**PRACTICAL - 4**

**AIM:-**

Extract and load data using Pentaho Data integration tool.

**THEORY:-**

* Pentaho Reporting is a suite (collection of tools) for creating relational and analytical reporting.
* Using Pentaho, we can transform complex data into meaningful reports and draw information out of them. Pentaho supports creating reports in various formats such as HTML, Excel, PDF, Text, CSV, and xml.
* Pentaho can accept data from different data sources including SQL databases, OLAP data sources, and even the Pentaho Data Integration ETL tool.

**The main components of Pentaho Data Integration are:**

1. **Spoon**:

It is a graphical tool which make the design of an ETTL process transformations easy to create. It performs the typical data flow functions like reading, validating, refining, transforming, writing data to a variety of different data sources and destinations. Transformations designed in Spoon can be run with Kettle Pan and Kitchen.

1. **Pan**:

It is an application dedicated to run data transformations designed in Spoon.

1. **Chef**:

It is a tool to create jobs which automate the database update process in a complex way.

1. **Kitchen**:

It is an application which helps execute the jobs in a batch mode, usually using a schedule which makes it easy to start and control the ETL processing.

1. **Carte**:

It is a web server which allows remote monitoring of the running Pentaho Data Integration ETL processes through a web browser.

**ETL Process Transformation:**

* As the name suggested ETL stands for Extract Transform and Load.Just like the name applies ETL tool Extracts data from the source.Transforms the data while in transit and then it loads the data in to Specified database.

1. **Sort Transformation:**

The Sort transformation sorts input data in ascending or descending order and copies the sorted data to the transformation output. You can apply multiple sorts to an input; each sort is identified by a numeral that determines the sort order. The column with the lowest number is sorted first, the sort column with the second lowest number is sorted next, and so on.

1. **Add Sequence:**

The Add sequence step adds a sequence to the stream. A sequence is an everchanging integer value with a specific start and increment value. You can either use a database sequence to determine the value of the sequence, or have it generated by Kettle.

1. **Concat:**

The Concat Fields step is used to concatenate multiple fields into one target field. The fields can be separated by a separator and the enclosure logic is completely compatible with the Text File Output step.

1. **Unique Rows:**

The Unique rows step removes duplicate rows from the input stream(s).

1. **Split:**

The Split Fields step allows you to split fields based on delimiter information.

**Features of Pentaho:**

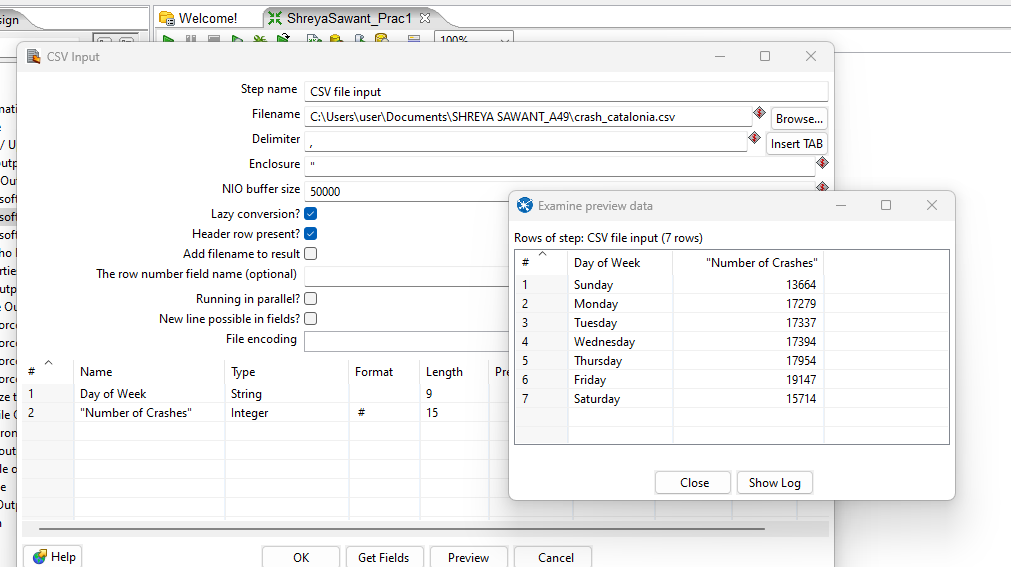
* Pentaho Reporting primarily includes a Reporting Engine, a Report Designer, a Business Intelligence (BI) Server.
* It comes loaded with the following features:
* Report Designer − Used for creating pixel perfect report.
* Metadata Editor − Allows to add user-friendly metadata domain to a data source.
* Report Designer and Design Studio − Used for fine-tuning of reports and ad-hoc reporting.
* Pentaho user console web interface − Used for easily managing reports and analyzing views.
* Ad-Hoc reporting interface − Offers a step-by-step wizard for designing simple reports. Output formats include PDF, RTF, HTML, and XLS.
* A complex scheduling sub-system − Allows users to execute reports at given intervals.
* Mailing − Users can email a published report to other users.

**PROGRAM 1:**

**Convert CSV To Excel And Apply Any 5 Tranformations.**

**STEP 1:**

Take input as csv file and select a csv file from your system.



**STEP 2:**

CONVERT CSV TO EXCEL: Take a microsoft excel output and connect hops.

Graphical user interface, text, application

Description automatically generated

1. **CONCAT FIELDS:**Graphical user interface, application

   Description automatically generated

**OUTPUT:**

Graphical user interface, text, application

Description automatically generated

1. **SORT FIELDS:**

Graphical user interface, application

Description automatically generated

**OUTPUT:**

Graphical user interface, application, table

Description automatically generated

1. **SEQUENCE:**

**Graphical user interface, application

Description automatically generated**

**OUTPUT:**

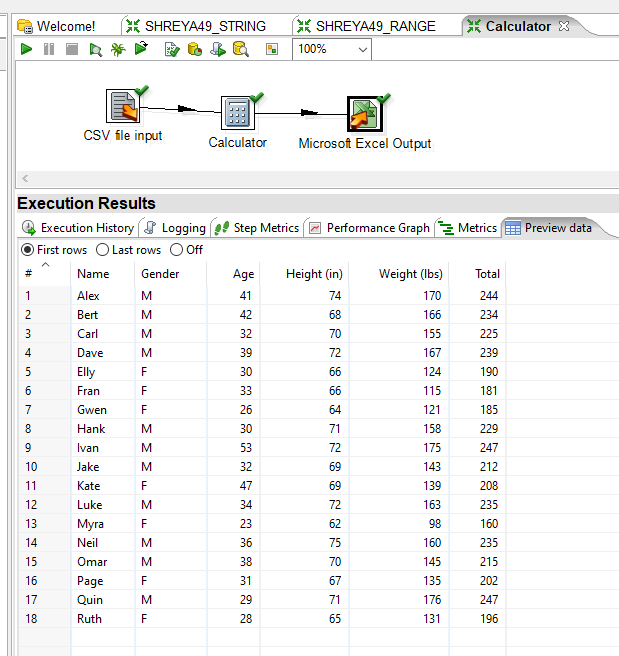
**Graphical user interface, text, application

Description automatically generated**

1. **CALCULATOR (SUM):**Graphical user interface, application

   Description automatically generated

**OUTPUT:**



1. **ADD A CHECKSUM:**Graphical user interface, application

   Description automatically generated

**OUTPUT:**

Graphical user interface, application, table

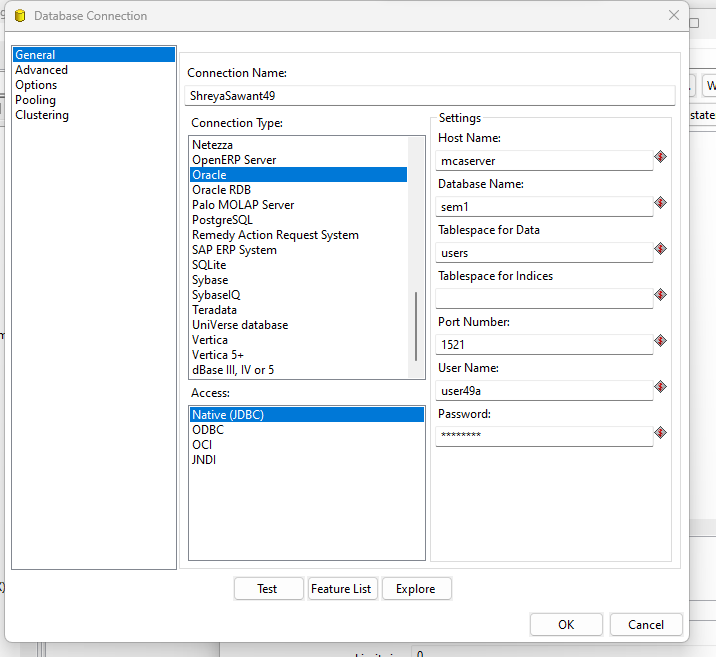
Description automatically generated

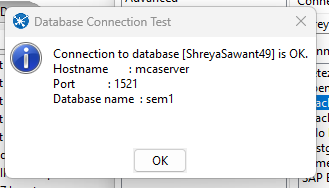
**PROGRAM 2:**

**Tranform Oracle Table Details Into Microsoft Excel File.**

**STEP 1:**

Connect your oracle database.





**STEP 2:**

Select a table from your database.

Graphical user interface, text, application

Description automatically generated

Graphical user interface, table

Description automatically generated with medium confidence

**STEP 3:**

Connect Hops from table input to microsoft excel output.

Graphical user interface, text, application

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

Table, Excel

Description automatically generated

1. **SORT FIELDS:**

Graphical user interface, application

Description automatically generated

**OUTPUT:**

Graphical user interface, application

Description automatically generated

Table, Excel

Description automatically generated

1. **CONCAT FIELDS:**

**Graphical user interface, application

Description automatically generated**

**Graphical user interface

Description automatically generated**

**OUTPUT:**

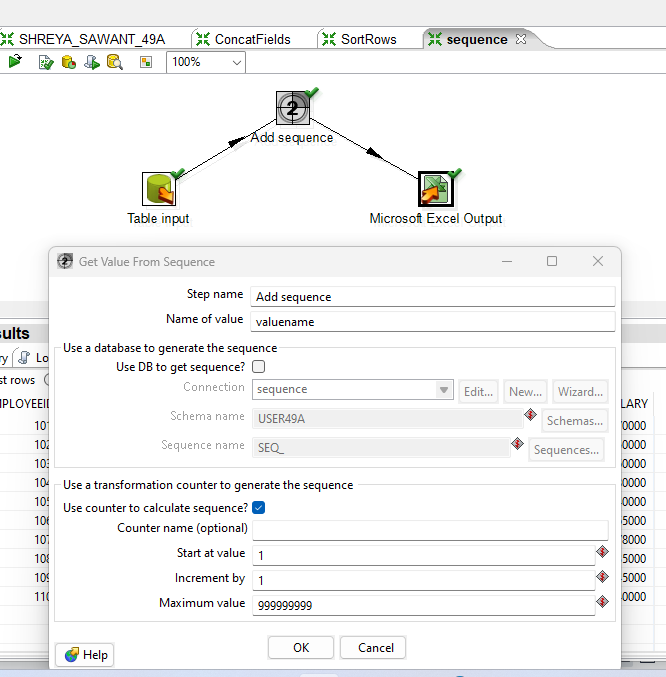
**Graphical user interface

Description automatically generated**

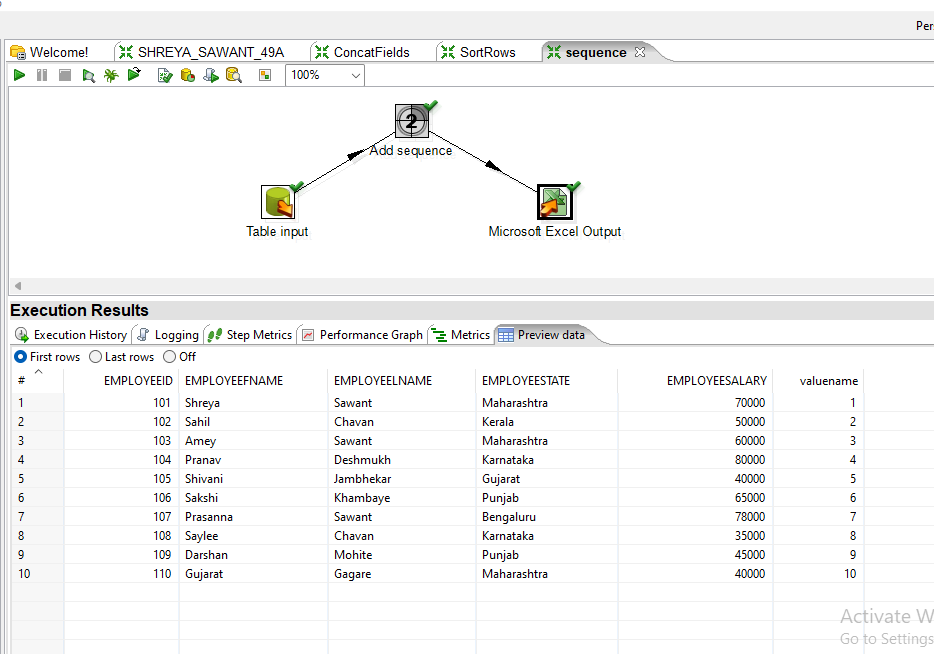
**Table

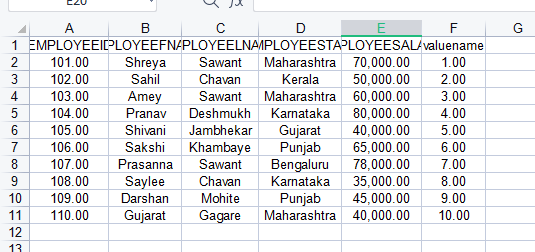
Description automatically generated**

1. **ADD A SEQUENCE:**



**OUTPUT:**





1. **STRING TRANSFORMATION (UPPER AND LOWER CASE):**Graphical user interface, application

   Description automatically generated

**OUTPUT:**

Graphical user interface, application

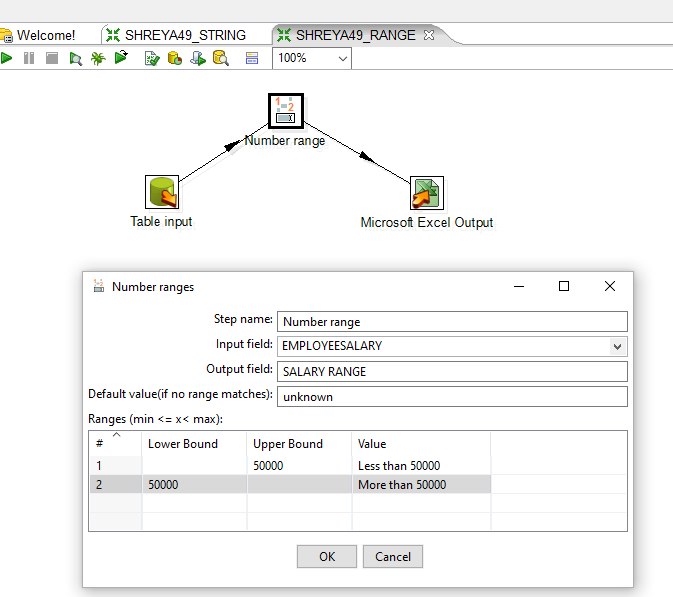
Description automatically generated

Graphical user interface, application

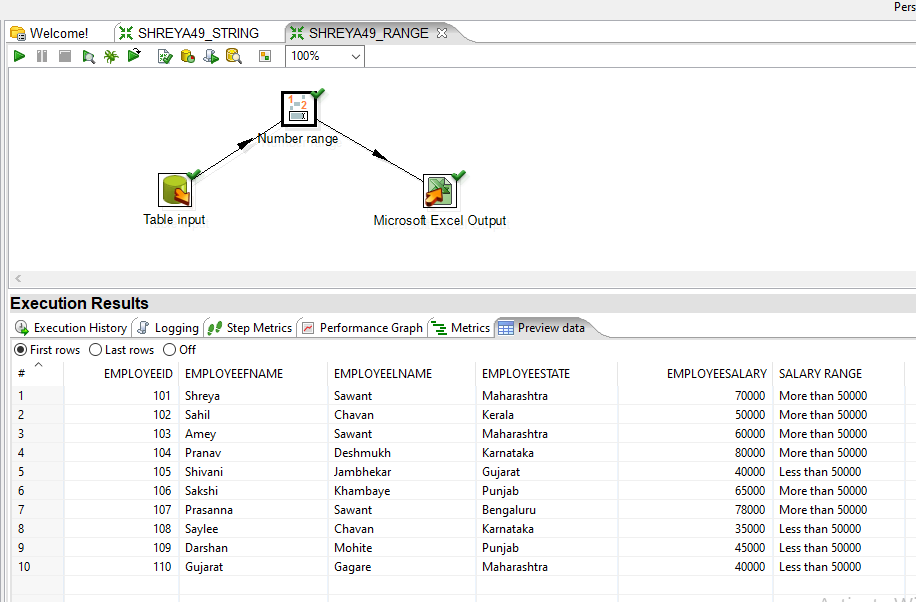
Description automatically generated

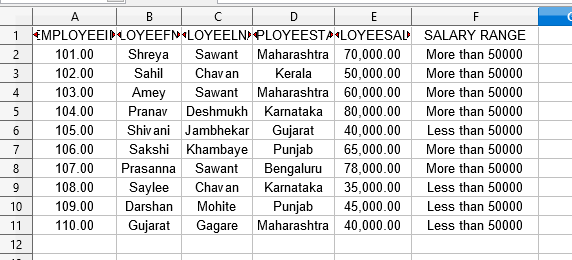
Graphical user interface, table

Description automatically generated

1. **NUMBER RANGE:**

**OUTPUT:**





**CONCLUSION:**

From this practical, I have learned how to do ETL (Extract, Transform, Load) in Pentaho.