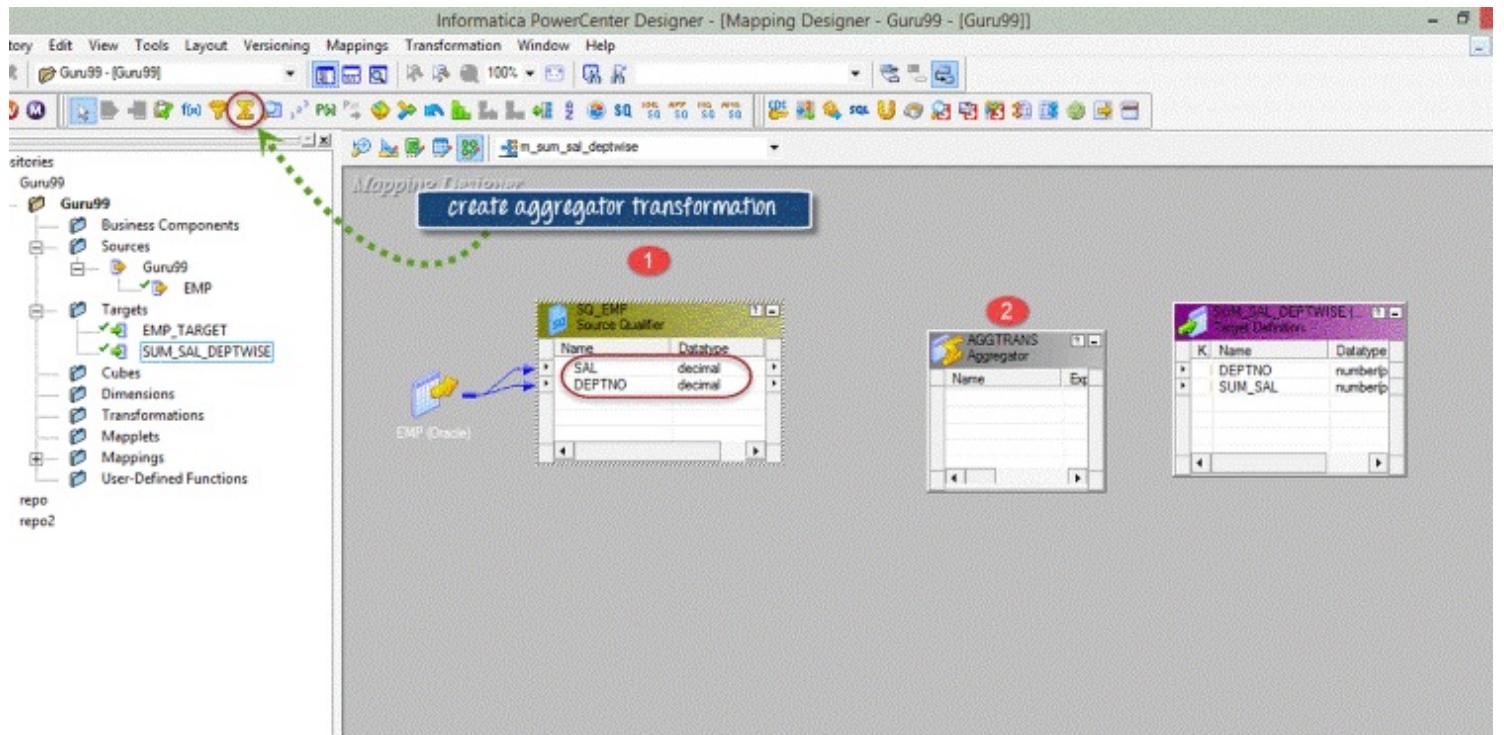
In order to create new mapping, we need source table (EMP) and target table (sum_sal_deptwise) both in mapping designer for that we need to

1. Import the target table "sum_sal_deptwise" in the mapping.
2. Import the source table "emp".

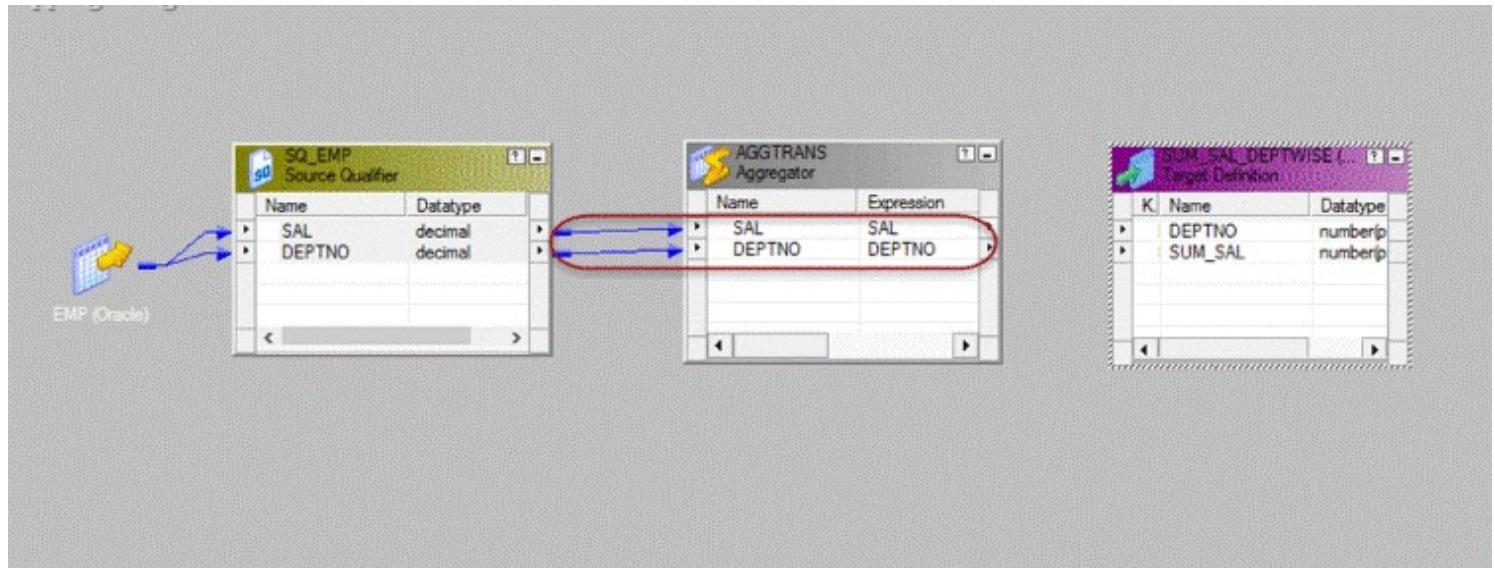


Step 3 – In the mapping,

1. From the Source Qualifier, delete the columns empno, ename, job, mgr, hiredate & comm so leaving only the columns deptno and sal.
2. Create a new aggregator transformation using the toolbox menu as shown in screen shot. When you click on the aggregator icon, a new aggregator transformation will be created.



Step 4 - Drag and drop SAL & DEPTNO columns from source qualifier (SQ_EMP) to the aggregator transformation



Step 5 – Double click on the aggregator transformation to open its properties, and then

1. Add a new port in the transformation
2. Rename the port name to SUM_SAL
3. Change the data type of this new port to double
4. Make this port as output port by selecting the checkbox of the output port.
5. Click on the expression option

Edit Transformations

Transformation Ports Properties Metadata Extensions

Select transformation: AGGTRANS

Transformation type: Aggregator

	Port Name	Datatype	Prec	Sc...	I	O	V	Expression	Group
1	SAL	decimal	7	2	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	SAL	<input type="checkbox"/>
2	DEPTNO	decimal	2	0	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	DEPTNO	<input type="checkbox"/>
3	SUM_SAL	double	15	0	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	SUM_SAL	<input type="checkbox"/>

Default value: ERROR(transformation error)

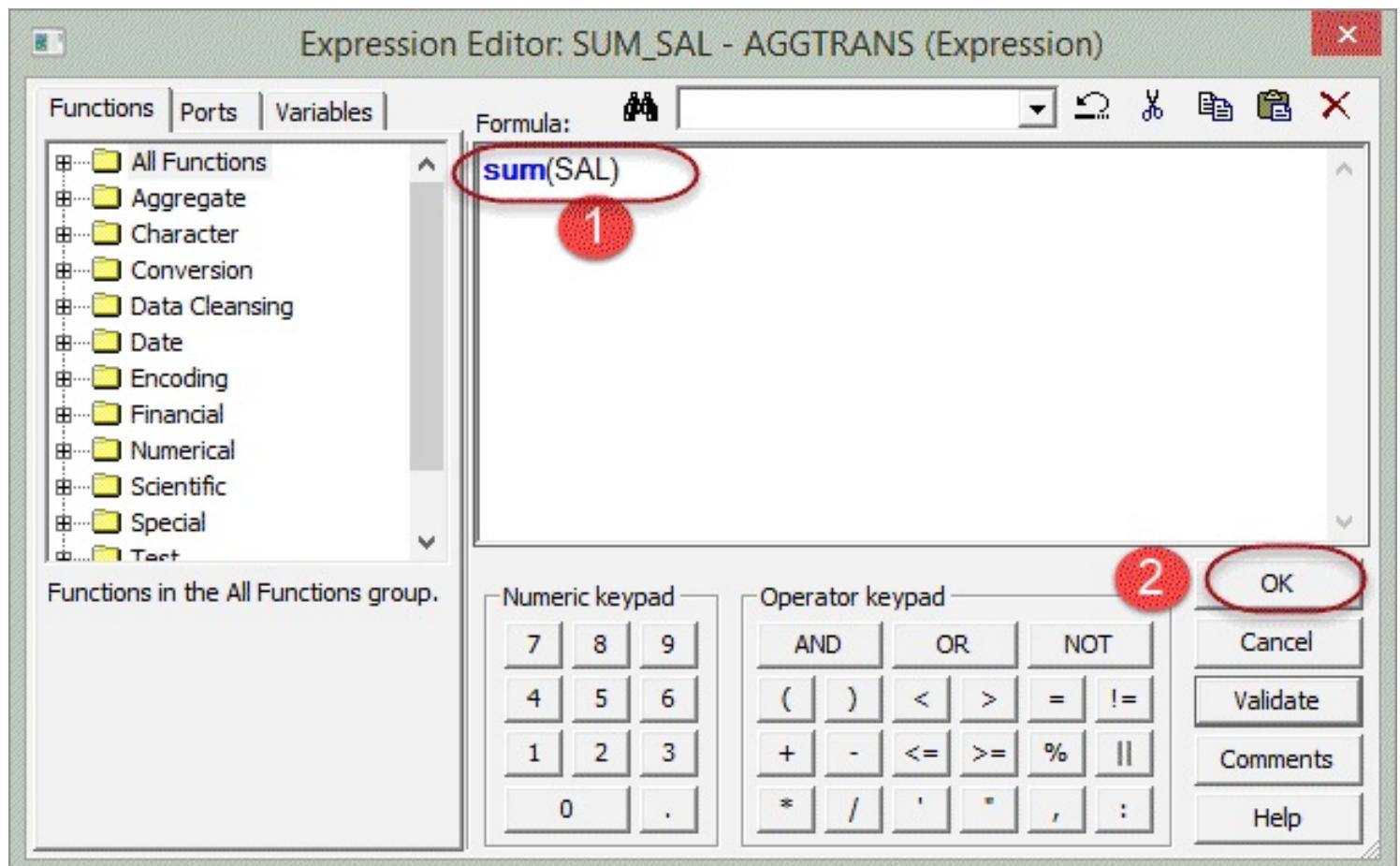
Description:

OK Cancel Apply Help

1 2 3 4 5

Step 6 – In the expression window

1. Add expression- sum(SAL), you have to write this expression.
2. Select Ok Button, this will bring back the edit transformation window.



Step 7 – In edit transformation window, select option "GroupBy" by marking the check box against the deptno column and Click Ok (by selecting group by against the deptno, we are instructing Informatica to group salaries by deptno)

Edit Transformations

Transformation Ports Properties Metadata Extensions

Select transformation: Σ AGGTRANS

Transformation type: Aggregator

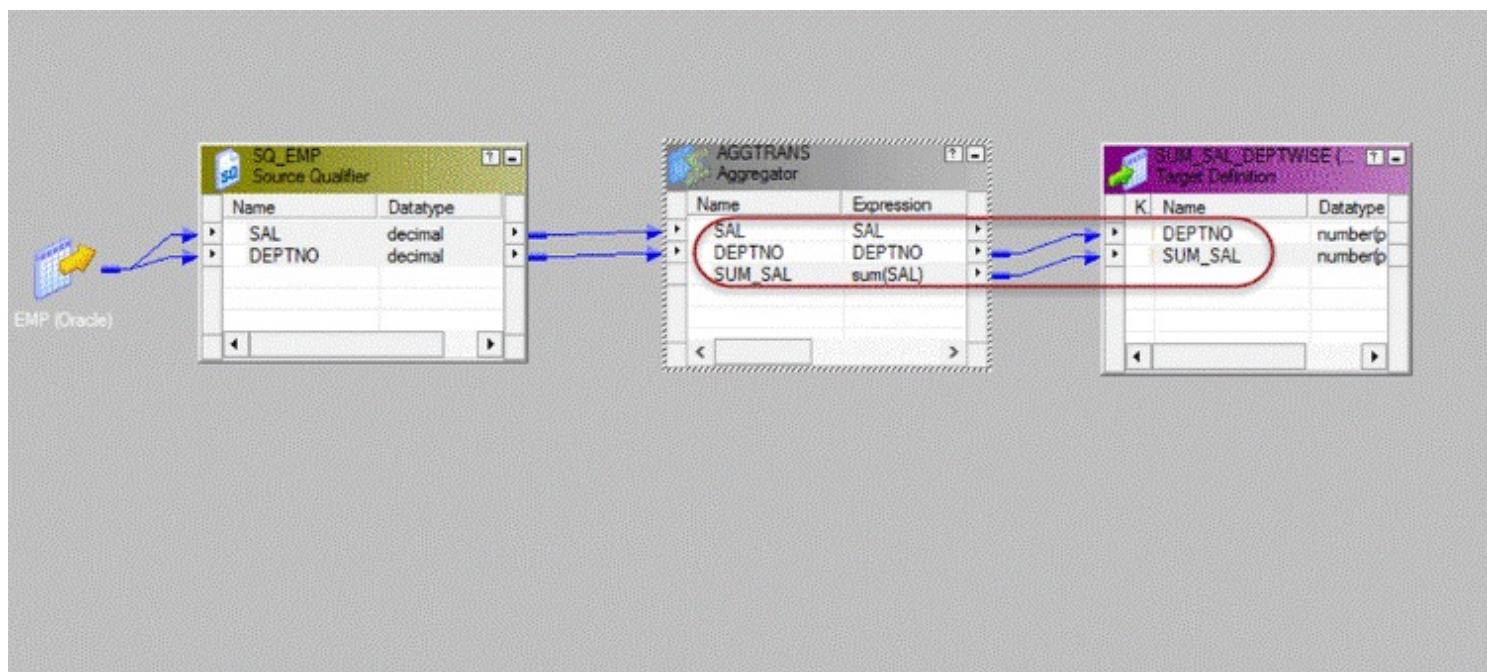
	Port Name	Datatype	Prec	Scale	I	O	V	Expression	GroupBy
1	SAL	decimal	7	2	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	SAL	<input type="checkbox"/>
2	DEPTNO	decimal	2	0	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	DEPTNO	<input checked="" type="checkbox"/>
3	SUM_SAL	double	15	0	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	sum(SAL)	<input type="checkbox"/>

Default value:

Description:

OK **Cancel** **Apply** **Help**

Step 8 – Link the deptno and sum_sal columns from aggregator transformation to the target table



Now save the mapping and execute it after creating a new session for this mapping. The target table would contain the sum of salaries department wise. In this way, we

can use aggregator transformation to calculate aggregate results.

Router Transformation

Similar to filter transformation the router transformation is also used to filter the source data.

The additional functionality provided beside filtering is that the discarded data (filtered out data) can also be collected in the mapping, as well as the multiple filter conditions can be applied to get multiple sets of data.

For example, when filtering the data from deptno =10, we can also get those records where deptno is not equal to 10. So, router transformation gives multiple output groups, and each output group can have its own filter condition.

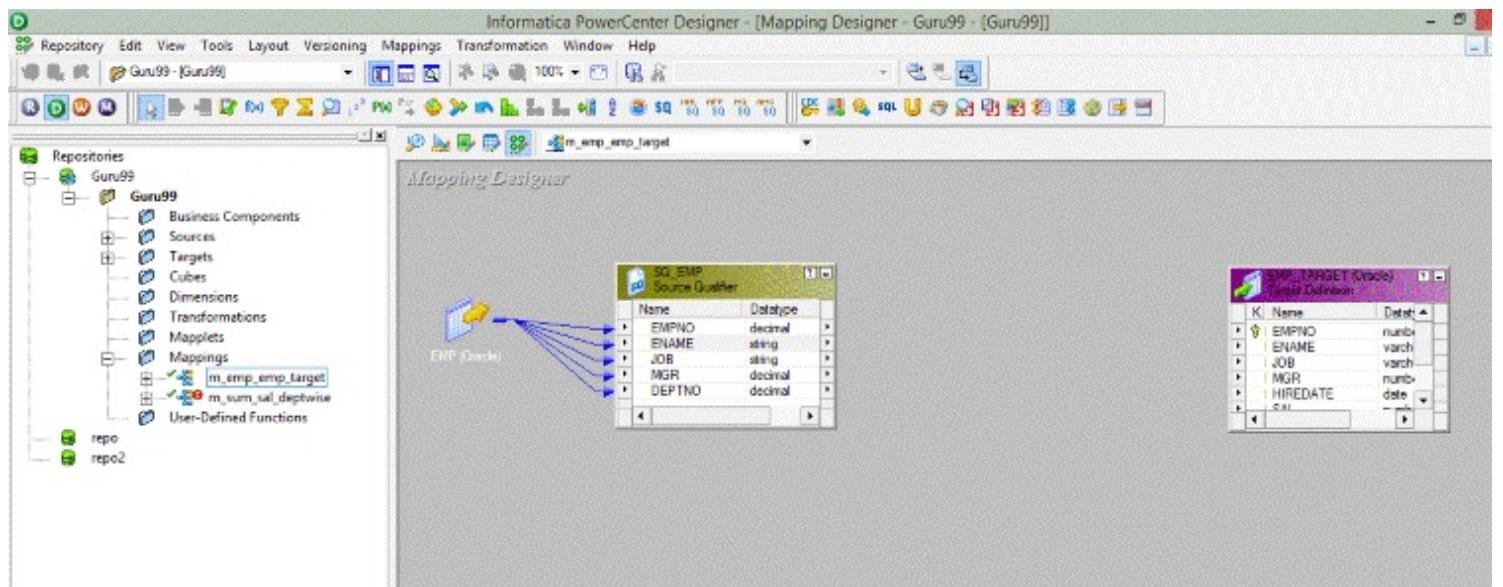
In addition there is also a default group, this default group has those record sets which doesn't satisfy any of the group conditions. For example, if you have created two groups for the filter conditions deptno=10 & dept=20 respectively, then those records which are not having deptno 10 and 20 will be passed into this default group. In short the data which is rejected by the filter groups will be collected by this default group and sometimes there can be a requirement to store these rejected data. In such scenarios, default output group can be useful.

To allow multiple filter condition, the router transformation provides group option.

- There is a default input group which takes input data
- There is also a default output group which provides all those data which is not passed by any filter condition
- For every filter condition, an output group is created in router transformation. You can connect different targets to these different groups.

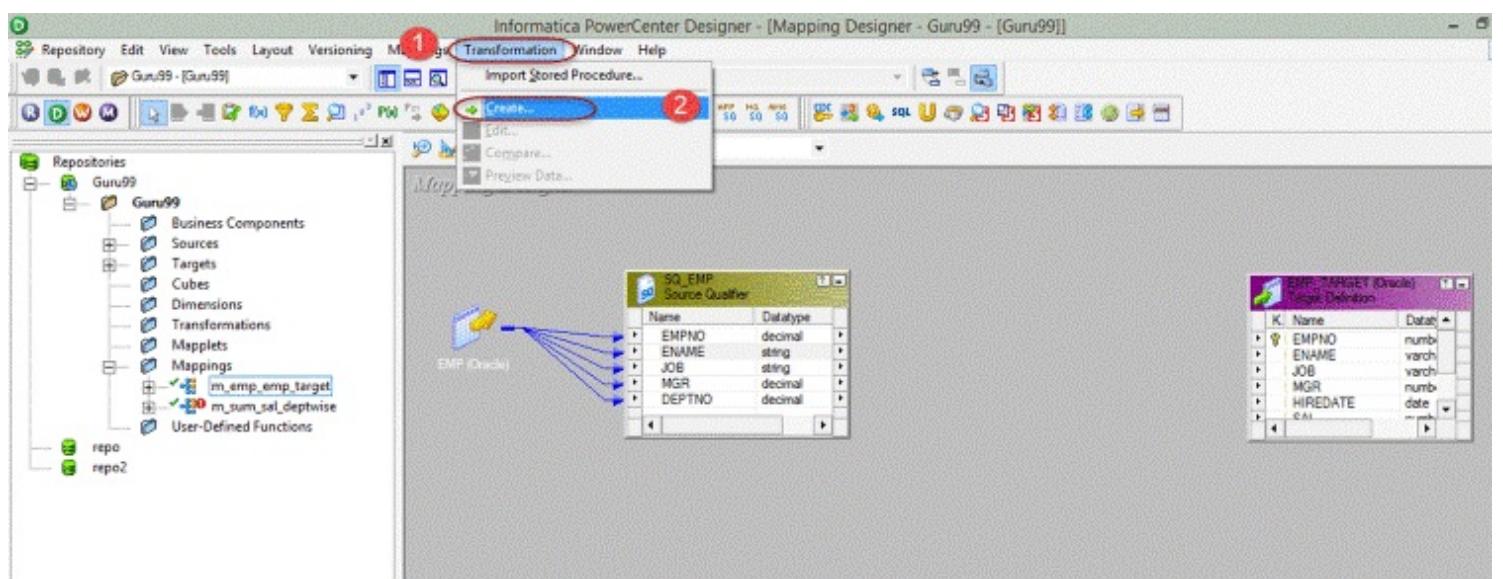
Creating Router Transformation

Step 1 – Create a mapping having source "EMP" and target "EMP_TARGET."



Step 2 – Then in the mapping

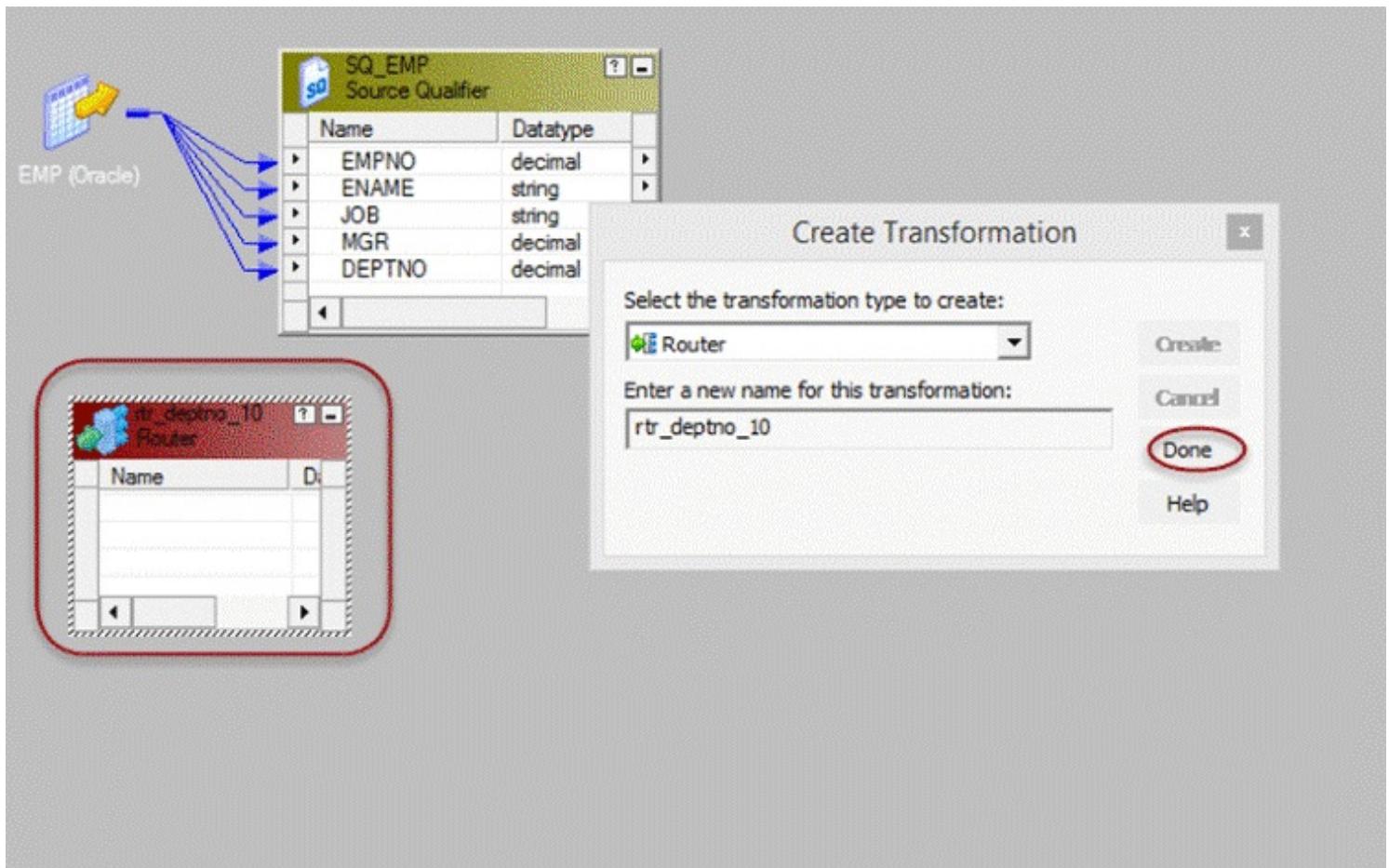
1. Select Transformation menu
2. Select create option



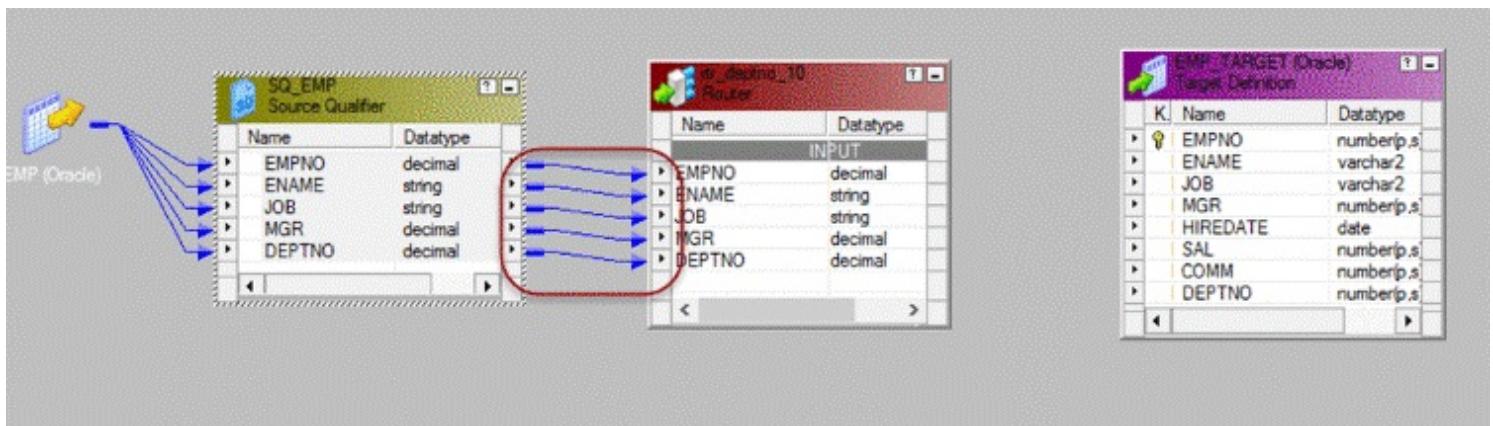
Step 3 – In the create transformation window

1. Select router transformation
2. Enter a name for the transformation "rtr_deptno_10"
3. Select Create option

Step 4 – The router transformation will be created in the mapping, select done option in the window

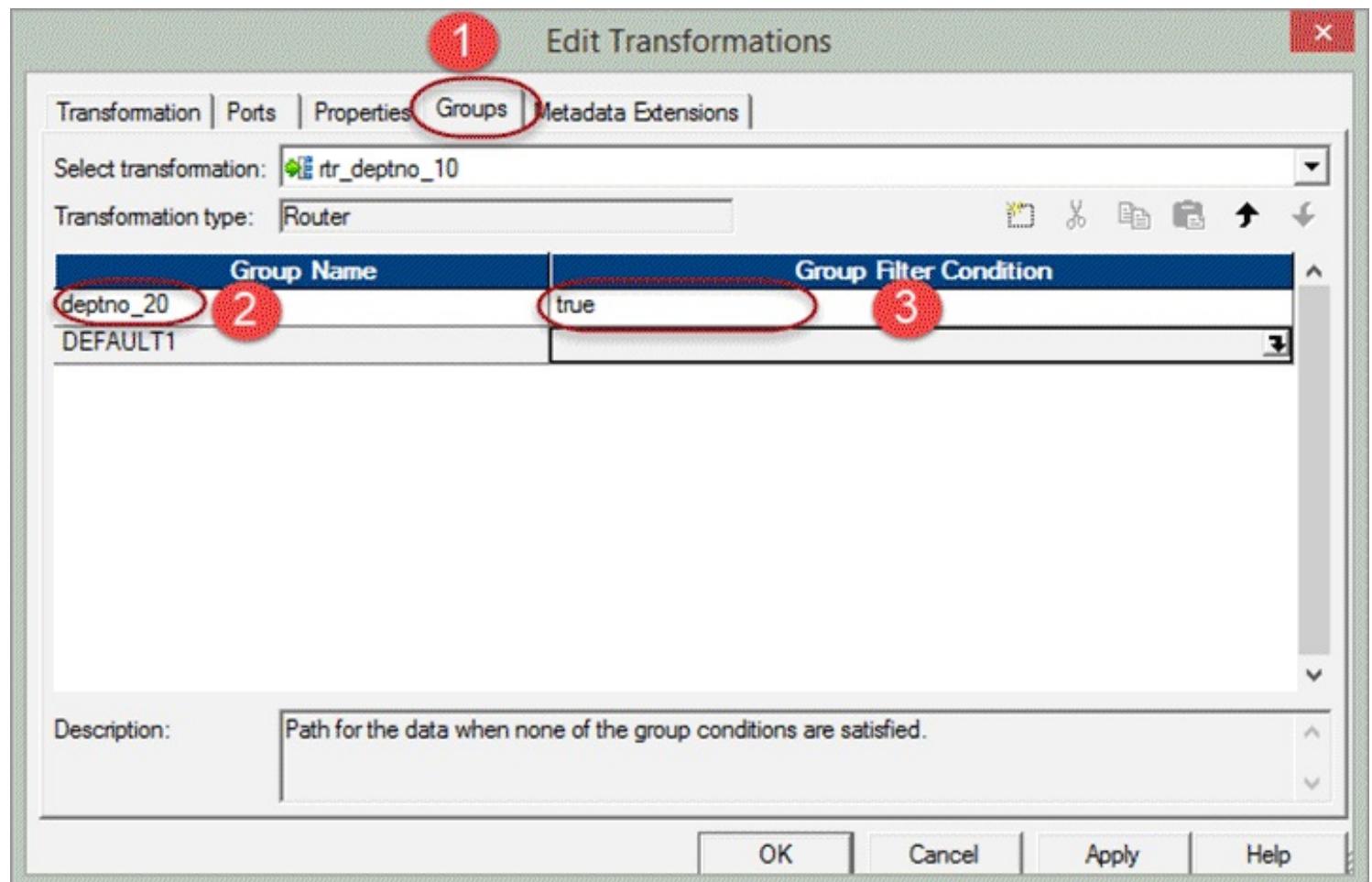


Step 5 – Drag and drop all the columns from Source qualifier to router transformation

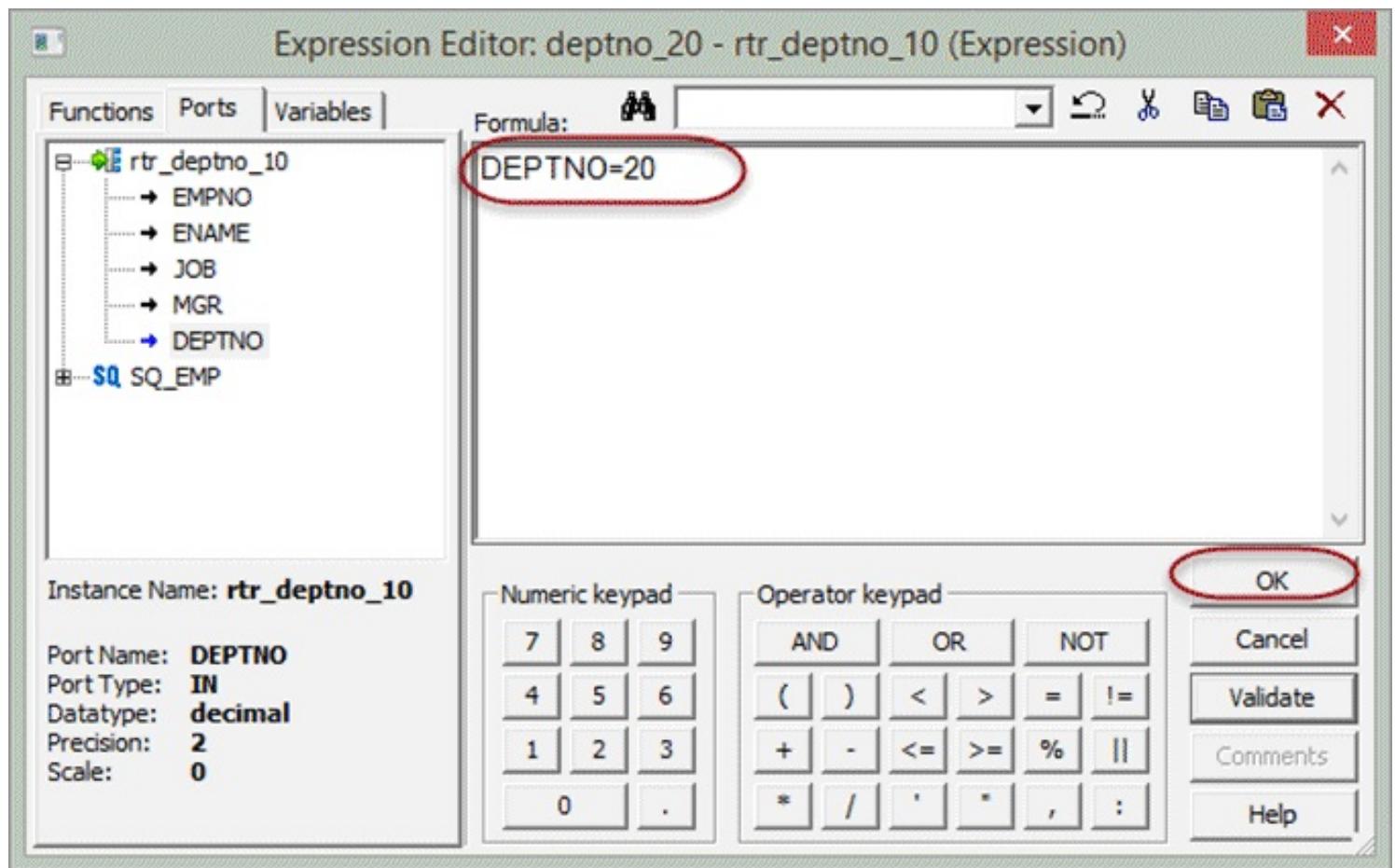


Step 6 – Double click on the router transformation, then in the transformation property of it

1. Select group tab
2. Enter group name "deptno_20"
3. Click on the group filter condition



Step 7 – In the expression editor, enter filter condition deptno=20 and select OK button.



Step 8 – Select OK button in the group window

Edit Transformations

Transformation | Ports | Properties | Groups | Metadata Extensions |

Select transformation: rtr_deptno_10

Transformation type: Router

Group Name	Group Filter Condition
deptno_20	DEPTNO=20
DEFAULT1	

Description:

Path for the data when none of the group conditions are satisfied.

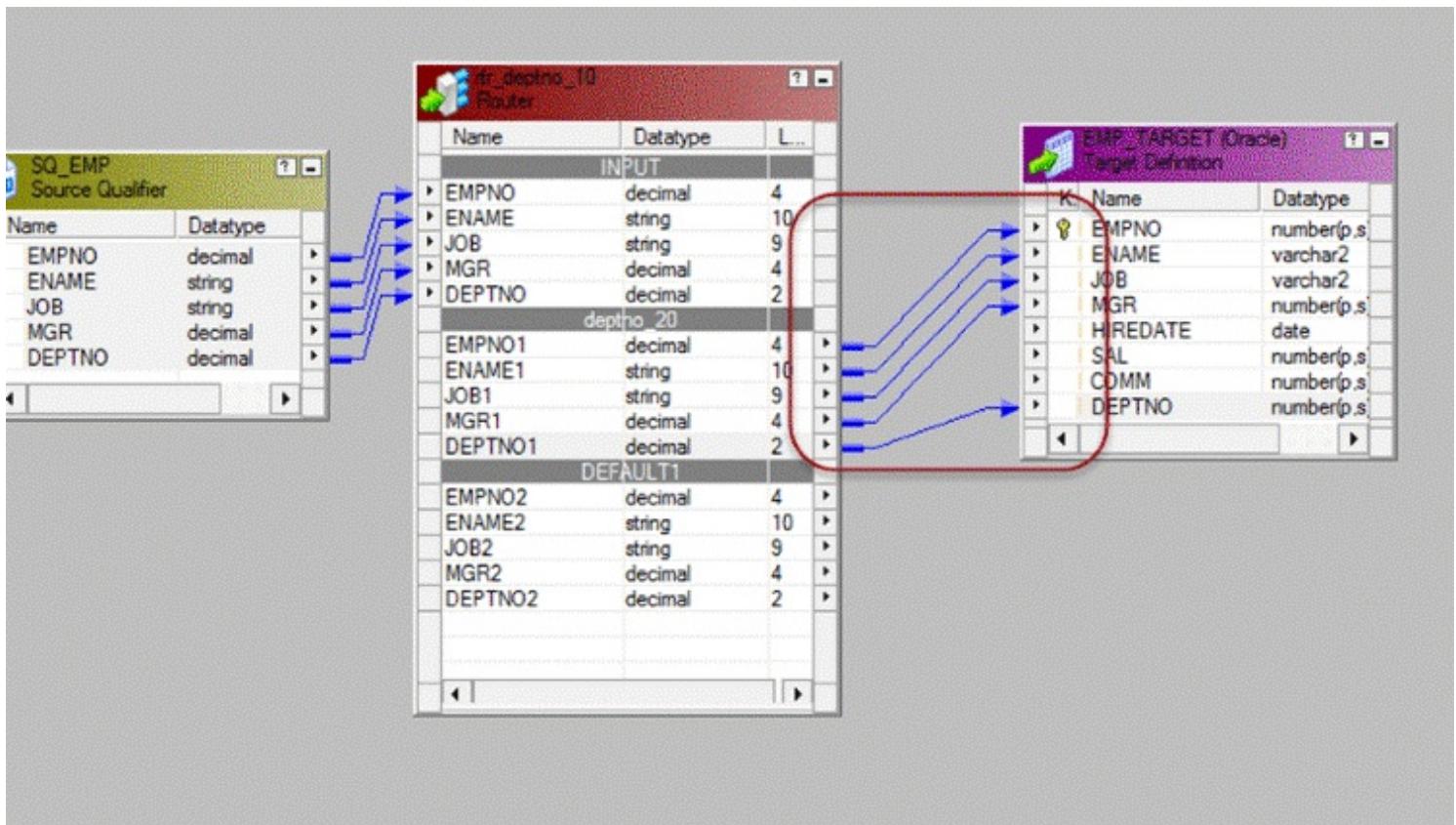
OK

Cancel

Apply

Help

Step 9 – Connect the ports from the group deptno_20 of router transformation to target table ports



Now, when you execute this mapping, the filtered records will get loaded into the target table.

Joiner transformation

The joiner transformation provides you the option to create joins in Informatica. The joins created using joiner transformation are similar to the joins in databases. The advantage of joiner transformation is that joins can be created for heterogeneous systems (different databases).

In joiner transformation, there are two sources which we are going to use it for joins. These two sources are called

- Master Source
- Detail Source

In the properties of joiner transformation, you can select which data source can be Master and which source can be detail source.

During execution, the master source is cached into the memory for joining purpose. So it is recommended to select the source with less number of records as the master

source.

The following joins can be created using joiner transformation

1. Master outer join

In Master outer join, all records from the Detail source are returned by the join and only matching rows from the master source are returned.

2. Detail outer join

In detail outer join only matching rows are returned from the detail source, and all rows from the master source are returned.

3. Full outer join

In full outer join, all records from both the sources are returned. Master outer and Detail outer joins are equivalent to left outer joins in SQL.

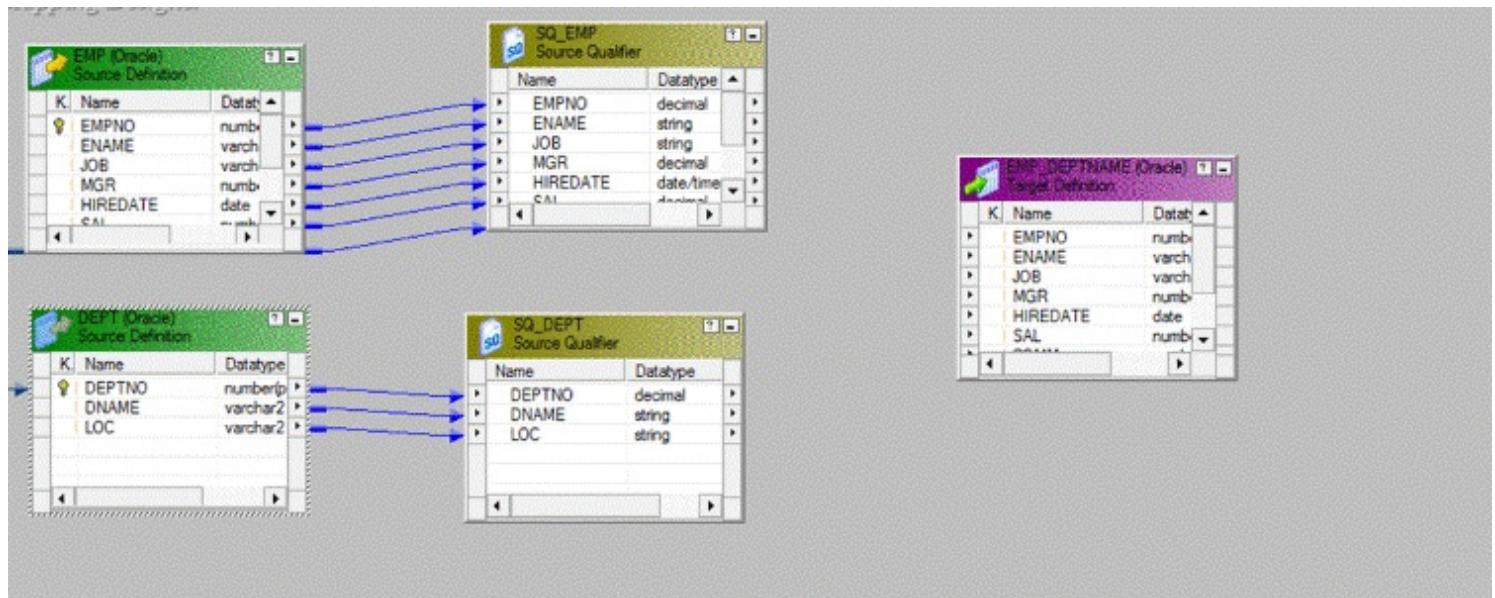
4. Normal join

In normal join only matching rows are returned from both the sources.

In this example, we will join emp and dept tables using joiner transformation

Step 1 – Create a new target table EMP_DEPTNAME in the database using the below script and import the table in Informatica targets.

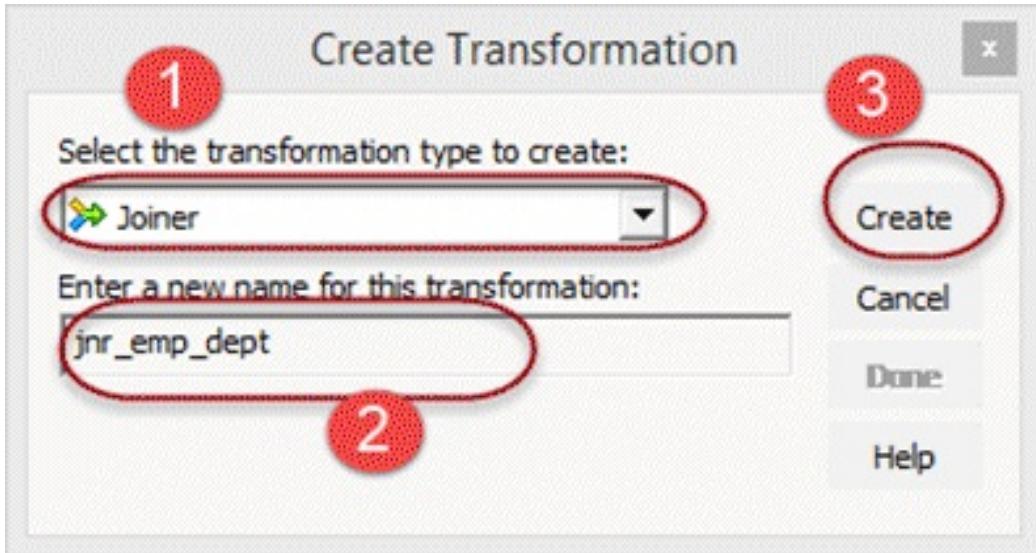
Step 2 - Create a new mapping and import source tables "EMP" and "DEPT" and target table which we created in the previous step



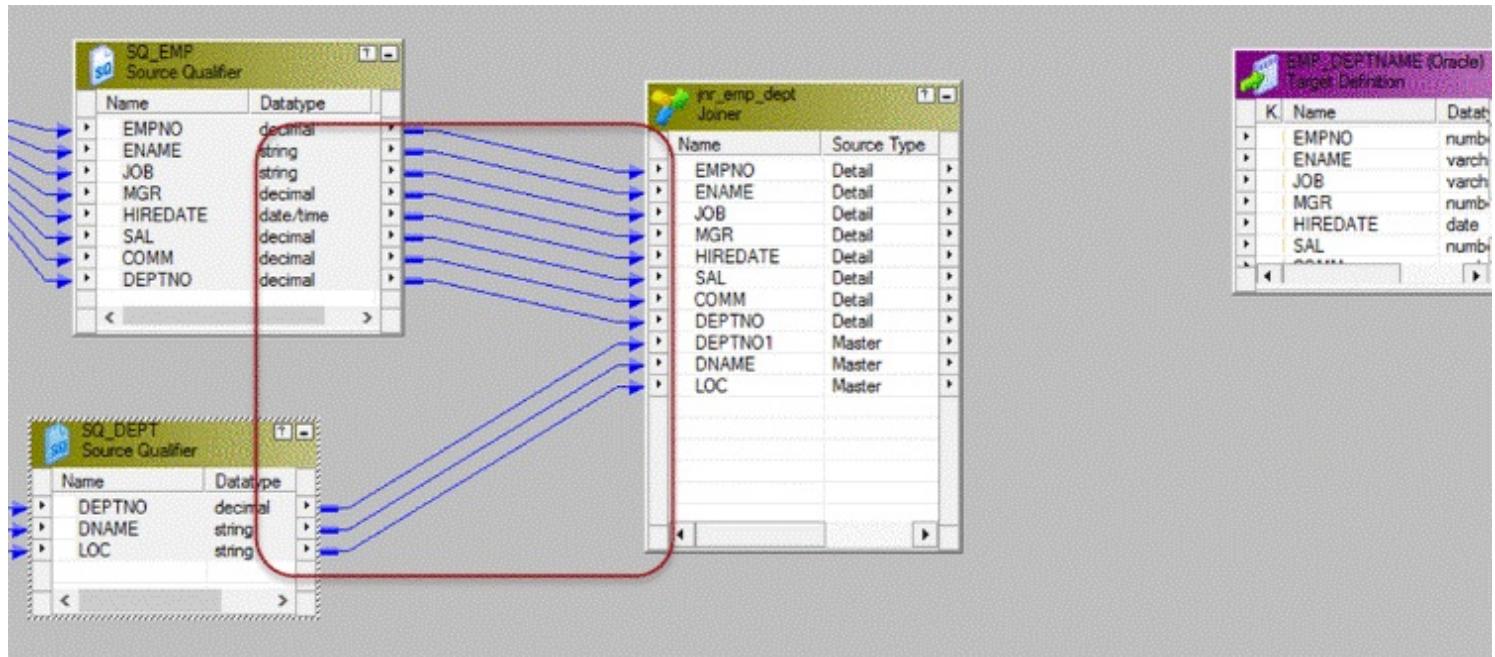
Step 3 – From the transformation menu, select create option.

1. Select joiner transformation

2. Enter transformation name "jnr_emp_dept"
3. Select create option

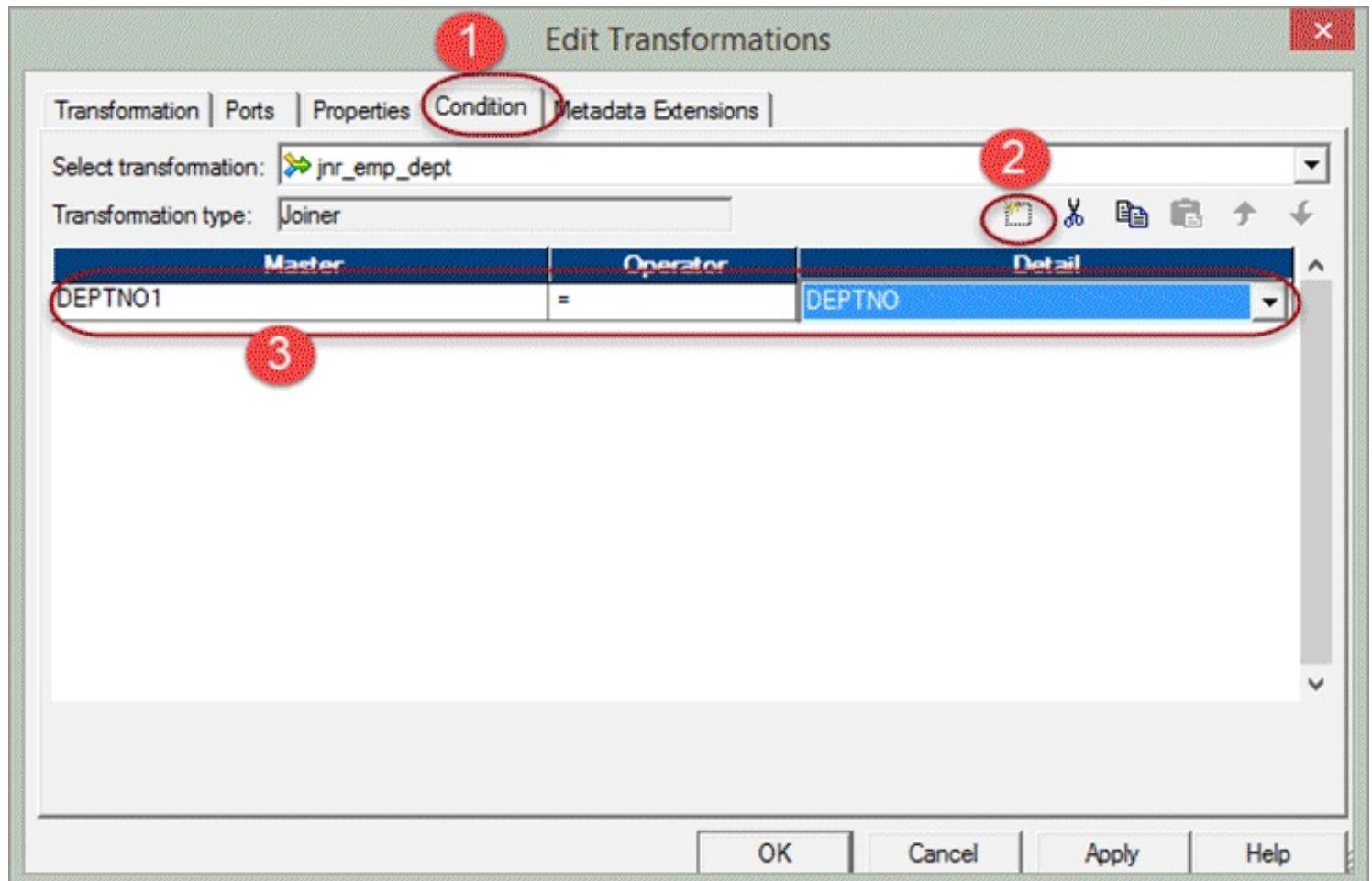


Step 4 – Drag and drop all the columns from both the source qualifiers to the joiner transformation



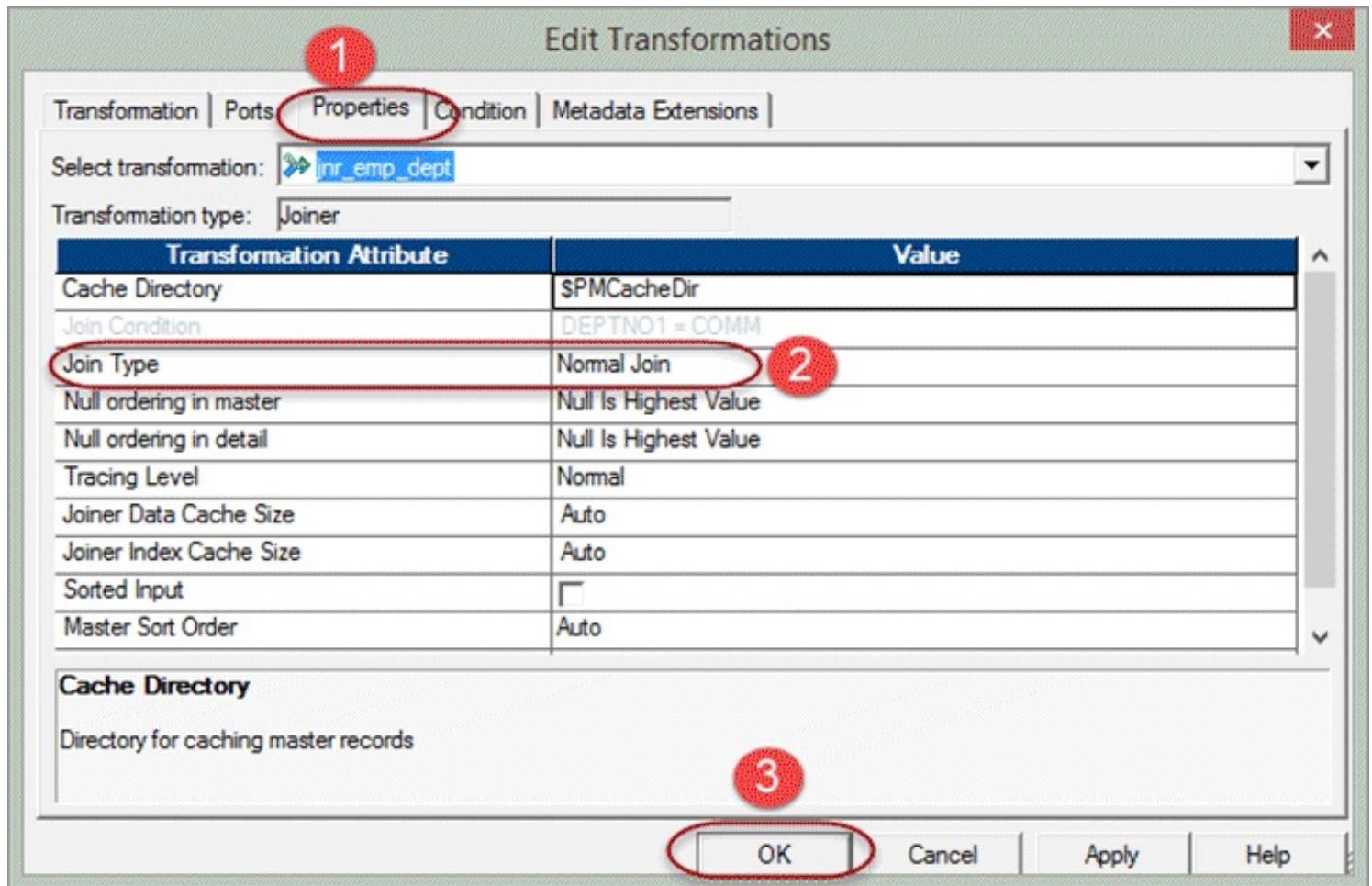
Step 5 - Double click on the joiner transformation, then in the edit transformation window

1. Select condition tab
2. Click on add new condition icon
3. Select deptno in master and detail columns list



Step 6 - Then in the same window

1. Select properties tab
2. Select normal Join as join type
3. Select OK Button



For performance optimization, we assign the master source to the source table pipeline which is having less no of records. To perform this task –

Step 7 – Double click on the joiner transformation to open edit properties window, and then

1. Select ports tab
2. Select any column of a particular source which you want to make a master
3. Select OK

Edit Transformations

1

Transformation Ports Properties Condition Metadata Extensions |

Select transformation: **inr_emp_dept**

Transformation type: Joiner

	Port Name	Datatype	Prec	Scale	I	O	M
1	EMPNO	decimal	4	0	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
2	ENAME	string	10	0	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
3	JOB	string	9	0	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
4	MGR	decimal	4	0	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
5	HIREDATE	date/time	29	9	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
6	SAL	decimal	7	2	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
7	COMM	decimal	7	2	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
8	DEPTNO	decimal	2	0	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
9	DEPTNO1	decimal	2	0	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
10	DNAME	string	14	0	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
11	LOC	string	13	0	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>

Default value:

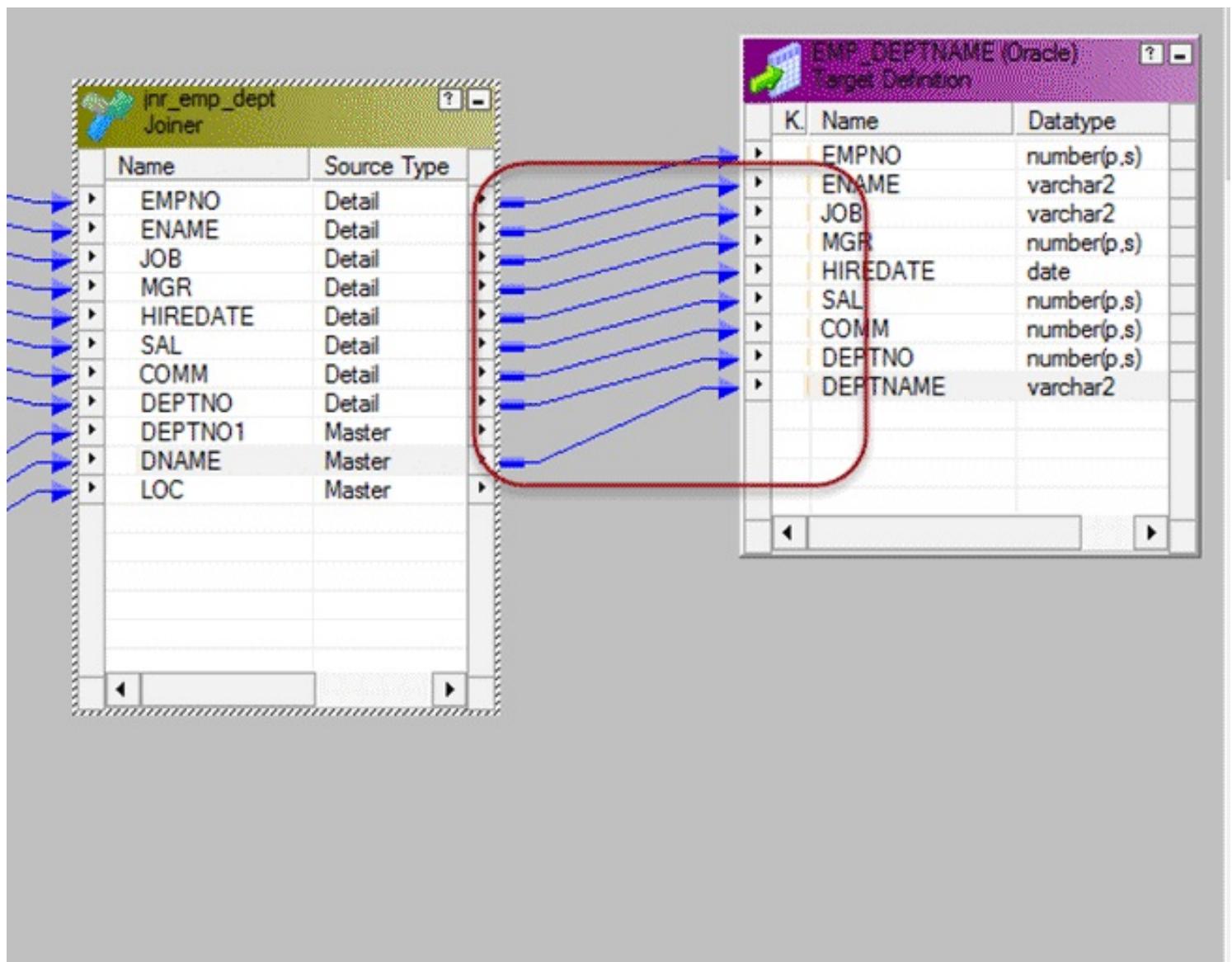
Description:

2

3

OK Cancel Apply Help

Step 8 – Link the relevant columns from joiner transformation to target table



Now save the mapping and execute it after creating session and workflow for it. The join will be created using Informatica joiner, and relevant details will be fetched from both the tables.

Rank Transformation

Rank Transformation performs the filtering of data based on group and ranks.

For example, you want to get ten records of employees having highest salary, such kind of filtering can be done by rank transformation.

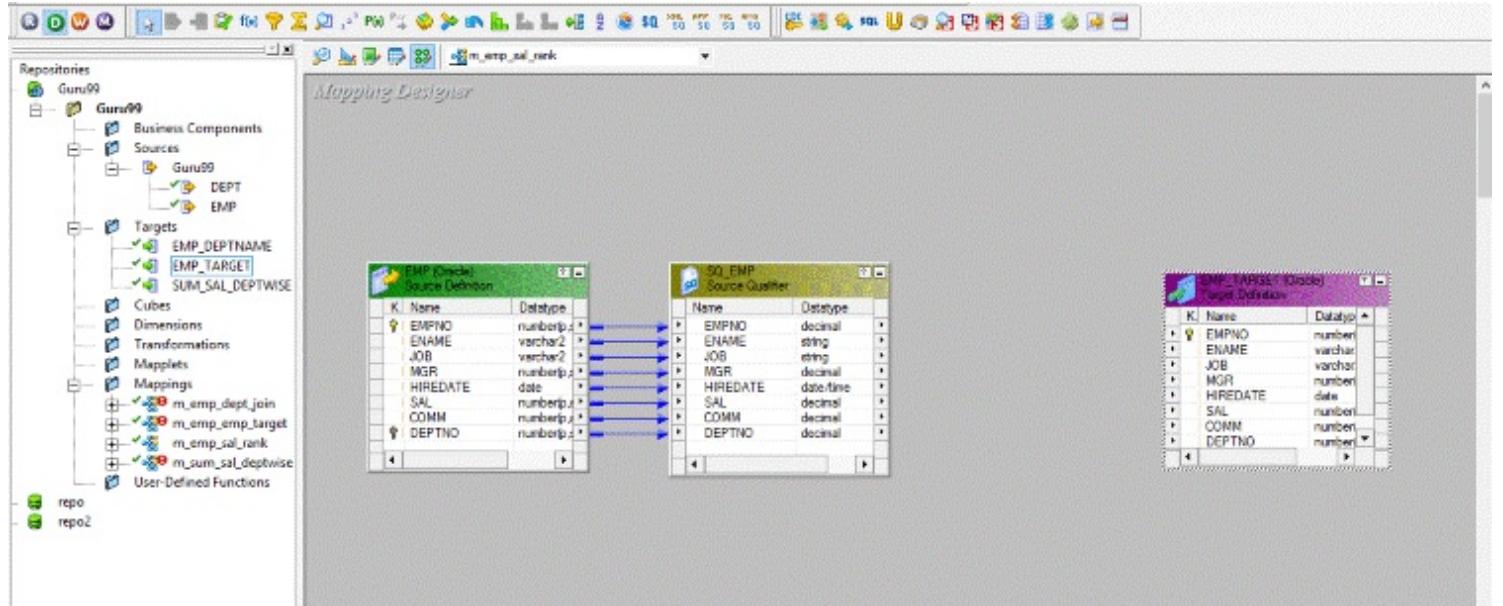
Rank transformation also provides the feature to do ranking based on groups. Like if you want to get top ten salaried employee department wise, then this grouping can be done with this transformation.

Rank transformation is an active transformation, as it affects the number of output rows.

The rank transformation has an output port by which it assigns a rank to the rows.

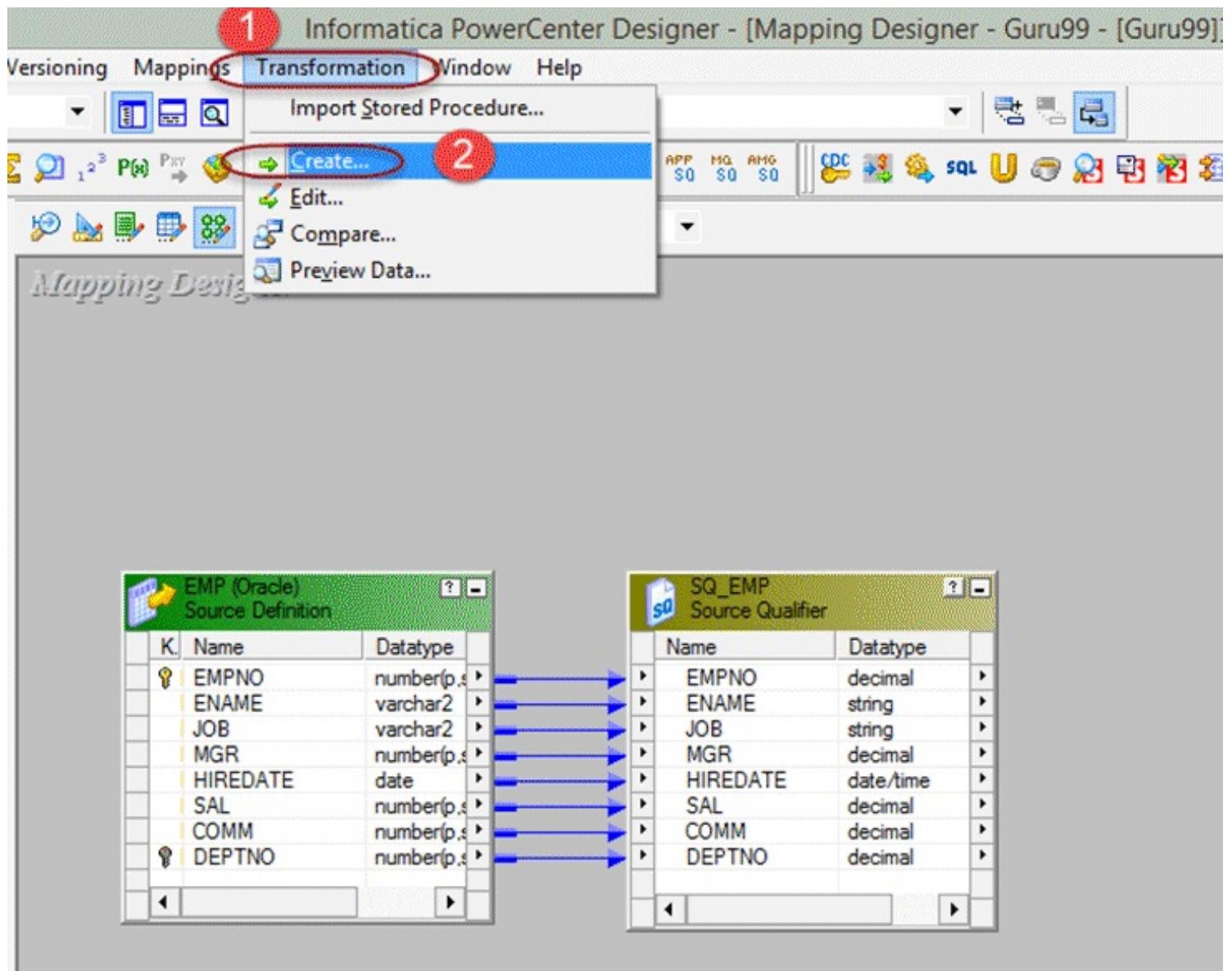
Our requirement is to load top 3 salaried employees for each department; we will implement this using rank transformation.

Step 1 - Create a mapping having source EMP and target EMP_TARGET



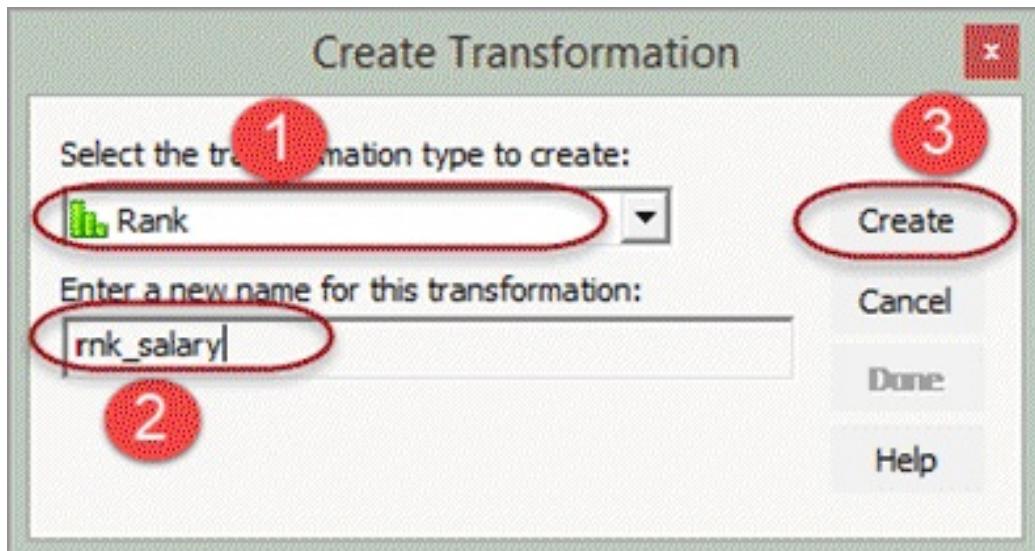
Step 2- Then in the mapping

1. Select transformation menu
2. Select create option

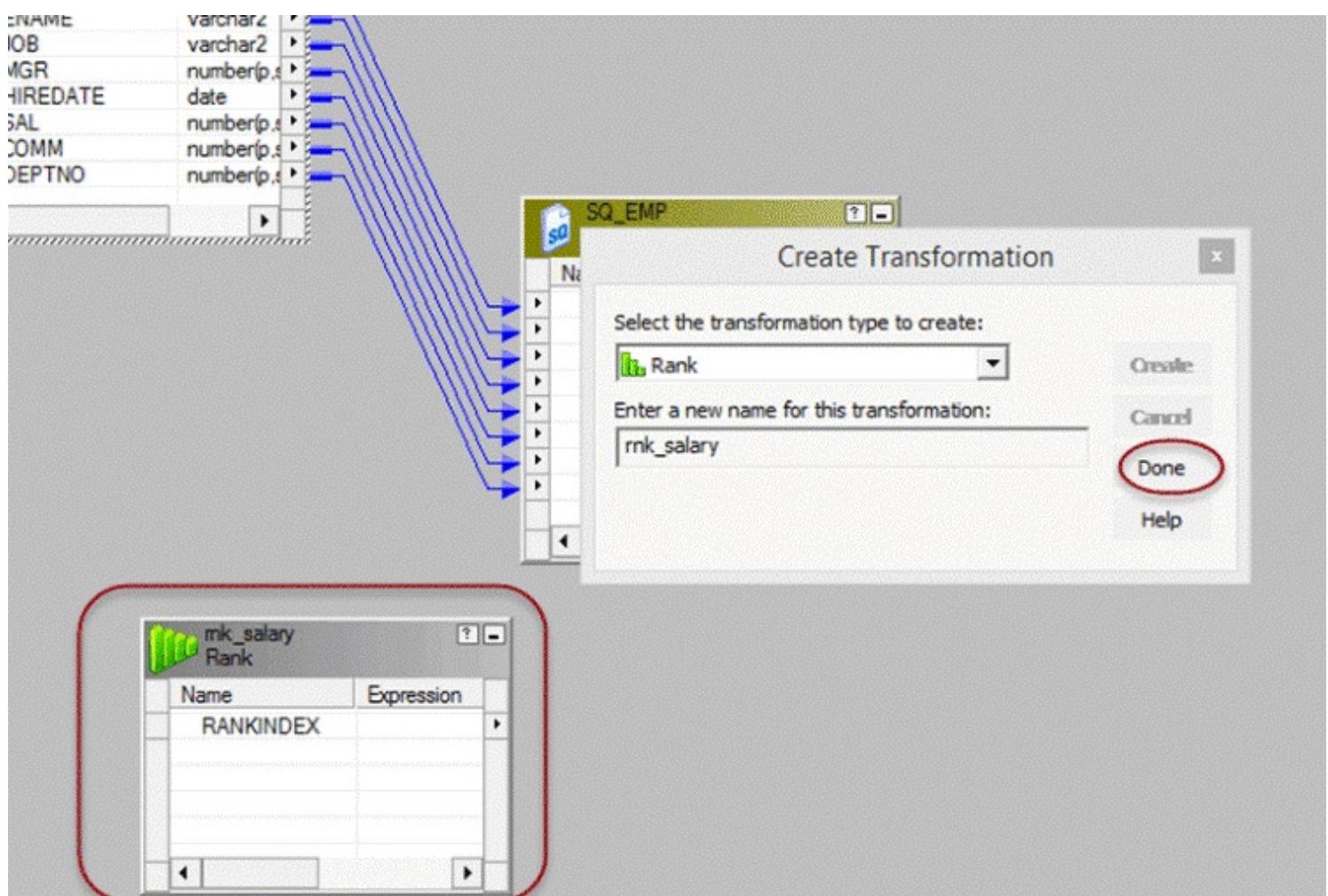


Step 3 – In the create transformation window

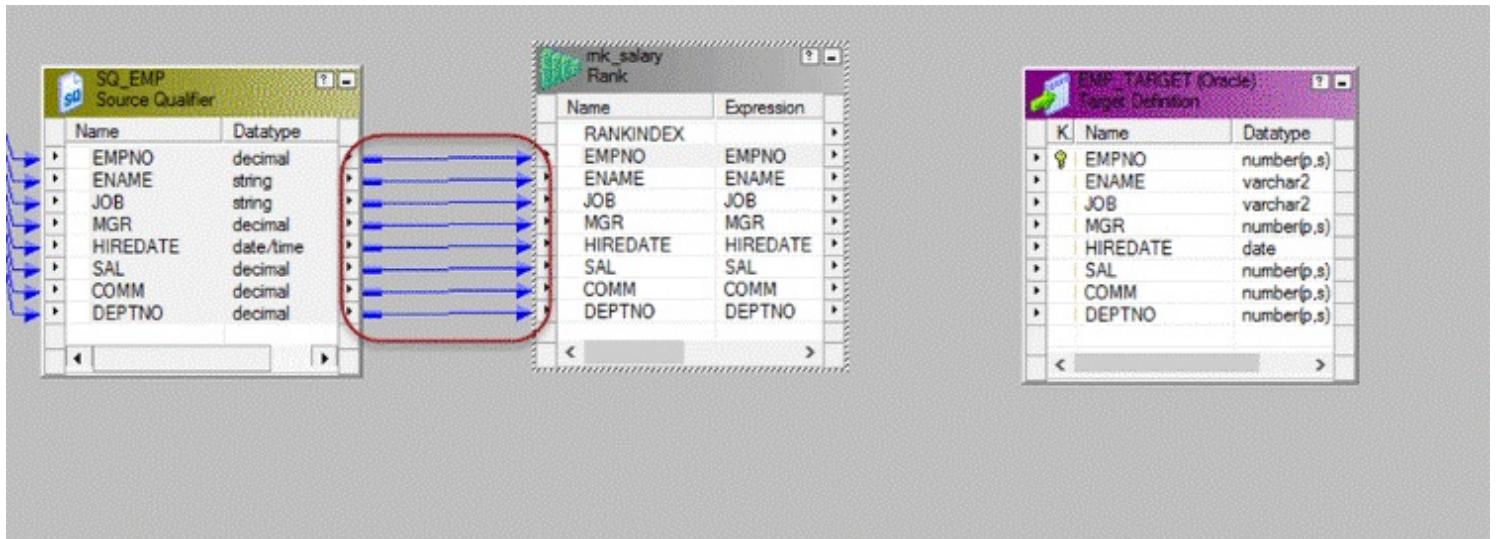
1. Select rank transformation
2. Enter transformation name "rnk_salary"
3. Select Create button



Step 4 – The rank transformation will be created in the mapping, select done button in the window

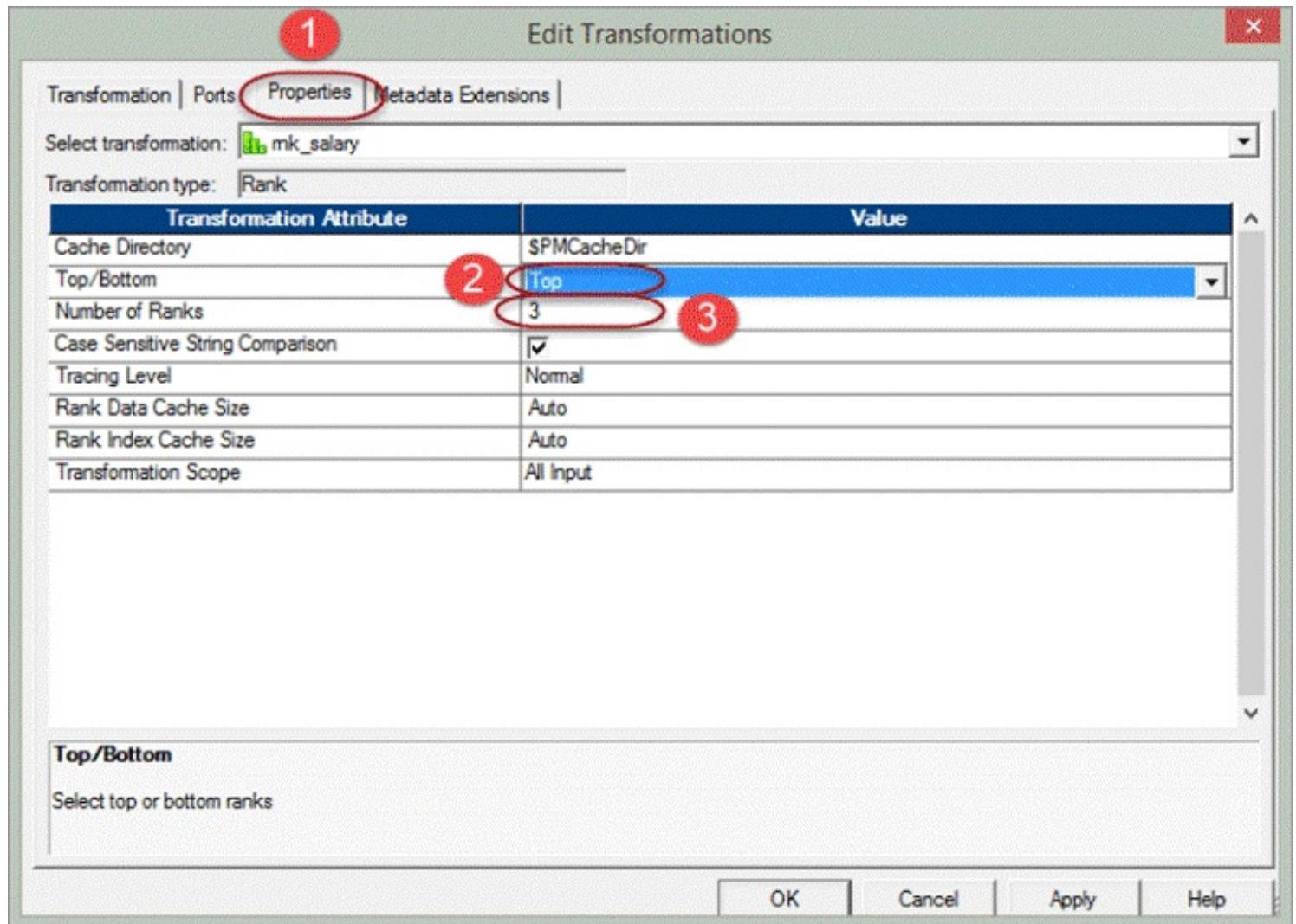


Step 5 – Connect all the ports from source qualifier to the rank transformation



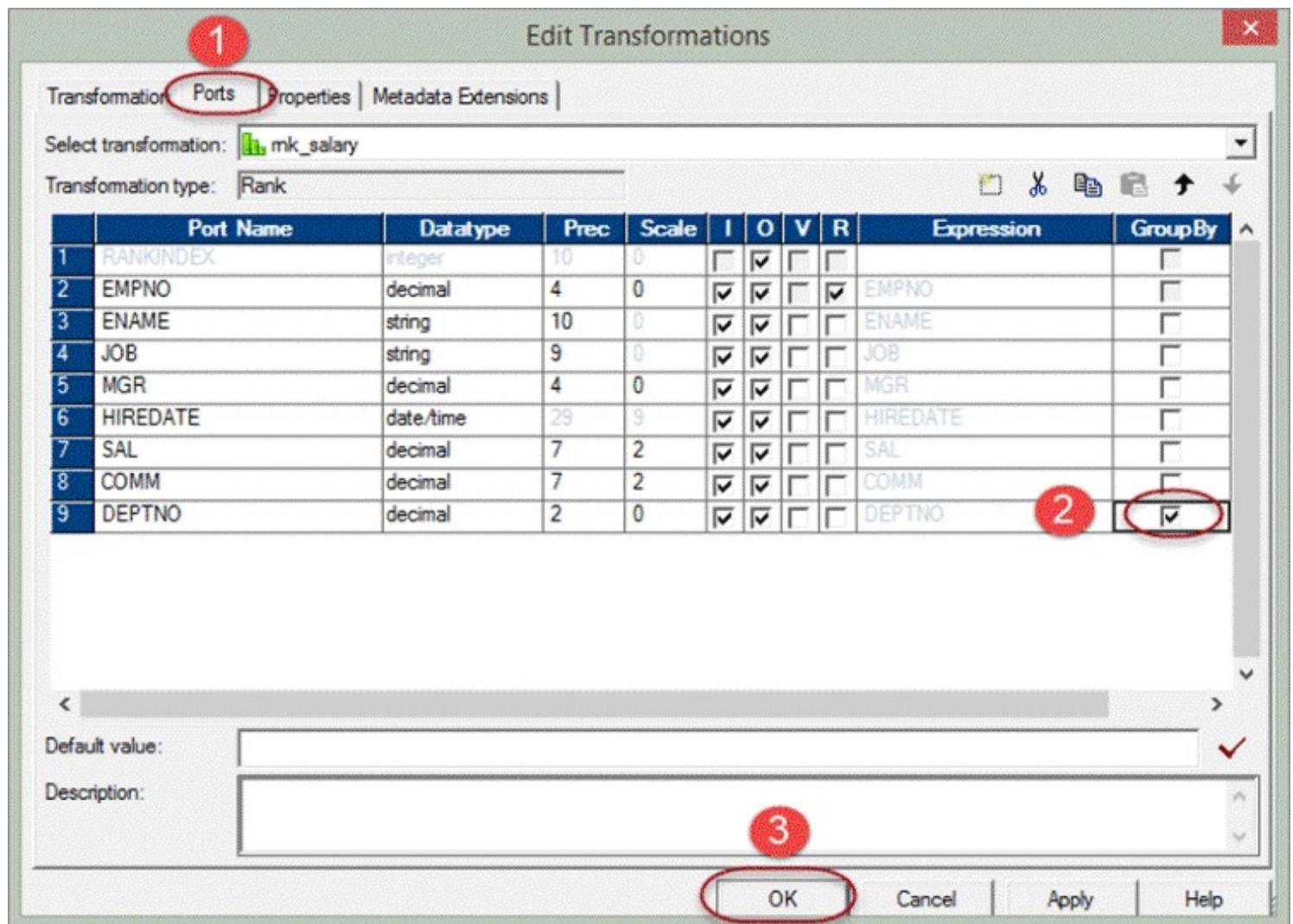
Step 6- Double click on the rank transformation and it will open "edit transformation window". In this window

1. Select properties menu
2. Select "Top" option from the Top/Bottom property
3. Enter 3 in the number of ranks

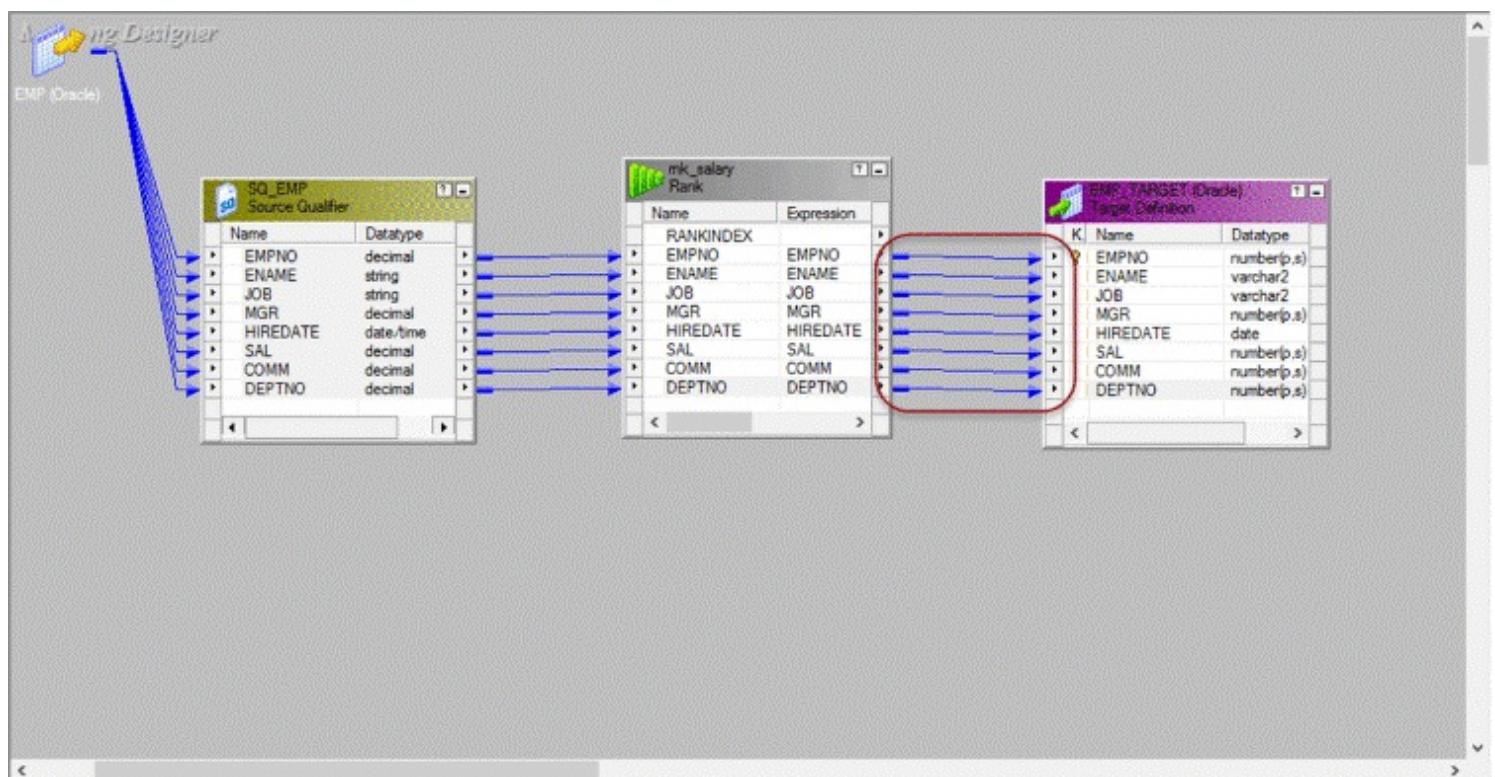


Step 7 – In the "edit transformation" window again

1. Select ports tab
2. Select group by option for the deptno column
3. Select ok button



Step 8 –Connect the ports from rank transformation to the target table



Now, save the mapping and execute it after creating session and workflow. The source qualifier will fetch all the records, but rank transformation will pass only records having three high salaries for each department.

Sequence Generator Transformation

Sequence generator transformation is used to generate numeric sequence values like 1, 2, 3, 4, 5 etc.

For example, you want to assign sequence values to the source records, then you can use sequence generator. The generated sequence values can be like 5, 10, 15, 20, 25 etc. or 10, 20, 30, 40, 50 etc. depending upon the configured properties of the transformation.

The sequence generator is also used to generate primary key values. It is a passive transformation & it does not affect the number of input rows.

It has two output ports

- CURRVAL
- NEXTVAL

CURRVAL port value is always NEXTVAL+1.

To generate the sequence numbers, we always use the NEXTVAL column.

Properties of Sequence Generator Transformation

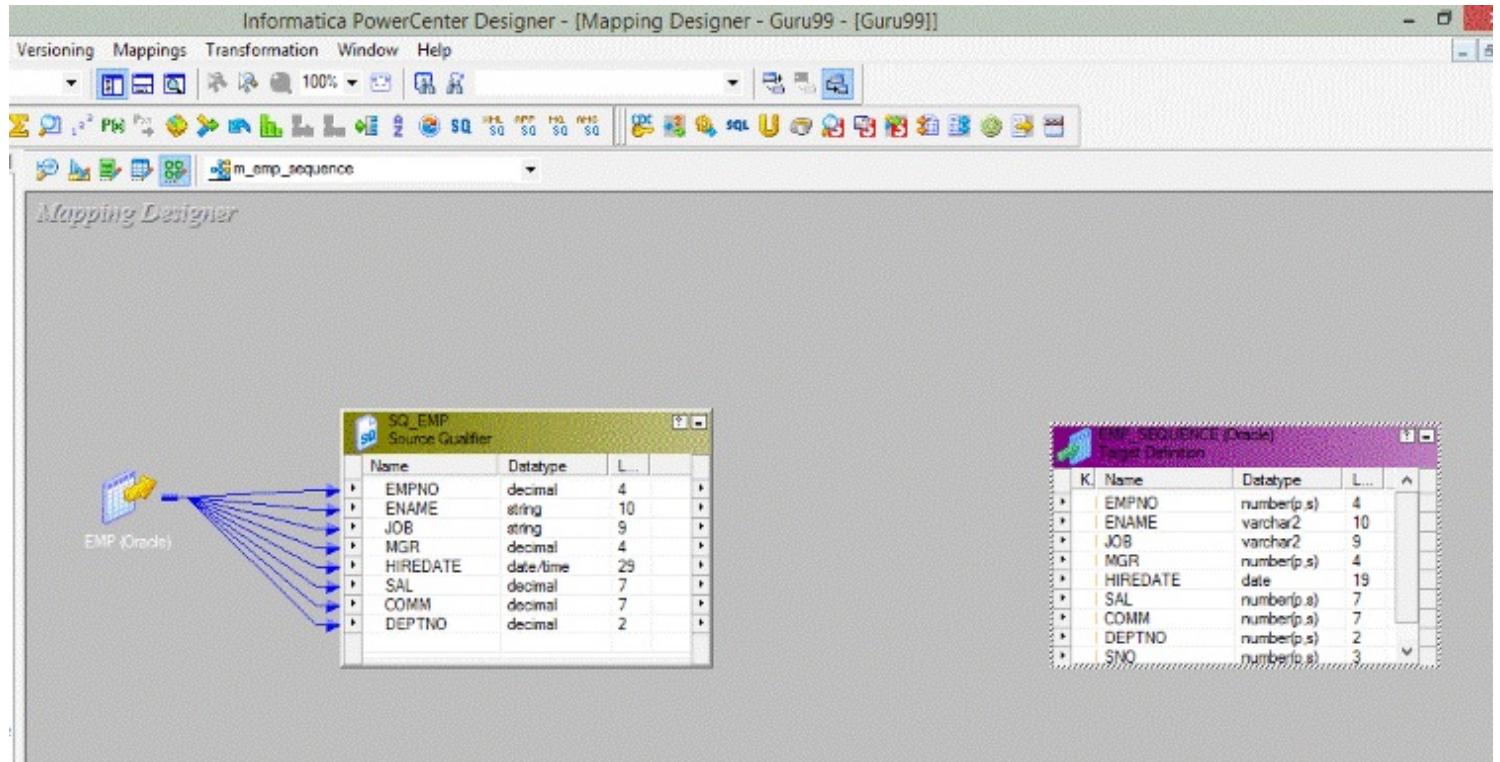
- Start Value – It is the first value that will be generated by the transformation, the default value is 0.
- Increment by – This is the number by which you want to increment the values. The default value is 1.
- End value – It is the maximum value that the transformation should generate.
- Cycle – if this option is set then after reaching the end of the value, the transformation restarts from the start value.

In this example, we will generate sequence numbers and store in the target.

Step 1 – Create a target table with the following script.

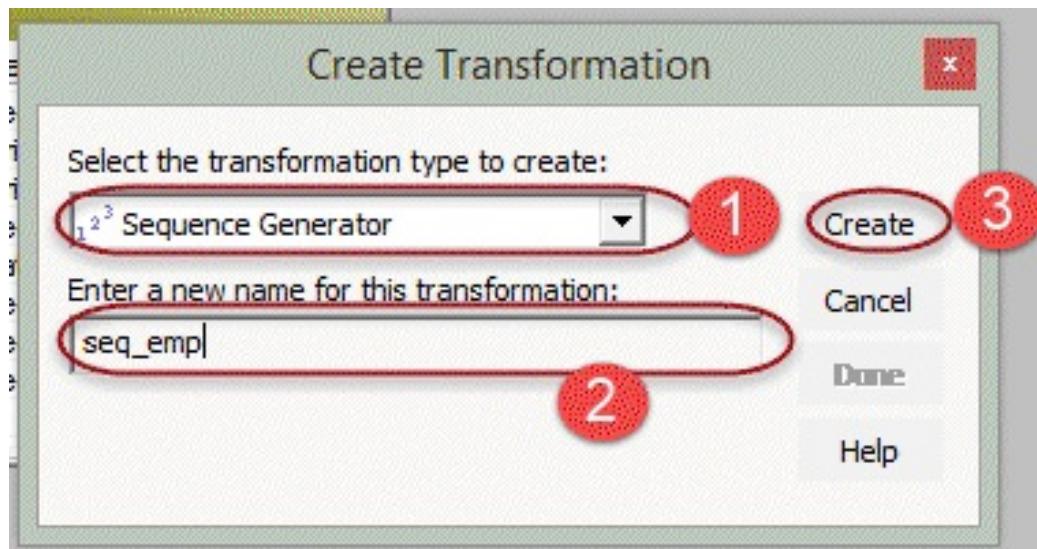
Step 2- Import the table in Informatica as target table

Step 3 – Create a new mapping and import EMP source and EMP_SEQUENCE target table

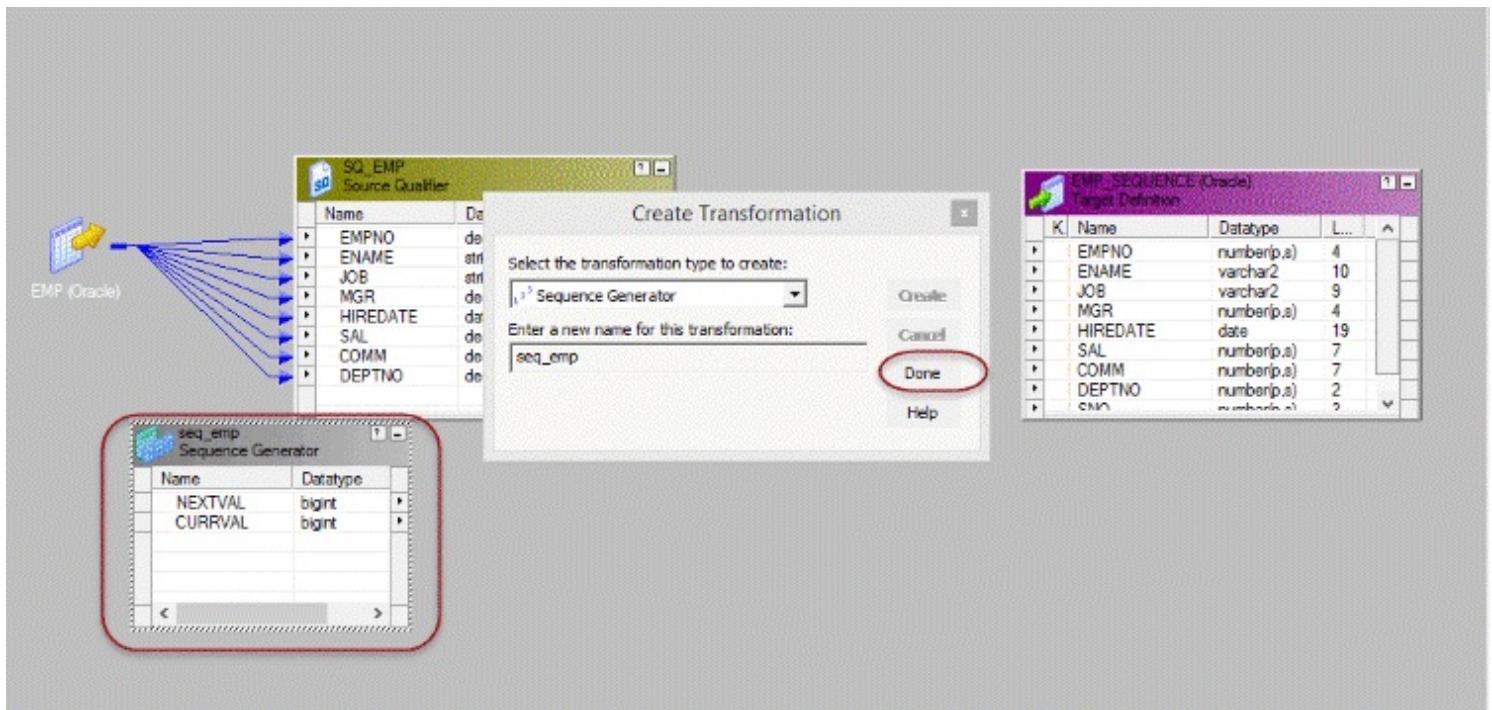


Step 4 – Create a new transformation in the mapping

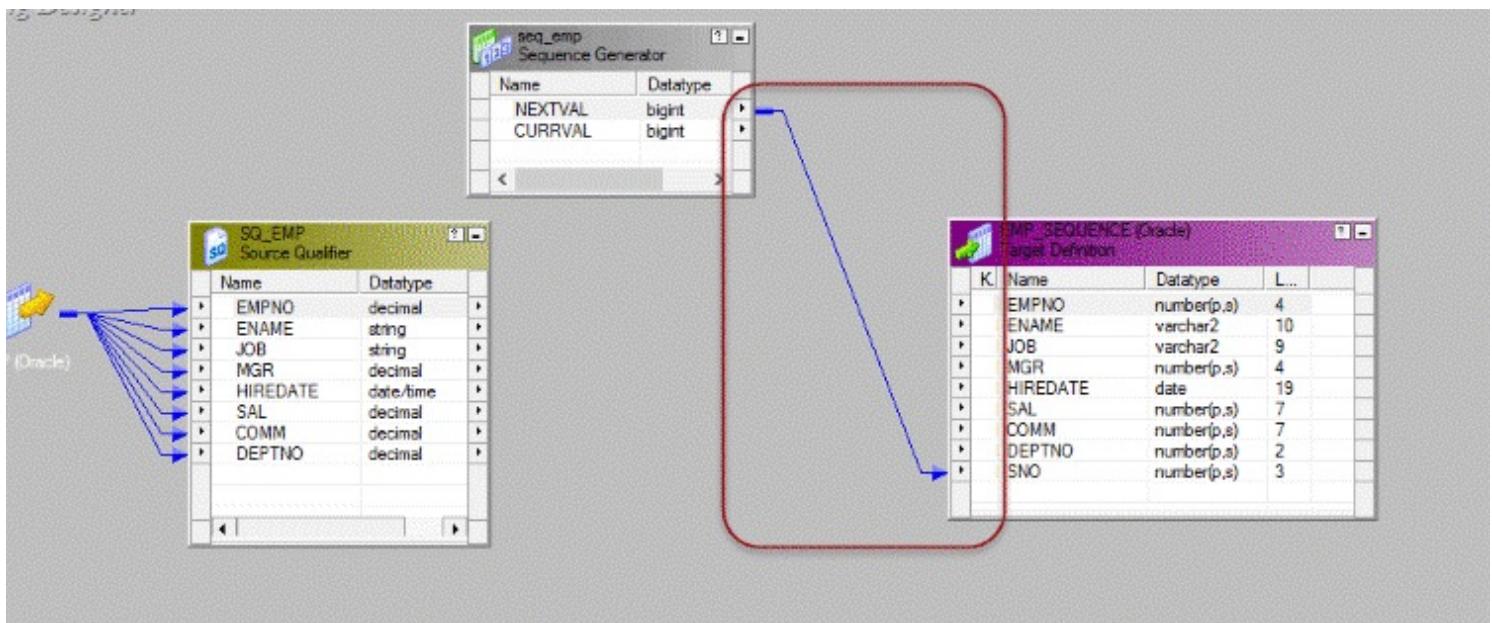
1. Select sequence transformation as the type
2. Enter transformation name "seq_emp"
3. Select Create option



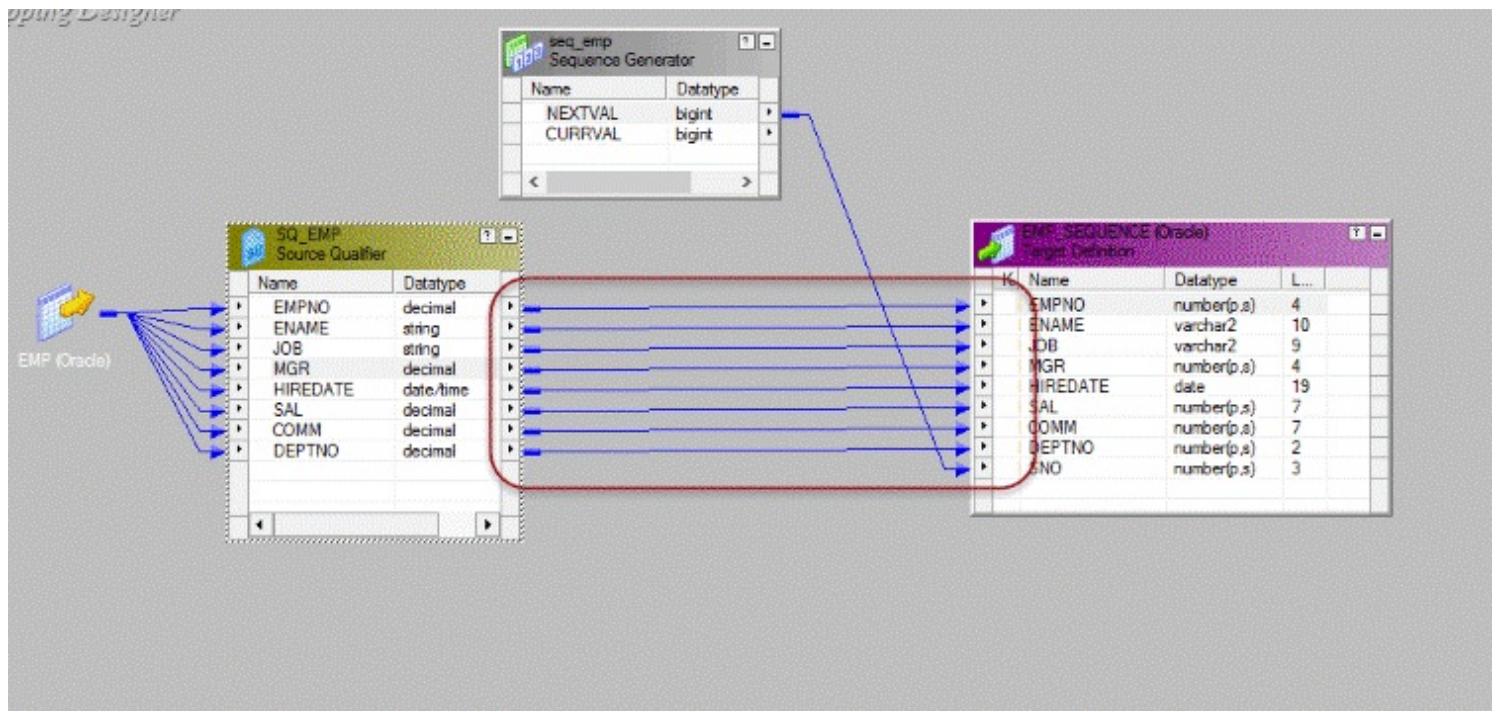
Step 5 - Sequence generator transformation will be created, select the done option



Step 6 - Link the NEXTVAL column of sequence generator to SNO column in target

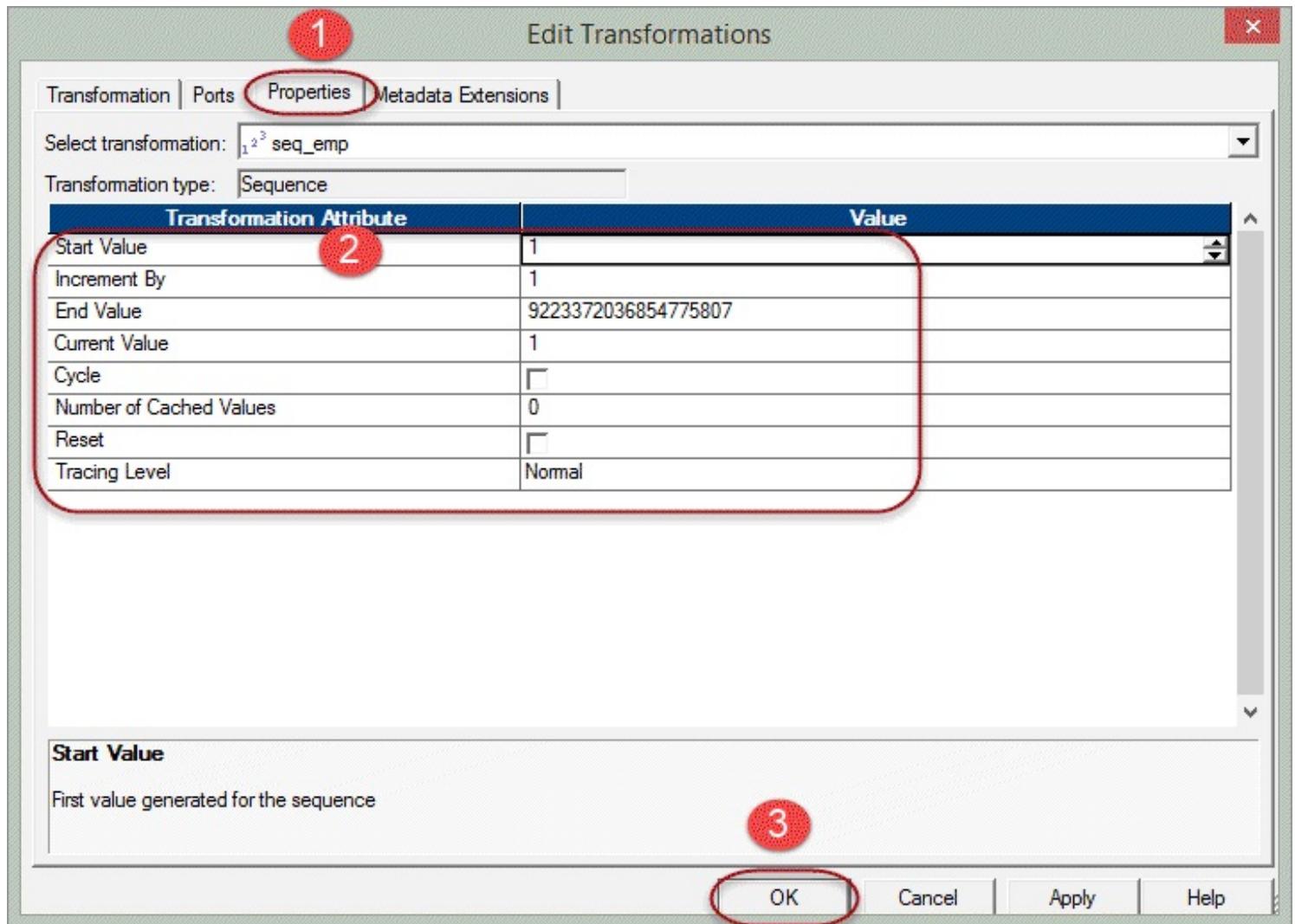


Step 7 – link the other columns from source qualifier transformation to the target table



Step 8 – Double click on the sequence generator to open property window, and then

1. Select the properties tab
2. Enter the properties with Start value =1, leave the rest properties as default
3. Select OK button



Now save the mapping and execute it after creating the session and workflow.

The sno column in the target would contain the sequence numbers generated by the sequence generator transformation.

In our example, the sequences will be like 1 – Scott, 2 – King, 3 – Adam, 4 – Miller, etc.

Transaction Control Transformation

Transaction control transformation allows us to commit or rollback transactions during the execution of the mapping.

Commit and rollback operations are of significant importance as it guarantees the availability of data.

When processing a high volume of data, there can be a situation when to commit the data to the target.

If a commit is performed too frequently, then it will be an overhead to the system. If a commit is performed too late then in case of failure there are chances of data loss.

So to provide flexibility Transaction control transformation is provided. There are five built variables available in this transformation to handle the operation.

1. **TC_CONTINUE_TRANSACTION**

In tc_continue_transaction there are no operations performed, the process of data load continues as it is

2. **TC_COMMIT_BEFORE**

In tc_commit_before when this flag is found set, a commit is performed before the processing of current row.

3. **TC_COMMIT_AFTER**

In tc_commit_after the current row is processed then a commit is performed.

4. **TC_ROLLBACK_BEFORE**

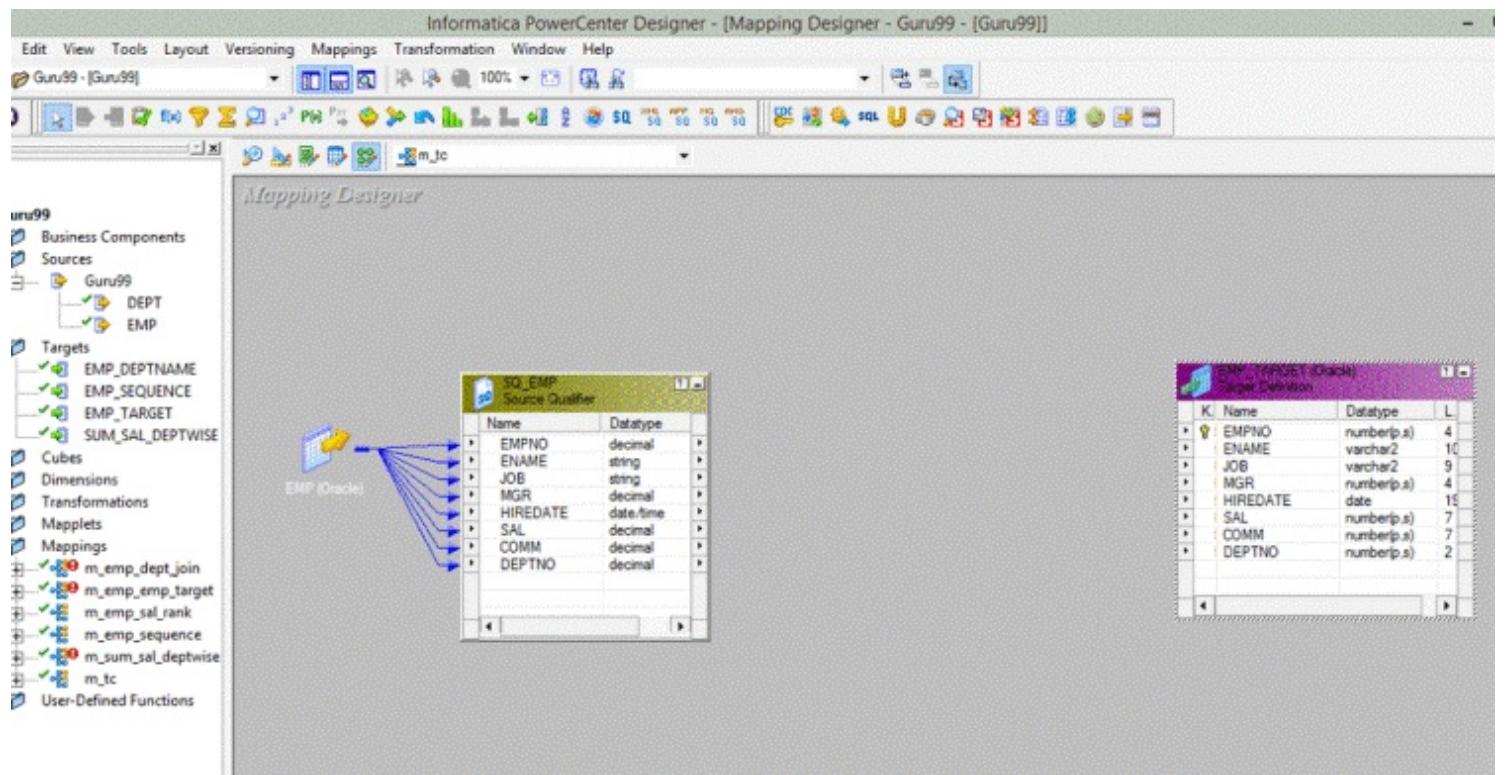
In tc_rollback_before, rollback is performed first then data is processed to write

5. **TC_ROLLBACK_AFTER**

In tc_rollback_after data is processed then the rollback is performed.

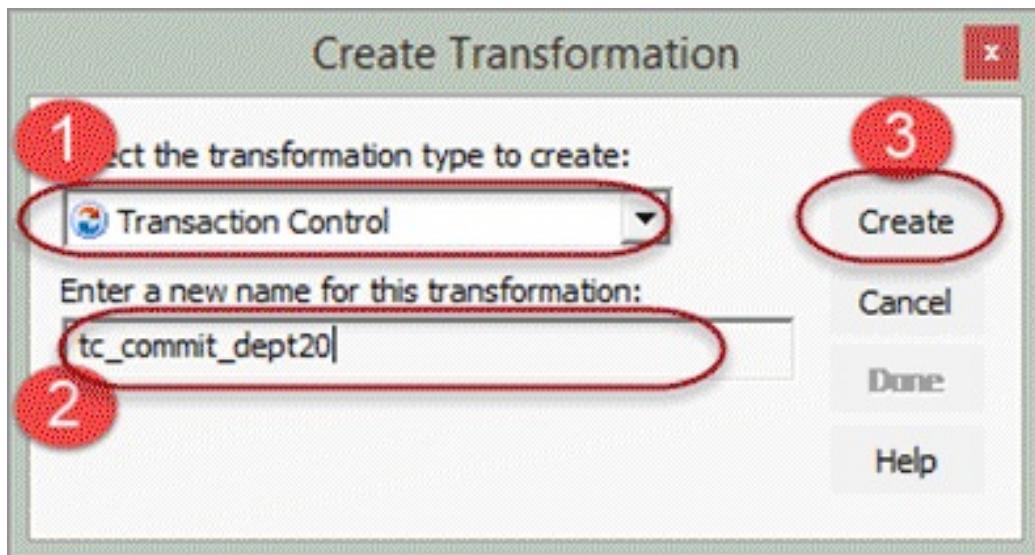
In this example, we will commit data to the target when dept no =20 condition is found true

Step 1 – Create a mapping with EMP as source and EMP_TARGET as target

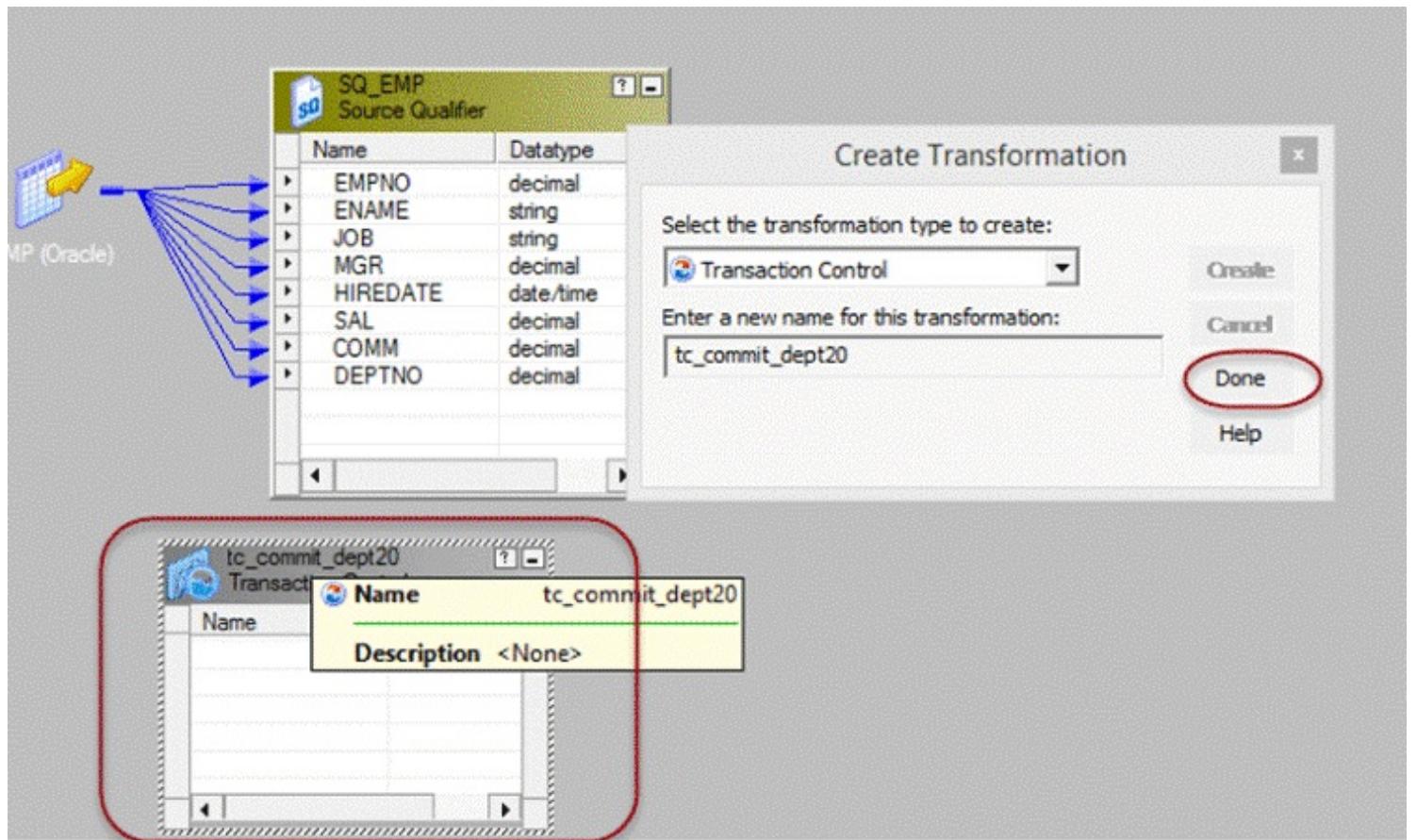


Step 2 – Create a new transformation using transformation menu, then

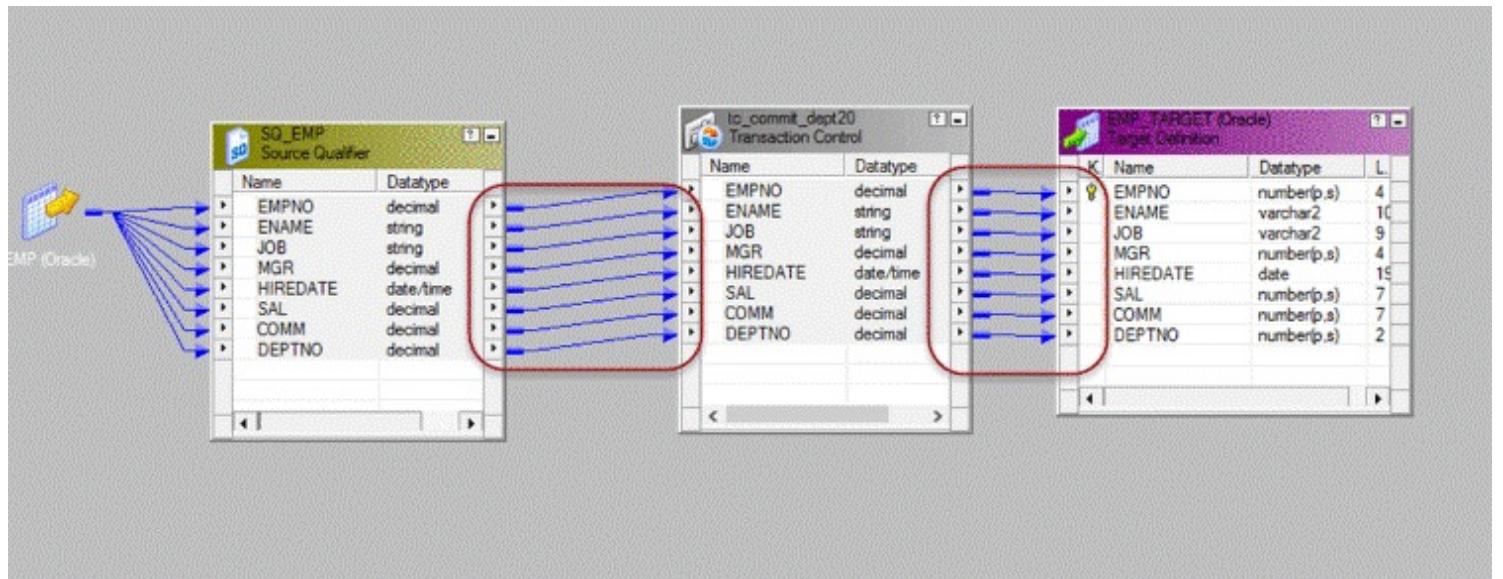
1. Select a transaction control as the new transformation
2. Enter transformation name "tc_commit_dept20"
3. Select create option



Step 3 – The transaction control transformation will be created, select done button

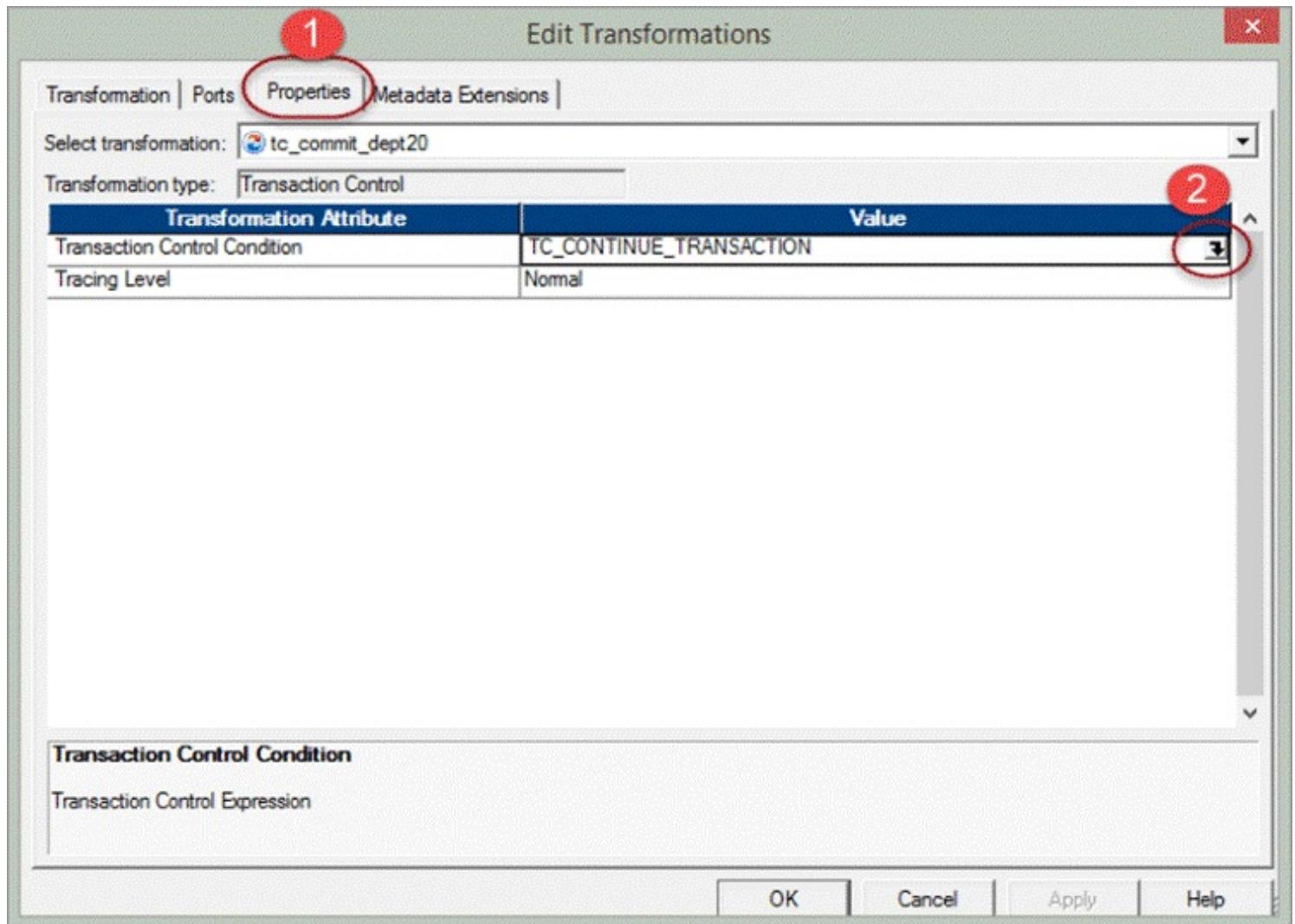


Step 4 - Drag and drop all the columns from source qualifier to the transaction control transformation then link all the columns from transaction control transformation to the target table



Step 5 – Double click on the transaction control transformation and then in the edit property window

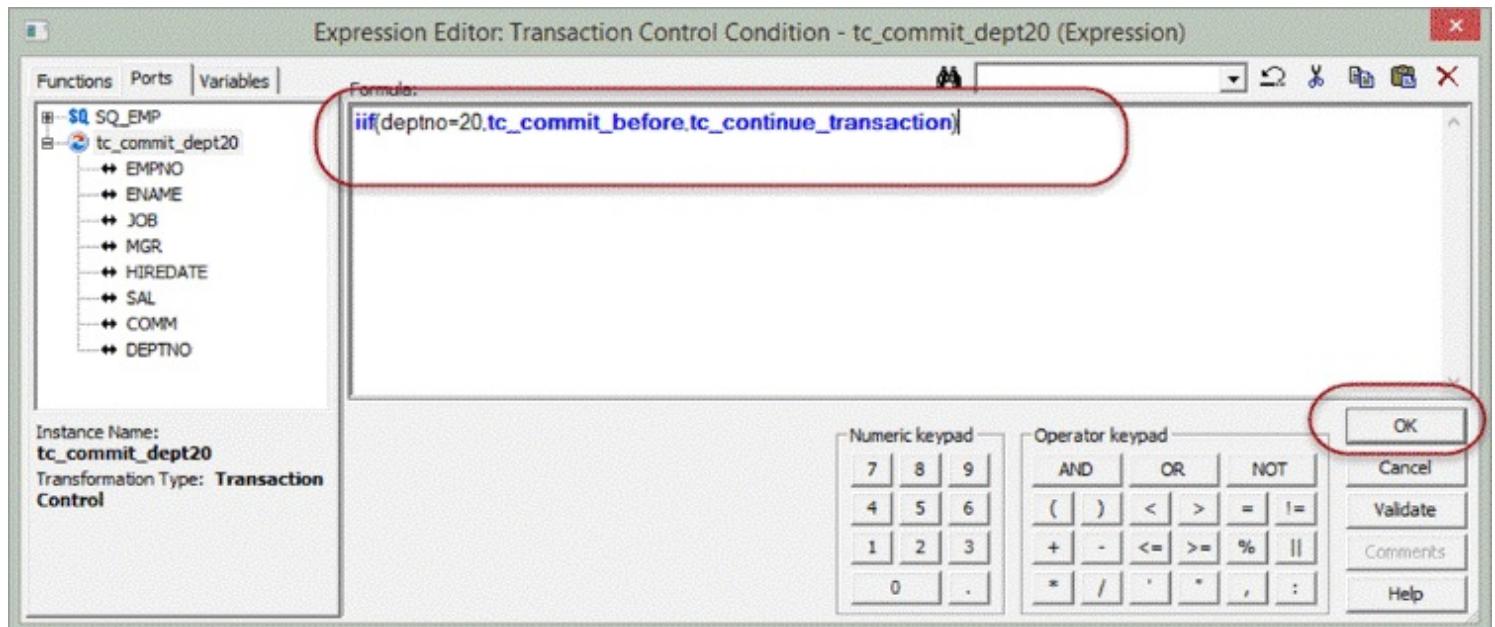
1. Select property tab
2. Click on the transaction control editor icon



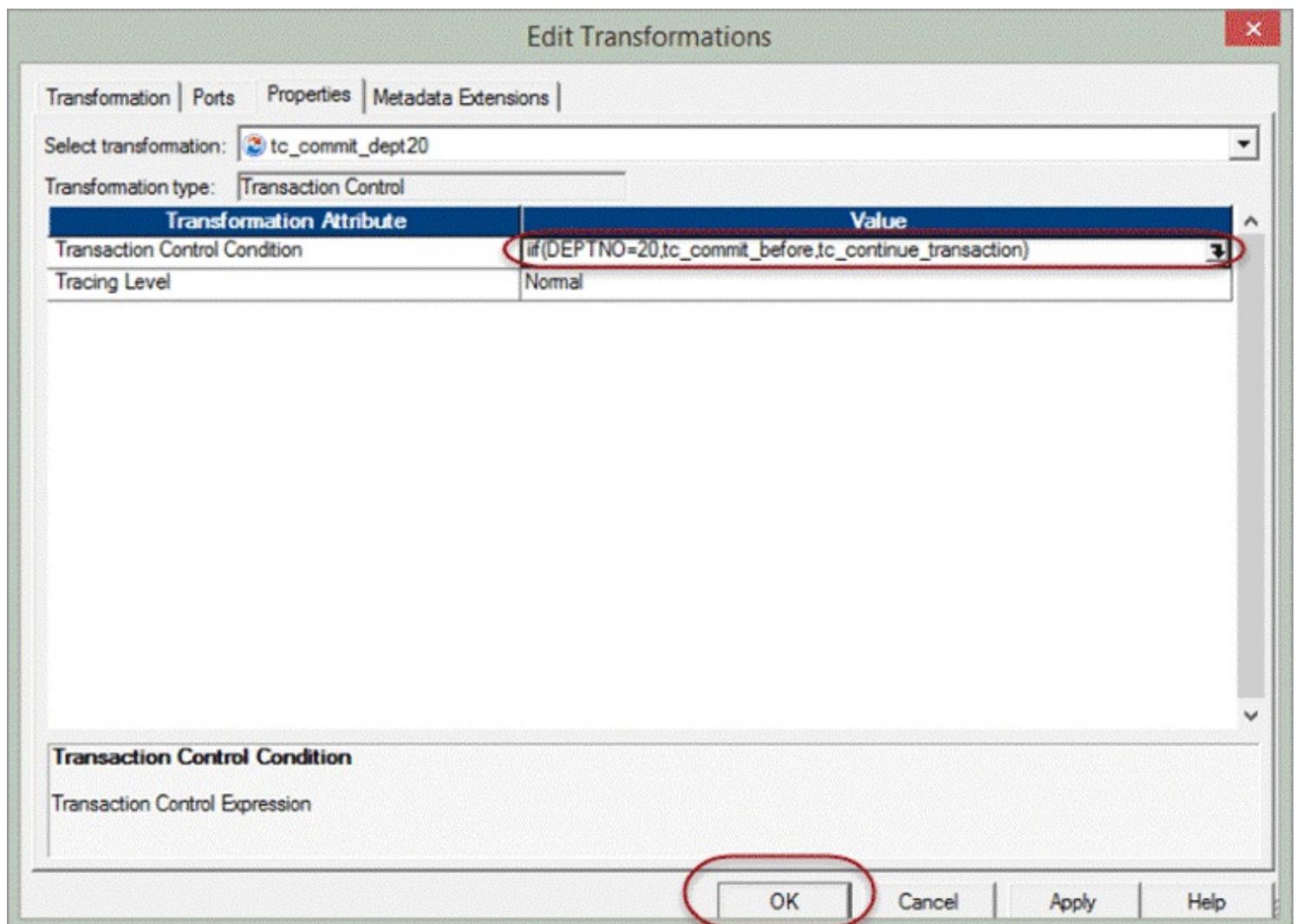
Step 6 –in the expression editor enter the expression –

"iif(deptno=20,tc_commit_before,tc_continue_transaction)" and select OK

It means if deptno 20 is found then commit transaction in target, else continue the current processing.



Step 7 – Select OK in the previous window



Now save the mapping and execute it after creating session and workflows. This mapping will commit the data to the target whenever department number 20 is found

in the data.

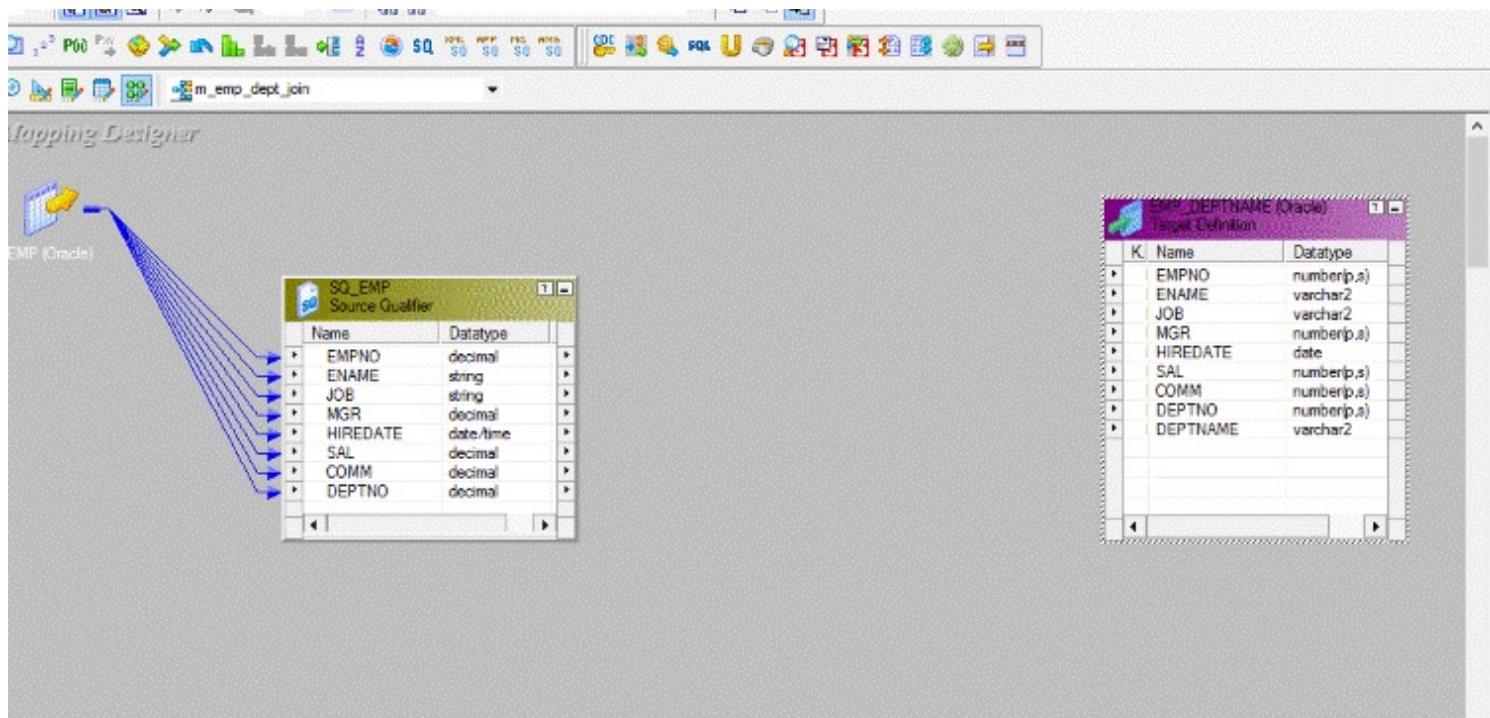
Lookup and Re-usable transformation

Lookup transformation provides the feature to lookup matching values in a table based on the values in source data. Basically, it's a kind of join operation in which one of the joining table is the source data, and the other joining table is the lookup table.

In previous topics, we used joiner transformation to join "emp" and "dept" table to bring department names.

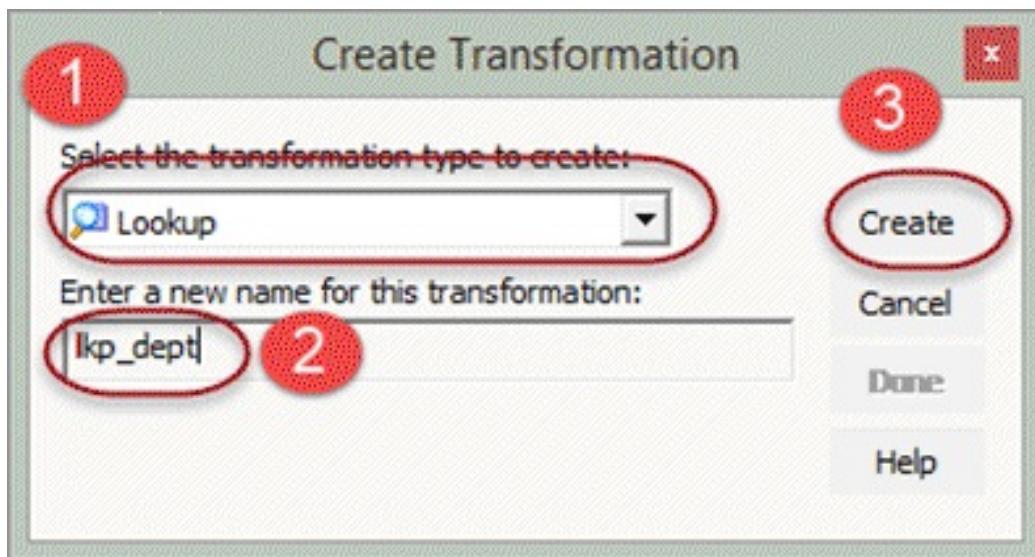
In this section, we will implement the same using lookup transformation.

Step 1 – Create a new mapping with EMP as source and EMP_DEPTNAME as target



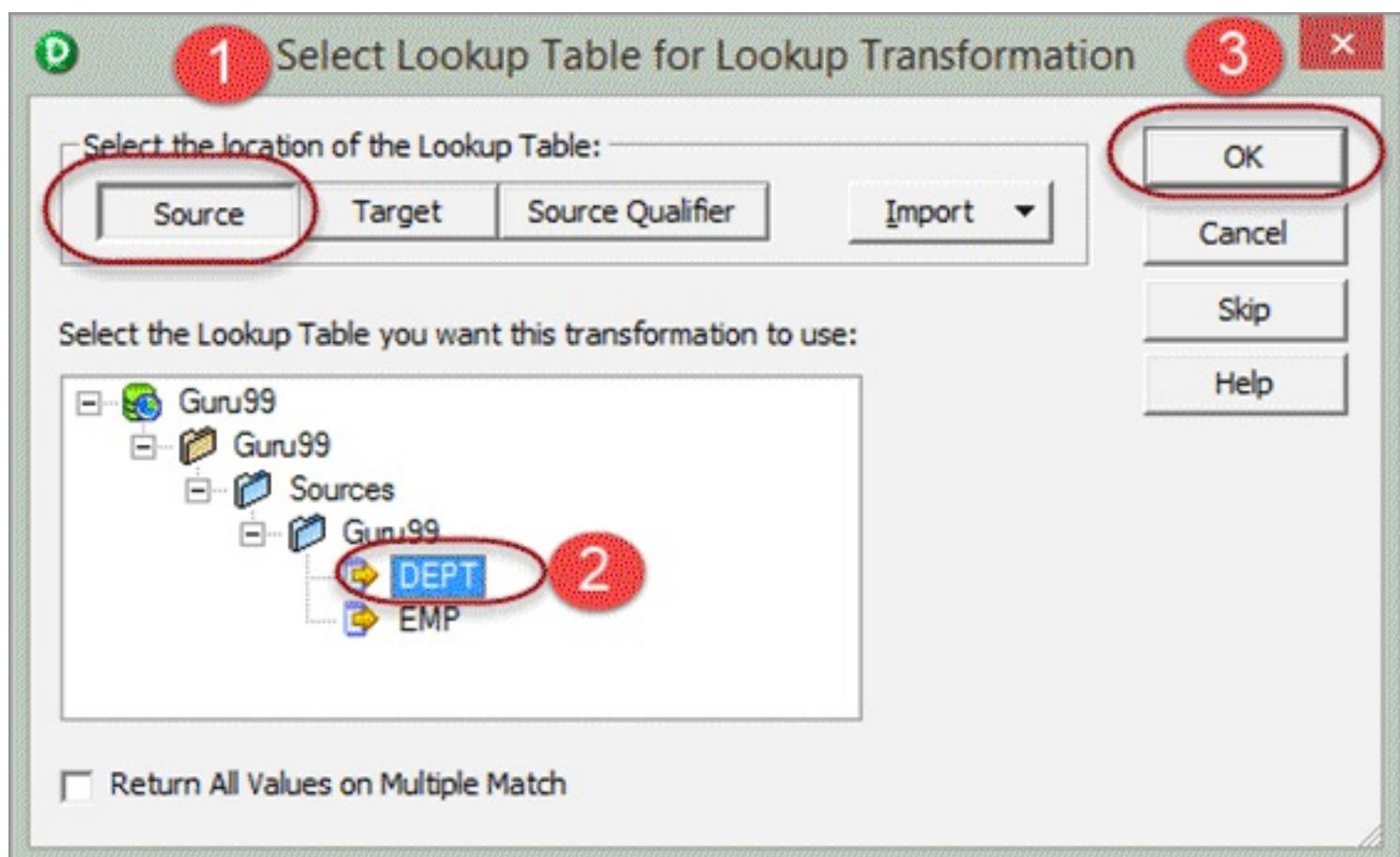
Step 2 – Create a new transformation using transformation menu then

1. Select lookup transformation as the transformation
2. Enter transformation name "lkp_dept"
3. Select create option

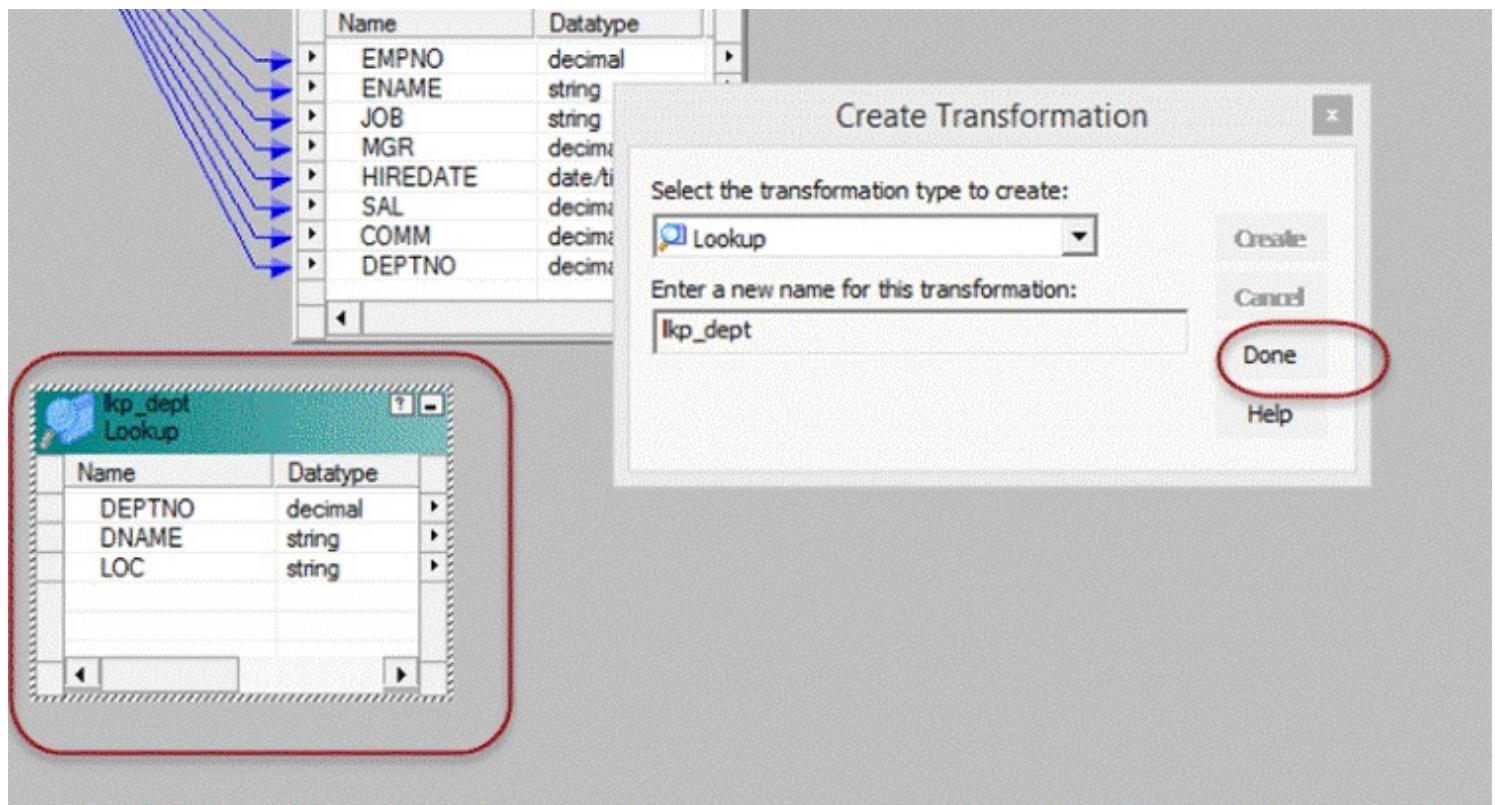


Step 3 – This will open lookup table window, in this window

1. Select source button
2. Select DEPT table
3. Select Ok Button

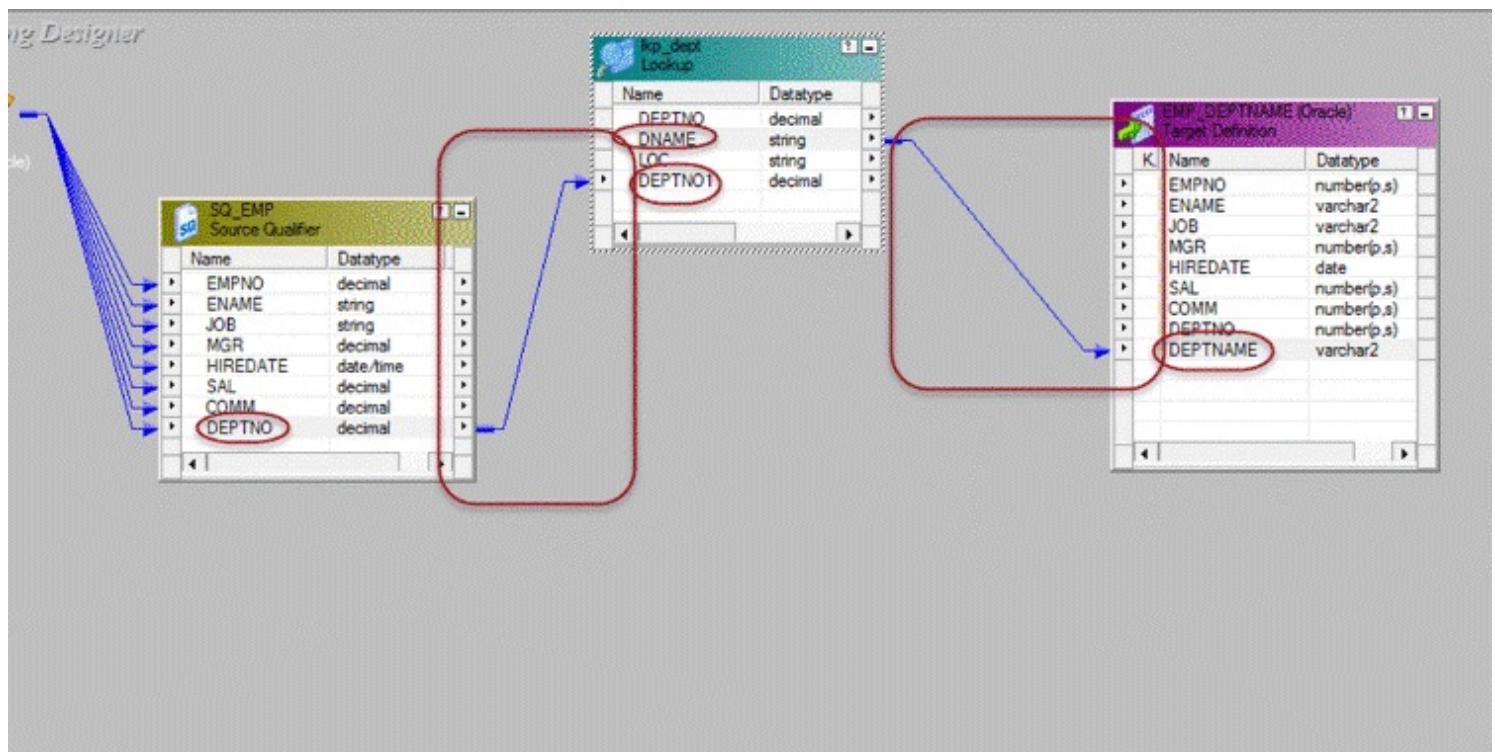


Step 4 - Lookup transformation will be created with the columns of DEPT table, now select done button



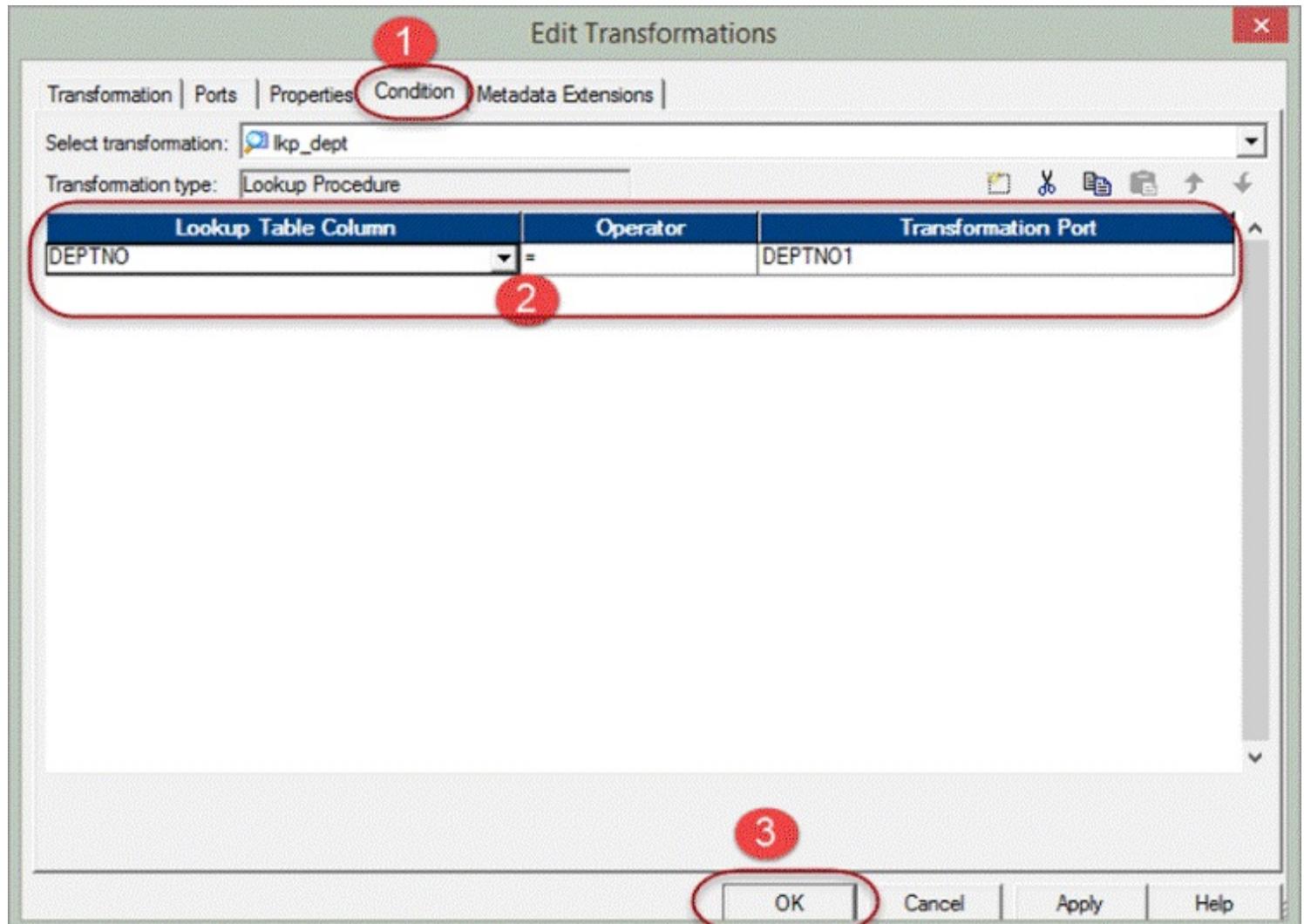
Step 5 - Drag and drop DEPTNO column from source qualifier to the lookup transformation, this will create a new column DEPTNO1 in lookup transformation. Then link the DNAME column from lookup transformation to the target table.

The lookup transformation will lookup and return department name based upon the DEPTNO1 value.

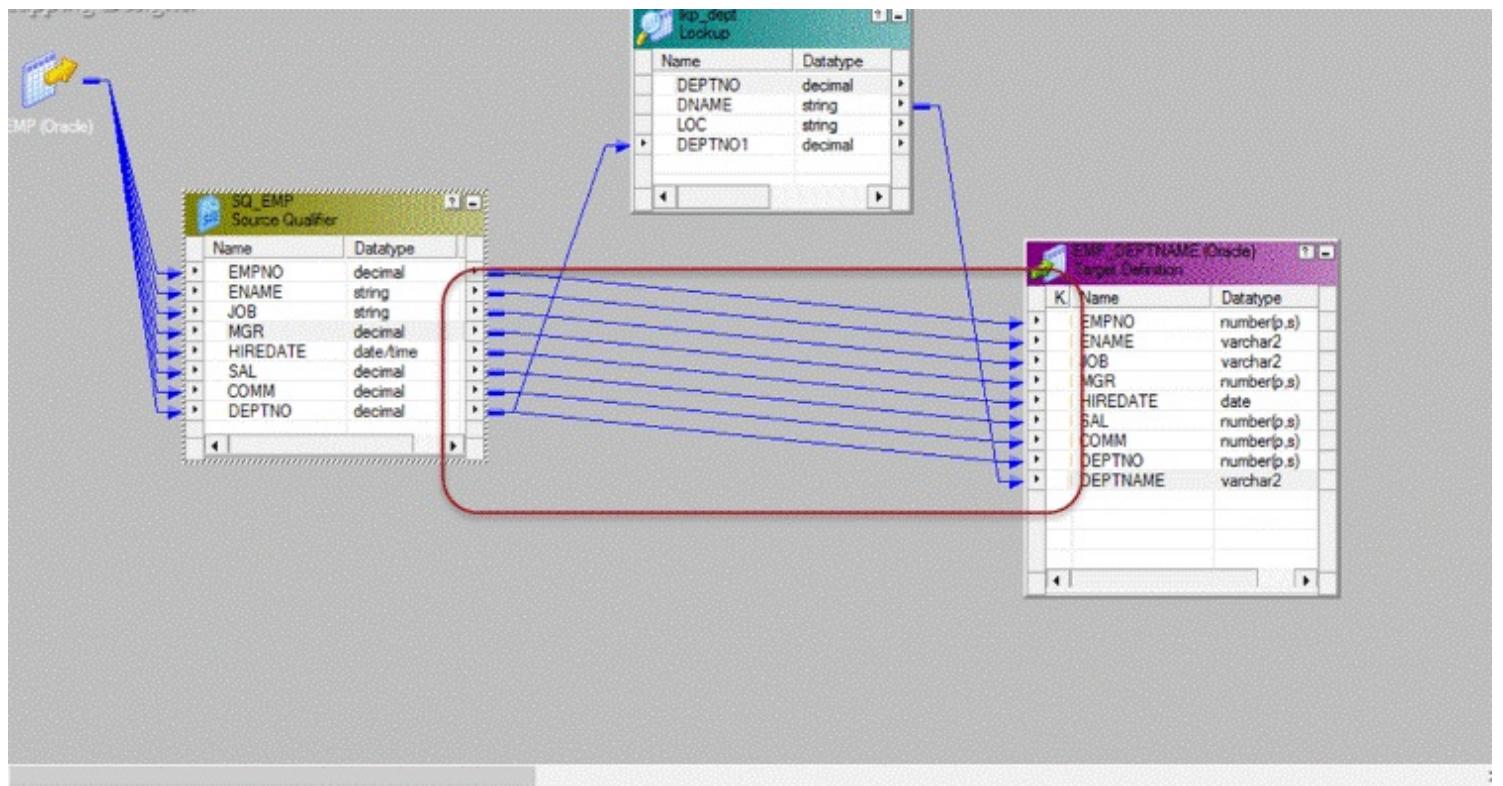


Step 6 – Double click on the lookup transformation. Then in the edit transformation window

1. Select condition tab
2. Set the condition column to DEPTNO = DEPTNO1
3. Select Ok Button



Step 7 – Link rest of the columns from source qualifier to the target table



Now, save the mapping and execute it after creating the session and workflow. This mapping will fetch the department names using lookup transformation.

The lookup transformation is set to lookup on dept table. And the joining condition is set based on dept number.

Reusable Transformation

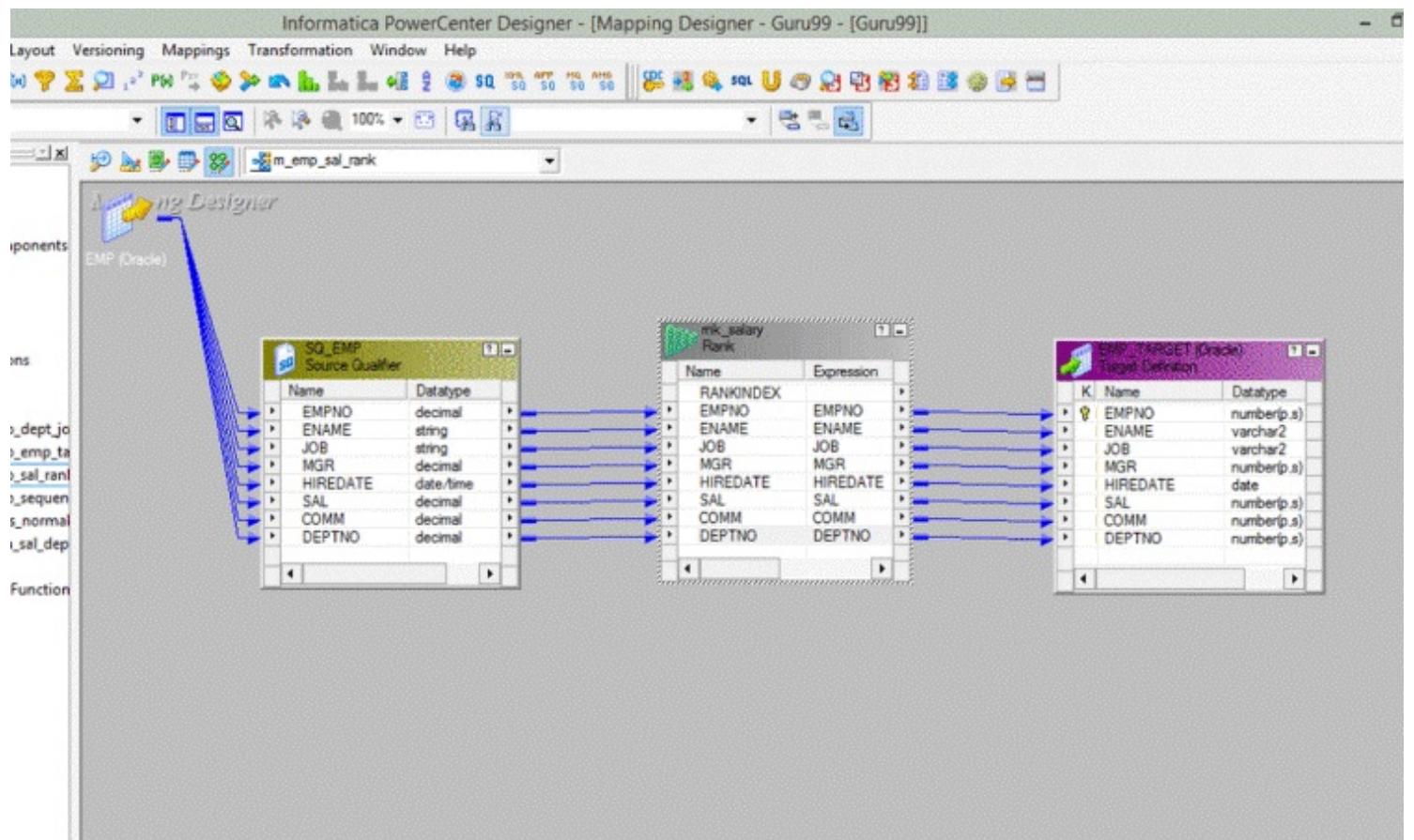
A normal transformation is an object that belongs to a mapping and can be used inside that mapping only. However, by making a transformation reusable it can be re-used inside several mappings.

For example, a lookup transformation which fetches employee details based on employee number can be used at multiple mappings wherever employee details are required.

By using reusable transformation, it reduces the overwork of creating same functionality again.

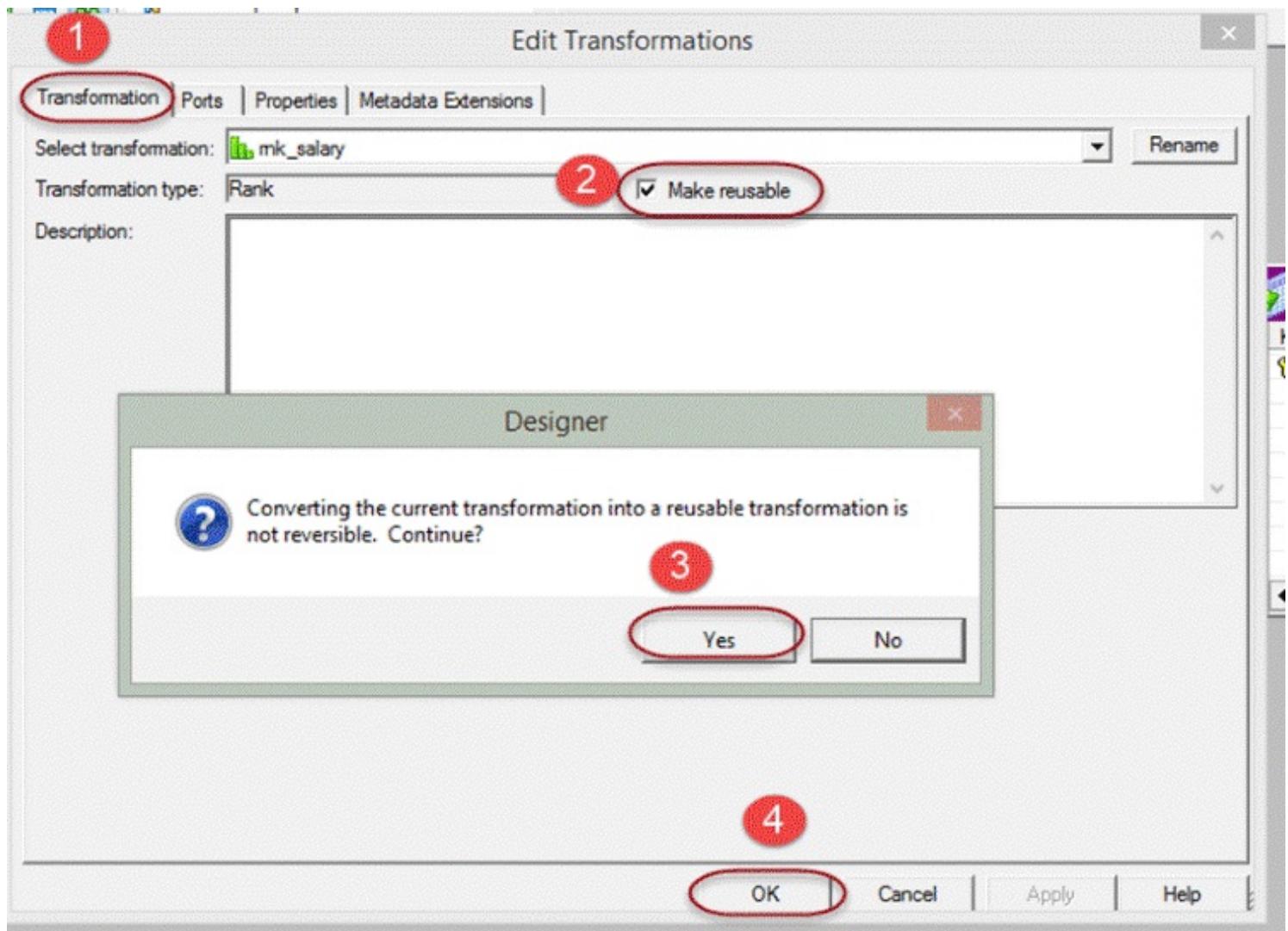
To make a transformation reusable –

Step1 – Open the mapping which is having the transformation, here we are making the rank transformation reusable.



Step 2 – Double click on the transformation to open edit transformation window.
Then

1. Select Transformation tab in the window
2. Select the check box to make transformation reusable
3. Select yes in the confirmation window
4. Select OK in the transformation properties window.



This will make the transformation reusable.

Normalizer Transformation

Normalizer transformation is a smart way of representing your data in more organized manner. It is used to convert a single row into multiple rows and vice versa. If in a single row there is repeating data in multiple columns, then it can be split into multiple rows. Sometimes we have data in multiple occurring columns. For example

Student Name	Class 9 Score	Class 10 Score	Class 11 Score	Class 12 Score
Student 1	50	60	65	80
Student 2	70	64	83	77

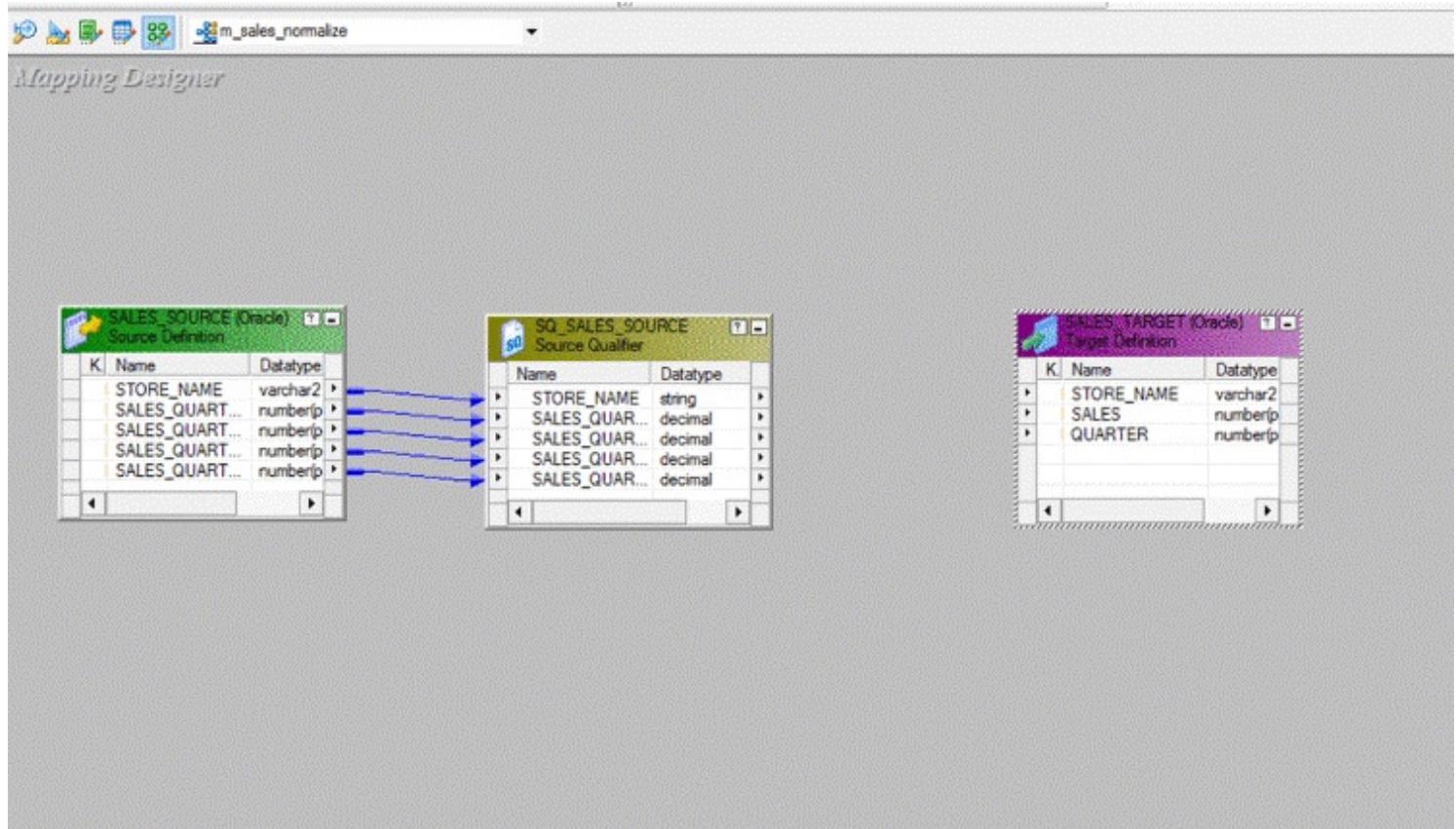
In this case, the class score column is repeating in four columns. Using normalizer, we can split these in the following data set.

Student Name	Class	Score

Student 1	9	50
Student 1	10	60
Student 1	11	65
Student 1	12	80
Student 2	9	70
Student 2	10	64
Student 2	11	83
Student 2	12	77

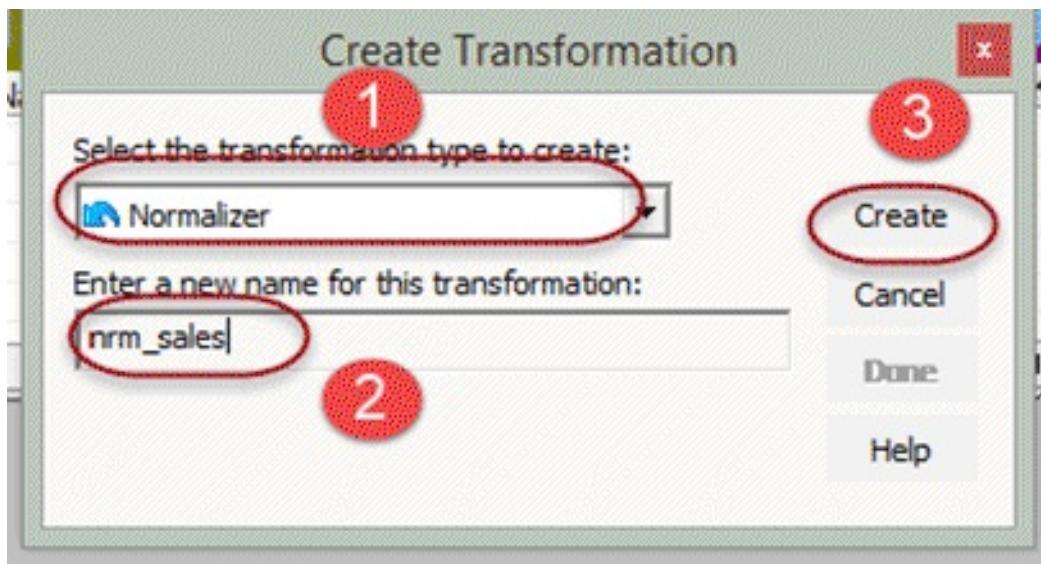
Step 1 – Create source table "sales_source" and target table "sales_target" using the script and import them in Informatica

Step 2 – Create a mapping having source "sales_source" and target table "sales_target"

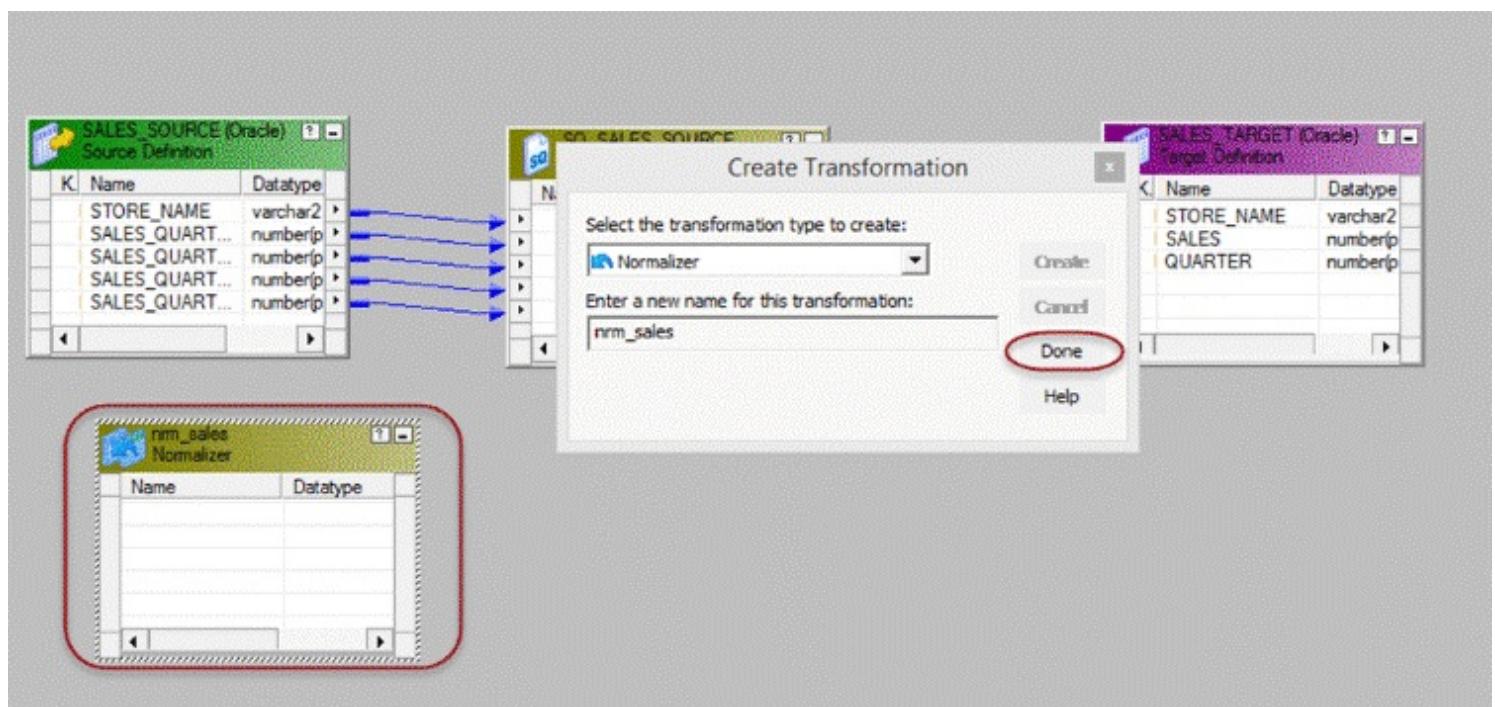


Step 3 – From the transformation menu create a new transformation

1. Select normalizer as transformation
2. Enter name, "nrm_sales"
3. Select create option



Step 4 – The transformation will be created, select done option



Step 5 – Double click on the normalizer transformation, then

1. Select normalizer tab
2. Click on icon to create two columns
3. Enter column names
4. Set number of occurrence to 4 for sales and 0 for store name
5. Select OK button

Edit Transformations

1

Transformation | Ports | Properties | **Normalizer** | Metadata Extensions |

Select transformation: **nm_sales**

2

	Column Name	Level	Occurs	Datatype	Prec	Scale
1	sales	0	4	number	10	0
2	store_name	0	0	string	40	0

3

4

2

Description:

5

OK

Cancel

Apply

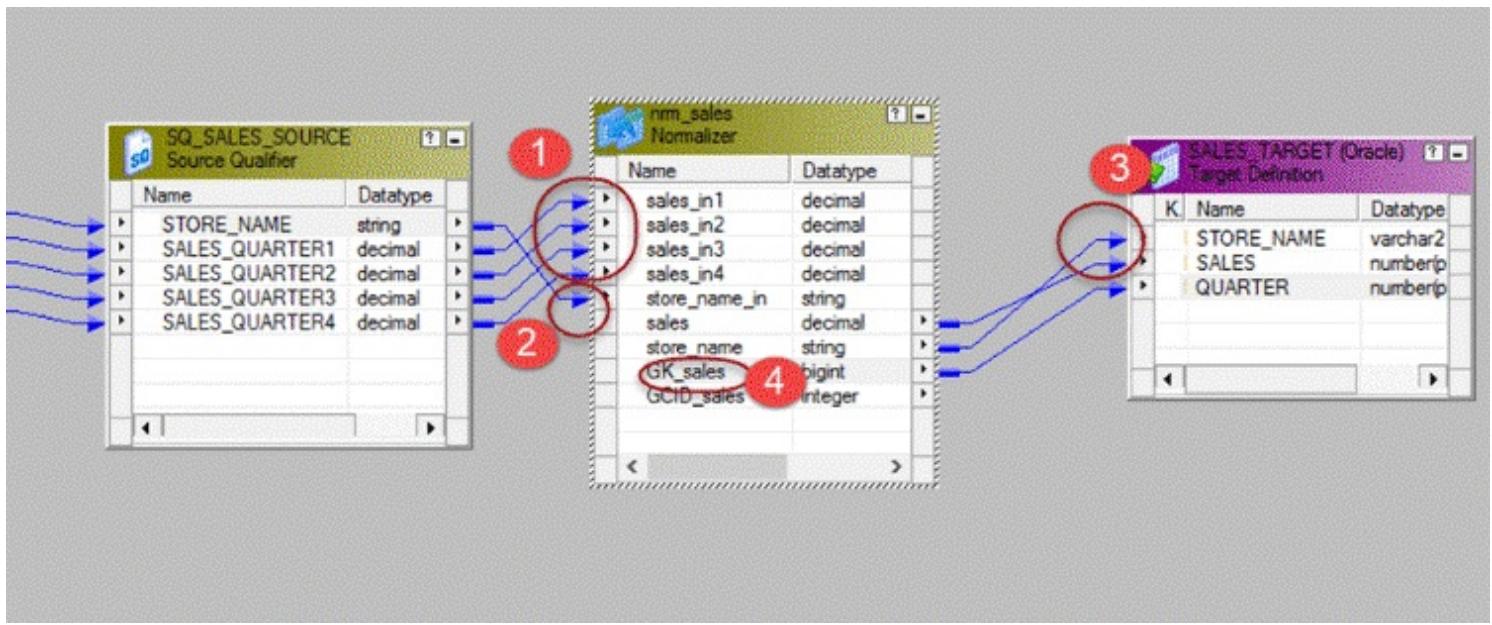
Help

Columns will be generated in the transformation. You will see 4 number of sales column as we set the number of occurrences to 4.

Name	Datatype
sales_in1	decimal
sales_in2	decimal
sales_in3	decimal
sales_in4	decimal
store_name_in	string
sales	decimal
store_name	string
GK_sales	bigint
GCID_sales	integer

Step 6 – Then in the mapping

1. Link the four column of source qualifier of the four quarter to the normalizer columns respectively.
2. Link store name column to the normalizer column
3. Link store_name & sales columns from normalizer to target table
4. Link GK_sales column from normalizer to target table



Save the mapping and execute it after creating session and workflow. For each quarter sales of a store, a separate row will be created by the normalizer transformation.

The output of our mapping will be like –

Store Name	Quarter	Sales
DELHI	1	150
DELHI	2	240
DELHI	3	455
DELHI	4	100
MUMBAI	1	100
MUMBAI	2	500
MUMBAI	3	350
MUMBAI	4	340

The source data had repeating columns namely QUARTER1, QUARTER2, QUARTER3, and QUARTER4. With the help of normalizer, we have rearranged the data to fit into a single column of QUARTER and for one source record four records are created in the target.

In this way, you can normalize data and create multiple records for a single source of data.

Performance Tuning for Transformation

Joiner Transformation -

- Always prefer to perform joins in the database if possible, as database joins are faster than joins created in Informatica joiner transformation.
- Sort the data before joining if possible, as it decreases the disk I/O performed during joining.
- Make the table with less no of rows as master table.

Lookup Transformation –

- Create an index for the column in a lookup table which is used in lookup condition. Since the lookup table will be queried for looking up the matching data, adding an index would increase the performance.
- If possible, instead of using lookup transformation use join in the database. As database joins are faster, performance will be increased.
- Delete unnecessary columns from the lookup table and keep only the required columns. This will bring down the overhead of fetching the extra columns from the database.

Filter Transformation –

- Use filter transformation as early as possible inside the mapping. If the unwanted data can be discarded early in the mapping, it would increase the throughput.'
- Use source qualifier to filter the data. You can also use source qualifier SQL override to filter the records, instead of using filter transformation.

Aggregator Transformation

- Filter the data before aggregating it. If you are using filter transformation in the mapping, then filter the data before using aggregator as it will reduce the unnecessary aggregation operation.
- Limit the no of ports used in the aggregator transformation. This will reduce the volume of data that aggregator transformation stores inside the cache.

Source Qualifier Transformation

- Bring only the required columns from the source. Most of the times not all the columns of the source table are required, so bring only the required fields by deleting the unnecessary columns.

- Avoid using order by clause inside the source qualifier SQL override. The order by clause requires additional processing and performance can be increased by avoiding it.