

SQL Server Integration Service (SSIS) is a component of the Microsoft SQL Server database software that can be used to execute a wide range of data migration tasks. SSIS is a fast & flexible data warehousing tool used for data extraction, loading and transformation like cleaning, aggregating, merging data, etc.

It makes it easy to move data from one database to another database. SSIS can extract data from a wide variety of sources like SQL Server databases, Excel files, Oracle and DB2 databases, etc.

ETL: Extract, transform ,load

Following are components of SSIS architecture:

- Control Flow (Stores containers and Tasks)
- Data Flow (Source, Destination, Transformations)
- Event Handler (sending of messages, Emails)
- Package Explorer (Offers a single view for all in package)
- Parameters (User Interaction)

1. Control Flow

Control flow is a brain of SSIS package. It helps you to arrange the order of execution for all its components. The components contain containers and tasks which are managed by precedence constraints.

Precedence Constraints

Precedence constrain are package component which direct tasks to execute in a predefined order. It also defines the workflow of the entire SSIS package. It controls the execution of the two linked tasks by executing the destination tasks based on the result of the earlier task — business rules which are defined using special expressions.

Task

A 'Task' is an individual unit of work. It is the same as a method/function used in a programming language. However, in SSIS, you don't use coding methods. Instead, you will use drag & drop technique to design surface and to configure them.

Containers

The container is units for grouping tasks together into units of work. Apart from offering visual consistency, it also allows you to declare variables and event handlers which should be in the scope of that specific container.

Three types of containers in SSIS are:

- A Sequence Container
- A For Loop Container
- Foreach Loop Container

Sequence Container: allows you to organize subsidiary tasks by grouping them, and allows you to you apply transactions or assign logging to the container.

For loop container: Provides the same functionality as the sequence Container except that it also lets you run the tasks multiple times. However, it is based on an evaluation condition, like a looping from 1 to 100.

For each Loop Container: It also allows looping. But the difference that instead of using a condition expression, loop s done over a set of objects, likes files in a folder.

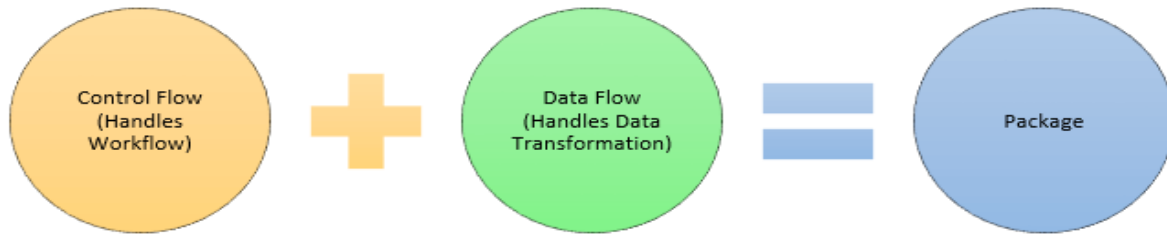
2. Data Flow

The main use of the SSIS tool is to extract data into the server's memory, transform it, and write it to another destination. If Control Flow is the brain, Data Flow is the heart of SSIS

SSIS Packages

Another core component of SSIS is the notion of a package. It is a collection of tasks which execute in an orderly fashion. Here, president constraints help manage the order in which the task will execute.

A package can help you to saves files onto a SQL Server, in the msdb or package catalog database. It can save as a .dtsx file, which is a structured file very similar to .rdl files are to Reporting Services.



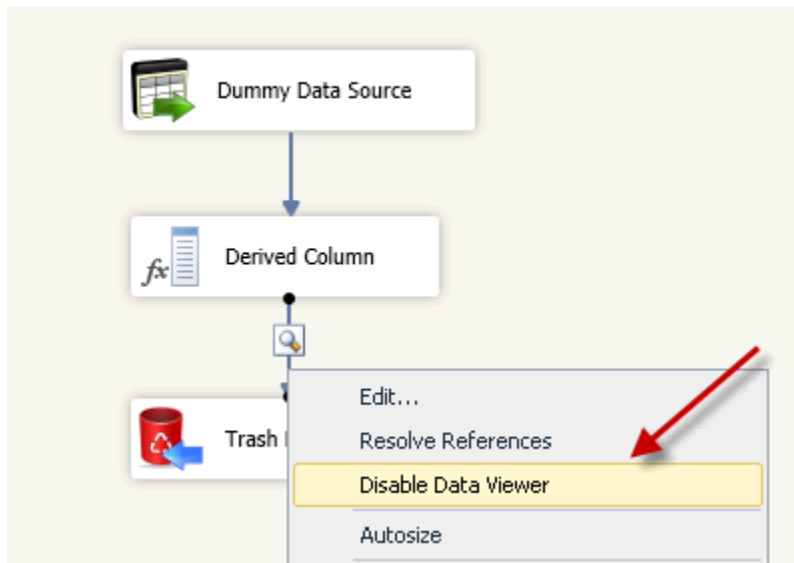
SSIS Connection Managers

The SSIS connection manager contains the physical connection string properties that will create or establish a connection at a run time. SSIS allows us to develop both the project level and package level connection managers. The SSIS project level connection managers shared among all the packages in that project. It is available at the SSIS project deployment. This section of SSIS connection manager connections between this SQL SSIS and different data sources.

- OLE DB Connection Manager
 - ADO Connection Manager
 - ADO.NET Connection Manager
 - Cache Connection Manager
 - EXCEL Connection Manager
 - File Connection Manager
 - FTP Connection Manager
 - SMO Connection Manager
-
- **OLE DB Source in SSIS** use OLE DB Connection Manager to connect and extract data from the Database in **SSIS**
 - **OLE DB Destination** uses the OLE DB Connection Manager to connect and write data into the Database
 - **Term Lookup Transformation** only supports OLE DB Connection Manager to connect with the term lookup table (or Reference table).
 - **Term Extraction Transformation** only supports the SSIS OLE DB Connection Manager to add the Exclusion List.
 - **Execute SQL task** use OLE DB Connection Manager to connect and extract data from the Database
 - The OLE DB Connection Manager is also helpful in accessing data from the unmanaged code.

Data Viewer:

SSIS Data Viewer to debug/troubleshoot SSIS DataFlow Pipeline issues at runtime. You can use SSIS Data Viewer any time you want to see what your data look like at given point in pipe line.



SSIS Data Viewer – Debug Pipeline

Data Conversion:

Data Conversion is a component of SSIS to convert data-type from one type to another type. Data Conversion means conversion of data-type of a column to another data-type before it reaches the destination of data-warehouse. If we want to convert any data-type of data which comes from source then Data Conversion is the best for such scenarios before loading it to destination.

Why we need Data Conversion:

Data comes from various sources in different formats, we will have different types of sources like csv, excel, notepad, sql all the formats we first take in ETL process i.e., Extracts, Transform and Loading. So, when we extracts this data it is obvious that the data types used will not be the same, we may come to know that some data is in wrong format. Let's say a numeric data is shown as string type. This happens in SSIS many times during ETL process when we extracts data from txt file or CSV file.

So to correct this problem in SSIS we need to do Data-Conversion so that data loads properly to destination data-warehouse in the proper format.

To understand more better let's do an example on the same. Here we will under our data-source as CSV file from where we will extract a data then before loading it Data -Warehouse i.e. SQL Server we will do the Data Conversion.

Conversion chart of SSIS data types to SQL Server data types

SSIS Data Type	SSIS Expression	SQL Server
single-byte signed integer	(DT_I1)	
two-byte signed integer	(DT_I2)	smallint
four-byte signed integer	(DT_I4)	int
four-byte signed integer	(DT_I4)	int
eight-byte signed integer	(DT_I8)	bigint
eight-byte signed integer	(DT_I8)	bigint
float	(DT_R4)	real
double-precision float	(DT_R8)	float
string	(DT_STR, «length», «code_page»)	char, varchar
Unicode text stream	(DT_WSTR, «length»)	nchar, nvarchar, sql_variant, xml
date	(DT_DATE)	date
Boolean	(DT_BOOL)	bit
numeric		decimal, numeric

	(DT_NUMERIC, «precision», «scale»)	
decimal	(DT_DECIMAL, «scale»)	decimal
currency	(DT_CY)	smallmoney, money
unique identifier	(DT_GUID)	uniqueidentifier
byte stream	(DT_BYTES, «length»)	binary, varbinary, timestamp
database date	(DT_DBDATE)	date
database time	(DT_DBTIME)	
database time with precision	(DT_DBTIME2, «scale»)	time(p)
database timestamp	(DT_DBTIMESTAMP)	datetime, smalldatetime
database timestamp with precision	(DT_DBTIMESTAMP2, «scale»)	datetime2
database timestamp with timezone	(DT_DBTIMESTAMPOFFSET, «scale»)	datetimeoffset(p)
file timestamp	(DT_FILETIME)	
image	(DT_IMAGE)	image
text stream	(DT_TEXT, «code_page»)	text
Unicode string	(DT_NTEXT)	ntext

Derived Column SSIS:

SSIS Derived Column Transformation is one of the SQL Server Integration Services that can be added within a Data Flow Task, it is **used to add a new column to the data pipeline by applying** SSIS expressions. The developer can choose whether to add a new derived column or to replace an existing column.

Aggregate:

The Aggregate Transformation in SSIS performs two main tasks:

1. Applying Aggregate Operations on your input data
2. It allows you to group your data based on values in your input columns.

Aggregate Transformation in SSIS performs a function similar to the [SQL Server Aggregate functions](#). The operations include the following:

1. GROUP BY: Just like the [GROUP BY Clause](#) in SQL [SELECT](#) queries.
2. COUNT: It will [Count](#) the number of values in this column. Null values are included in the count if you select (*) as the input column. Otherwise, null values ignored.
3. COUNT DISTINCT: It will count the number of [Distinct](#) values in this column.
4. SUM: Calculate the [Sum](#) of the Column values.
5. AVERAGE: Calculate the [Average](#) of the Column values.
6. MINIMUM: Calculate the [Minimum](#) of Column values.
7. MAXIMUM: Calculate the [Maximum](#) of Column values.

The Aggregate transformation in [SSIS](#) can configure in Basic mode or Advanced mode. In Basic mode, the Aggregate transformation has one output. And in Advanced mode, it can have more than one output, each containing the result of a different aggregation.

Multi-Cast:

Multicast Transformation in SSIS sends input data to multiple destination paths without applying any conditions or transformations. OR, Takes ONE Input and makes the logical COPY of data and passes the same data to multiple outputs.

SSIS Multicast Transformation is very useful when we want to apply different transformations on the same dataset in parallel. For instance, If we're going to perform two types of transformations on the same dataset. Then use SSIS Multicast Transformation to make two copies of the same data. Next, on each copy, we can perform different transformations.

Conditional Split:

A Conditional Split Transformation in SSIS is just like the IF condition or CASE statement. It checks the given condition and based on the condition result; the output will send to the appropriate destination path. It has ONE input and MANY outputs.

Error handling:

SSIS Error Handling: In SQL Server Integration Services or SSIS, errors might occur when we are extracting data from a source, or loading data into a destination, or when you are performing a transformation. It is because of the unexpected data, and it is a bad practice to fail the package. To handle these situations, SSIS introduced the concept called Error Handling.

Failed Component

If the data flow task fails, it fails the package. It is the default option.

Ignore failure

If we change the error mode to ignore failure, it ignores the error message and completes the execution. For example, in our case, it will ignore the rows with bad data and continue working with other rows. It does not throw any error message. We should not use it in the production environment.

Redirect rows

It is a useful configuration for this article. We can redirect the failed rows to a configured destination and view the bad data rows in that file. We can configure

it for the truncation or the error message.

Specify how row-level errors are handled by this component. You can handle errors in the row, or truncation errors in columns. Errors can fail the component, or they can be ignored, or they can be redirected to an error output.

Input or Output	Column	Error	Truncation	Description
Data Conversion				
	Copy of Order...	Fail component	Fail component	Conversion
	Copy of Region	Fail component	Fail component	Conversion
	Copy of Name	Fail component	Fail component	Conversion
	Copy of Item	Fail component	Fail component	Conversion
	Copy of Units	Fail component	Fail component	Conversion
	Copy of Unit ...	Fail component	Fail component	Conversion
	Copy of Total ...	Fail component	Fail component	Conversion

Set this value to selected cells: Fail component Apply

OK Cancel

We can use all the columns list and error configurations for error handling in SSIS package. We have following configuration available.

Row Sampling Transformation in SSIS

Row Sampling Transformation in SSIS provides an option to specify the number of rows you would like to retrieve from the data source. SSIS Row Sampling Transformation Data is randomly selected unless you use this Use the following random seed option.

Row Sampling Transformation in SSIS will take the whole data from a source, and it randomly extracts the selected number of rows. SSIS Row Sampling gives you two outputs. The first is the selected data, and the second one is unselected data. This SSIS Row Sampling transformation is useful in training the data mining model.

Percentage Sampling Transformation in SSIS

In general, the Percentage sampling transformation in SSIS is for data modeling. This SSIS Percentage Sampling transformation selects a given percentage of rows from the data source. Next, it transfers them to selected output rows option, and the remaining rows will move to unselected output rows.

For example, if the input has 1000 rows, and if we specify ten as a percentage sample, then the SSIS Percentage Sampling transformation returns approximately 10% of the RANDOM records from the input data.

Sort Transformation in SSIS

The Sort Transformation in SSIS is used to sort the source data in either Ascending or Descending order, which is similar to the T-SQL command ORDER BY statement.

- If the Sort Order value is a positive number, then Sort Transformation sort the data in Ascending order
- If the Sort Order value is Negative number, SSIS Sort Transformation sort the data in descending order

Some transformations like [Merge Transformation](#) and [Merge Join Transformation](#) needs data to sort before using them. In these situations, we use SSIS Sort Transformation to sort the data.

Excel Source in SSIS:

The EXCEL Source in SSIS (SQL Server Integration Services) is used to extract data from the Excel worksheet. The SSIS Excel Source uses Excel Connection Manager in SSIS to connect with the Excel files.

Union All Transformation in SSIS

Union All Transformation in SSIS is used to combine data from multiple sources (excel files, flat files, etc.). Or multiple SQL tables and produce one output to store in the destination table. Union All Transformation in SSIS does not follow any particular order while merging the data and storing in the destination table

Audit Transformation in SSIS

Audit Transformation in SSIS allows you to add extra columns with auditing information (Machine name, package name, id, date, etc.) required in auditing. We can use this SSIS audit transformation to save that information in the target table.

1. ExecutionInstanceGUID – The GUID that identifies the execution instance of the package.
2. PackageID – This is the unique identifier of the package.
3. PackageName – Shows the name of the package.
4. VersionID – The unique version number of the package.
5. ExecutionStartTime – The time the package started to run.
6. MachineName – The Name of the computer.
7. UserName – The login name of the person who started the package.
8. TaskName – Name of the Data Flow task with which the Audit transformation is associated.
9. TaskId – The unique identifier of the Data Flow task.

SSIS Script Component as Transformation

The Script Component is one of the most important and powerful items in SQL Server Integration Services. We can use this SSIS script component as a Source, Destination, and Transformation.

we can use this SSIS Script Component as a Transformation to combine the work of multiple transformations in one place.