**What do we mean by dataflow in SSIS?**  
Data flow is nothing but the flow of data from the corresponding sources to the referred destinations. In this process, the data transformations make changes to the data to make it ready for the data warehouse.

**• What is a breakpoint in SSIS? How is it setup? How do you disable it?**  
A breakpoint is a stopping point in the code. The breakpoint can give the Developer\DBA an  
opportunity to review the status of the data, variables and the overall status of the SSIS package.  
10 unique conditions exist for each breakpoint.  
Breakpoints are setup in BIDS. In BIDS, navigate to the control flow interface. Right click on the  
object where you want to set the breakpoint and select the ‘Edit Breakpoints…’ option.

**• Can you name 5 or more of the native SSIS connection managers?**

1) OLEDB connection – Used to connect to any data source requiring an OLEDB connection (i.e.,  
SQL Server 2000)  
2) Flat file connection – Used to make a connection to a single file in the File System. Required for reading information from a File System flat file  
3) ADO.Net connection – Uses the .Net Provider to make a connection to SQL Server 2005 or other  
connection exposed through managed code (like C#) in a custom task  
4) Analysis Services connection – Used to make a connection to an Analysis Services database or project. Required for the Analysis Services DDL Task and Analysis Services Processing Task  
5) File connection – Used to reference a file or folder. The options are to either use or create a file or folder  
6) Excel

**• What is the use of Bulk Insert Task in SSIS?**  
Bulk Insert Task is used to upload large amount of data from flat files into Sql Server. It supports only OLE DB connections for destination database.

**• What is Conditional Split transformation in SSIS?**  
This is just like IF condition which checks for the given condition and based on the condition evaluation, the output will be sent to the appropriate OUTPUT path. It has ONE input and MANY outputs. Conditional Split transformation is used to send paths to different outputs based on some conditions. For example, we can organize the transform for the students in a class who have marks greater than 40 to one path and the students who score less than 40 to another path.

**• How do you eliminate quotes from being uploaded from a flat file to SQL Server?**  
This can be done using TEXT QUALIFIER property. In the SSIS package on the Flat File Connection Manager Editor, enter quotes into the Text qualifier field then preview the data to ensure the quotes are not included.

**• Can you explain how to setup a checkpoint file in SSIS?**  
The following items need to be configured on the properties tab for SSIS package:  
**CheckpointFileName** – Specify the full path to the Checkpoint file that the package uses to save the value of package variables and log completed tasks. Rather than using a hard-coded path as shown above, it’s a good idea to use an expression that concatenates a path defined in a package variable and the package name.  
**CheckpointUsage** – Determines if/how checkpoints are used. Choose from these options: Never(default), IfExists, or Always. Never indicates that you are not using Checkpoints. IfExists is the typical setting and implements the restart at the point of failure behavior. If a Checkpoint file is found it is used to restore package variable values and restart at the point of failure. If a Checkpoint file is not found the package starts execution with the first task. The Always choice raises an error if the Checkpoint file does not exist.  
**SaveCheckpoints** – Choose from these options: True or False (default). You must select True to implement the Checkpoint behavior.

**• What are the different values you can set for CheckpointUsage property ?**  
There are three values, which describe how a checkpoint file is used during package execution:  
1) **Never**: The package will not use a checkpoint file and therefore will never restart.  
2) **If Exists**: If a checkpoint file exists in the place you specified for the CheckpointFilename property, then it will be used, and the package will restart according to the checkpoints written.  
3) **Always**: The package will always use a checkpoint file to restart, and if one does not exist, the package will fail.

**• What is the ONLY Property you need to set on TASKS in order to configure CHECKPOINTS to RESTART package from failure?**  
The one property you have to set on the task is **FailPackageOnFailure**. This must be set for each task or container that you want to be the point for a checkpoint and restart. If you do not set this property to true and the task fails, no file will be written, and the next time you invoke the package, it will start from the beginning again.

**• Where can we set the CHECKPOINTS, in DataFlow or ControlFlow ?**  
Checkpoints only happen at the Control Flow; it is not possible to checkpoint transformations or restart inside a Data Flow. The Data Flow Task can be a checkpoint, but it is treated as any other task.

**• Can you explain different options for dynamic configurations in SSIS?**  
1) XML file  
2) custom variables  
3) Database per environment with the variables  
4) Use a centralized database with all variables

**• What is the use of Percentage Sampling transformation in SSIS?**  
Percentage Sampling transformation is generally used for data mining. This transformation builds a random sample of set of output rows by choosing specified percentage of input rows. For example if the input has 1000 rows and if I specify 10 as percentage sample then the transformation returns 10% of the RANDOM records from the input data.

**• What is the use of Term Extraction transformation in SSIS?**  
Term Extraction transformation is used to extract nouns or noun phrases or both noun and noun phrases only from English text. It extracts terms from text in a transformation input column and then writes the terms to a transformation output column. It can be also used to find out the content of a dataset.

**• What is Data Viewer and what are the different types of Data Viewers in SSIS?**  
A Data Viewer allows viewing data at a point of time at runtime. If data viewer is placed before and after the Aggregate transform, we can see data flowing to the transformation at the runtime and how it looks like after the transformation occurred. The different types of data viewers are:

1. Grid  
2. Histogram  
3. Scatter Plot  
4. Column Chart.

**• What is Ignore Failure option in SSIS?**  
In Ignore Failure option, the error will be ignored and the data row will be directed to continue on the next transformation. Let’s say you have some JUNK data(wrong type of data or JUNK data) flowing from source, then using this option in SSIS we can REDIRECT the junk data records to another transformation instead of FAILING the package. This helps to MOVE only valid data to destination and JUNK can be captured into separate file.

**• Which are the different types of Control Flow components in SSIS?**  
The different types of Control Flow components are: Data Flow Tasks, SQL Server Tasks, Data Preparation Tasks, Work flow Tasks, Scripting Tasks, Analysis Services Tasks, Maintenance Tasks, Containers.

**• What are containers? What are the different types of containers in SSIS?**  
Containers are objects that provide structures to packages and extra functionality to tasks. There are four types of containers in SSIS, they are: Foreach Loop Container, For Loop Container, Sequence Container and Task Host Container.

**• What are the different types of Data flow components in SSIS?**  
There are 3 data flow components in SSIS.  
1. Sources  
2. Transformations  
3. Destinations

**• What are the different types of data sources available in SSIS?**  
There are 7 types of data sources provided by SSIS: a.) Data Reader source b.) Excel source c.) Flat file source d.) OLEDB source e.) Raw file source f.) XML source g.) Script component

**• What is SSIS Designer?**  
It is a graphical tool for creating packages. It has 4 tabs: Control Flow, Data Flow, Event Handlers and Package Explorer.

**• What is Control Flow tab?**  
It is the tab in SSIS designer where various Tasks can be arranged and configured. This is the tab where we provide and control the program flow of the project.

**• What is Data Flow tab?**  
This is the tab where we do all the work related to ETL job. It is the tab in SSIS Designer where we can extract data from sources, transform the data and then load them into destinations.

**• What is the function of control flow tab in SSIS?**  
On the control flow tab, the tasks including dataflow task, containers and precedence constraints that connect containers and tasks can be arranged and configured.

**• What is the function of Event handlers tab in SSIS?**  
On the Event handlers tab, workflows can be configured to respond to package events.  
For example, we can configure Work Flow when ANY task Failes or Stops or Starts ..

**• What is the function of Package explorer tab in SSIS?**  
This tab provides an explorer view of the package. You can see what is happening in the package. The Package is a container at the top of the hierarchy.

**• What is Solution Explorer?**  
It is a place in SSIS Designer where all the projects, Data Sources, Data Source Views and other miscellaneous files can be viewed and accessed for modification.

**• How do we convert data type in SSIS?**  
The Data Conversion Transformation in SSIS converts the data type of an input column to a different data type.

**• How are variables useful in ssis package?**  
Variables can provide communication among objects in the package. Variables can provide communication between parent and child packages. Variables can also be used in expressions and scripts. This helps in providing dynamic values to tasks.

**• Explain Aggregate Transformation in SSIS?**  
It aggregates data, similar you do in applying TSQL functions like Group By, Min, Max, Avg, and Count. For example you get total quantity and Total line item for each product in Aggregate Transformation Editor. First you determine input columns, then output column name in Output Alias table in datagrid, and also operations for each Output Alias in Operation columns of the same datagrid. Some of operation functions listed below :  
• Group By  
• Average  
• Count  
• Count Distinct : count distinct and non null column value  
• Min, Max, Sum  
In Advanced tab, you can do some optimization here, such as setting up Key Scale option (low, medium, high), Count Distinct scale option (low, medium, high), Auto Extend factor and Warn On Division By Zero. If you check Warn On Division By Zero, the component will give warning instead of error. Key Scale option will optimize transformation cache to certain number of key threshold. If you set it low, optimization will target to 500,000 keys written to cache, medium can handle up to 5 million keys, and high can handle up to 25 million keys, or you can specify particular number of keys here. Default value is unspecified. Similar to number of keys for Count Distinct scale option. It is used to optimize number of distinct value written to memory, default value is unspecified. Auto Extend Factor is used when you want some portion of memory is used for this component. Default value is 25% of memory.

**• Explain Audit Transformation ?**  
It allows you to add auditing information as required in auditing world specified by HIPPA and Sarbanes-Oxley (SOX). Auditing options that you can add to transformed data through this transformation are :  
1. Execution of Instance GUID : ID of execution instance of the package  
2. PackageID : ID of the package  
3. PackageName  
4. VersionID : GUID version of the package  
5. Execution StartTime  
6. MachineName  
7. UserName  
8. TaskName  
9. TaskID : uniqueidentifier type of the data flow task that contains audit transformation.

**• Explain Character Map Transformation?**  
It transforms some character. It gives options whether output result will override the existing column or add to new column. If you define it as new column, specify new column name. Operations available here are:  
1. Uppercase  
2. Lowercase  
3. Byte reversal : such as from 0×1234 to 0×4321  
4. Full width  
5. Half width  
6. Hiragana/katakana/traditional Chinese/simplified Chinese  
7. Linguistic casing

**• Explain Conditional split Transformation ?**  
It functions as if…then…else construct. It enables send input data to a satisfied conditional branch. For example you want to split product quantity between less than 500 and greater or equal to 500. You can give the conditional a name that easily identifies its purpose. Else section will be covered in Default Output Column name.  
After you configure the component, it connect to subsequent transformation/destination, when connected, it pops up dialog box to let you choose which conditional options will apply to the destination transformation/destination.

**• Explain Copy column Transformation?**  
This component simply copies a column to another new column. Just like ALIAS Column in T-Sql.

**• Explain Data conversion Transformation?**  
This component does conversion data type, similar to TSQL function CAST or CONVERT. If you wish to convery the data from one type to another then this is the best bet. But please make sure that you have COMPATABLE data in the column.

**• Explain Data Mining query Transformation?**  
This component does prediction on the data or fills gap on it. Some good scenarios uses this component is:  
1. Take some input columns as number of children, domestic income, and marital income to predict whether someone owns a house or not.  
2. Take prediction what a customer would buy based analysis buying pattern on their shopping cart.  
3. Filling blank data or default values when customer doesn’t fill some items in the questionnaire.

**• Explain Derived column Transformation?**  
Derived column creates new column or put manipulation of several columns into new column. You can directly copy existing or create a new column using more than one column also.

**• Explain Merge Transformation?**  
Merge transformation merges two paths into single path. It is useful when you want to break out data into path that handles errors after the errors are handled, the data are merge back into downstream or you want to merge 2 data sources. It is similar with Union All transformation, but Merge has some restrictions :  
1. Data should be in sorted order  
2. Data type , data length and other meta data attribute must be similar before merged.

**• Explain Merge Join Transformation?**  
Merge Join transformation will merge output from 2 inputs and doing INNER or OUTER join on the data. But if you the data come from 1 OLEDB data source, it is better you join through SQL query rather than using Merge Join transformation. Merge Join is intended to join 2 different data source.

**• Explain Multicast Transformation?**  
This transformation sends output to multiple output paths with no conditional as Conditional Split does. Takes ONE Input and makes the COPY of data and passes the same data through many outputs. In simple Give one input and take many outputs of the same data.

**• Explain Percentage and row sampling Transformations?**  
This transformation will take data from source and randomly sampling data. It gives you 2 outputs. First is selected data and second one is unselected data. It is used in situation where you train data mining model. These two are used to take the SAMPLE of data from the input data.

**• Explain Sort Transformation?**  
This component will sort data, similar in TSQL command ORDER BY. Some transformations need sorted data.

**• Explain Union all Transformation?**  
It works in opposite way to Merge transformation. It can take output from more than 2 input paths and combines into single output path.

**• What r the possible locations to save SSIS package?**  
You can save a package wherever you want.  
SQL Server  
Package Store  
File System

**• What is a package?**  
A discrete executable unit of work composed of a collection of control flow and other objects, including data sources, transformations, process sequence, and rules, errors and event handling, and data destinations.

**• What is a workflow in SSIS?**  
A workflow is a set of instructions on how to execute tasks.  
(It is a set of instructions on how to execute tasks such as sessions, emails and shell commands. a workflow is created form work flow mgr.

**• What is the diff between control flow Items and data flow Items?**  
The control flow is the highest level control process. It allows you to manage the run-time process activities of data flow and other processes within a package.  
When we want to extract, transform and load data within a package. You add an SSIS dataflow task to the package control flow.

**• What are the main component of SSIS(project-architecture)?**  
SSIS archItecture has 4 main components  
1.ssis service  
2.ssis runtime engine & runtime executables  
3.ssis dataflow engine & dataflow components  
4.ssis clients

**• Different components in SSIS package?**  
Control flow  
Data flow  
Event handler  
Package explorer

**• What are Connection Managers?**  
It is a bridge b/w package object and physical data. It provides logical representation of a connection at design time the properties of the connection mgr describes the physical connection that integration services creates when the package is run.

**• What is environment variable in SSIS?**  
An environment variable configuration sets a package property equal to the value in an environment variable.  
Environmental configurations are useful for configuring properties that are dependent on the computer that is executing the package.

**• How to provide securIty to packages?**  
We can provide security in two ways  
1. Package encryption  
2. Password protection.

**• What are Precedence constraints?**  
Constraints that link executable, container, and tasks wIthin the package control flow and specify condItion that determine the sequence and condItions for determine whether executable run.

**• What is Design time Deployment in SSIS ?**

Whenyou run a package from with in BIDS,it is built and temporarily deployed to the folder. By default the package will be deployed to the BIN folder in the Package’s Project folder and you can configure for custom folder for deployment. When the Package’s execution is completed and stopped in BIDS,the deployed package will be deleted and this is called as Design Time Deployment.

**What is SQL Server Integration Services (SSIS)?**

* SQL Server Integration Services (SSIS) is component of SQL Server 2005 and later versions. SSIS is an enterprise scale ETL (Extraction, Transformation and Load) tool which allows you to develop data integration and workflow solutions. Apart from data integration, SSIS can be used to define workflows to automate updating multi-dimensional cubes and automating maintenance tasks for SQL Server databases.

**How does SSIS differ from DTS?**

* SSIS is a successor to DTS (Data Transformation Services) and has been completely re-written from scratch to overcome the limitations of DTS which was available in SQL Server 2000 and earlier versions. A significant improvement is the segregation of the control/work flow from the data flow and the ability to use a buffer/memory oriented architecture for data flows and transformations which improve performance.

**What is the Control Flow?**

* When you start working with SSIS, you first create a package which is nothing but a collection of tasks or package components.  The control flow allows you to order the workflow, so you can ensure tasks/components get executed in the appropriate order.

**What is the Data Flow Engine?**

* The Data Flow Engine, also called the SSIS pipeline engine, is responsible for managing the flow of data from the source to the destination and performing transformations (lookups, data cleansing etc.).  Data flow uses memory oriented architecture, called buffers, during the data flow and transformations which allows it to execute extremely fast. This means the SSIS pipeline engine pulls data from the source, stores it in buffers (in-memory), does the requested transformations in the buffers and writes to the destination. The benefit is that it provides the fastest transformation as it happens in memory and we don't need to stage the data for transformations in most cases.

**What is a Transformation?**

* A transformation simply means bringing in the data in a desired format. For example you are pulling data from the source and want to ensure only distinct records are written to the destination, so duplicates are  removed.  Anther example is if you have master/reference data and want to pull only related data from the source and hence you need some sort of lookup. There are around 30 transformation tasks available and this can be extended further with custom built tasks if needed.

**What is a Task?**

* A task is very much like a method of any programming language which represents or carries out an individual unit of work. There are broadly two categories of tasks in SSIS, Control Flow tasks and Database Maintenance tasks. All Control Flow tasks are operational in nature except Data Flow tasks. Although there are around 30 control flow tasks which you can use in your package you can also develop your own custom tasks with your choice of .NET programming language.

**What is a Precedence Constraint and what types of Precedence Constraint are there?**

* SSIS allows you to place as many as tasks you want to be placed in control flow. You can connect all these tasks using connectors called Precedence Constraints. Precedence Constraints allow you to define the logical sequence of tasks in the order they should be executed. You can also specify a condition to be evaluated before the next task in the flow is executed.
* These are the types of precedence constraints and the condition could be either a constraint, an expression or both
  + Success (next task will be executed only when the last task completed successfully) or
  + Failure (next task will be executed only when the last task failed) or
  + Complete (next task will be executed no matter the last task was completed or failed).

**What is a container and how many types of containers are there?**

* A container is a logical grouping of tasks which allows you to manage the scope of the tasks together.
* These are the types of containers in SSIS:
  + Sequence Container - Used for grouping logically related tasks together
  + For Loop Container - Used when you want to have repeating flow in package
  + For Each Loop Container - Used for enumerating each object in a collection; for example a record set or a list of files.
* Apart from the above mentioned containers, there is one more container called the Task Host Container which is not visible from the IDE, but every task is contained in it (the default container for all the tasks).

**What are variables and what is variable scope?**

* A variable is used to store values. There are basically two types of variables, System Variable (like ErrorCode, ErrorDescription, PackageName etc) whose values you can use but cannot change and User Variable which you create, assign values and read as needed. A variable can hold a value of the data type you have chosen when you defined the variable.
* Variables can have a different scope depending on where it was defined. For example you can have package level variables which are accessible to all the tasks in the package and there could also be container level variables which are accessible only to those tasks that are within the container.

**What are SSIS Connection Managers?**

* When we talk of integrating data, we are actually pulling data from different sources and writing it to a destination. But how do you get connected to the source and destination systems? This is where the connection managers come into the picture. Connection manager represent a connection to a system which includes data provider information, the server name, database name, authentication mechanism, etc. For more information check out the [SQL Server Integration Services (SSIS) Connection Managers](http://www.mssqltips.com/sqlservertutorial/209/sql-server-integration-services-ssis-connection-managers/) and [Connection Managers in SQL Server 2005 Integration Services SSIS](http://www.mssqltips.com/sqlservertip/1147/connection-managers-in-sql-server-2005-integration-services-ssis/) tips.

**What is the RetainSameConnection property and what is its impact?**

* Whenever a task uses a connection manager to connect to source or destination database, a connection is opened and closed with the execution of that task. Sometimes you might need to open a connection, execute multiple tasks and close it at the end of the execution. This is where RetainSameConnection property of the connection manager might help you. When you set this property to TRUE, the connection will be opened on first time it is used and remain open until execution of the package completes.

**What are a source and destination adapters?**

* A source adaptor basically indicates a source in Data Flow to pull data from. The source adapter uses a connection manager to connect to a source and along with it you can also specify the query method and query to pull data from the source.
* Similar to a source adaptor, the destination adapter indicates a destination in the Data Flow to write data to. Again like the source adapter, the destination adapter also uses a connection manager to connect to a target system and along with that you also specify the target table and writing mode, i.e. write one row at a time or do a bulk insert as well as several other properties.
* Please note, the source and destination adapters can both use the same connection manager if you are reading and writing to the same database.

**What is the Data Path and how is it different from a Precedence Constraint?**

* Data Path is used in a Data Flow task to connect to different components of a Data Flow and show transition of the data from one component to another. A data path contains the meta information of the data flowing through it, such as the columns, data type, size, etc. When we talk about differences between the data path and precedence constraint; the data path is used in the data flow, which shows the flow of data. Whereas the precedence constraint is used in control flow, which shows control flow or transition from one task to another task.

**What is a Data Viewer utility and what it is used for?**

* The data viewer utility is used in Business Intelligence Development Studio during development or when troubleshooting an SSIS Package. The data viewer utility is placed on a data path to see what data is flowing through that specific data path during execution. The data viewer utility displays rows from a single buffer at a time, so you can click on the next or previous icons to go forward and backward to display data. Check out the Data Viewer enhancements in [SQL Server 2012](http://www.mssqltips.com/sqlservertip/2448/sql-server-denali-ssis-enhancements-part-2-of-2/).

**What is an SSIS breakpoint? How do you configure it? How do you disable or delete it?**

* A breakpoint allows you to pause the execution of the package in Business Intelligence Development Studio during development or when troubleshooting an SSIS Package. You can right click on the task in control flow, click on Edit Breakpoint menu and from the Set Breakpoint window, you specify when you want execution to be halted/paused. For example OnPreExecute, OnPostExecute, OnError events, etc. To toggle a breakpoint, delete all breakpoints and disable all breakpoints go to the Debug menu and click on the respective menu item. You can event specify different conditions to hit the breakpoint as well. To learn more about breakpoints, refer to[Breakpoints in SQL Server 2005 Integration Services SSIS](http://www.mssqltips.com/sqlservertip/1110/breakpoints-in-sql-server-2005-integration-services-ssis/).

**What is SSIS event logging?**

* Like any other modern programming language, SSIS also raises different events during package execution life cycle. You can enable or write these events to trace the execution of your SSIS package and its tasks. You can also can write your custom message as a custom log. You can enable event logging at the package level as well as at the tasks level. You can also choose any specific event of a task or a package to be logged. This is essential when you are troubleshooting your package and trying to understand a performance problem or root cause of a failure. Check out this tip about [Custom Logging in SQL Server Integration Services SSIS](http://www.mssqltips.com/sqlservertip/1417/custom-logging-in-sql-server-integration-services-ssis/).

**What are the different SSIS log providers?**

* There are several places where you can log execution data generated by an SSIS event log:
  + SSIS log provider for Text files
  + SSIS log provider for Windows Event Log
  + SSIS log provider for XML files
  + SSIS log provider for SQL Profiler
  + SSIS log provider for SQL Server, which writes the data to the msdb..sysdtslog90 or msdb..sysssislog table depending on the SQL Server version.

**How do you enable SSIS event logging?**

* SSIS provides a granular level of control in deciding what to log and where to log. To enable event logging for an SSIS Package, right click in the control flow area of the package and click on Logging. In the Configure SSIS Logs window you will notice all the tasks of the package are listed on the left side of the tree view. You can specifically choose which tasks you want to enable logging. On the right side you will notice two tabs; on the Providers and Logs tab you specify where you want to write the logs, you can write it to one or more log providers together. On the Details tab you can specify what events do you want to log for the selected task.
* Please note, enabling event logging is immensely helpful when you are troubleshooting a package, but also incurs additional overhead on SSIS in order to log the events and information. Hence you should only enabling event logging when needed and only choose events which you want to log. Avoid logging all the events unnecessarily.

**What is the LoggingMode property?**

* SSIS packages and all of the associated tasks or components have a property called LoggingMode. This property accepts three possible values: Enabled - to enable logging of that component, Disabled - to disable logging of that component and UseParentSetting - to use parent's setting of that component to decide whether or not to log the data.

**What is the transaction support feature in SSIS?**

* When you execute a package, every task of the package executes in its own transaction. What if you want to execute two or more tasks in a single transaction? This is where the transaction support feature helps. You can group all your logically related tasks in single group.  Next you can set the transaction property appropriately to enable a transaction so that all the tasks of the package run in a single transaction. This way you can ensure either all of the tasks complete successfully or if any of them fails, the transaction gets roll-backed too.

**What properties do you need to configure in order to use the transaction feature in SSIS?**

* Suppose you want to execute 5 tasks in a single transaction, in this case you can place all 5 tasks in a Sequence Container and set the TransactionOption and IsolationLevel properties appropriately.
  + The TransactionOption property expects one of these three values:
    - Supported - The container/task does not create a separate transaction, but if the parent object has already initiated a transaction then participate in it
    - Required - The container/task creates a new transaction irrespective of any transaction initiated by the parent object
    - NotSupported - The container/task neither creates a transaction nor participates in any transaction initiated by the parent object
* Isolation level dictates how two more transaction maintains consistency and concurrency when they are running in parallel. To learn more about Transaction and Isolation Level, refer to [this tip](http://www.mssqltips.com/sqlservertip/1585/how-to-use-transactions-in-sql-server-integration-services-ssis/).

**When I enabled transactions in an SSIS package, it failed with this exception: "The Transaction Manager is not available. The DTC transaction failed to start."  What caused this exception and how can it be fixed?**

* SSIS uses the MS DTC (Microsoft Distributed Transaction Coordinator) Windows Service for transaction support.  As such, you need to ensure this service is running on the machine where you are actually executing the SSIS packages or the package execution will fail with the exception message as indicated in this question.

**What is event handling in SSIS?**

* Like many other programming languages, SSIS and its components raise different events during the execution of the code. You can write an even handler to capture the event and handle it in a few different ways. For example consider you have a data flow task and before execution of this data flow task you want to make some environmental changes such as creating a table to write data into, deleting/truncating a table you want to write, etc. Along the same lines, after execution of the data flow task you want to cleanup some staging tables. In this circumstance you can write an event handler for the OnPreExcute event of the data flow task which gets executed before the actual execution of the data flow. Similar to that you can also write an event handler for OnPostExecute event of the data flow task which gets executed after the execution of the actual data flow task.  Please note, not all the tasks raise the same events as others. There might be some specific events related to a specific task that you can use with one object and not with others.

**How do you write an event handler?**

* First, open your SSIS package in Business Intelligence Development Studio (BIDS) and click on the Event Handlers tab.  Next, select the executable/task from the left side combo-box and then select the event you want to write the handler in the right side combo box. Finally, click on the hyperlink to create the event handler. So far you have only created the event handler, you have not specified any sort of action. For that simply drag the required task from the toolbox on the event handler designer surface and configure it appropriately. To learn more about event handling, [click here.](http://www.sql-server-performance.com/2009/SSIS-Features-And-Properties-Part2/)

**What is the DisableEventHandlers property used for?**

* Consider you have a task or package with several event handlers, but for some reason you do not want event handlers to be called. One simple solution is to delete all of the event handlers, but that would not be viable if you want to use them in the future. This is where you can use the DisableEventHandlers property.  You can set this property to TRUE and all event handlers will be disabled. Please note with this property you simply disable the event handlers and you are not actually removing them.  This means you can set this value to FALSE and the event handlers will once again be executed.

**What is SSIS validation?**

* SSIS validates the package and all of it's tasks to ensure it has been configured correctly.  With a given set of configurations and values, all the tasks and package will execute successfully. In other words, during the validation process, SSIS checks if the source and destination locations are accessible and the meta data about the source and destination tables are stored with the package are correct, so that the task will not fail if executed. The validation process reports warnings and errors depending on the validation failure detected. For example, if the source/destination tables/columns get changed/dropped it will show as error.  Whereas if you are accessing more columns than used to write to the destination object this will be flagged as a warning. To learn about validation [click here.](http://www.sql-server-performance.com/2009/ssis-features-and-properties-part4/)

**Define design time validation versus run time validation.**

* Design time validation is performed when you are opening your package in BIDS whereas run time validation is performed when you are actually executing the package.

**Define early validation (package level validation) versus late validation (component level validation).**

* When a package is executed, the package goes through the validation process.  All of the components/tasks of package are validated before actually starting the package execution. This is called early validation or package level validation. During execution of a package, SSIS validates the component/task again before executing that particular component/task.  This is called late validation or component level validation.

**What is DelayValidation and what is the significance?**

* As I said before, during early validation all of the components of the package are validated along with the package itself. If any of the component/task fails to validate, SSIS will not start the package execution. In most cases this is fine, but what if the second task is dependent on the first task?  For example, say you are creating a table in the first task and referring to the same table in the second task? When early validation starts, it will not be able to validate the second task as the dependent table has not been created yet.  Keep in mind that early validation is performed before the package execution starts. So what should we do in this case?  How can we ensure the package is executed successfully and the logically flow of the package is correct?  This is where you can use the DelayValidation property.  In the above scenario you should set the DelayValidation property of the second task to TRUE in which case early validation i.e. package level validation is skipped for that task and that task would only be validated during late validation i.e. component level validation. Please note using the DelayValidation property you can only skip early validation for that specific task, there is no way to skip late or component level validation.

**What are the different components in the SSIS architecture?**

* The SSIS architecture comprises of four main components:
  + The SSIS runtime engine manages the workflow of the package
  + The data flow pipeline engine manages the flow of data from source to destination and in-memory transformations
  + The SSIS object model is used for programmatically creating, managing and monitoring SSIS packages
  + The SSIS windows service allows managing and monitoring packages
* To learn more about the architecture [click here.](http://technet.microsoft.com/en-us/library/ms141709(SQL.90).aspx)

**How is SSIS runtime engine different from the SSIS dataflow pipeline engine?**

* The SSIS Runtime Engine manages the workflow of the packages during runtime, which means its role is to execute the tasks in a defined sequence.  As you know, you can define the sequence using precedence constraints. This engine is also responsible for providing support for event logging, breakpoints in the BIDS designer, package configuration, transactions and connections. The SSIS Runtime engine has been designed to support concurrent/parallel execution of tasks in the package.
* The Dataflow Pipeline Engine is responsible for executing the data flow tasks of the package. It creates a dataflow pipeline by allocating in-memory structure for storing data in-transit. This means, the engine pulls data from source, stores it in memory, executes the required transformation in the data stored in memory and finally loads the data to the destination. Like the SSIS runtime engine, the Dataflow pipeline has been designed to do its work in parallel by creating multiple threads and enabling them to run multiple execution trees/units in parallel.

**How is a synchronous (non-blocking) transformation different from an asynchronous (blocking) transformation in SQL Server Integration Services?**

* A transformation changes the data in the required format before loading it to the destination or passing the data down the path. The transformation can be categorized in Synchronous and Asynchronous transformation.
* A transformation is called synchronous when it processes each incoming row (modify the data in required format in place only so that the layout of the result-set remains same) and passes them down the hierarchy/path. It means, output rows are synchronous with the input rows (1:1 relationship between input and output rows) and hence it uses the same allocated buffer set/memory and does not require additional memory. Please note, these kinds of transformations have lower memory requirements as they work on a row-by-row basis (and hence run quite faster) and do not block the data flow in the pipeline. Some of the examples are : Lookup, Derived Columns, Data Conversion, Copy column, Multicast, Row count transformations, etc.
* A transformation is called Asynchronous when it requires all incoming rows to be stored locally in the memory before it can start producing output rows. For example, with an Aggregate Transformation, it requires all the rows to be loaded and stored in memory before it can aggregate and produce the output rows. This way you can see input rows are not in sync with output rows and more memory is required to store the whole set of data (no memory reuse) for both the data input and output. These kind of transformations have higher memory requirements (and there are high chances of buffer spooling to disk if insufficient memory is available) and generally runs slower. The asynchronous transformations are also called "blocking transformations" because of its nature of blocking the output rows unless all input rows are read into memory. To learn more about it [click here.](http://www.sql-server-performance.com/2009/ssis-an-inside-view-part-2/)

**What is the difference between a partially blocking transformation versus a fully blocking transformation in SQL Server Integration Services?**

* Asynchronous transformations, as discussed in last question, can be further divided in two categories depending on their blocking behavior:
  + Partially Blocking Transformations do not block the output until a full read of the inputs occur.  However, they require new buffers/memory to be allocated to store the newly created result-set because the output from these kind of transformations differs from the input set. For example, Merge Join transformation joins two sorted inputs and produces a merged output. In this case if you notice, the data flow pipeline engine creates two input sets of memory, but the merged output from the transformation requires another set of output buffers as structure of the output rows which are different from the input rows. It means the memory requirement for this type of transformations is higher than synchronous transformations where the transformation is completed in place.
  + Full Blocking Transformations, apart from requiring an additional set of output buffers, also blocks the output completely unless the whole input set is read. For example, the Sort Transformation requires all input rows to be available before it can start sorting and pass down the rows to the output path. These kind of transformations are most expensive and should be used only as needed. For example, if you can get sorted data from the source system, use that logic instead of using a Sort transformation to sort the data in transit/memory. To learn more about it [click here.](http://www.sql-server-performance.com/2009/ssis-an-inside-view-part-2/)

**What is an SSIS execution tree and how can I analyze the execution trees of a data flow task?**

* The work to be done in the data flow task is divided into multiple chunks, which are called execution units, by the dataflow pipeline engine.  Each represents a group of transformations. The individual execution unit is called an execution tree, which can be executed by separate thread along with other execution trees in a parallel manner. The memory structure is also called a data buffer, which gets created by the data flow pipeline engine and has the scope of each individual execution tree. An execution tree normally starts at either the source or an asynchronous transformation and ends at the first asynchronous transformation or a destination. During execution of the execution tree, the source reads the data, then stores the data to a buffer, executes the transformation in the buffer and passes the buffer to the next execution tree in the path by passing the pointers to the buffers. To learn more about it [click here.](http://www.sql-server-performance.com/2009/ssis-an-inside-view-part-2/)
* To see how many execution trees are getting created and how many rows are getting stored in each buffer for a individual data flow task, you can enable logging of these events of data flow task: PipelineExecutionTrees, PipelineComponentTime, PipelineInitialization, BufferSizeTunning, etc. To learn more about events that can be logged [click here](http://msdn.microsoft.com/en-us/library/ms141122.aspx).

**How can an SSIS package be scheduled to execute at a defined time or at a defined interval per day?**

* You can configure a [SQL Server Agent Job](http://www.mssqltips.com/sql-server-tip-category/27/sql-server-agent/) with a job step type of SQL Server Integration Services Package, the job invokes the [dtexec command line utility](http://www.mssqltips.com/sqlservertip/1812/dtexec-command-line-parameters-using-command-files/) internally to execute the package. You can run the job (and in turn the SSIS package) on demand or you can create a schedule for a one time need or on a reoccurring basis. Refer to [this tip](http://www.mssqltips.com/sqlservertip/1180/sql-server-integration-services-package-scheduling-with-sql-agent/) to learn more about it.

**What is an SSIS Proxy account and why would you create it?**

* When we try to execute an SSIS package from a SQL Server Agent Job it fails with the message "Non-SysAdmins have been denied permission to run DTS Execution job steps without a proxy account". This error message is generated if the account under which SQL Server Agent Service is running and the job owner is not a sysadmin on the instance or the job step is not set to run under a proxy account associated with the SSIS subsystem. Refer to[this tip](http://www.mssqltips.com/sqlservertip/2163/running-a-ssis-package-from-sql-server-agent-using-a-proxy-account/) to learn more about it.

**How can you configure your SSIS package to run in 32-bit mode on 64-bit machine when using some data providers which are not available on the 64-bit platform?**

* In order to run an SSIS package in 32-bit mode the SSIS project property Run64BitRuntime needs to be set to "False".  The default configuration for this property is "True".  This configuration is an instruction to load the 32-bit runtime environment rather than 64-bit, and your packages will still run without any additional changes. The property can be found under SSIS Project Property Pages -> Configuration Properties -> Debugging.