

# wIntroduction to Business Intelligence

The test for this connection was successful.

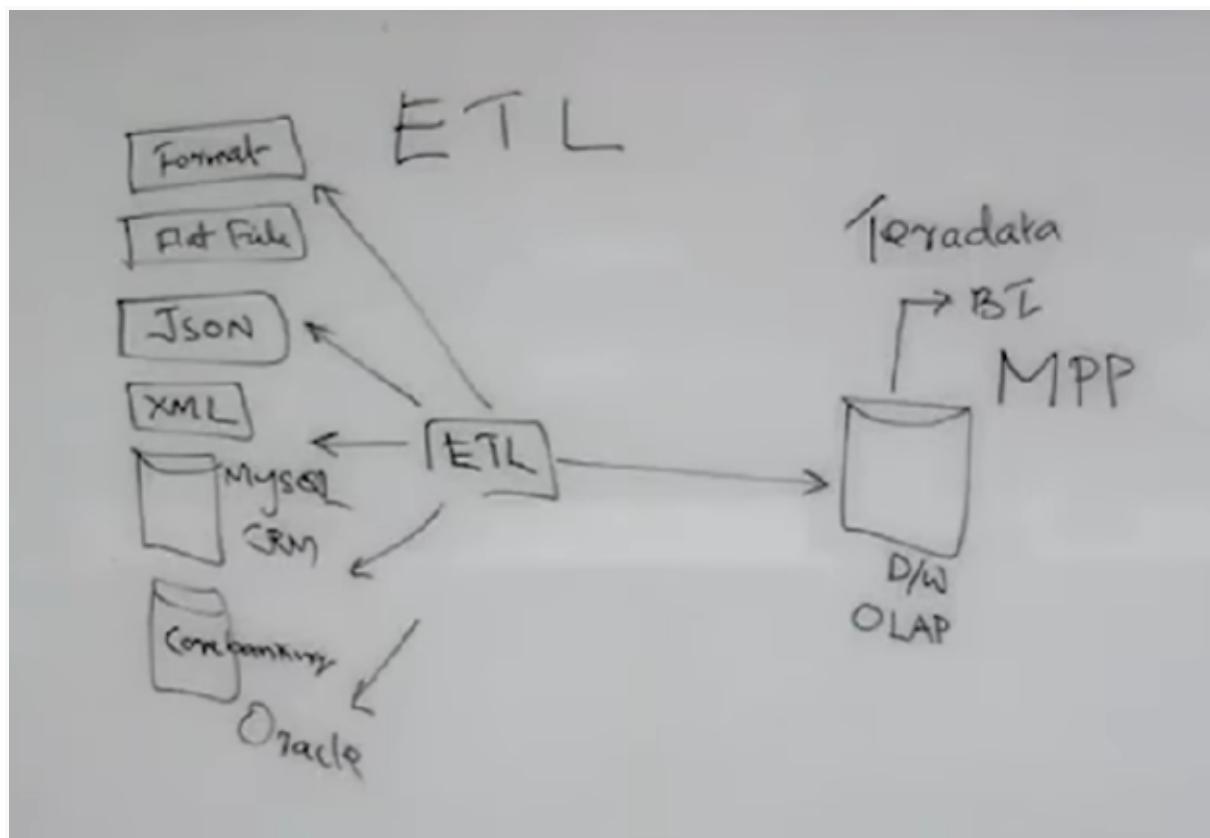
**Connection Details**

Connection Name: <sup>*</sup>	oracle_connection
Description:	
Type: <sup>*</sup>	Oracle

**Oracle Connection Properties**

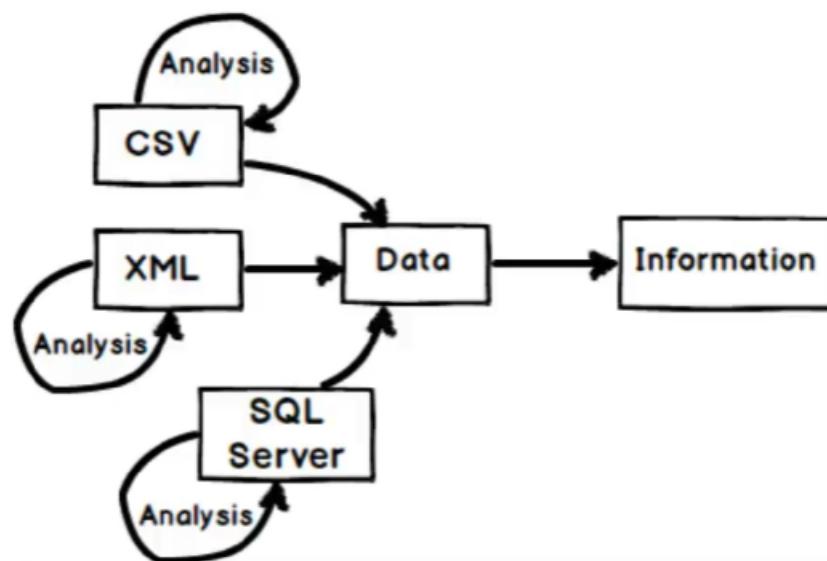
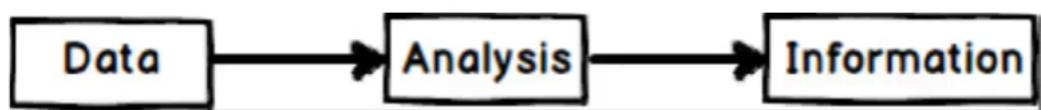
Runtime Environment: <sup>*</sup>	DESKTOP-QG6V1FB
User Name: <sup>*</sup>	HR
Password: <sup>*</sup>	• •
Host: <sup>*</sup>	localhost
Port: <sup>*</sup>	1521
Service Name: <sup>*</sup>	xe
Schema:	
Code Page: <sup>*</sup>	UTF-8
Encryption Method:	No Encryption
Crypto Protocol Version:	TLSv1
Validate Server Certificate:	False
Trust Store:	
Trust Store Password:	
Host Name in Certificate:	
Key Store:	

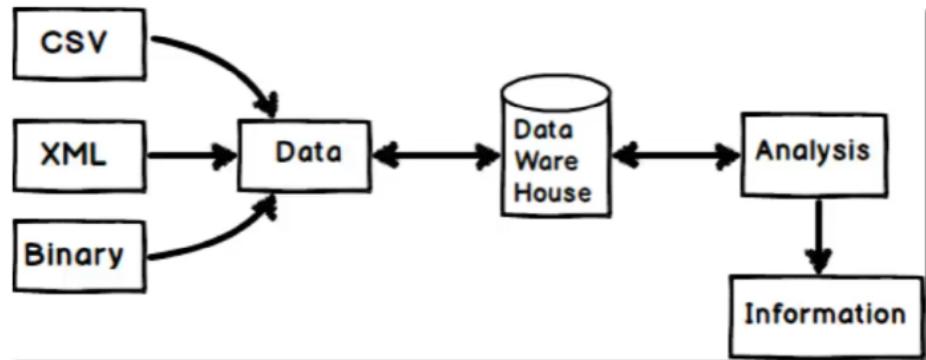
What is BI



It is the process of converting data into information

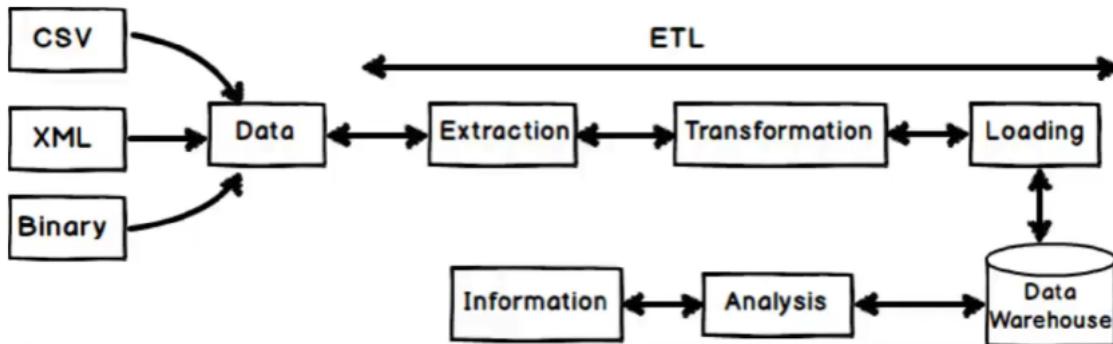


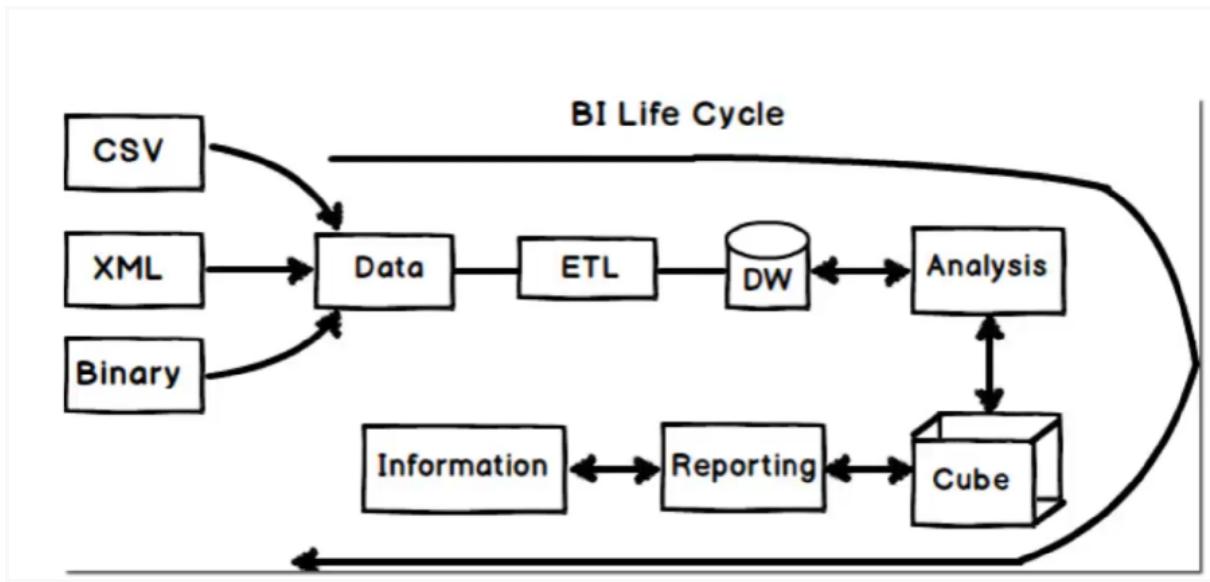




**Data Warehousing (DW)** is a process for collecting and managing data from varied sources to provide meaningful business insights. A Data warehouse is typically used to connect and analyse business data from heterogeneous sources. The data warehouse is the core of the BI system which is built for data analysis and reporting.

<https://weld.app/blog/top-5-data-warehouses>





BI(Business Intelligence) is a set of processes, architectures, and technologies that convert raw data into meaningful information that drives profitable business actions.

It is a suite of software and services to transform data into actionable intelligence and knowledge.

## Why is BI important?

- Measurement: creating KPI (Key Performance Indicators) based on historical data
- Identify and set benchmarks for varied processes.
- With BI systems organizations can identify market trends and spot business problems that need to be addressed.
- BI helps on data visualization that enhances the data quality and thereby the quality of decision making.
- BI systems can be used not just by enterprises but SME (Small and Medium Enterprises)

Difference between Database and Datawarehouse

## What is Database?

A database is a collection of related data which represents some elements of the real world. It is designed to be built and populated with data for a specific task. It is also a building block of your data solution.

## KEY DIFFERENCE

- Database is a collection of related data that represents some elements of the real world whereas Data warehouse is an information system that stores historical and commutative data from single or multiple sources.
- Database is designed to record data whereas the Data warehouse is designed to analyze data.
- Database is application-oriented-collection of data whereas Data Warehouse is the subject-oriented collection of data.
- Database uses Online Transactional Processing (OLTP) whereas Data warehouse uses Online Analytical Processing (OLAP).
- Database tables and joins are complicated because they are normalized whereas Data Warehouse tables and joins are easy because they are denormalized.
- ER modeling techniques are used for designing Database whereas data modeling techniques are used for designing Data Warehouse.

**Online Analytical Processing (OLAP):** Online Analytical Processing consists of a type of software tools that are used for data analysis for business decisions. OLAP provides an environment to get insights from the database retrieved from multiple database systems at one time. **Examples -** Any type of Data warehouse system is an OLAP system. The uses of OLAP are as follows:

- Spotify analyzed songs by users to come up with a personalized homepage of their songs and playlist.
- Netflix movie recommendation system.

**Online transaction processing (OLTP):** Online transaction processing provides transaction-oriented applications in a 3-tier architecture. OLTP administers the day-to-day transactions of an organization.

**Examples:** Uses of OLTP are as follows:

- ATM center is an OLTP application.
- OLTP handles the ACID properties during data transactions via the application.
- It's also used for Online banking, Online airline ticket booking, sending a text message, add a book to the shopping cart.

What is Data lake

- Data Lake is a centralised storage system to store all the raw data from different source/ingestion systems.

- It is scalable enough to store all the data for the whole enterprise.
- It can store data in various formats and types. It can store structured (rows and columns), semi-structured (CSV, XML, JSON), unstructured (documents, emails, pdf), and binary (audio, images, videos) data.
- Later, various big data processing techniques are used to transform this data to extract the intelligence for the business.

## What is Data Mart

A **Data Mart** is a subset of a directorial information store, generally oriented to a specific purpose or primary data subject which may be distributed to provide business needs. Data Marts are analytical record stores designed to focus on particular business functions for a specific community within an organization. Data marts are derived from subsets of data in a data warehouse, though in the bottom-up data warehouse design methodology, the data warehouse is created from the union of organizational data marts.

The fundamental use of a data mart is **Business Intelligence (BI)** applications. BI is used to gather, store, access, and analyze record. It can be used by smaller businesses to utilize the data they have accumulated since it is less expensive than implementing a data warehouse.

## Reasons for creating a data mart

- Creates collective data by a group of users
- Easy access to frequently needed data
- Ease of creation
- Improves end-user response time
- Lower cost than implementing a complete data warehouses
- Potential clients are more clearly defined than in a comprehensive data warehouse
- It contains only essential business data and is less cluttered.

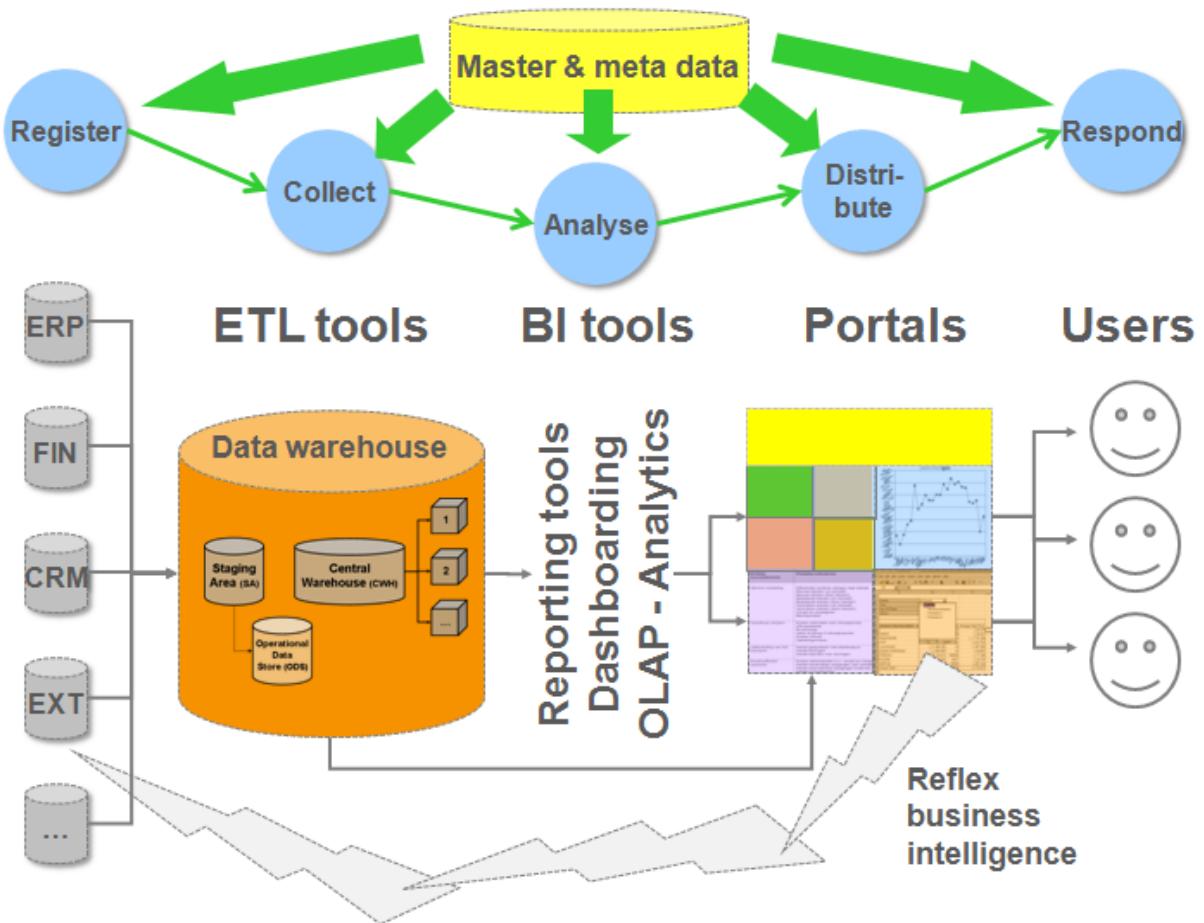
Then fill we will get data to tools like power bi and Tableau

Where Informatica Comes into picture

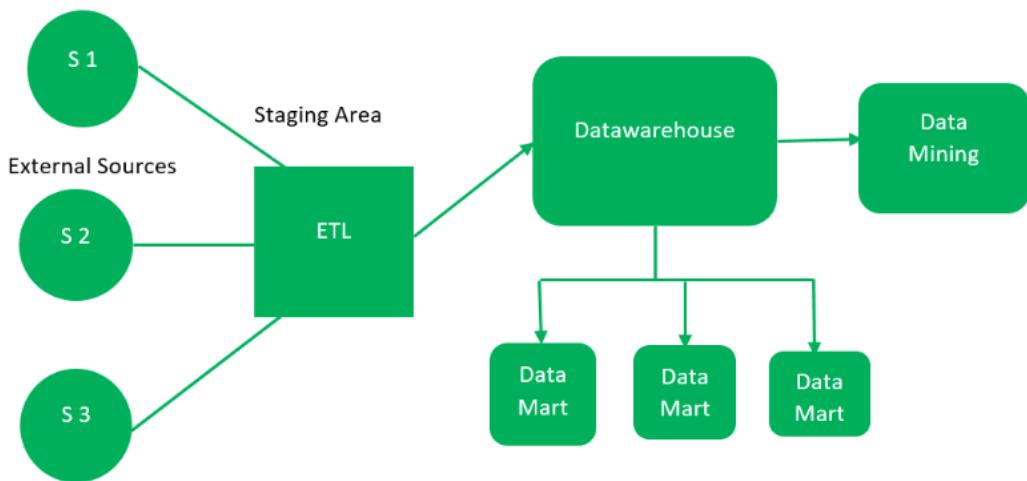
Informatica is a data processing tool that is widely used for ETL to extract, transform and load processing.

Informatica Intelligent Cloud Services (IICS) is a cloud-based data integration platform that provides a variety of features such as business data integration, application integration, and API management between cloud and local applications

**Conclusion of what is bi Architecture**



## Data Warehouse Architecture



<https://www.geeksforgeeks.org/data-warehouse-architecture/>

Explain Top down and bottom down

### Who needs Data warehouse?

DWH (Data warehouse) is needed for all types of users like:

- Decision makers who rely on mass amount of data
- Users who use customized, complex processes to obtain information from multiple data sources.
- It is also used by the people who want simple technology to access the data
- It is also essential for those people who want a systematic approach for making decisions.
- If the user wants fast performance on a huge amount of data which is a necessity for reports, grids or charts, then Data warehouse proves useful.
- Data warehouse is a first step if you want to discover ‘hidden patterns’ of data-flows and groupings.

<https://www.guru99.com/data-warehousing.html>

Youtube analytics as an example

Informatica MDM (amazon address example)

**Featured snippet from the web**

Informatica MDM full form is **Master Data Management** that primarily focuses on the betterment of business data. A Master Data Solution refers to a management tool that binds the system and the information in sync with an organisation. It is essential when it comes to managing company data.

What is Dimension and fact table

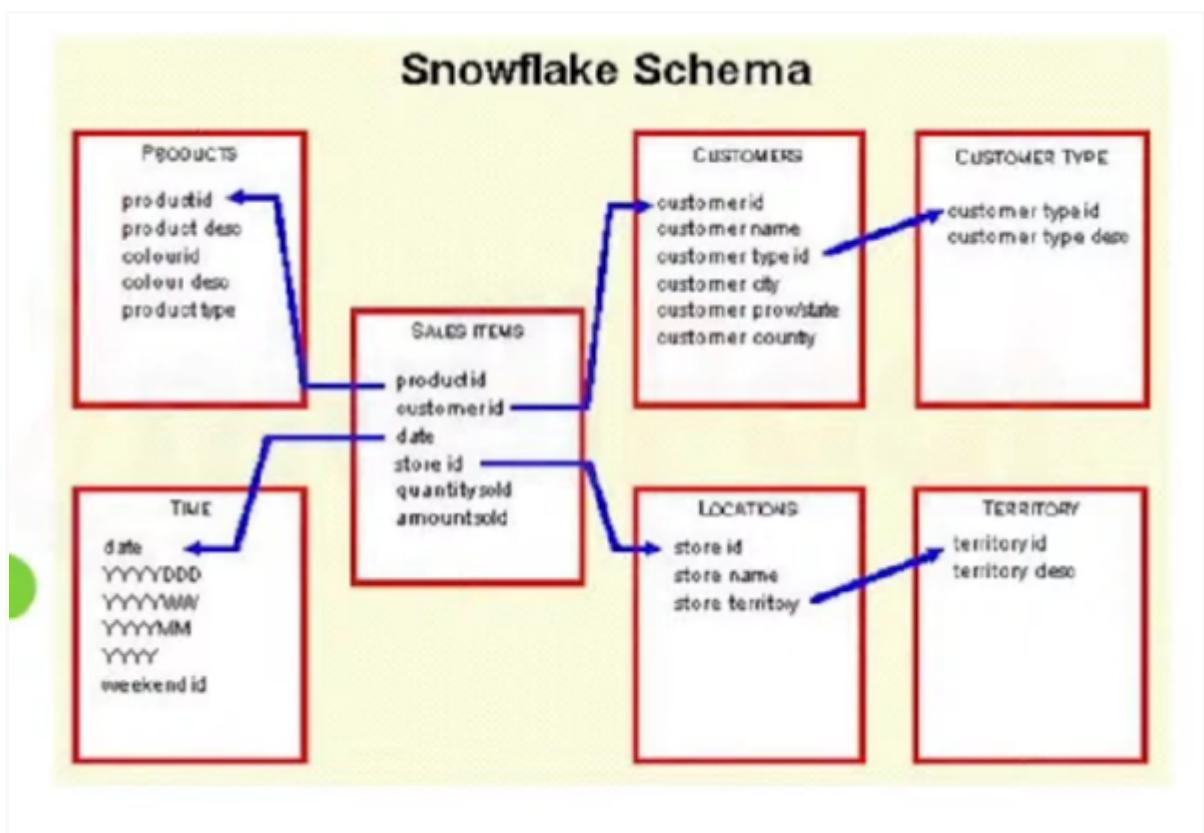
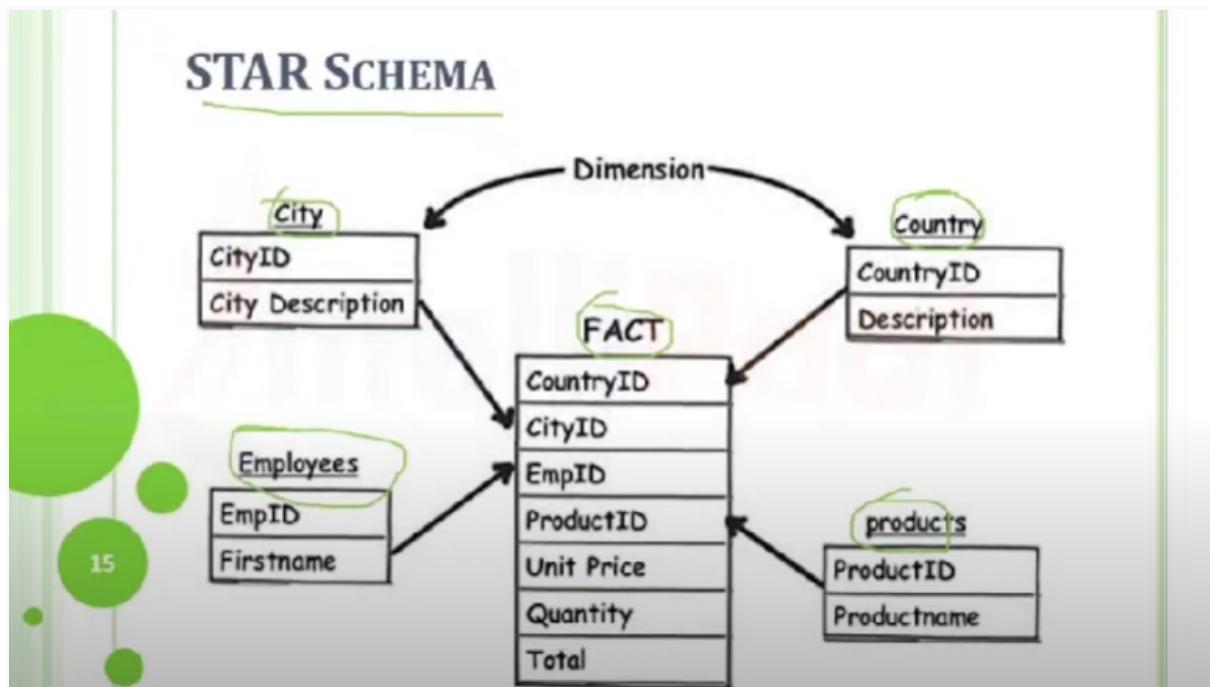
<https://www.geeksforgeeks.org/difference-between-fact-table-and-dimension-table/>

#### KEY DIFFERENCE

- Fact table contains measurements, metrics, and facts about a business process while the Dimension table is a companion to the fact table which contains descriptive attributes to be used as query constraining.
- Fact table is located at the center of a star or snowflake schema, whereas the Dimension table is located at the edges of the star or snowflake schema.
- Fact table is defined by their grain or its most atomic level whereas Dimension table should be wordy, descriptive, complete, and quality assured.
- Fact table helps to store report labels whereas Dimension table contains detailed data.
- Fact table does not contain a hierarchy whereas the Dimension table contains hierarchies.

## Star Schema and Snowflake

<https://www.geeksforgeeks.org/difference-between-star-schema-and-snowflake-schema/>



We normalise tabela to avoid data redundancy (to avoid duplication)

## DATA WAREHOUSING ROLES:

- Business Users --- People with knowledge of the business and needs for reporting
- Report Writers --- People who write reports (Developers); tools for writing reports
- Data Modelers --- People who create data models from Business Requirements
- ETL Developers --- People who write / develop ETL processes
- DBA --- People who manage and support databases

## ETL Tools

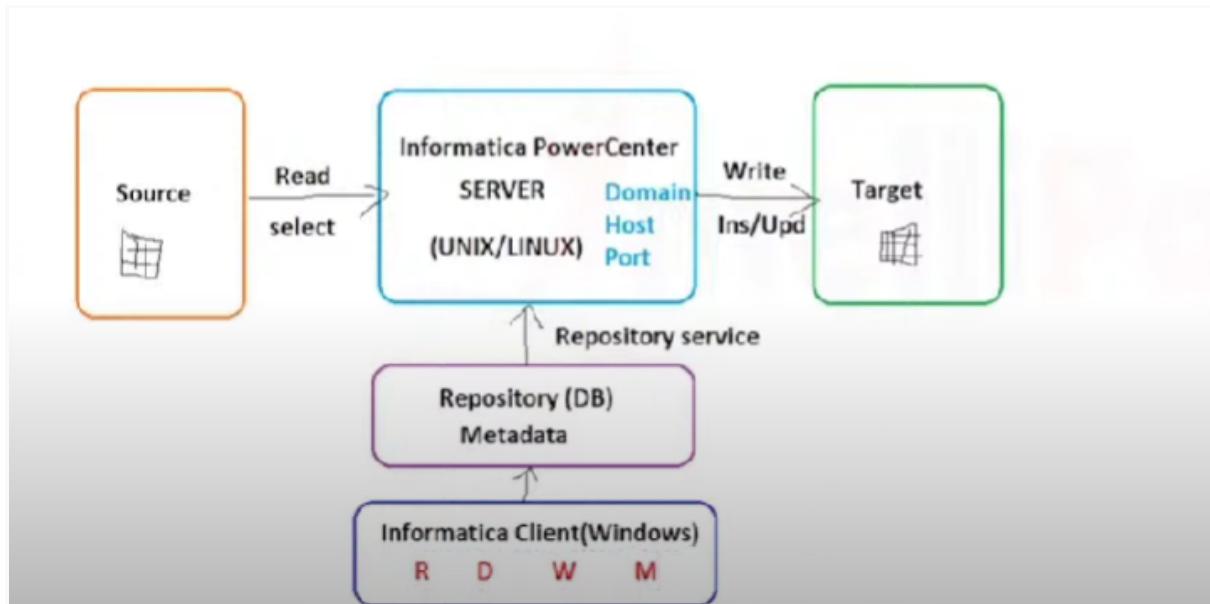
## POPULAR ETL TOOLS

- ❖ **Informatica - PowerCenter** ✓
- ❖ **IBM - Websphere DataStage**
- ❖ **Microsoft – SQL Server Integrated Services (SSIS)**
- ❖ **SAP – Business Objects Data Services (BODS)**
- ❖ **IBM - Cognos Data Manager**
- ❖ **Oracle - Data Integrator(ODI)**
- ❖ **SAS - Data Integration Studio**
- ❖ **Oracle - Warehouse Builder**
- ❖ **AB Initio**
- ❖ **Talend Open Studio**
- ❖ **Apache NiFi**
- ❖ **Pentaho Data Integration**
- ❖ **Xplenty - cloud-based ETL**

19

## POPULAR DATA VISUALIZATION TOOLS

- ❖ **Tableau**
- ❖ **Microsoft Power BI**
- ❖ **Qlikview/Qliksense**
- ❖ **IBM Cognos**
- ❖ **SAP BO**
- ❖ **Zoomdata**
- ❖ **Google Looker**
- ❖ **SSRS**
- ❖ **OBIEE**
- ❖ **TIBCO Spotfire**
- ❖ **Micro strategy Analytics**



The PowerCenter repository is a relational database managed by the Repository Service

The repository consists of database tables that store metadata. Metadata describes different types of objects, such as mappings and transformations, that you can create or modify using the PowerCenter Client tools.

If your informatica developer gets the domain,host, and port from the admin team

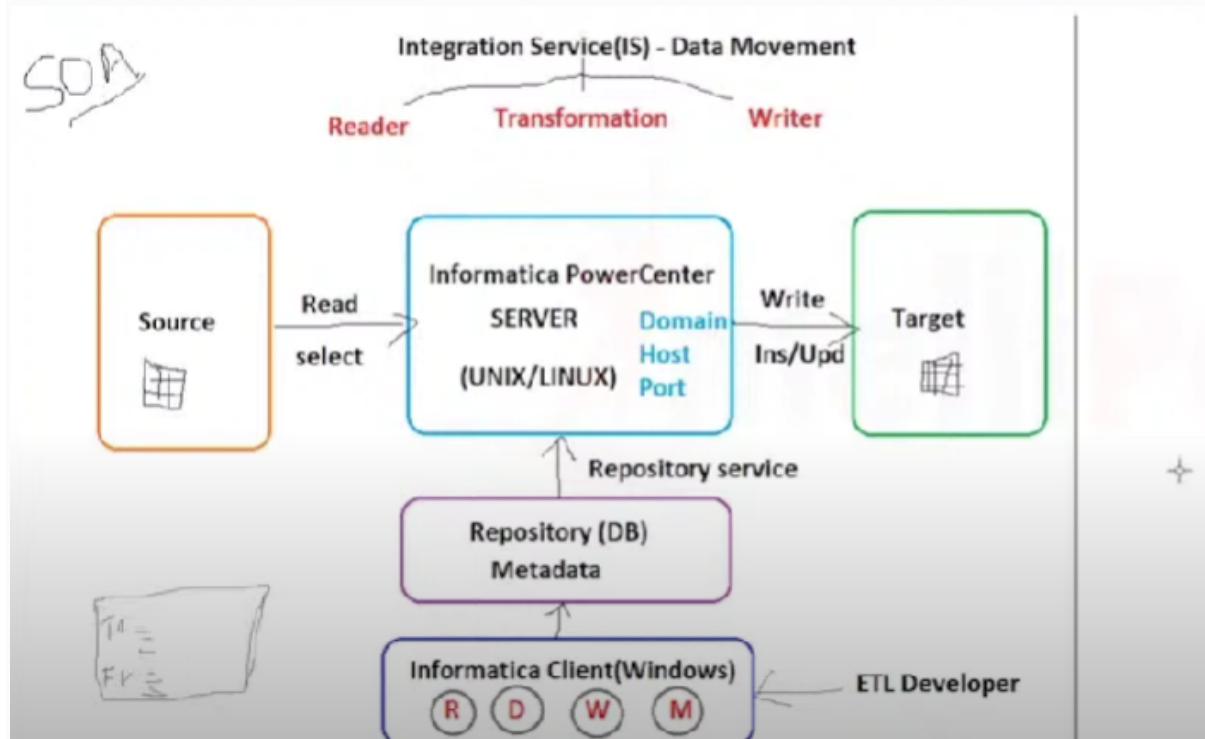
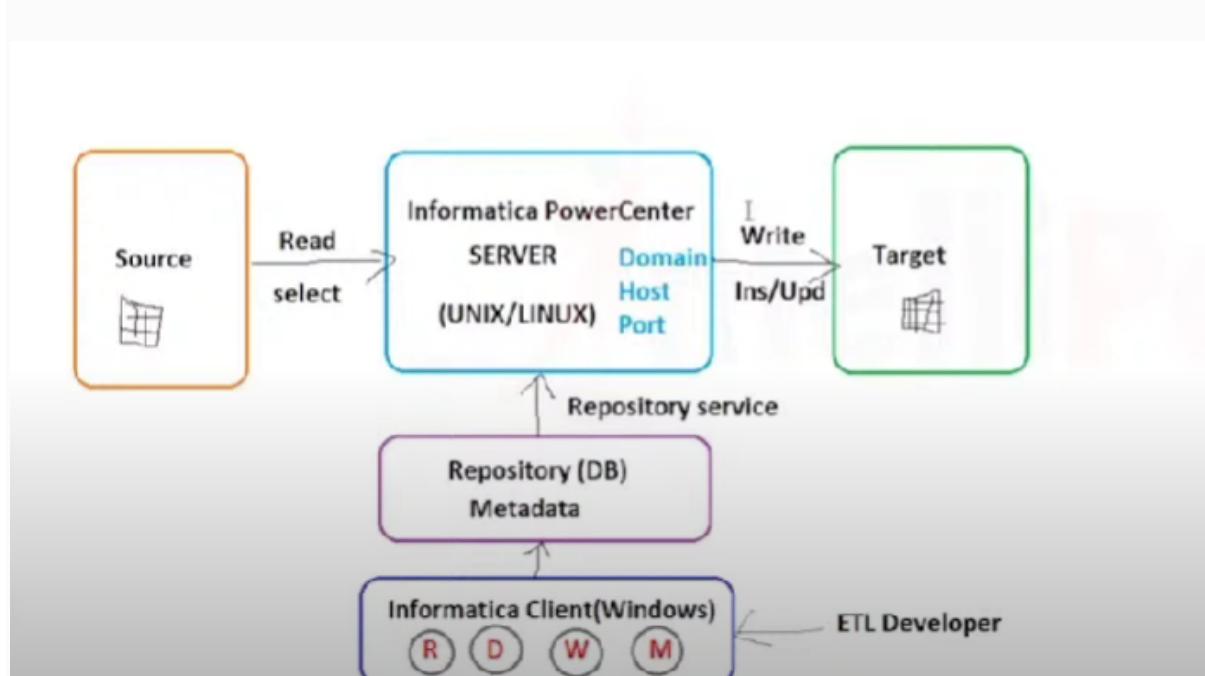
It has four client tools

The PowerCenter Client application consists of the tools to manage the repository and to design mappings, mapplets, and sessions to load the data. The PowerCenter Client application has the following tools:

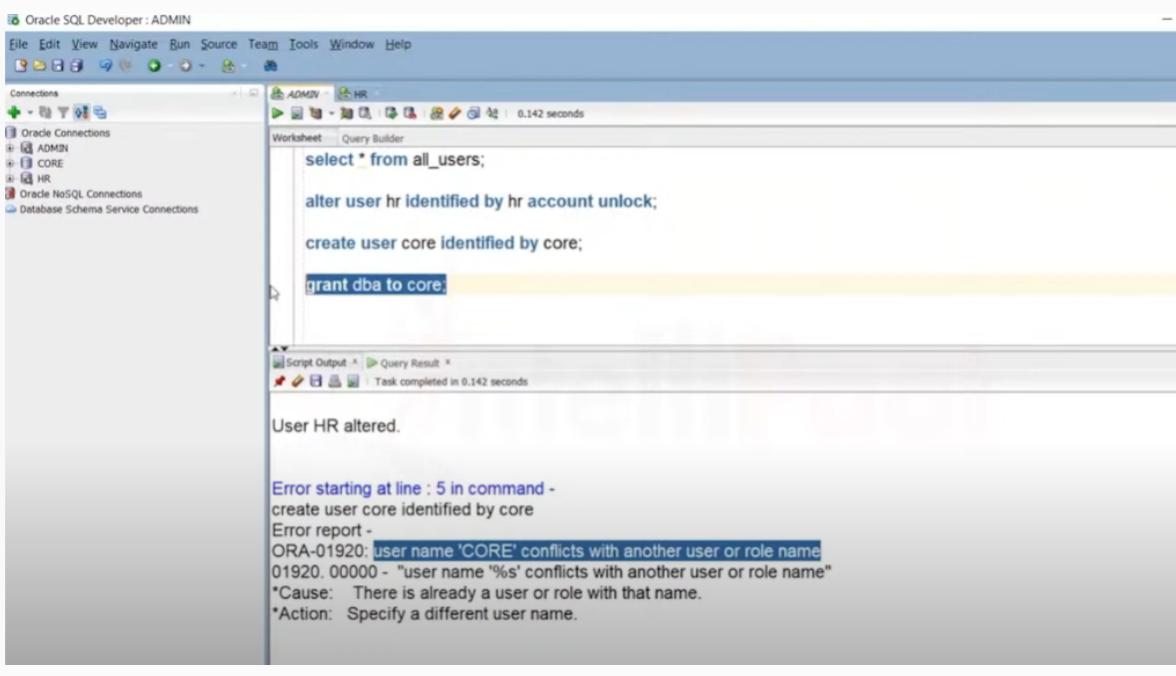
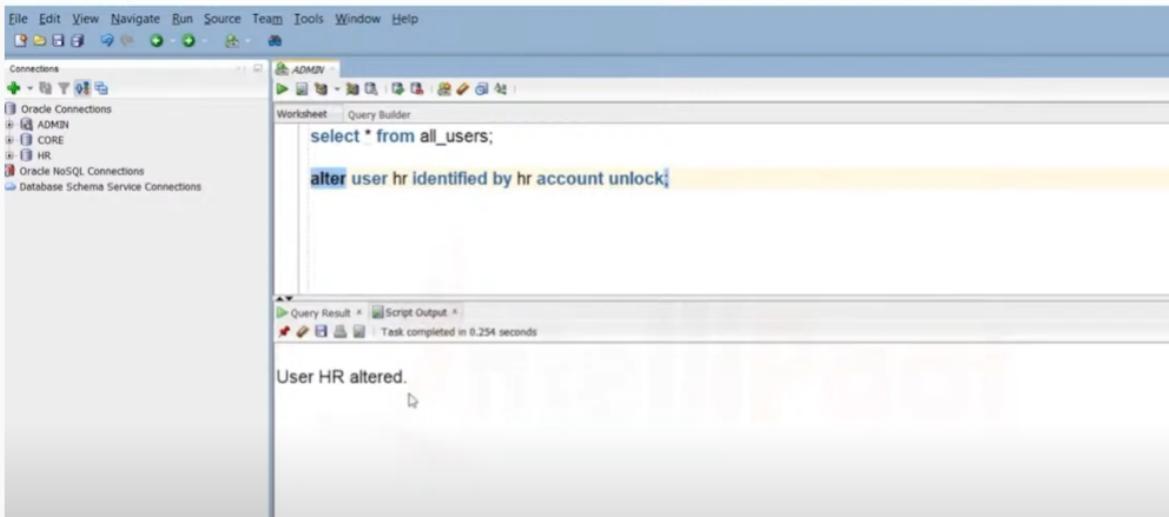
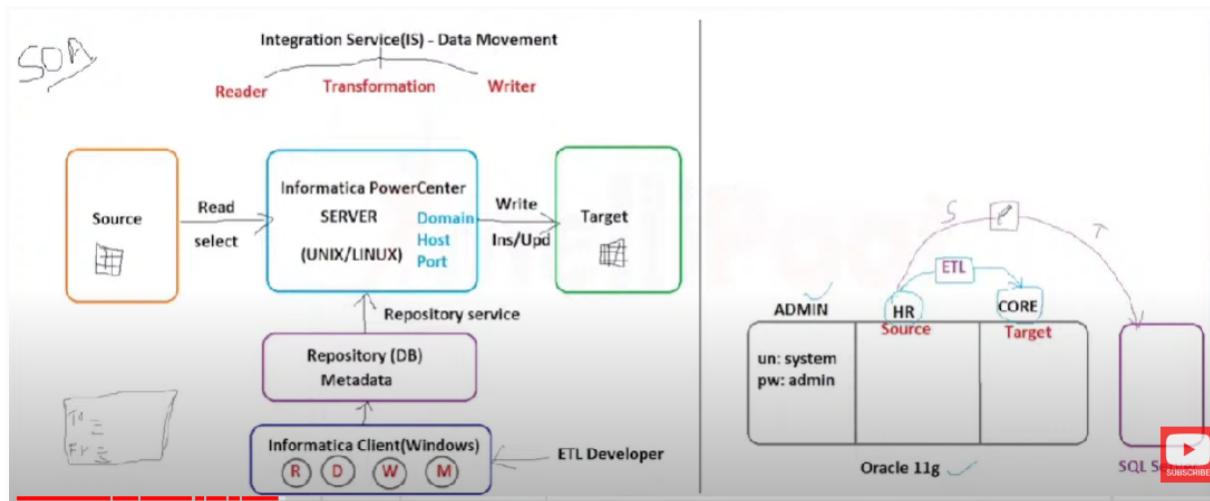
- **Designer.** Use the Designer to create mappings that contain transformation instructions for the Integration Service.
- **Mapping Architect for Visio.** Use the Mapping Architect for Visio to create mapping templates that generate multiple mappings.
- **Repository Manager.** Use the Repository Manager to assign permissions to users and groups and manage folders.
- **Workflow Manager.** Use the Workflow Manager to create, schedule, and run workflows. A workflow is a set of instructions that describes how and when to run tasks related to extracting, transforming, and loading data.

- **Workflow Monitor.** Use the Workflow Monitor to monitor scheduled and running workflows for each Integration Service.

Install the client application on a Microsoft Windows computer.



Schema logical division of database



Oracle SQL Developer : ADMIN

File Edit View Navigate Run Source Team Tools Window Help

Connections

- Oracle Connections
  - ADMN
  - CORE
    - Tables (Filtered)
    - Views
    - Indexes
    - Packages
    - Procedures
    - Functions
    - Operators
    - Queues
    - Queues Tables
    - Triggers
    - Types
    - Sequences
    - Materialized Views
    - Materialized View Logs
    - Synonyms
    - Public Synonyms
    - Database Links
    - Directories
    - Editors
    - Application Express
    - XML Schemas
    - XML DB Repository
    - Scheduler
    - Recycle Bin
    - Other Users
  - HR
- Oracle NoSQL Connections
- Database Schema Service Connections

Worksheet Query Builder

```
select * from all_users;

alter user hr identified by hr account unlock;

create user core identified by core;

grant dba to core;

create table core.t_employees as select * from hr.employees where 1=2;
```

Script Output X Query Result X

All Rows Fetched: 0 in 0.009 seconds

EMPLOYEE_ID	FIRST_NAME	LAST_NAME	EMAIL	PHONE_NUMBER	HIRE_DATE
1	100 Steven	King	SKING	515.123.4567	17-JUN-03
2	101 Neena	Kochhar	NKOCHHAR	515.123.4568	21-JUN-03
3	102 Lex	De Haan	LDEHAAN	515.123.4569	13-JUN-03
4	103 Alexander	Hunold	AHUNOLD	590.423.4567	03-JUN-03
5	104 Bruce	Ernst	BERNST	590.423.4568	21-JUN-03
6	105 David	Austin	DAUSTIN	590.423.4569	25-JUN-03
7	106 Valli	Pataballa	VPATABAL	590.423.4560	05-JUN-03
8	107 Diana	Lorentz	DLORENTZ	590.423.5567	07-JUN-03
9	108 Nancy	Greenberg	NGREENBE	515.124.4569	17-JUN-03
10	109 Daniel	Faviet	DFAVIE	515.124.4169	16-JUN-03
11	110 John	Chen	JCHEN	515.124.4269	28-JUN-03
12	111 Kenneral	Sriramra	LSRIARRA	515.124.4769	30-JUN-03

Oracle SQL Developer : CORE

File Edit View Navigate Run Source Team Tools Window Help

Connections

- Oracle Connections
  - ADMN
  - CORE
    - Tables (Filtered)
    - Views
    - Indexes
    - Packages
    - Procedures
    - Functions
    - Operators
    - Queues
    - Queues Tables
    - Triggers
    - Types
    - Sequences
    - Materialized Views
    - Materialized View Logs
    - Synonyms
    - Public Synonyms
    - Database Links
    - Directories
    - Editors
    - Application Express
    - XML Schemas
    - XML DB Repository
    - Scheduler
    - Recycle Bin
    - Other Users
  - HR
- Oracle NoSQL Connections
- Database Schema Service Connections

Worksheet Query Builder

```
select * from t_employees; I
```

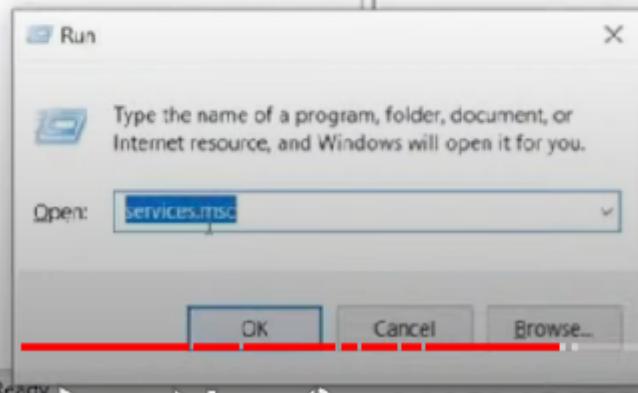
trunc

Query... X SQL Fetched 30 rows in 0.01 seconds

EMPLOYEE_ID	FIRST_NAME	LAST_NAME	EMAIL	PHONE_NUMBER	HIRE_DATE
1	100 Steven	King	SKING	515.123.4567	17-JUN-03
2	101 Neena	Kochhar	NKOCHHAR	515.123.4568	21-JUN-03
3	102 Lex	De Haan	LDEHAAN	515.123.4569	13-JUN-03
4	103 Alexander	Hunold	AHUNOLD	590.423.4567	03-JUN-03
5	104 Bruce	Ernst	BERNST	590.423.4568	21-JUN-03
6	105 David	Austin	DAUSTIN	590.423.4569	25-JUN-03
7	106 Valli	Pataballa	VPATABAL	590.423.4560	05-JUN-03
8	107 Diana	Lorentz	DLORENTZ	590.423.5567	07-JUN-03
9	108 Nancy	Greenberg	NGREENBE	515.124.4569	17-JUN-03
10	109 Daniel	Faviet	DFAVIE	515.124.4169	16-JUN-03
11	110 John	Chen	JCHEN	515.124.4269	28-JUN-03
12	111 Kenneral	Sriramra	LSRIARRA	515.124.4769	30-JUN-03

## 1) Repository Manager - R - Admin related task

- \* **Folder creation/deletion**
- \* **Metadata**
- \* **Access /Privileges**
- \* **Code Migration (xml import/export)**
- \* **Historical run status**



## 2) Designer - D - Developer

- \* Source Definition
- \* Target Definition
- \* Mapping (Business logic)

## 3) Workflow Manager - W

- \* Session
- \* Connection
- \* Workflow

```
File Edit Format View Help
Steps to create a workflow
=====
1) Source Definition - Source Analyzer - Desinger D
2) Target Definition - Target Designer - Desinger D
3) Mapping - Mapping Designer - Desinger D
4) Session(Reusable) - Task Developer - Workflow Manager W
5) Connections - Task Developer - Workflow Manager W
6) Workflow -- Workflow Designer - Workflow Manager W
7) Monitoring - Workflow Monitor - M
```

## SOURCE ANALYZER

Source Analyzer in Informatica is a tool that allows you to create and modify **source definitions**. With the help of Source analyzer, you can create or import different types of sources in Informatica like flat files, relational databases, Excel worksheets, XML definitions, etc.

## TARGET DESIGNER

Target Designer in Informatica – Importing target table from database. Target Designer in Informatica allows us to modify the existing target table, create new target table and can import different target definitions like flat files, XML, . xls and so on.

## MAPPING DESIGN

Mapping in Informatica is a structural flow of data from source to target through transformations (or) it is the pipeline which tells how to flow data from source to target. Mapping is one of the basic elements in Informatica code

## DEMO

1. CREATE SIMPLE ETL WITH EXPRESSION TRANSFORMATION SOURCE ORACLE AND TARGET ORACLE
2. CREATE USING EXPRESSION CONCAT

Customer\_id:38

Full\_name: 30

Phone number: 38

Email id: 50

Oder\_id: 38

CONCAT(CUST\_FIRST\_NAME,CUST\_LAST\_NAME)

FIRST WITH NO SPACE

SECOND WITH SPACE

THIRD WITH VARIABLE PORT

3. Project in IICS

- 4.