# Practical – 04

**Title: - Implementation of ETL transformation with Pentaho Aim: -** ETL Transformation with Pentaho.

## Lab Objectives: -

Students will understand following concepts:

1. Copy data from Source (Table/Excel/ Oracle) and store it to Target (Table/Excel/ Oracle)
2. Adding sequence, Adding Calculator, Concatenation of two fields, Splitting of two fields
3. String Operations, Sorting data, Implement the merge join transformation on tables.

## Description: -

**Pentaho Data Integration(PDI)**

It is a business Intelligence system (BI) Also known as KETTLE.

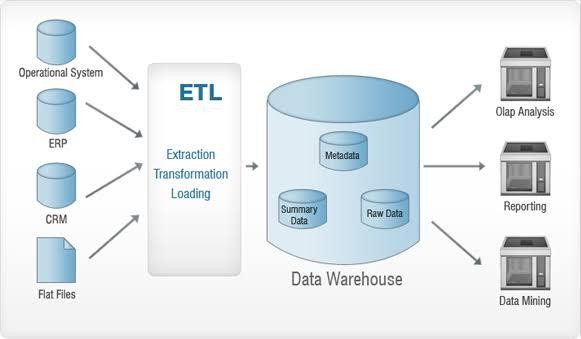
Pentaho, a subsidiary of Hitachi Vantara, is free and open- source platform for data integration and analytics. The software comes in a free community edition and a subscription-based enterprise edition.

Pentaho Data Integration (PDI) is one of the most powerful tool for building ETL processes. Founded in 2004 and Stable released on 9.1.0.0-324 / September 7, 2020.

Available for Windows, Linux, MAC OSX.

PDI is a java-based tool (Uses the Apache Java application server)

**PDI Architecture**



**Components of PDI (KETTLE)**

|  |  |  |
| --- | --- | --- |
| Tool | Windows command | Linux/ Mac OS command |
| Spoon is the graphical interface used to create transformations and jobs. | spoon.bat | spoon.sh |
| Pan is a batch-style command line tool used to run transformations | pan.bat | pan.sh |
| Kitchen is a batch-style command line tool used to run jobs. | Kitchen.bat | Kitchen.sh |
| Carte is a web server that can be used to run jobs on remote servers. | Carte.bat | Carte.sh |