Chapter-3

Creating ETL Solution With SSIS Implementation Control Flow with SSIS

❖ What Is Server?

- All advance database management system includes some kind of service application as a part of database system.
- **For Ex:** You use the SQL server database you must have an application called SQL server engine.
- When you install SQL server service on your computer system the SQL server services will be automatically installed on your computer system.
- Your programs which need to access the database will not access the database file directory.
- Your program will communicate with the database service application.

❖ How Does Microsoft SQL Server Work? (In SQL Server)

- SQL Server is provided by Microsoft (MSSQL), SQL server is one of the most popular & advance database system currently available.
- Microsoft SQL Server is some time known as **"Sequel Server"**.
- Microsoft SQL was the result of a partnership between **Microsoft** & **Sybase**.
- SQL stands of **Structure Query Language** became standardize in 1980.
- It was designed by **Dr. E. F. Codd** in **1970** to be the first relational database management system.

- SQL server supports large application with millions of users on multiple databases.
- SQL server is much powerful then access & provides other advance features & much better security.

❖Introduction To ETL With SSIS

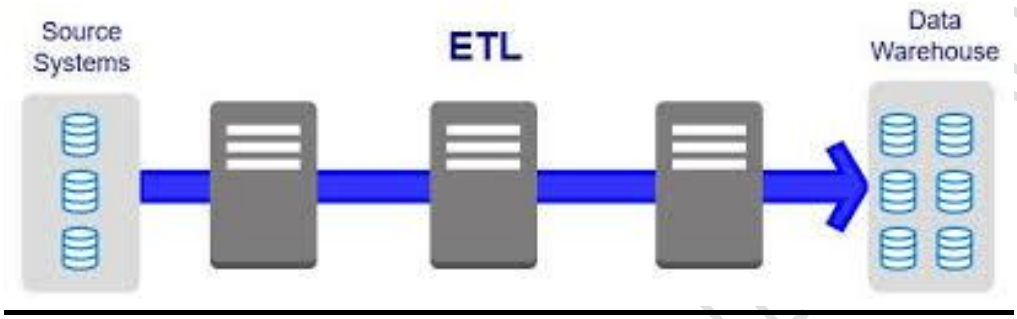
❖ What Is ETL?

- ETL stands for **Extract, Transformation & Load**.
- **Extracting** data from source operational or system which are primary source of data for the Data Warehouse.
- **Transforming** the data which may involve cleaning, filtering, validating& applying business rules.
- **Loading** the data into a data warehouse or any other database or application that houses data.
- The ETL process it is also very often refer to as data integration process & an ETL tool as a database integration platform.

❖ How ETL Works?

- Data from one or more sources is extracted then copied to the data warehouse.
- When dealing with large volume of data & multiple source system the data is separate.

- ETL process is used to migrate data from one database to another database & the specific process requires loading data to & from data marts & data warehouse.
- It is a process that is also used to transform (Convert) large database from one format to another.



❖ What is SQL Services?

- A service is a program that runs in the background that does not require any user interaction.
- The SQL Server Service is the executable process that the SQL server database.
- The SQL services is not dependent on any other services to run.
- SQL service can be configured to run as a domain user, local user, manage service account, virtual account & built in system account.
- Microsoft SQL server provides the following some services which is mandatory (compulsory) for database creation & maintenance.
  1. **SQL Server Agent.**
  2. **SQL Server browser.**
  3. **SQL Server Full Text Search.**
  4. **SQL Server Integration Services (SSIS).**
  5. **SQL Server Reporting Services (SSRS).**
  6. **SQL Server Analysis Services (SSAS).**

➢ SQL Server Integration Services (SSIS)

- SQL Server Integration Services is a powerfulETL tool which is used for building enterprise level data transformation & integration solution.
- SSIS is a business intelligent tool that provides data transformation solution for various organizations.
- SSIS can be used to extract data from a wide range of sources such as excel files, flat files, XML files, RDBMS files & as per your requirement finally load the data into the destination.

➢ ETL With SSIS

- SSIS is a tool from Microsoft which helps to perform operation.
- SSIS is a component of Microsoft SQL server database software that can be used to perform a broad range of data integration & data transformation task.
- It is basically ETL tool part of the Microsoft Business Intelligence (BI).
- It is mainly used to achieve data integration.

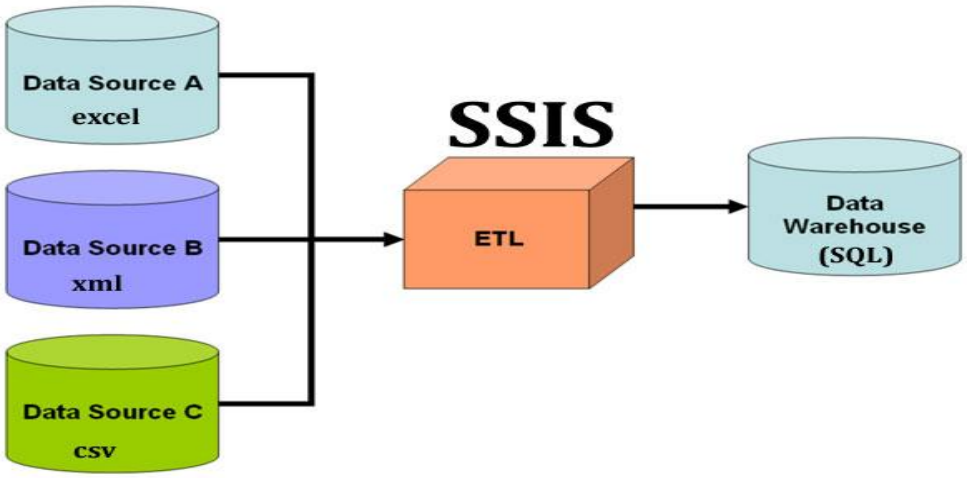


Figure: ETL With SSIS

- SSIS is a platform for data transformation & work flow application. It features a data warehousing tool used data extraction, transformation & loading process.
- This tool may also be used to automate maintenance of SQL server databases & update to multi-dimensional data.
- SSIS package data warehouse think & document all the data sources & transformation.
- Data source to connect to source database & target data warehouse database.
- SSIS package different format file to stay destination unique maintain.
- ETL is very important part of data warehouse when you first a lot of data of your work goes into designing & building the table structure that will hold the data.
- ETL process often does have some changes & some maintenance that needs to be done.
- Microsoft SQL Server Integration Services include many built in task & transformation to solve the complex business process by building performance data integration package.
- You can use SSIS for updating data warehouse, data mining, downloading, copy files, extract file & transfer data from different format to SQL.
- SSIS package to use ETL process in database usage & especially in data warehousing.
- SSIS data flow between your selected source & destination (the source, transformation destination).
- ETL package that include looping, configuration, error flow logic & logging.

❖Exploring Data Sources In SSIS

- A data source connection that are OLEDB & ADO.NET data sources such as SQL server, oracle, DB (2) & even not traditional data sources such as services & outlook.
- The data sources can be in scopes to a single SSIS package or share across multiple package in a project.
- A flat file sources provide a data source for connection such as text file for data.
- They can be shared across multiple packages in a project or single package.
- A data source view or unified view of the metadata from the specified tables & view that the data source defines in the project.
- Storing the metadata in the data source view enable you to work with the metadata during development without an open connection to any underline maintenance or data source.
- Microsoft has truly established SSIS has a major work in the extraction, transformation, loading (ETL).
- Not only is the SSIS technology a complete code we write from SQL server.
- In memory architecture is helps SSIS scale & run faster than staging data & running stored procedure.
- Data sources are this data pipelines & they are represented by connection that can be used by sources or destination once they have been define.

❖Introduction To Control Flow

- A package consists of control & optionally one or more data flow.
- SQL server integrated service provides 3 different types of control flow elements.
  1. **Structuring Package.**
  2. **Functionality.**
  3. **Constraints that connect the execution.**
- A package part can only contain one control flow task or constraints.
- A package part appears in the control flow in the SSIS toolbox.
- There are some control flow tasks in SSIS.
  1) Bulk Insert Task.
  2) Data Profiling Task.
  3) Execute T-SQL Statement Task.
  4) Execute SQL Task.
  5) Execute Package Task.
  6) File System Task.
  7) FTP Task.
  8) Script Task.
  9) Web Services Task.
- The integrated service support the nesting of container & a control flow can include multiple level of nested container.
- **For Ex:** A package could contain a container such as a foreach loop container & which in turn could contain another foreach loop container & so on.
- The data flow is a special task of the control flow.
- The data flow is a constraint where you can read data from various sources into the memory of the machine that is executing the SSIS package.
- The data is in memory you can perform different kinds of transformation.

- A data flow between your selected entities like sources, transformation &destination.
- A control flow defines a work flow of task to be executed after a particular order.
- SSIS provides the building of blocks we need to define the control flow.
- The form of task & container that from the SSIS toolbox to the control flow design surface.
- A task is a component that performs a specific function such as a T-SQL block or send email messaging.

❖ Implementing Data Flow Using SSIS

- The data flow task encapsulate the data flow engine that move data between source & destination & lets the user transfer, clean & modify data as it is moved.
- Additional of a dataflow task to a package control flow makes it possible for the package to extract, transfer & load data.
- A data flow consists of at least one of data flow component but is a typically a set of connected data flow component like source that extract data, transformation that modify, summarized data & destination that load data.
- SSIS supports many control flow item that manage a package work flow but the most often used is data flow task.
- The data flow is a constraint where you can read data from various sources into the memory of the machine that is executing the SSIS package.
- SSIS data flow transformation like:
  - ✓ **Aggregate.**
  - ✓ **Condition Split.**
  - ✓ **Data Conversion.**

- ✓ **Derived Column.**
- ✓ **Looked Up.**
- ✓ **Merge.**
- ✓ **Merge Join.**
- ✓ **Multicast.**
- ✓ **Script Component.**
- ✓ **Sort.**
- ✓ **Union All.**

- A package can include multiple data flow task it with its own data flow.
- **For Ex.:** If a package requires the data flow to run in a specific sequence or that other task perform between data flow.
- You must use a separate data flow task in each data flow.
- All ETL task related to data are done by data flow elements.
- It is not necessary to have a data flow element in the SSIS package.
- The data flow task is executable within the package that creates order, runs in data flow.
- A separate instance of the data flow engine is open for each data flow task in the package.
- In SSIS a work flow is called a control flow & this flow links together over modular data flow as a series of operation in order to reach at desired result.
- SSIS component, data flow component provides is to set up cost effective integration using SSIS ETL engine.

❖Creating Dynamic Package In SSIS

- One of the challenges to applying this technique to more complex scenario is that it can be difficult to create the initial for the SSIS package.
- Being able to generate SSIS package dynamically can greatly reduce SSIS package development time & administrative effort.
- Additionally your interface standards across the entire package in your project such as standardize SSIS execution & task name.
- SSIS packages there is some time needed to only run path in the package execution either based on may be a parameter value that is passed in the package or query from the database.
- SSIS includes multiple objects that can be used to create dynamic packages.
- SSIS can dynamically set a priority in a task or component using an expression to using built a set of building blocks including variable, parameter, function, literals and more.
- When package runs the expression is evaluated in used the property each task or component is access.
- The BIHL (Business Intelligent Markup Language) allow the automatic creation of the SSIS package by simplified the set of repetitive task perform in the extraction process of different source system.
- If you need to create package dynamically or to manage & execute integration service package outside the development environment you can manipulate packages programmatically.
- You can use the integration services object model to write code that is create configurable & execute packet in manage programming language.

❖ Implementing Package Configuration

- A package deployment package is to look at its configuration & override the package with the new configuration.
- There are three package configurations.
  1) Connection Properties.
  2) Package Variable Properties.
  3) Package Properties.

1) Connection Properties: This property set the string into the server name, the username & password.

2) Package variable Properties: You can set variable values, variable description & the raised event properly.

3) Package Properties: This includes any property on the package level such as checkpoint & security settings.

- Before you can enable package configuration you must convert your SSIS project to the deployment model.
- In new package in the SQL server 2012 of SSIS is set up for the project deployment model.
- By default each package has its package configuration turned off.
- The package itself does not move or transfer data.
- The package release on the task it contains to perform the work.
- A package is created & executed with various task & connection manager objects can add to give additional functionality to your package.

❖Introduction Of Container

- Container is defined by **Docker File** which defines a sequence to build a container.
- A container can include score of database including secondary & log file needed.
- Database can be copied & run within the container file system using the **MOUNT DB**command.
- Container with database in container can be sure as customer image to support creation of identical container.
- Containers with mounted database require a slightly different approach.
- Each container includes a private file system & users are limited to their container & file system.
- Containers are fast & light weight & deliver improve resource utilization.
- Container defines a new method of application packaging combine with user & process isolation for application multi tenancy.

❖ Using Container

- Container provides structure encapsulation & scope when real world concepts are implemented in SSIS.
- Task representation logical units of work can be groped in containers.
- Loop containers allowed a unit of work to be executed repeatedly.
- The sequence is defined by using sequence container.
- SSIS variable can be used by control flow task, containers by data flow components & they available to event handler.
- Containers in SSIS to understand the differences between so that you can use the appropriate in the SSIS solution.

- Before SSIS package can be executed in debug mode as can individual task a group of task enclosed in a container.
- Container executes the encapsulated task repeatedly based on expression.
- Container has no programmatic logic other than providing a structure to in encapsulate taste that from a logical during override a scope of SSIS variables.
- SSIS box a data flow task into used the for-each loop container.