# SQL Server Integration Services: -

## What is ETL?

ETL (Extract, Transform and Load) is a process in data warehousing responsible for pulling data out of the various source systems and placing it into a data warehouse (Destination).

**Extracting the data** from source systems (SAP, ERP, and other operational systems), data from different source systems is converted into one consolidated data warehouse format which is ready for transformation processing.   
  
**Transforming the data** may involve the following tasks:

  applying business rules (so-called derivations, e.g., calculating new measures and dimensions),

   cleaning (e.g., mapping NULL to 0 or "Male" to "M" and "Female" to "F" etc.),

   filtering (e.g., selecting only certain columns to load),

   splitting a column into multiple columns and vice versa,

   joining together data from multiple sources (e.g., lookup, merge),

   transposing rows and columns,

   applying any kind of simple or complex data validation (e.g., if the first 3 columns in a row are empty then reject the row from processing)

**Loading the data** into a data warehouse or data repository other reporting applications.

## What is Business Intelligence (BI)

Business Intelligence (BI) refers to technologies, applications and practices for the collection, integration, analysis, and presentation of business information. The purpose of BI is to support business decision making. Essentially, BI systems are data-driven Decision Support Systems (DSS).

## Difference between SSIS 2012 and SSIS 2008

## **Undo Change in SSIS Package:** Undo was not available in 2008 SSIS Package. Once you have deleted any Task or Transformation, there was no way to undo the change. SSIS 2012 has added Undo facility. You can use Ctrl+Z or Click on the Button as shown in below picture.

## **Group Tasks and Transformations:** The Tasks and Transformations can be grouped together. You will not be using this Group to connect to other Containers or Tasks. This grouping is just for arranging the Tasks or Transformations but you will not be able to set the properties like Containers.

## **Change the Scope of Variable/s:** In earlier versions of SSIS, to change the scope of variable from one Task to another Task or from Task level to Package level, we had to delete the variable and then create with new scope. In SSIS 2012, we can change the scope of variable without deleting and recreating it as shown below.

# **Decimal Data Type Variable:** In previous version of SQL Server Integration Services (SSIS) Decimal data type was not available for Variables. In SSIS 2012 Microsoft introduced Decimal data type

## **Expression Character Limit:** We often have to write expression in Execute SQL Task for source query. The character limit for SQL Server 2008R2 and old version are 4000. This limit is removed in SQL Server 2012 Integration Services.

## **SSIS Toolbox and Variable Button:** SSIS Toolbox and Variables buttons are added to Package Design Surface as shown below.

## **Zoom In and Zoom Out:** For Zoom in and Zoom out a scale is added at the right bottom corner of Package design surface.

## **New Added Functions for Expressions:** Below are the newly added functions which can be used while we write expressions. (LEFT, TOKEN, TOKENCOUNT, REPLACENULL).

## **Row Count New GUI:** Row Count GUI is simplified, the purpose of Row Count was to save the Count in user variable but Row Count Transformation in SSIS 2008R2 and earlier versions were also displaying all system variables. In SSIS 2012, it only displays the Variables which are Integer Type and does not provide the list of all variables that helps to configure this Transformation quickly.

## **Data Viewer:** Extra features of Data Viewer are removed in SSIS 2012 such as see the data as Histogram, Scatter Plot and Column Chart. I have been always using Grid and never used other types and specially for debugging other charts really does not help.

## **Pivot Transformation GUI Improvements:** New Graphical User Interface is provided in SSIS 2012 for Pivot Transformation to configure it quickly and easily.

## 

## **ODBC Source and ODBC Destination:** ODBC Source and ODBC Destination is added in SQL Server Integration Services 2012.

## Difference between Control Flow and Data Flow task?

|  |  |
| --- | --- |
| Control Flow | Data Flow |
| 1. Control flow is where you define operations and order of execution of those operations. For example you put two operations: execute t-sql command on a database and send mail. Then you define order of them with Precedence Constraint for example the t-sql statement should be executed and then if it succeeded mail will be send. 2. The smallest unit in control flow is called Task. 3. In control flow, tasks require completion (Success, Failure, and Completion) before moving to next task. | 1. Data flow is where you define data stream; where data comes from (data sources), how data should be transformed (transformations) and where data should be loaded (data destinations). You can pass data from a component to another components with Data Paths. 2. The smallest unit in data flow task is called Component. 3. In Data flow, one component will not wait for other component to finish, all of them will work together in processing and managing data in streaming way. |

## What are the difference between Merge and Union All transformations in SSIS?

1. **Merge** can only accept two datasets while **Union All** can accept more than two datasets for input.
2. **Merge** requires both datasets to be sorted while **Union All** does not require sorted datasets.

## **Transformations: -**

**Sort: -**

The sort transformation used to sort the source data in either Ascending or Descending order, which is similar to T-SQL command ORDER BY statement. Some transformations like Merge and Merge Join Transformation needs data to be sorted using them. In these situations we use sort transformation to sort the data.

**Note:** The sort transformation is a full blocking operation because it needs to read all the rows in the data flow (buffer) in order to sort the data. No data will be sent downstream until all rows have been read first by the sort transformation.