

MES Wadia College of Engineering, Pune-01
Department of Computer Engineering

Name of Student:	Class:
Semester/Year:	Roll No:
Date of Performance:	Date of Submission:
Examined By:	Subject: LPVI (E-VI BI)

Assignment No. 2

Aim: Perform the Extraction Transformation and Loading (ETL) process to construct the database in the SQL server.

Software requirements: SQL SERVER 2012 FULL VERSION
(SQLServer2012SP1-FullSlipstream-ENU-x86)

Theory:

ETL, which stands for extract, transform and load, is a data integration process that combines data from multiple data sources into a single, consistent data store that is loaded into a data warehouse or other target system.

As the databases grew in popularity in the 1970s, ETL was introduced as a process for integrating and loading data for computation and analysis, eventually becoming the primary method to process data for data warehousing projects.

ETL provides the foundation for data analytics and machine learning workflows. Through a series of business rules, ETL cleanses and organizes data in a way which addresses specific business intelligence needs, like monthly reporting, but it can also tackle more advanced analytics, which can improve back-end processes or end user experiences. ETL is often used by an organization to:

- Extract data from legacy systems
- Cleanse the data to improve data quality and establish consistency
- Load data into a target database

Steps to be followed:

Step 1: Open SQL Server Management Studio to restore backup file

Step 2: Right click on Databases → Restore Database

Step 3: Select Device → click on icon towards end of device box

Step 4: Click on Add → Select path of backup files

Step 5: Select both files at a time

Step 6 : Click ok and in select backup devices window Add both files of AdventureWorks

Step 7: Open SQL Server Data Tools

Select File → New → Project → Business Intelligence → Integration Services

Project & give appropriate project name. Environment consists of SQL Server Integration Services(SSIS)

Step 8: Right click on Connection Managers in solution explorer and click on New Connection Manager. Add SSIS connection manager window appears.

Step 9: Select OLEDB Connection Manager and Click on Add

Step 10: Configure OLE DB Connection Manager window appears → Click on New

Step 11: Select Server name(as per your machine) from drop down and database name and click on Test connection. If test connection succeeded click on OK.

Step 12: Click on OK Connection is added to connection manager

Step 13: Drag and drop Data Flow Task in Control Flow tab

Step 14: Drag OLE DB Source from Other Sources and drop into Data Flow Tab

Step 15: Double click on OLE DB source → OLE DB Source Editor appears → click on New to add connection manager. Select [Sales].[Store] table from drop down → ok

Step 16: Drag ole db destination in data flow tab and connect both

Step 17: Double click on OLE DB destination Click on New to run the query to get [OLE DB Destination] in Name of the table or the view. Click on ok

Step 18: Click on start.

Step 19: Go to SQL Server Management Studio In database tab → Adventureworks → Right click on [dbo].[OLE DB Destination] → Scrip Table as → SELECT To → New Query Editor Window

Step 20: Execute following query to get output.

```
USE [AdventureWorks2012]
```

```
GO
```

```
SELECT [BusinessEntityID]
```

```
,[Name]
```

```
,[SalesPersonID]
```

```
,[Demographics]
```

```
,[rowguid]
```

```
,[ModifiedDate]
```

```
FROM [dbo].[OLE DB Destination]
```

```
GO
```

Questions: (<https://www.ibm.com/in-en/topics/etl>)

1. Explain How ETL works?

2. What are the benefits and challenges of ETL?