Learn Informatica in 1 Day

By Krishna Rungta

Copyright 2019 - 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 to Informatica

1. What is Informatica?
2. Why do we need Informatica?

Chapter 2: Informatica Architecture Tutorial

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: How to Download & Install Informatica PowerCenter

Chapter 4: How to Configure Client and Repository in Informatica

1. What is Informatica domain?
2. How to Open the Administrator Home Page
3. How to Create RepositoryService
4. How to create repositorycontents
5. How to create IntegrationService
6. How to Configuring client and domain
7. How to create user

Chapter 5: Source Analyzer & Target Designer in Informatica

1. How to open Source Analyzer
2. How to import Source table in Source Analyzer
3. How to import target in target designer
4. How to create folder in Informatica

Chapter 6: Mappings in Informatica: Create, Components, Parameter, Variable

1. What is a Mapping?
2. Why do you need Mapping?
3. Components of Mapping
4. What is Stage Mapping?
5. How to Create a Mapping
6. Mapping Parameters and Variables
7. How to Create Mapping Parameter
8. How to Create Mapping Variable

Chapter 7: Workflow in Informatica: Create, Task, Parameter, Reusable, Manager

1. What is Workflow?
2. How to open Workflow Manager
3. How to Create Connections for Workflow Manager
4. To Create a Relational Connection
5. Components of Workflow manager
6. How to create command task
7. How to create workflow to execute command task
8. How to execute workflow
9. Session Task

Chapter 8: Workflow Monitor in Informatica: Task & Gantt Chart View Examples

1. How to open Workflow Monitor
2. Task View
3. Gantt Chart View
4. Example- How to monitor and view details

Chapter 9: Debugger in Informatica: Session, Breakpoint, Verbose Data & Mapping

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

Chapter 10: Session Properties in Informatica

1. Properties Of Session
2. Treat Source Rows As Property
3. How To Make Treat Source Rows – Delete
4. Commit Interval - Property
5. Session Log File Name & Session Log File directory
6. Enable Test Load
7. Memory Properties
8. Log options
9. Error Handling
10. Mapping and Source/Target Properties

Chapter 11: Introduction to Transformations in Informatica and Filter Transformation

1. Classification of Transformation
2. Following is the List of Transformations in Informatica
3. What is Filter Transformation?

Chapter 12: Source Qualifier Transformation in Informatica with EXAMPLE

1. What is Source Qualifier Transformation?
2. Properties of Source Qualifier

Chapter 13: Aggregator Transformation in Informatica with Example

Chapter 14: Router Transformation in Informatica with EXAMPLE

Chapter 15: Joiner Transformation in Informatica with EXAMPLE

Chapter 16: Rank Transformation in Informatica with EXAMPLE

Chapter 17: Sequence Transformation in Informatica with EXAMPLE

Chapter 18: Transaction Control Transformation in Informatica with EXAMPLE

Chapter 19: Lookup Transformation in Informatica & Re- usable Transformation Example

1. What is Lookup Transformation?
2. Reusable Transformation

Chapter 20: Normalizer Transformation in Informatica with EXAMPLE

Chapter 21: Performance Tuning in Informatica

Chapter 1: Introduction to Informatica

What is Informatica?

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?

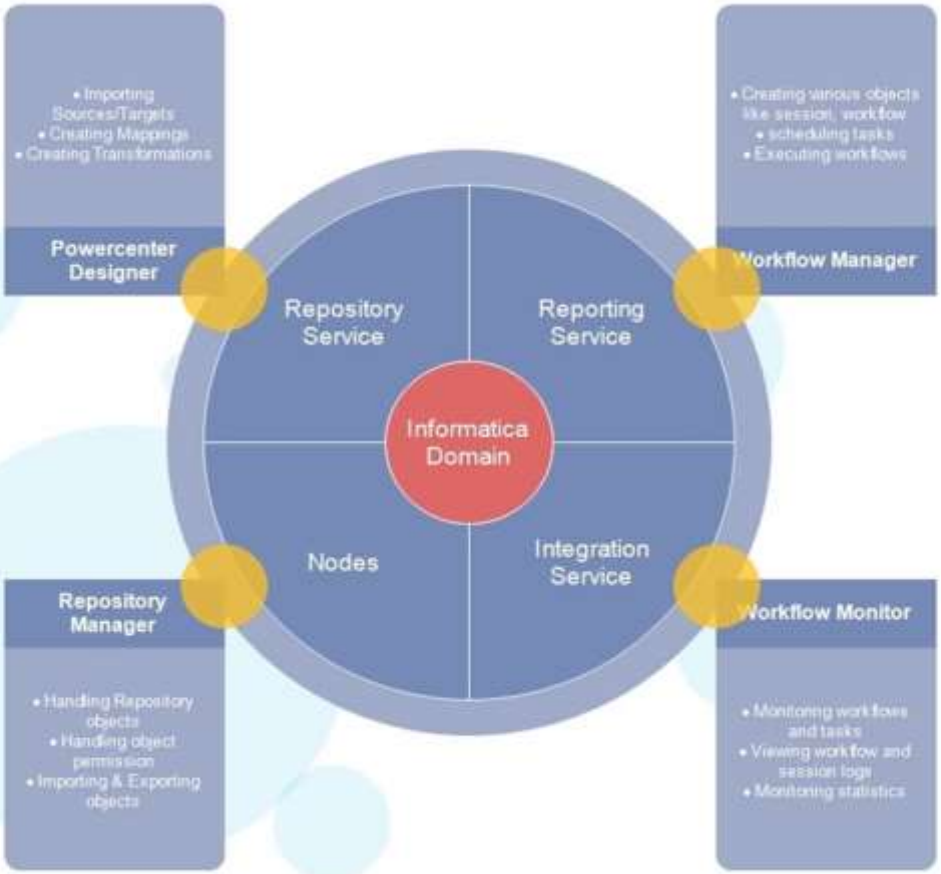
Informatica comes to the picture wherever we have a data system available and at the backend we want to perform certain operations on the data. It can be like cleaning up of data, modifying the data, etc. based on certain set of rules or simply loading of bulk data from one system to another.

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: Informatica Architecture

Informatica ETL 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

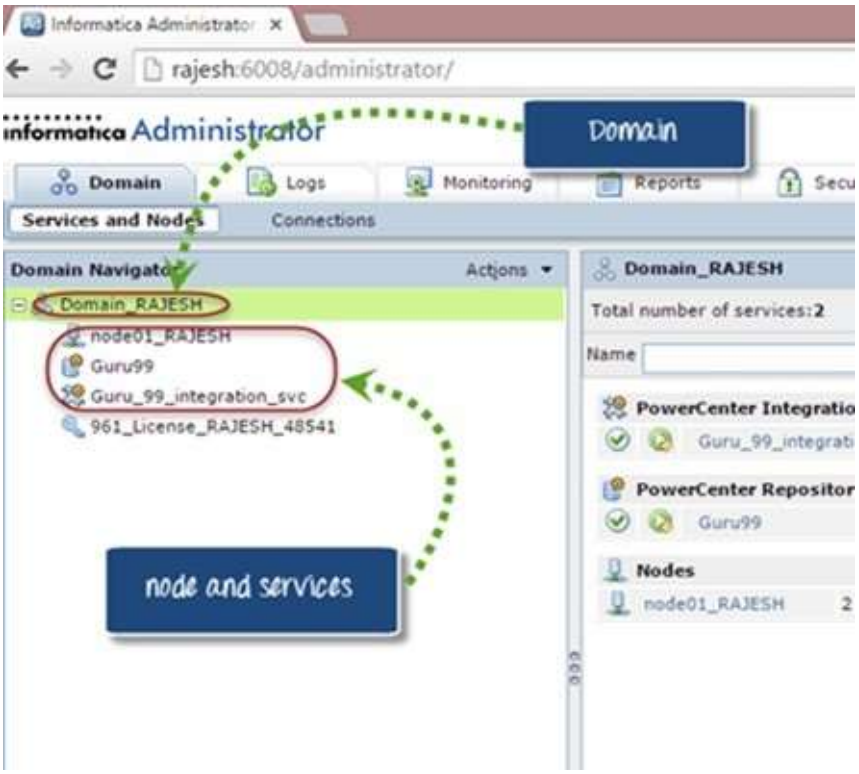


Informatica Domain

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

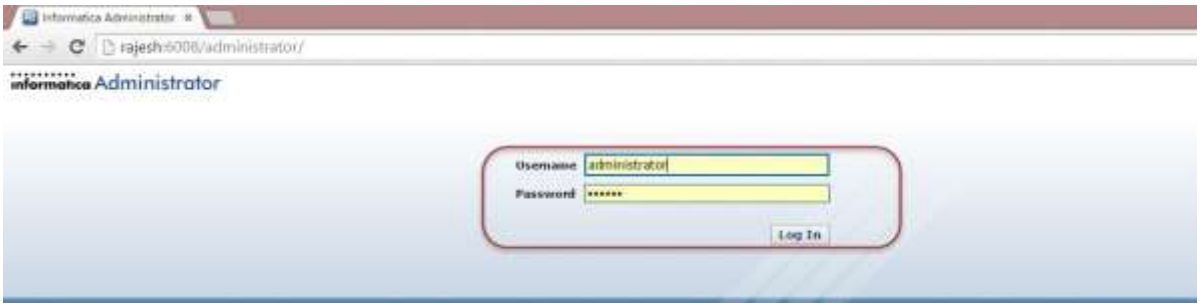
- **Valid**: Valid objects are those objects whose syntax is correct according to Informatica. These objects can be used in the execution of workflows.
- **Invalid**: Invalid objects are those who does not adhere to the standard or rules specified. When any object is saved in Informatica, it is checked 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.

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.

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.

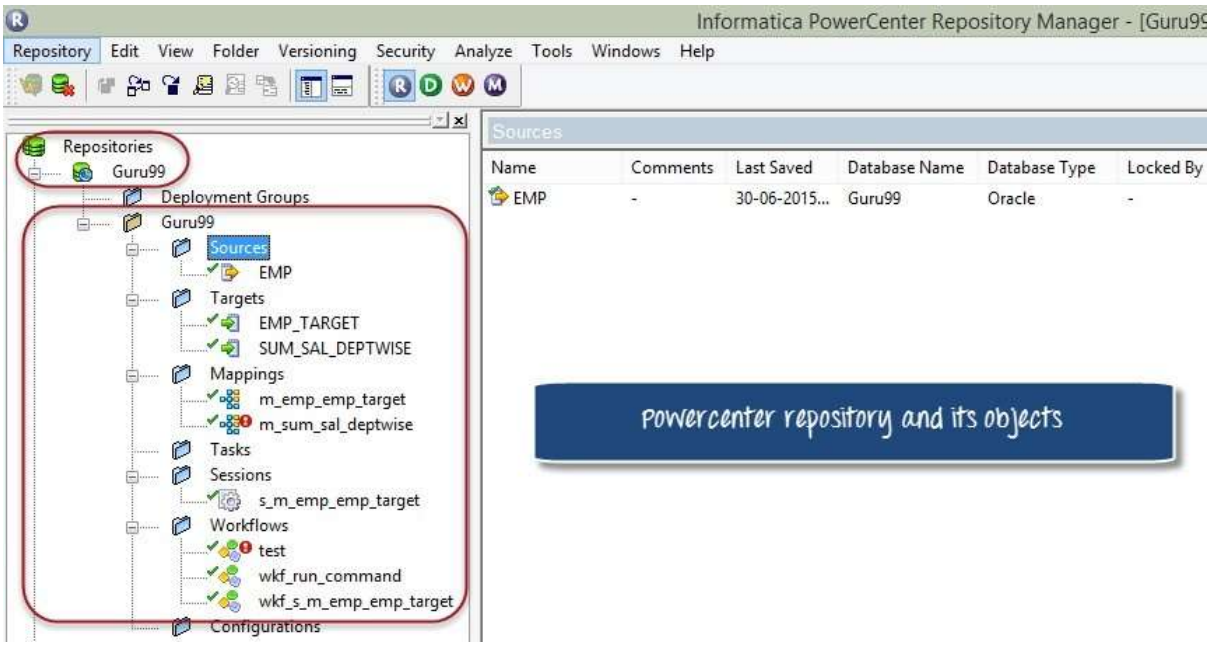
The mappings and objects that we create in these client tools are saved in the Informatica repository which resides on the Informatica server. So the client tools must have network connectivity to the server.

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.

Buy Now $9.99