|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Difference between control flow and data flow**  Control flow deals with orderly processing of individual, isolated tasks, these tasks are linked through precedence constraints in random order. Also the output for task has finite outcome i.e., Success, Failure, or Completion. A subsequent task does not initiate unless its predecessor has completed. Data flow, on the other hand, streams the data in pipeline manner from its source to a destination and modifying it in between by applying transformations. Another distinction between them is the absence of a mechanism that would allow direct transfer of data between individual control flow tasks. On the other hand, data flow lacks nesting capabilities provided by containers.   |  |  |  | | --- | --- | --- | |  | **Control Flow** | **Data Flow** | |  | Process Oriented | Data Oriented | | **Made up of** | Tasks and Container | Source, Transformation and Destination | | **Connected through** | Precedence constraint | Paths | | **Smallest unit** | Task | Component | | **Outcome** | Finite- Success, Failure, Completion | Not fixed | |
| **If you want to send some data from Access database to SQL server database. What are different component of SSIS will you use?**  In the data flow, we will use one OLE DB source, data conversion transformation and one OLE DB destination or SQL server destination. OLE DB source is data source is useful for reading data from Oracle, SQL Server and Access databases. Data Conversion transformation would be needed to remove datatype abnormality since there is difference in datatype between the two databases (Access and SQL Server) mentioned. If our database server is stored on and package is run from same machine, we can use SQL Server destination otherwise we need to use OLE DB destination. The SQL Server destination is the destination that optimizes the SQL Server. |
| **Difference and similarity between merge and merge join transformation**   |  |  |  | | --- | --- | --- | |  | **Merge Transofrmations** | **Merge Join Transformation** | |  | The data from 2 input paths are merged into one | The data from 2 inputs are merged based on some common key. | | **Works as** | UNION ALL | JOIN (LEFT, RIGHT OR FULL) | | **Supports** | 2 Datasets | 1 Dataset | | **Columns** | Metadata for all columns needs to be same | Key columns metadata needs to be same. | | **Pre-requisites** | Data must be sorted.  Merged columns should have same datatype i.e. if merged column is EmployeeName with string of 25 character in Input 1, it can be of less than or equal to 25 characters for merging to happen. | Data must be sorted.  Merged columns should have same datatype i.e. if merged column is EmployeeName with string of 25 character in Input 1, it can be of less than or equal to 25 characters for merging to happen. | | **Limitations** | Only 2 input paths can be merged.  Does not support error handling. | Does not support error handling. | | **Use** | Merging of data from 2 data source  Can create complex datasets using nesting merge transformation, | When data from 2 tables having foreign key relationship needs to present based on common key. | |
| **What is precedence constraint?**  A precedence constraint is a link between 2 control flow tasks and lays down the condition on which the second task is run. They are used to control the workflow of the package. There are 3 kinds of precedence constraint – success (green arrow), failure (red arrow) or Completion script task (blue arrow). By default, when we add 2 tasks, it links by green arrow. The way the precedence constraint is evaluated can be based on outcome of the initial task. Also, we can add expression to evaluate such outcome. Any expression that can be judged as true or false can be used for such purpose. The precedence constraint is very useful in error handling in SSIS package. |
| **Explain why variables called the most powerful component of SSIS.**  Variable allows us to dynamically control the package at runtime. Example: You have some custom code or script that determines the query parameter’s value. Now, we cannot have fixed value for query parameter. In such scenarios, we can use variables and refer the variable to query parameter. We can use variables for like: 1. updating the properties at runtime,  2. populating the query parameter value at runtime,  3. used in script task,  4. Error handling logic and  5. With various looping logic. |
| **Can we add our custom code in SSIS?**  We can customize SSIS through code by using Script Task. The main purpose of this task is to control the flow of the package. This is very useful in the scenario where the functionality you want to implement is not available in existing control flow item.  **To add your own code:-** 1. In control flow tab, drag and drop Script Task from toolbox.  2. Double click on script task to open and select edit to open Script task editor.  3. In script task editor, there are 3 main properties  i.) General – Here you can specify name and description  ii.) Script – through this we can add our code by clicking on Design Script button. The scripting language present is VB.Net only.  iii.) Expression |
| **What is conditional split?**  As the name suggest, this transformation splits the data based on condition and route them to different path. The logic for this transformation is based on CASE statement. The condition for this transformation is an expression. This transformation also provides us with default output, where rows matching no condition are routed. Conditional split is useful in scenarios like Telecom industry data you want to divide the customer data on gender, condition would be:  GENDER == ‘F’ |
| **Explain the use of containers in SSIS and also their types.**  Containers can be defined as objects that stores one or more tasks. The primary purpose of container is grouping logically related tasks. Once the task is placed into the containers, we can perform various operations such as looping on container level until the desired criterion is met. Nesting of container is allowed. Container is placed inside the control flow.   **There are 4 types of Container:-** 1. Task Host container- Only one task is placed inside the container. This is default container.  2. Sequence Container – This container can be defined as subset of package control flow.  3. For loop container – Allows looping based on condition. Runs a control flow till condition is met.  4. For each loop container - Loop through container based on enumerator. |
| **Why is the need for data conversion transformations?**  This transformation converts the datatype of input columns to different datatype and then route the data to output columns.   **This transformation can be used to:** 1. Change the datatype  2. If datatype is string then for setting the column length  3. If datatype is numeric then for setting decimal precision.  This data conversion transformation is very useful where you want to merge the data from different source into one. This transformation can remove the abnormality of the data. Example à The Company’s offices are located at different part of world. Each office has separate attendance tracking system in place. Some offices stores data in Access database, some in Oracle and some in SQL Server. Now you want to take data from all the offices and merged into one system. Since the datatypes in all these databases vary, it would be difficult to perform merge directly. Using this transformation, we can normalize them into single datatype and perform merge. |
| **Error Handling in SSIS?**  An error handler allows us to create flows to handle errors in the package in quite an easy way. Through event handler tab, we can name the event on which we want to handle errors and the task that needs to be performed when such an error arises. We can also add sending mail functionality in event of any error through SMTP Task in Event handler. This is quite useful in event of any failure in office non-working hours. In Data flow, we can handle errors for each connection through following failure path or red arrow. |

**Name a few SSIS components?**

**Ans**:

* Integration Services Projects
* Integration Services Packages
* Control Flow Elements
* Data Flow Elements
* Integration Services Connections
* Integration Services Variables
* Integration Services Event Handlers
* Integration Services Log Providers

**What is a project and Package in SSIS?**

**Ans:**

Project is a container for developing packages. Package is nothing but an object. It implements the functionality of ETL — Extract, Transform and Load — data.

**What are the 4 elements (tabs) that you see on a default package designer in BIDS?**

**Ans:**

Control Flow, Data Flow, event Handler and package explorer. (Parameters – 2012 Data Tools)

**What is a Control flow and Data Flow elements in SSIS?**

**Ans:**

**Control Flow:**

Control flow element is one that performs any function or provides structure or control the flow of the elements. There must be at least one control flow element in the SSIS package. In SSIS a workflow is called a control-flow. A control-flow links together our modular data-flows as a series of operations in order to achieve a desired result.

A control flow consists of one or more tasks and containers that execute when the package runs. To control order or define the conditions for running the next task or container in the package control flow

**Data Flow:**

All ETL tasks related to data are done by data flow elements. It is not necessary to have a data flow element in the SSIS package. A data flow consists of the sources and destinations that extract and load data, the transformations that modify and extend data, and the paths that link sources, transformations, and destinations. Before you can add a data flow to a package, the package control flow must include a Data Flow task. The Data Flow task is the executable within the SSIS package that creates, orders, and runs the data flow. A separate instance of the data flow engine is opened for each Data Flow task in a package.

**What are the 3 different types of control flow elements in SSIS?**

**Ans:**

* Structures provided by Containers
* Functionality provided by Tasks
* Precedence constraints that connect the executables, containers, and tasks into an ordered control flow.

**What are the 3 data flow components in SSIS?**

**Ans:**

* Source
* Transformation
* Destination

**What are connections and connection managers in SSIS?**

**Ans:**

Connection as its name suggests is a component to connect to any source or destination from SSIS — like a sql server or flat file or lot of other options that SSIS provides. Connection manager is a logical representation of a connection.

**What is the use of Check Points in SSIS?**

**Ans:**

SSIS provides a Checkpoint capability which allows a package to restart at the point of failure.

**What are the command line tools to execute SQL Server Integration Services packages?**

**Ans:**

**DTSEXECUI –** When this command line tool is run a user interface is loaded in order to configure each of the applicable parameters to execute an SSIS package.

**DTEXEC –** This is a pure command line tool where all of the needed switches must be passed into the command for successful execution of the SSIS package.

**Can you explain the SQL Server Integration Services functionality in Management Studio?**

**Ans:**

You have the ability to do the following:

* Login to the SQL Server Integration Services instance
* View the SSIS log
* View the packages that are currently running on that instance
* Browse the packages stored in MSDB or the file system
* Import or export packages
* Delete packages
* Run packages

**Can you name some of the core SSIS components in the Business Intelligence Development Studio you work with on a regular** **basis** **when** **building an SSIS package?**

**Ans:**

* Connection Managers
* Control Flow
* Data Flow
* Event Handlers
* Variables window
* Toolbox window
* Output window
* Logging
* Package Configurations

**Name Transformations available in SSIS?**

**Ans:**

**DATACONVERSION:** Converts columns data types from one to another type. It stands for Explicit Column Conversion.

**DATAMININGQUERY:** Used to perform data mining query against analysis services and manage Predictions Graphs and Controls.

**DERIVEDCOLUMN:** Create a new (computed) column from given expressions.

**EXPORTCOLUMN:** Used to export a Image specific column from the database to a flat file.

**FUZZYGROUPING:** Used for data cleansing by finding rows that are likely duplicates.

**FUZZYLOOKUP:** Used for Pattern Matching and Ranking based on fuzzy logic.

**AGGREGATE:** It applies aggregate functions to Record Sets to produce new output records from aggregated values.

**AUDIT:** Adds Package and Task level Metadata: such as Machine Name, Execution Instance, Package Name, Package ID, etc..

**CHARACTERMAP:** Performs SQL Server column level string operations such as changing data from lower case to upper case.

**MULTICAST:** Sends a copy of supplied Data Source onto multiple Destinations.

**CONDITIONALSPLIT:** Separates available input into separate output pipelines based on Boolean Expressions configured for each output.

**COPYCOLUMN:** Add a copy of column to the output we can later transform the copy keeping the original for auditing.

**IMPORTCOLUMN:** Reads image specific column from database onto a flat file.

**LOOKUP:** Performs the lookup (searching) of a given reference object set to a data source. It is used for exact matches only.

**MERGE:** Merges two sorted data sets into a single data set into a single data flow.

**MERGEJOIN:** Merges two data sets into a single dataset using a join junction.

**ROWCOUNT:** Stores the resulting row count from the data flow / transformation into a variable.

**ROWSAMPLING:** Captures sample data by using a row count of the total rows in dataflow specified by rows or percentage.

**UNIONALL:** Merge multiple data sets into a single dataset.

**PIVOT:** Used for Normalization of data sources to reduce anomalies by converting rows into columns

**UNPIVOT:** Used for de-normalizing the data structure by converts columns into rows in case of building Data Warehouses.

**List out different types of connection or files that support SSIS?**

Different types of connection that work within SSIS are

* ODBC
* OLEDB
* .net SQLClient
* Flat File
* Excel
* XML

**Explain what is a container? How many types of containers are there in SSIS?**

In SSIS, a container is a logical grouping of tasks, and it allows to manage the scope of a task together.

Types of containers in SSIS are

* Sequence container
* For loop container
* Foreach loop container
* Task host container

**Explain what is Precedence Constraint in SSIS?**

Precedence Constraint in SSIS enables you to define the logical sequence of tasks in the order they should be executed.  You can connect all the tasks using connectors- Precedence Constraints.

**Explain what variables in SSIS and what are the types of variables in SSIS?**

Variable in SSIS is basically used to store values.  In SSIS, there are two types of variables**system variable** and **user variable**.

**Explain what is a checkpoint in SSIS?**

Checkpoint in SSIS allows the project to restart from the point of failure. Checkpoint file stores the information about the package execution, if the package run successfully the checkpoint file is deleted or else it will restart from the point of failure.

**Explain what is connection managers in SSIS?**

While gathering data from different sources and writing it to a destination, connection managers are helpful.  Connection manager facilitates the connection to the system that include information’s like data provider information, server name, authentication mechanism, database name, etc.

**Explain what is SSIS breakpoint?**

A breakpoint enables you to pause the execution of the package in business intelligence development studio during troubleshooting or development of an SSIS package.

**Explain what is event logging in SSIS?**

In SSIS, event logging allows you to select any specific event of a task or a package to be logged. It is very helpful when you are troubleshooting your package to understand the performance package.

**Explain what is logging mode property?**

SSIS packages and all the associated tasks have a property called LoggingMode.   This property accepts three possible values

* Disabled: To enable logging of the component
* Enabled: To disable logging of the component
* UseParentSetting: To use parent’s setting of the component

**Explain what is a data flow buffer?**

SSIS operates using buffers; it is a kind of an in-memory virtual table to hold data.

**For what data checkpoint data is not saved?**

Checkpoint data is not saved for ForEach Loop and ForLoop containers.

**Explain what is conditional split transactions in SSIS?**

Conditional split transformation in SSIS is just like IF condition, which checks for the given condition based on the condition evaluation.

**List out the different types of Data viewers in SSIS?**

Different types of data viewers in SSIS include

* Grid
* Histogram
* Scatter Plot
* Column Chart

**Mention what are the possible locations to save SSIS package?**

You can save SSIS package at

* SQL Server
* Package Store
* File System

**What will be your first approach if the package that runs fine in Business Intelligence Development Studio (BIDS) but fails when running from an SQL agent job?**

The account that runs SQL Agent Jobs might not have the required permission for one of the connections in your package. In such cases, either you can create a proxy account or elevate the account permissions.

**Explain what is the role of Event Handlers tab in SSIS?**

On the event handlers tab, workflows can be configured to respond to package events.  For instance, you can configure workflow when any task stops, fails or starts.

**Explain how you can notify the staff members about package failure?**

Either inside the package you could add a Send Mail Task in the event handlers, or you can even set notification in the SQL Agent when the package runs.

**Explain how would you do logging in SSIS?**

Logging in SSIS can be done by logging various events like onError, onWarning, etc. to the various options like a flat file, XML, SQL server table, etc.

**Mention how would you deploy an SSIS package on production?**

To deploy SSIS package we need to execute the manifest files and need to determine whether to deploy this into File System or onto SQL Server.  Alternatively you can also import package from SSMS from SQL Server or File System.

**Explain how to handle Early Arriving Facts o**r **Late Arriving Dimension?**

Late Arriving Dimension are unavoidable, to handle these we can create a dummy dimensions with natural/business key and keep the rest of the attributes as null or default. So when actual dimension arrives, the dummy dimension is updated with Type 1 change. This is also referred as Inferred Dimensions.

**Explain how can you do an incremental load?**

The best and fastest way to do incremental load is by using Timestamp column in the source table and storing the last ETL timestamp.