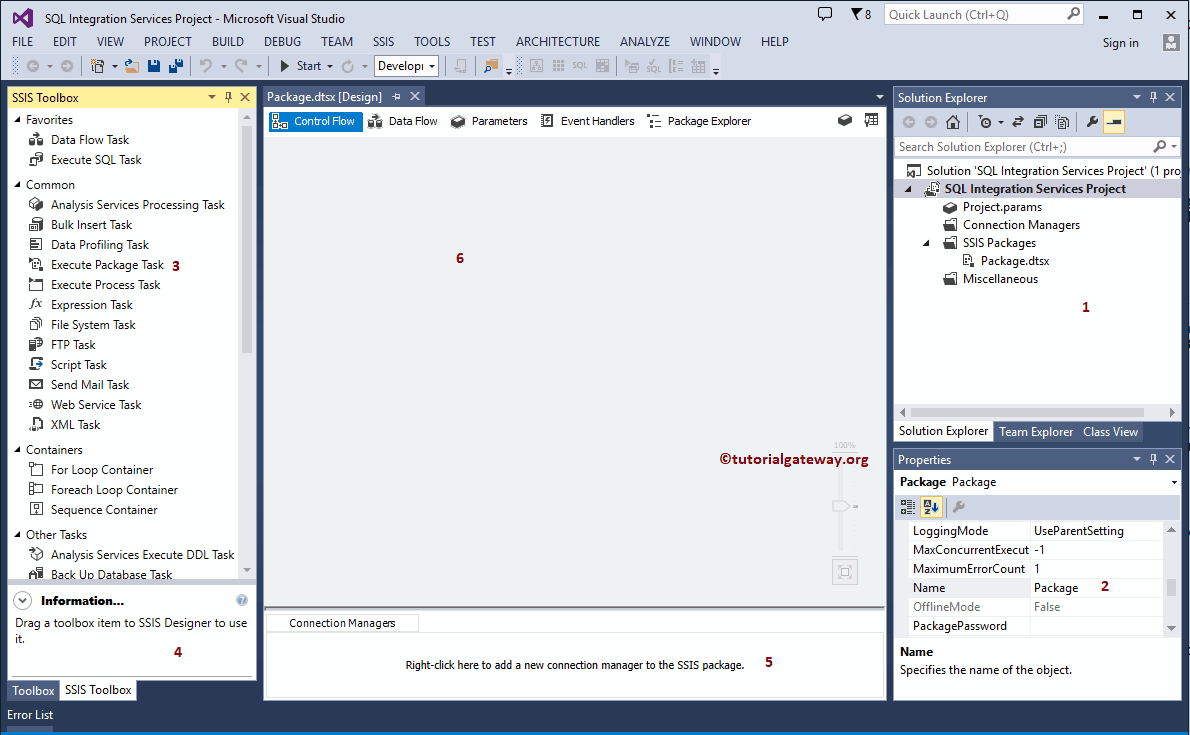
1. **SSIS:-**

SQL Server is the new platform that was introduced in SQL Server 2005, for data transformations by applying ETL operations and data integration solutions.

SQL Server Integration Services is an ETL (Extract, Transform and Load) tool. It means SSIS can be used to extract data from a wide variety of sources such as Excel Files, Flat Files, XML Files, Relational databases, and transform (slice and dice) them as per your requirements and finally load the data into the destination.

**SSIS Work environment:-**



1. **Solution Explorer:** This is a combination of project level connection managers, actual packages, and project parameters.
2. **Properties:** Use this window to change the properties of each and every task.
3. **Toolbox:** SSIS Toolbox provides a lot of built-in tasks, containers, transformations, sources, destinations, and administrative tasks to solve complex business problems. You can use these graphical SSIS tools by drag and drop those tasks in the work environment. It means you do not have to write a single line of code to perform most of the operations.
4. **Information:** You can see the information of each and every item in a toolbox by clicking the toolbox items
5. **Connection Managers:** This window is to create a package level connection managers
6. Package: Design your package

**SSIS Project Development: -** SQL Server Integration services project is a combination of connections managers, packages, and project parameters (parameters)

1. **SSIS Component:-**

Integration Services Projects

Integration Services Packages

Control Flow Elements

Data Flow Elements

Integration Services Event Handlers

Integration Services Connections

Integration Services Variables

Integration Services Log Providers

1. **What are the elements (tabs) that you see on a default package designer?**

Control Flow, Data Flow, Event Handlers, Package Explorer, and Parameters.

1. **What is Control flow?**

Control flow defines the workflow of the tasks to be executed in a particular order. And provide the logic for when dataflow components are executed and how they are run. Performing different types of operations like looping containers, call stored procedures using Execute SQL task, move files, check a condition and call a different tasks depending result…

Control Flow components are Containers, Tasks, Precedence constraints that connects the executables, and tasks into an ordered control flow.

1. **What is Dataflow?**

All ETL tasks related to data are done in dataflow, A Dataflow consists of the sources, transformations and destinations. We extract the data from sources, on that applying transformations on extracted data and finally load the data into destination.

Dataflow components are Sources, Transformation, and Destination.

1. **Synchronous & Asynchronous Transformation: -**

**Synchronous (non-blocking) transformations** always offer the best performance. They do not change the shape of the data. You can think of it as data one row in and then one row out. Note that most transformations are non-blocking.

Synchronous transformations are either Stream based or row based. **Streaming transformation** are calculated in memory and do not required another resources to transform the data. This are the fastest transformations around. **Row based transformation** run a little bit slower because they require calling a service or looking up data from another source to calculate their values. They still move data one row at a time but the call to the resource outside of the SSIS package costs time.

**Non-Blocking Synchronous Streaming Transformations (Fastest)**

* Audit
* Cache Transform
* Character Map
* Conditional Split
* Copy Column
* Data Conversion
* Derived Column
* Lookup, if configured to Full Cache is enabled to store search results
* Multicast
* Percent Sampling
* Row Count

### Non-Blocking Synchronous Row-Based Transformations (Fast)

* DQS Cleansing
* Export Column
* Import Column
* Lookup, if configured to “Partial Cache” or “No Cache” to store search results
* OLE DB Command
* Script Component, if configured to use an outside resource
* Slowly Changing Dimension (SCD)

**Asynchronous (semi-blocking) transformations** require a sub set of the data to be collected before they can be sent to the destinations. The shape of the data can change. A subtotal or sampling of data may be extracted from the sources.

### Semi-Blocking Asynchronous Transformations (Medium)

* Data Mining Query
* Merge
* Merge Join
* Pivot
* Term Lookup
* Unpivot
* Union All

**Asynchronous (Full-Blocking) transformation** are the slowest transformations. They require all the data to be pulled from the sources before they can be sent to the destination. All source data must be loaded into the memory first. As much as we should try to avoid these, they can sometimes still be required. Such as sorting data pulled from a flat file source.

### Fully Blocking Asynchronous Transformations (Slowest)

* Aggregate
* Fuzzy Grouping
* Fuzzy Lookup
* Row Sampling
* Sort
* Term Extraction
* Script Component, if configured to collect all data before send it to a destination. Set the “SynchronousInputID” property to “None” on the Output column to make it asynchronous.

If there is more data than memory available, it will use the %TEMP% directory to cache some of the data. If you want to use a different location (such as a fast drive), you can set the **BufferTempStoragePath** property of the Data Flow Task to point to a different folder location. Be sure packages running these transformations are running on x64 platforms with plenty of memory so it won’t be cached to the hard drive.

1. **Parallel Execution:-**

Parallel execution in SSIS improves performance on computers that have multiple physical and logical processors.

To support parallel execution of different tasks in package, SSIS uses two properties **MaxConcurrentExecutables** and **EngineThreads**.

**MaxConcurrentExecutables**: - The MaxConcurrentExecutables Property is a property of the package. The property defines how many tasks can run simultaneously by specifying the maximum number of executables that can execute in parallel per package. The default value is

**-1**, which equates to the number of physical or logical processors plus 2.

Please note that if your box has hyperthreading turned on, it is the logical processor rather than the physically present processor that is counted.

**EngineThreads: -** The Engine Threads property is a property of the data flow task, which defines how many work threads the scheduler will create and run in parallel. Its default value is 10 in SQL SERVER 2008 onwards

**Example: -** If we set Engine Threads to 10 on all 3 Data flow tasks, then all the 30 flows will start off at once.

One thing we want to clear about Engine Threads is that it governs both source threads (for source components) and work threads (for transformation and destination components). Source threads and work threads of both Engine Threads created by the Data Flow’s Scheduler. So, the above example the value of 10 for Engine Threads means up to 10 source threads and 10 work threads.

1. **DelayValidation: -**

**DelayValidation** Property is available on Task level, Connection Manager, Container and on Package level. By default the value of this property is set to false that means that when the package start execution, It validates all the Tasks, Containers, Connection Managers and objects( Tables, Views, Stored Procedures etc.) used by them. If any object such as table or destination file etc. is not available then Package validation fails and Package stop execution.  
  
By setting this property to True, We enforce our SSIS Package not to validate that Task, Connection Manager or entire Package at start but validate at run time.

1. **List of Transformations: -**
   1. Aggregate Transformation
   2. Audit Transformation
   3. Cache Transform Transformation
   4. Character Map Transformation
   5. Conditional Split Transformation
   6. Copy Column Transformation
   7. Data Conversion Transformation
   8. Data Mining Query Transformation
   9. Derived Column Transformation
   10. DQS Cleansing Transformation
   11. Export Column Transformation
   12. Fuzzy Grouping Transformation
   13. Fuzzy LookUp Transformation
   14. Import Column Transformation
   15. Look Up Transformation
   16. Merge Transformation
   17. Merge Join Transformation
   18. Multicast Transformation
   19. OLEDB Command Transformation
   20. Percentage Sampling Transformation
   21. PIVOT Transformation
   22. Row Count Transformation
   23. Row Sampling Transformation
   24. Script Component Transformation
   25. Slowly Changing Dimension Transformation
   26. Sort Transformation
   27. Term Extraction Transformation
   28. Term Look Up Transformation
   29. Union All Transformation
   30. UnPivot Transformation