**Building ETL Processes (Vendor)**

**Software Requirements Document**

* 1. Introduction

# Informatica

# Informatica is introduced as a software development company in the market. It provides a complete data integration solution and data management system. It launched multiple products that mainly focused on data integration.

# Informatica is used to extracting required data form operation all systems and transforms the same data on its server and load it to the data warehouse.

# Informatica is also introduced as a data integration tool. This tool is based on the ETL architecture. It provides data integration software and services for different industries, businesses, government organizations, as well as telecommunication, health care, insurance, and financial services.

# It has a unique property to connect, process, and fetch the data from a different type of mixed sources.

# **For example**, we can connect with more than one server database.

**1.2 Need of Informatica**

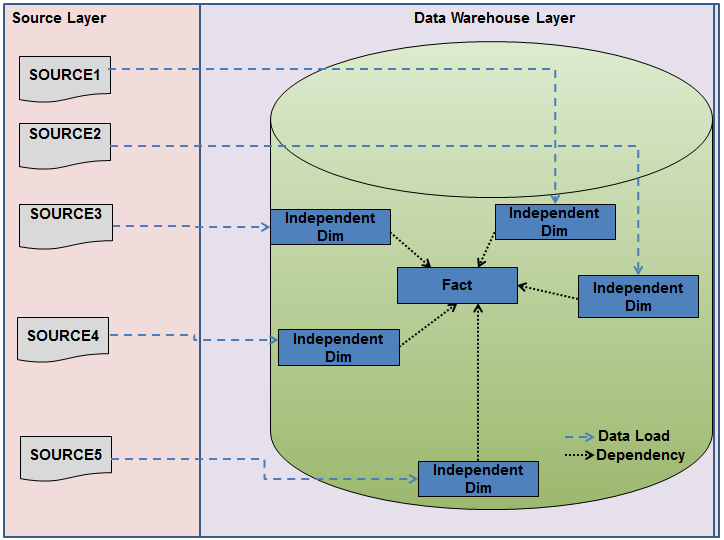
* To perform some operations on the data at the backend in a data system, then we need the Informatica.
* To modify, cleaning up the data based on some set of rules, we need the Informatica.
* By using the Informatica, it is accessible to the loading of bulk data from one system to another.
* It provides a broad set of features such as integration of the data from multiple unstructured, semi-structured or structured systems, operations at row level on data, and scheduling operation of the data operation.
* It also supports the features of metadata, so it keeps preserved the information of the process and data operations.

1. **Architecture diagram**

**2.1 Physical Architecture**

**Physical Architecture:**

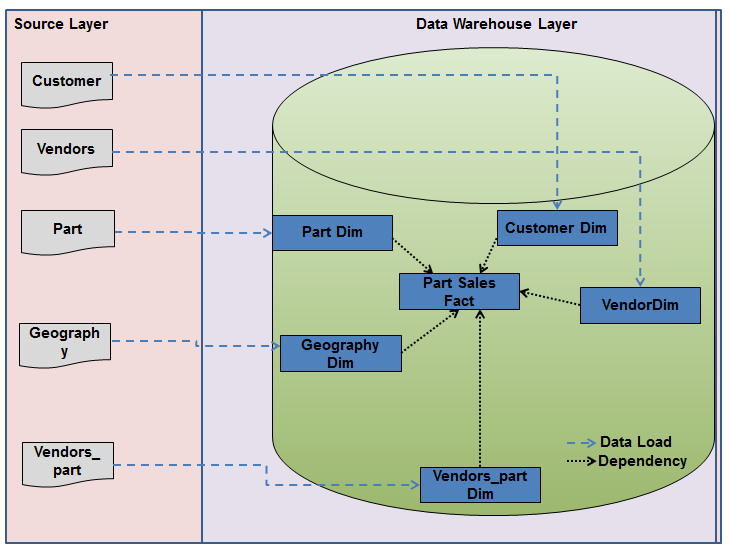
A physical architecture is an arrangement of physical elements, (system elements and physical interfaces) that provides the designed solution for a product, service, or enterprise. It is intended to satisfy logical architecture elements and system requirements. Auto Identification Process follows a three layered architecture namely presentation layer, business logic layer and data access layer.



**2.2 Logical Architecture**

**Logical Architecture:**

The Logical Architecture defines the Processes (the activities and functions) that are required to provide the required User Services. Many different Processes must work together and share information to provide a User Service. The Processes can be implemented via software, hardware, or firmware. The Logical Architecture is independent of technologies and implementations.



**3 Table Definitions & Mappings.**

|  |  |  |
| --- | --- | --- |
| **Table Name** | **Table Type** | **Table Description** |
| Geography | Type 1 Dimension | **Geography** |
| Vendors | Type 2 Dimension | **Vendors** |
| Parts | Reference table | **Parts** |
| Vendor\_parts | Type 2 Dimension | **Vendor\_Parts** |
| Customer | Type 1 Dimension | **Customers** |
| Part\_sales\_fact | fact | **Part Sales Fact** |

**4 Source Files:**

CDW\_SAPP\_D\_SALES



CDW\_SAPP\_D\_VENDOR\_PART



CDW\_SAPP\_D\_VENDOR



CDW\_SAPP\_D\_PARTS



CDW\_SAPP\_D\_GEOGRAPHY



CDW\_SAPP\_D\_CUSTOMER



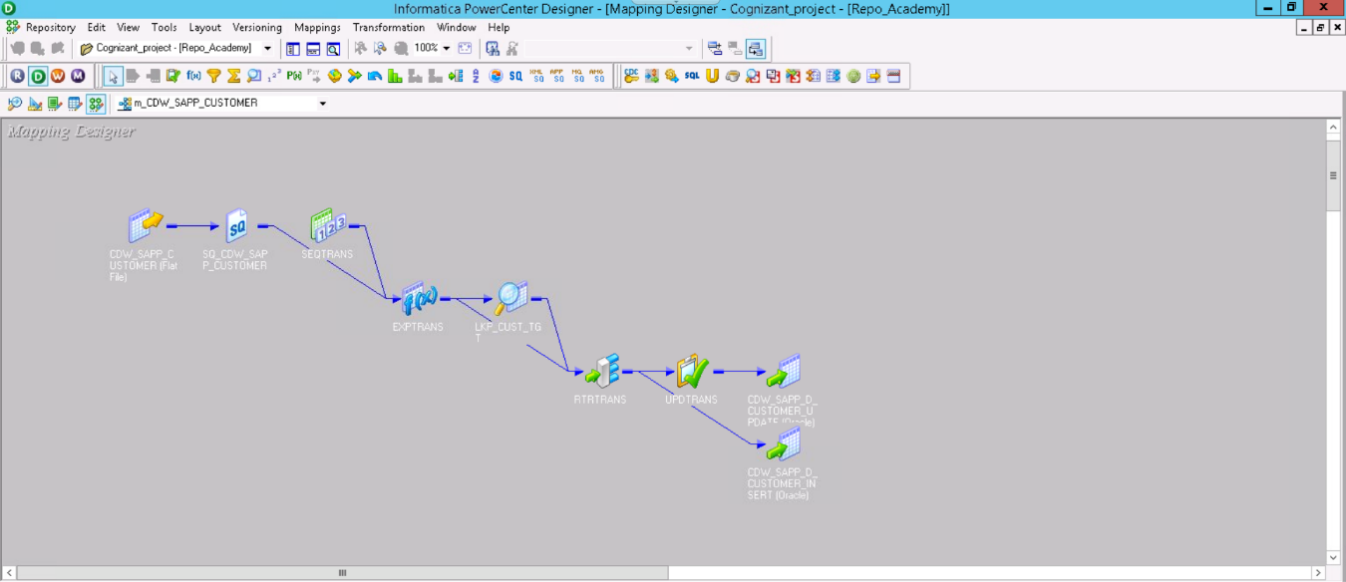
**5 Mapping Document :**



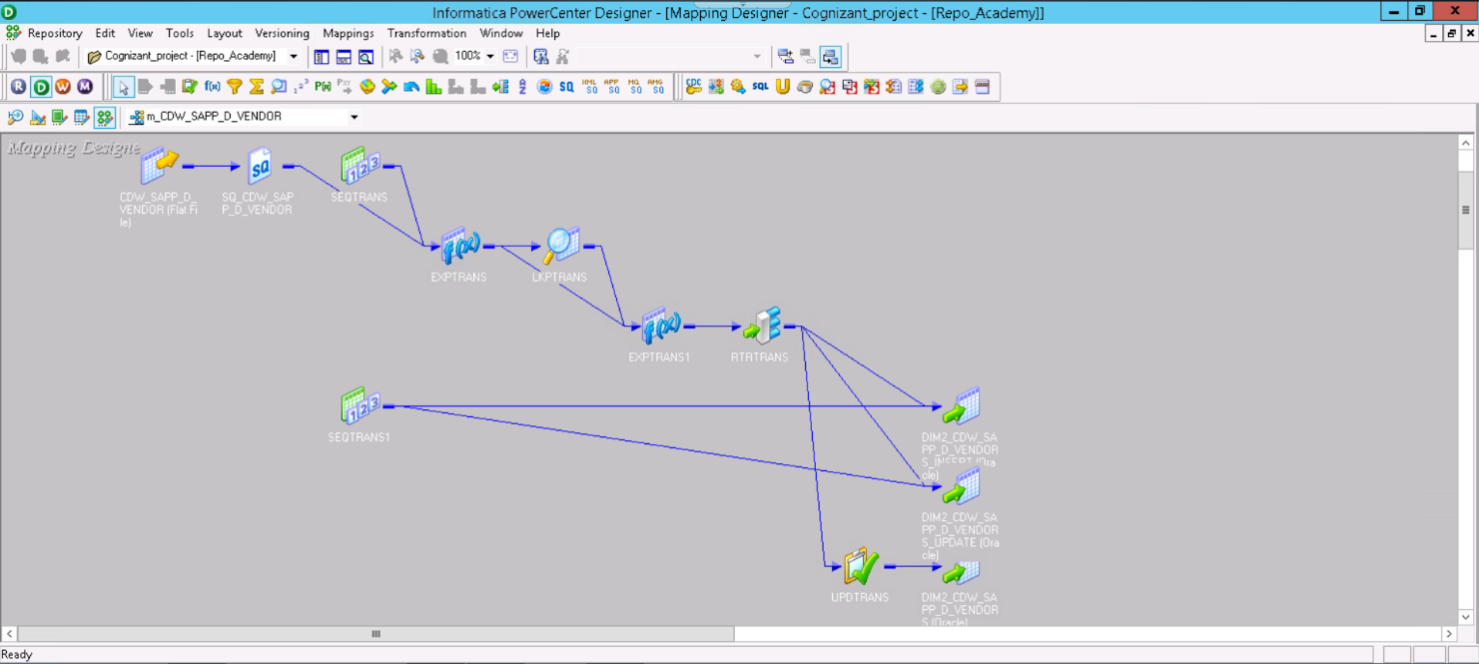
**6 Mappings :**

### 

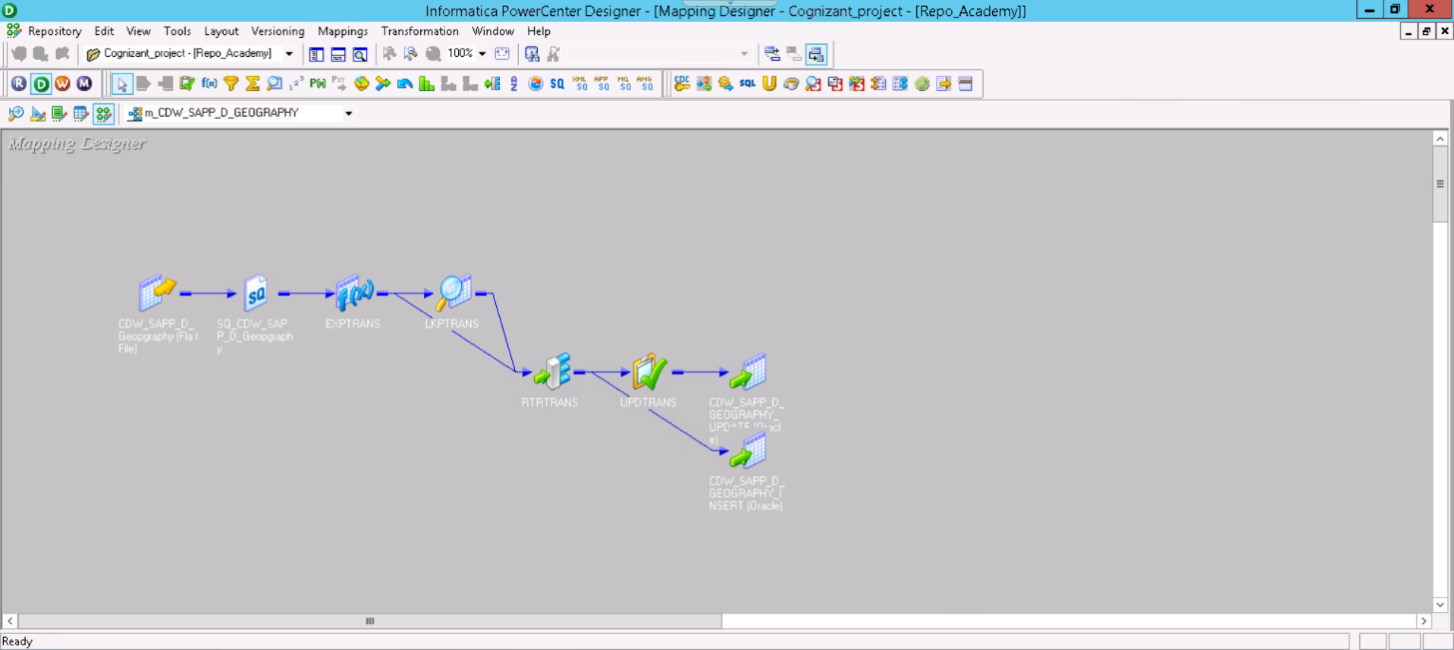
1. CDW\_SAPP\_CUSTOMER



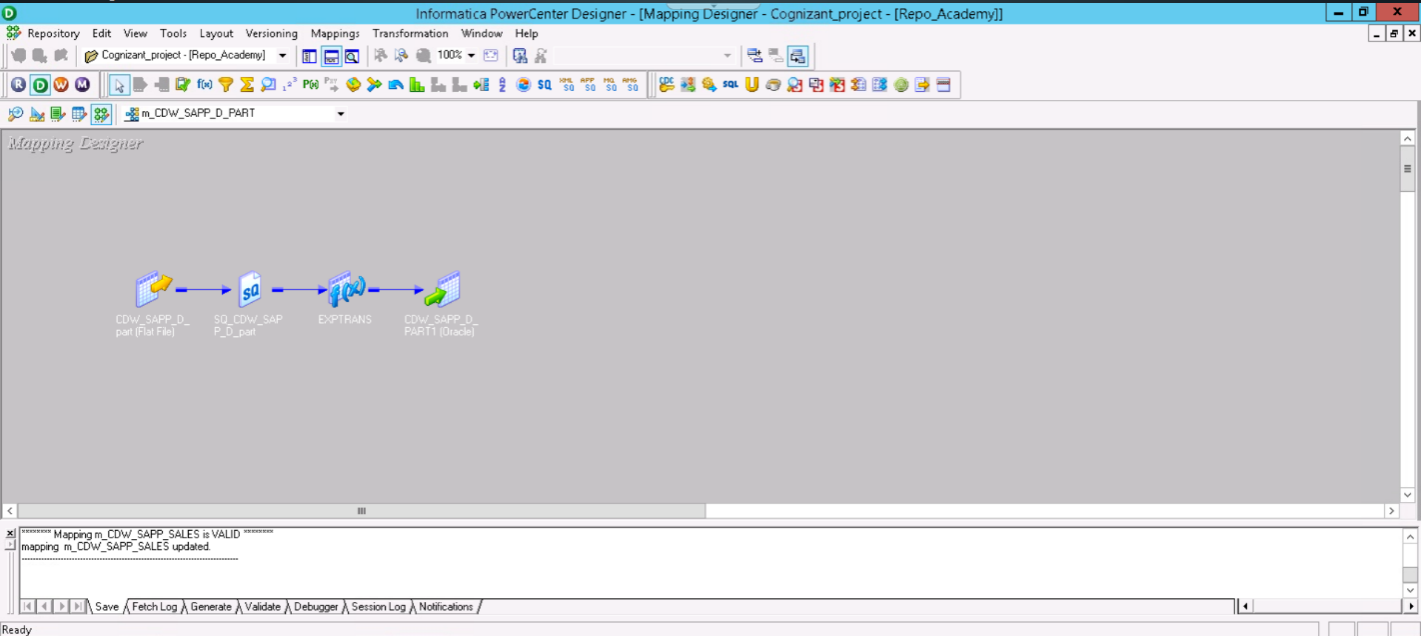
1. CDW\_SAPP\_D\_VENDOR



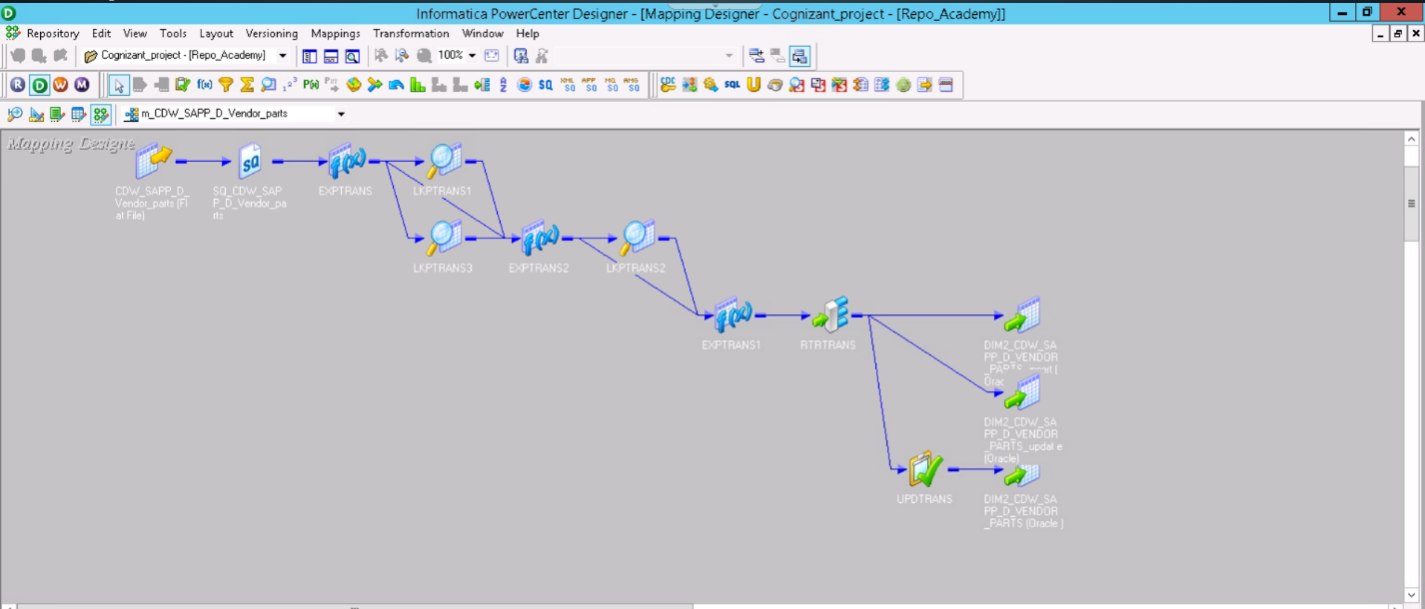
1. CDW\_SAPP\_D\_GEOGRAPHY



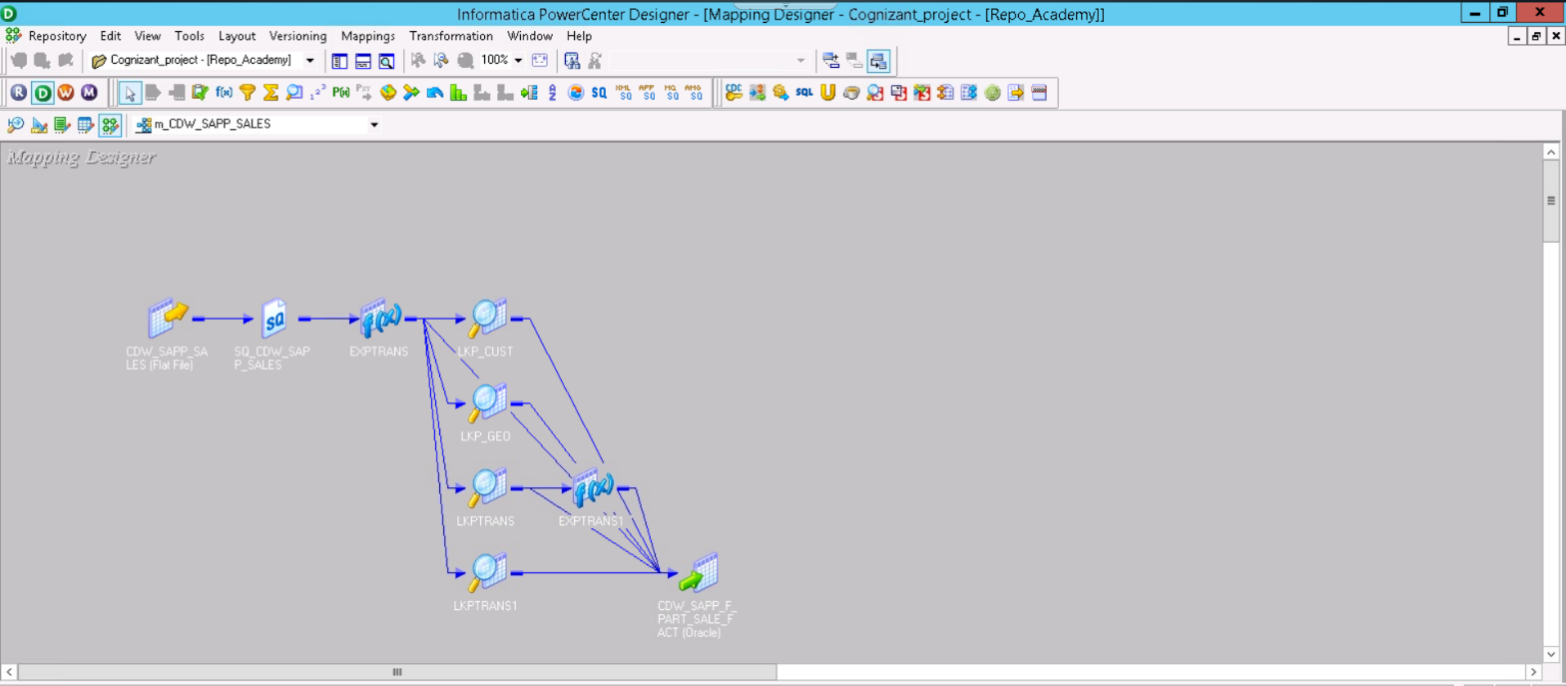
1. CDW\_SAPP\_D\_PART



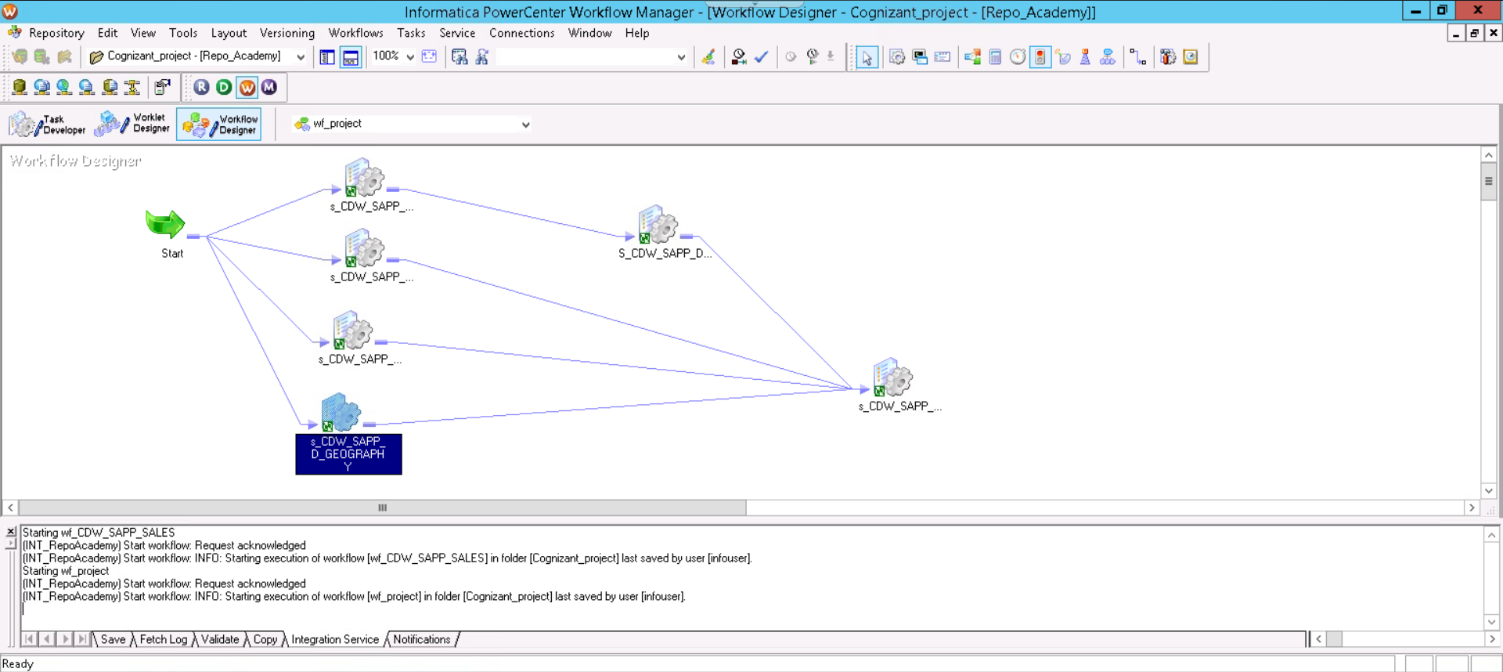
1. CDW\_SAPP\_D\_Vendor\_parts



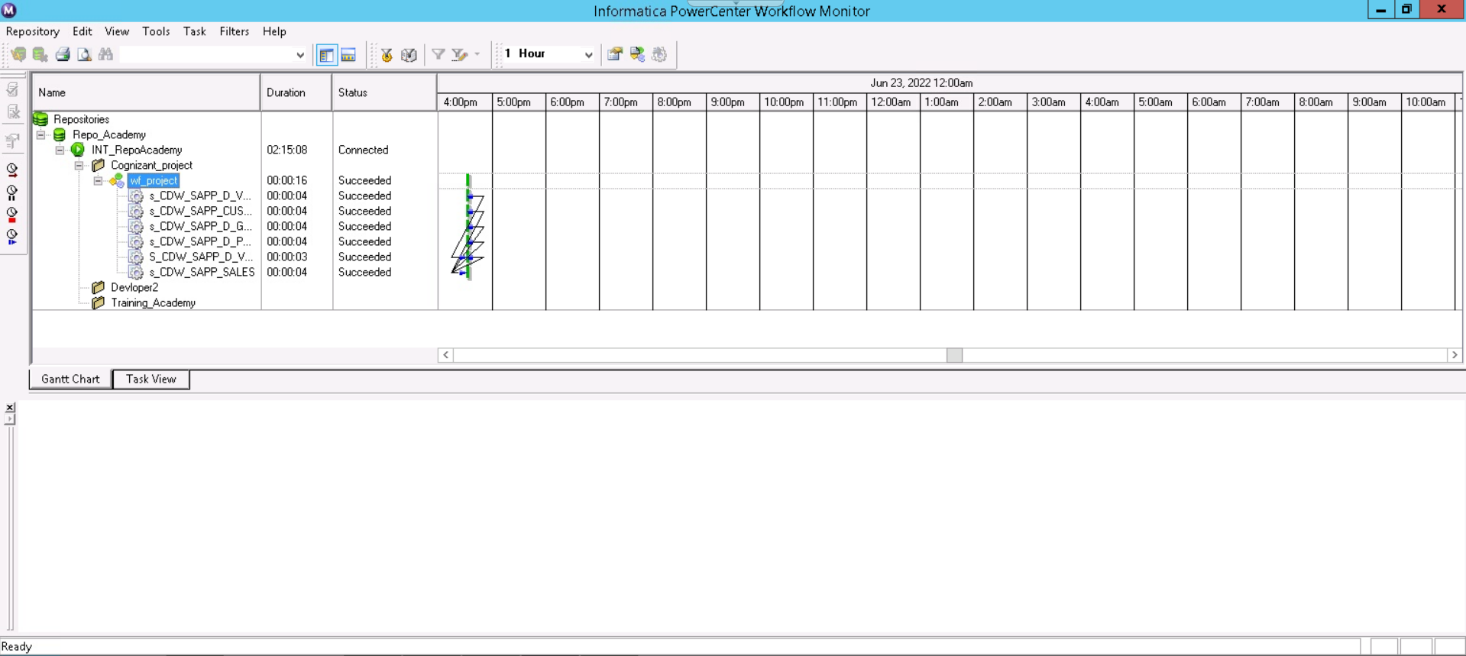
1. CDW\_SAPP\_SALES



**6 Workflow :**



**6 Monitor :**



**7 Target files:**

CDW\_SAPP\_SALES



CDW\_SAPP\_D\_VENDOR\_PARTS



CDW\_SAPP\_D\_VENDOR



CDW\_SAPP\_D\_PART



CDW\_SAPP\_D\_GEOGRAPHY



CDW\_SAPP\_D\_CUSTOMER



**5. Conclusion:**

The Informatica PowerCenter is a powerful tool used for data integration operations and used to create data warehouses for the industries. The tool takes no time for data transformation operation and delivers the data to industries so that the data warehouse can be created and used in projects.