**SQL Server Integration Services (SSIS) Architecture (material taken from [1])**

In SSIS, there are some important concepts that we need to discuss in advance to proceed further

1. **Packages**

Packages are one of the core components of SSIS. They are just like executable. In a typical SSIS project, you will develop packages to execute them or you may store them to later deploy it with SQL Server for scheduled ETL tasks.

In a typical package, you basically define or design tasks to be executed in an order or sequence. The packages are stored with a .DTSX extension which is a typical XML file.

1. **Control Flow**

The control flow is the brain of SSIS. Basically it orchestrates the order of execution of components. The components could be tasks or containers.

**2.1 Tasks:**

Tasks are basically individual unit of works. They are similar to the functions in a programming language. However, in SSIS package you just drag, drop and configure tasks. You can also develop your own tasks but many of them have already been provided to you. For example:

Bulk Insert Task: Loads data into a table by using the BULK INSERT SQL command.

CDC Control Task: Maintains and interacts with the change data capture (CDC) feature from SQL Server.

Data Flow Task: This very specialized task loads and transforms data into an OLE DB and ADO.NET destination

Execute Package Task: Allows you to execute a package from within a package, making your SSIS packages modular.

Execute Process Task: Executes a program external to your package, such as one to split your extract file into many files before processing the individual files.

Execute SQL Task: Executes a SQL statement or stored procedure.

Expression Task: Sets a variable to an expression at runtime.

File System Task: This task can handle directory operations such as creating, renaming, or deleting a directory. It can also manage file operations such as moving, copying, or deleting files.

FTP Task: Sends or receives files from an FTP site

Script Task: This task enables you to perform .NET-based scripting in the Visual Studio Tools for Applications programming environment.

Send Mail Task: Sends a mail message through SMTP

XML Task: Parses or processes an XML file. It can merge, split, or reformat an XML file.

Backup Database Task: To take backup of SQL Server database

Shrink Database Task: To shrink the database

Transfer Database Task: To transfer database to other destinations

**2.2 Precedence Constraints:**

The precedence constraints are basically connectors that not only connect tasks but also help in defining the workflow of your tasks. By using precedence constraints, we can control the execution of destination tasks on the basis of the state of the prior task. The SSIS expression language is used to enforce precedence constraints.

**2.3 Containers:**

Containers are also core to the SSIS packages. They are used to logically organize tasks together to work as unit. There are four types of containers

Task Host Container: Not a visible element that you’ll find in the Toolbox, but rather an abstract concept like an interface.

Sequence Container: Allows you to group tasks into logical subject areas. Within the development environment, you can then collapse or expand this container for usability.

For Loop Container: Loops through a series of tasks until a condition is met.

Foreach Loop Container: Loops through a series of files or records in a data set, and then executes the tasks in the container for each record in the collection.

1. **Data Flow**

The data flow is the heart of SSIS. Basically it helps in designing data flows from sources to destinations. While transferring data, you can also use different transformation to transform your data in different forms and formats.

The data flow is supported by connection managers to connect with sources and destinations.

**3.1 Sources**

Sources are the data sources which you want to fetch into destinations. There are different types of sources supported by SSIS

OLE DB Source: Connects to nearly any OLE DB data source, such as SQL Server, Access, Oracle, or DB2, to name just a few.

Excel Source: Specializes in receiving data from Excel spreadsheets. This source also makes it easy to run SQL queries against your Excel spreadsheet to narrow the scope of the data that you wish to pass through the flow.

Flat File Source: Connects to a delimited or fixed-width file.

Raw File Source: Produces a specialized binary file format for data that is in transit; it is especially quick to read by SSIS. This component is one of the only components that does not use a Connection Manager.

Xml Source: Retrieves data from an XML document. This source does not use a Connection Manager to configure it.

ADO.NET Source: This source is just like the OLE DB Source but only for ADO.NET based sources. The internal implementation uses an ADO.NET Data Reader as the source. The ADO.NET connection is much like the one you see in the .NET Framework when hand-coding a connection and retrieval from a database.

CDC Source: Reads data out of a table that has change data capture (CDC) enabled. Used to retrieve only rows that have changed over a duration of time.

ODBC Source: Reads data out of table by using an ODBC provider instead of OLE DB. When you are given the choice between OLE DB and ODBC, it is still recommended in SSIS packages that you use OLE DB.

**3.2 Transformations**

Transformations are used to transform data from one form/format to another form/format. Typical transformations include:

Character Map: Makes common string data changes for you, such as changing data from lowercase to uppercase.

Conditional Split: Splits the data based on certain conditions being met. For example, this transformation could be instructed to send data down a different path if the State column is equal to Florida.

Copy Column: Adds a copy of a column to the transformation output. You can later transform the copy, keeping the original for auditing purposes.

Data Conversion: Converts a column’s data type to another data type.

Derived Column: Creates a new derived column calculated from an expression.

Fuzzy Grouping: Performs data cleansing by finding rows that are likely duplicates.

Fuzzy Lookup: Matches and standardizes data based on fuzzy logic. For example, this can transform the name Jon to John.

Merge: Merges two sorted data sets into a single data set in a Data Flow.

Merge Join: Merges two data sets into a single data set using a join function.

Multicast: Sends a copy of the data to an additional path in the workflow

**3.3 Destinations**

Destinations are basically consumer of data where data after transformations are stored. SSIS can send the data to nearly any OLE DB– compliant, flat file, or ADO.NET data source. All destinations are managed by connection manager. Typical destinations are:

ADO.NET Destination: Exposes data to other external processes, such as a .NET application.

Dimension Processing: Loads and processes an Analysis Services dimension. It can perform a full, update, or incremental refresh of the dimension.

Excel Destination: Outputs data from the Data Flow to an Excel spreadsheet.

Flat File Destination: Enables you to write data to a comma-delimited or fixed-width file.

ODBC Destination: Outputs data to an ODBC data connection like SQL Server, DB2, or Oracle.

OLE DB Destination: Outputs data to an OLE DB data connection like SQL Server, Oracle, or Access.

SQL Server Destination: The destination that you use to write data to SQL Server.

1. **Variables**

Variables are important part of SSIS and they can be used to be evaluated in scripts and expressions to make the package dynamic

1. **Parameters**

Parameters are same like variables that make your package dynamic but they are read-only. They need to be passed to the package for its execution

1. **Error Handling and Logging**

By using the event handler design surface, your can design a sequence of workflow to be performed when an even occurs. For example, for an error in the execution of package or for an onError event you can define tasks to notify operators.

SSIS keeps logs of errors by default to support troubleshooting. You can configure logging of your package by selecting and deselecting the types of logs you want to monitor.

References:

[1] Knight, B., Knight, D., Moss, J. M., Davis, M., and Rock, C.: “Professional Microsoft SQL Server 2014 Integration Services”. Publisher: Wrox