

SQL Server Integration Services (SSIS)

Lesson 00: Training Kit(Part 1)



People matter, results count.

Document History

Date	Course Version No.	Software Version No.	Developer / SME	Reviewer(s)	Approver	Change Record Remarks
01-Feb-12	1.0	1.0	Ajit jog			Initial Document
June 2016	2.0	NA	Sarita Kaloya	Pramodh Daniel	Mahima Sharma	Material Revamp as per Integrated ToC for I & D LoT



Copyright © Capgemini 2015. All Rights Reserved 2

Course Audience and Course Goals

- This course is designed for
 - Analysts, Designers and Developers involved in Design, Development, and Maintenance of Data Warehousing Applications
- Course Goals
 - To understand SSIS as an ETL tool



Copyright © Capgemini 2015. All Rights Reserved 3

Pre-requisites

- Good knowledge of RDBMS
- Fair knowledge of Data Warehousing Concepts



Copyright © Capgemini 2015. All Rights Reserved 4

Intended Audience

- Professional Data warehousing experts, or anyone requiring a foundation knowledge in ETL process



Day Wise Schedule

- Day 1

- Lesson 1: Introduction to ETL and Integration Services
- Lesson 2: SSIS Background
- Lesson 3: Typical Uses of Integration Services
- Lesson 4: Tools

- Day 2

- Lesson 5: SSIS Objects
- Lesson 6: Data Flow Task
- Lesson 7: Data Flow Task -Source

- Day 3

- Lesson 8: Data Flow Tasks-Transformation
- Lesson 9: Data Flow Tasks-Destination



Copyright © Capgemini 2015. All Rights Reserved 6

Day Wise Schedule

- Day 4
 - Lesson 10: Operating system-level tasks
 - Lesson 11: SQL Server Tasks
 - Lesson 12: Scripting Tasks
- Day 5
 - Lesson 13: Profiling Task
 - Lesson 14: Workflow Tasks
 - Lesson 15: Processing XML
 - Lesson 16: Analysis Services Tasks
 - Lesson 17: Maintenance Tasks
- Day 6
 - Lesson 18: Container
 - Lesson 19: Precedence Constraint
 - Lesson 20: Variables and Expressions



Copyright © Capgemini 2015. All Rights Reserved 7

Day Wise Schedule

- Day 7

- Lesson 21: SSIS Architecture
- Lesson 22: Error Handling, Logging and Transactions
- Lesson 23: Execution , debugging and Protection

- Day 8

- Lesson 24: Deployment
- Lesson 25: Performance Tuning
- Lesson 26: Best Practices



Copyright © Capgemini 2015. All Rights Reserved 8

Course Content

- Lesson 1:Introduction to ETL and Integration Services
 - Overview of BI
 - ETL Concept
 - Different ETL Tool providers
 - Defining SQL Server Integration Services (SSIS)
- Lesson 2: SSIS Background
 - DTS Limitation
 - Difference between DTS , 2005, 2008 R2 ,2012,2014 and 2016
 - Lesson 3: Typical Uses of Integration Services
 - Merging Data from Heterogeneous Data Stores
 - Populating Data Warehouses and Data Marts Cleaning and Standardizing Data
 - Automating Administrative Functions and Data Loading
- Lesson 4:Tools
 - Business Intelligence Development Studio and SQL Server data Tool
 - SQL Server Management Studio
 - Import and Export Wizard
 - Package Configuration Wizard
 - Package Deployment Wizard



Copyright © Capgemini 2015. All Rights Reserved 9

Course Content

- Lesson 5: SSIS Objects
 - Package
 - Control Flow
 - Data Flow
 - Connection Manager
- Lesson 6: Data Flow Task
 - Data Flow Task
- Lesson 7: Data Flow Task -Source
 - ADO NET Source
 - Excel Source
 - Flat File Source
 - OLE DB Source
 - Raw File Source
 - Script Component
 - XML Source
- Lesson 8: Data Flow Tasks-Transformation
 - Aggregate
 - Audit
 - Cache Transform
 - Character Map



Copyright © Capgemini 2015. All Rights Reserved 10

Course Content

- Conditional Split
- Copy Column
- Data Conversion
- Data Mining Query
- Derived Column
- Export Column
- Fuzzy Grouping
- Fuzzy Lookup
- Import Column
- Lookup
- Merge
- Merge Join
- Multicast
- OLE DB Command
- Percentage Sampling
- Pivot
- Row Count
- Row Sampling
- Script Component
- Slowly Changing Dimension



Copyright © Capgemini 2015. All Rights Reserved 11

Course Content

- Sort
- Term Extraction
- Term Lookup
- Union All
- Unpivot
- Lesson 9: Data Flow Tasks-Destination
 - ADO NET Destination
 - Data Mining Model Training Destination
 - DataReader Destination
 - Dimension Processing Destination
 - Excel Destination
 - Flat File Destination
 - OLE DB Destination
 - Partition Processing Destination
 - Raw File Destination
 - Recordset Destination
 - Script Component
 - SQL Server Compact Edition Destination
 - SQL Server Destination



Copyright © Capgemini 2015. All Rights Reserved 12

Course Content

- Lesson 10: Operating system-level tasks
 - File System Task
 - FTP Task
 - Send Mail Task
- Lesson 11: SQL Server Tasks
 - Bulk Insert Task
 - Execute SQL Task
 - Transfer Database Task
 - Transfer Error Messages Task
 - Transfer Jobs Task
 - Transfer Logins Task
 - Transfer Master Stored Procedures Task
 - Transfer SQL Server Objects Task
- Lesson 12: Scripting Tasks
 - Script Task
- Lesson 13: Profiling Task
 - Data Profiling Task
- Lesson 14: Workflow Tasks
 - Execute Package Task
 - Execute Process Task
 - Message Queue Task



Copyright © Capgemini 2015. All Rights Reserved 13

Course Content

- Lesson 15: Processing XML
 - XML Task
- Lesson 16: Analysis Services Tasks
 - Analysis Services Processing Task
 - Analysis Services Execute DDL Task
 - Data Mining Query Task
- Lesson 17: Maintenance Tasks
 - Back Up Database Task
 - Check Database Integrity Task
 - Execute SQL Server Agent Job Task
 - Execute T-SQL Statement Task
 - History Cleanup Task
 - Notify Operator Task
 - Rebuild Index Task
 - Reorganize Index Task
 - Shrink Database Task
 - Update Statistics Task
- Lesson 18: Container
 - For each loop
 - For loop
 - Sequence



Copyright © Capgemini 2015. All Rights Reserved 14

Course Content

- Lesson 19: Precedence Constraint
 - Configuring the Precedence Constraint
- Lesson 20: Variables and Expressions
 - System Variables
 - User Defined Variables
 - Configuring Variables
- Lesson 21: SSIS Architecture
 - SSIS Architecture overview
 - Distinguishing between data flow pipeline and package runtime
 - Executing packages on the client side or hosted in the SSIS service
- Lesson 22: Error Handling, Logging and Transactions
 - Event Handler
 - SSIS Logging
 - Check point
- Lesson 23: Execution , debugging and Protection
 - Package execution
 - data Viewer
 - Breakpoints
 - Package Protection



Copyright © Capgemini 2015. All Rights Reserved 15

Course Content

- Lesson 24: Deployment
 - Switching between project and package deployment modes
 - Deploying packages to the SSISDB
 - Running packages from SQL Server
 - Leveraging Project/package parameters
- Lesson 25: Performance Tuning
 - Designing SSIS Packages for High Performance
 - Monitoring the SSIS Performance with Performance Counters
 - Tuning Tips
 - Configure Rows per Batch and Maximum Insert Commit Size in OLEDB destination
- Lesson 26: Best Practices
 - Best Practices for SSIS



Copyright © Capgemini 2015. All Rights Reserved 16

Introduction to ETL and Integration Services

Lesson 1

Overview of BI and Data Warehousing

Business Intelligence

- Business Intelligence covers different technologies for gathering, storing, analyzing and providing access to data.
- BI is a decision support because it simplifies information discovery and analysis, making .

Data Warehousing

- DW contains historical data derived from transaction data.
- Data warehouse is a relational database that is designed for query and analysis rather than for transaction processing
- DW includes ETL,OLAP client analysis tools, and other application



Copyright © Capgemini 2015. All Rights Reserved 18

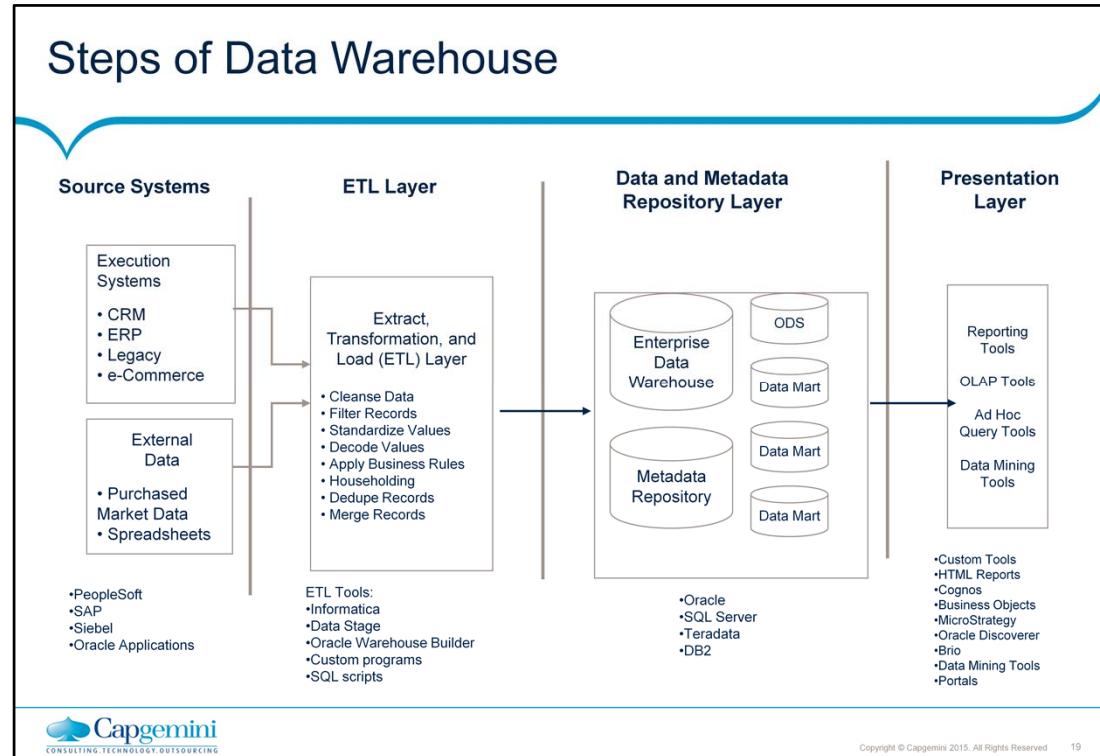
Business Intelligence is an area that covers a number of different technologies for gathering, storing, analyzing and providing access to data that will help a large company make better business decisions.

BI is a decision support system because it simplifies information discovery and analysis, making it possible for decision-makers at all levels of an organization to more easily access, understand, analyze, collaborate, and act on information

DW contains historical data derived from transaction data.

Data warehouse is a relational database that is designed for query and analysis rather than for transaction processing

Data warehouse environment includes an extraction, transportation, transformation, and loading (ETL) solution, an online analytical processing (OLAP) engine, client analysis tools, and other applications that manage the process of gathering data and delivering it to business users.



- **Source System:** From where data is coming to DW system
- **ETL Layer:** Clean ,Validate, Transform data
- **Data and Metadata Repository Layer:** Load data into table
- **Presentation Layer:** Report generation for Business user

ETL Concept

- ETL stands for extraction, transformation and loading. Etl is a process that involves the following tasks:

Extract

- **extracting data** from source operational or archive systems which are the primary source of data for the data warehouse

Transform

- **transforming the data** - which may involve cleaning, filtering, validating and applying business rules

Load

- **loading the data** into a data warehouse or any other database or application that houses data



Copyright © Capgemini 2015. All Rights Reserved 20

➤ **ETL** stands for extraction, transformation and loading. Etl is a process that involves the following tasks:

➤ **extracting data** from source operational or archive systems which are the primary source of data for the data warehouse

➤ **transforming the data** - which may involve cleaning, filtering, validating and applying business rules

➤ **loading the data** into a data warehouse or any other database or application that houses data

ETL Tools & Different providers

- ETL Tools:
- Extract, transform, load tools are software packages that facilitate the performing of ETL tasks.
- At present the most popular and widely used ETL tools and applications on the market are:

IBM Websphere DataStage	Informatica Power center	Oracle warehouse builder
AB Initio	Pentaho Data Integration - Kettle Project (open source ETL)	SAS ETL studio
Cognos Decisionstream	Business Objects Data Integrator (BODI)	Microsoft SQL Server Integration Services (SSIS)



Copyright © Capgemini 2015. All Rights Reserved 21

IBM Websphere DataStage (Formerly known as Ascential DataStage and Ardent DataStage)
Informatica Power center
Oracle warehouse builder
AB Initio
Pentaho Data Integration - Kettle Project (open source ETL)
SAS ETL studio
Cognos Decisionstream
Business Objects Data Integrator (BODI)
Microsoft SQL Server Integration Services (SSIS)

SSIS AND INFORMATICA

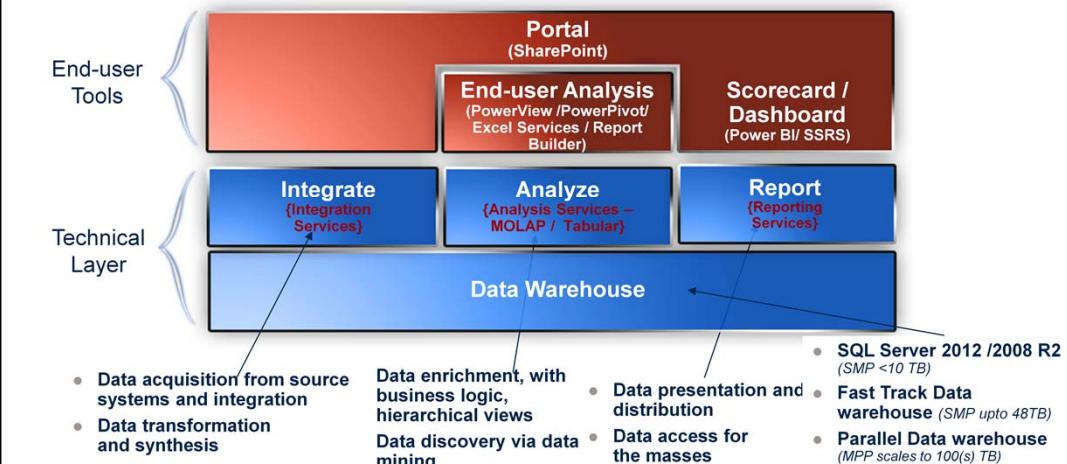
Informatica	SSIS	Description
Aggregator Transformation	Aggregate Transformation	<i>Collates data with alike attributes.</i>
Expression Transformation	Derived Column Transformation	<i>Manipulates data in the transformation pipeline using a data expression language.</i>
Filter Transformation	Conditional Split Transformation	<i>Filters data in the transformation pipeline.</i>
Joiner Transformation	Merge Join Transformation	<i>Amalgamates two streams of data.</i>
Lookup Transformation	Lookup Transformation	<i>Fetches data from a source external to the transformation pipeline.</i>
Mapping	Data flow	<i>Defines how data moves from one place to another. A mapping contains a definition of the source and the target, but doesn't care where they are.</i>
Maplet	n/a	<i>A reusable collection of transformations that operate as a single module of work.</i>
Router Transformation	Conditional Split Transformation	<i>Sends data in the transformation pipeline to different places depending on attributes of that data.</i>
Session	A session is an instantiation of a mapping. There is no direct correlation of a data flow in SSIS.	<i>A session is the instantiation of a mapping and is where the source & target locations are defined. Workflows contain sessions, not mappings.</i>
Sorter Transformation	Sort Transformation	<i>Orders data.</i>
Source	Connection	<i>Represents a source of data.</i>
Source Qualifier	Source Adapter	<i>The component used to fetch data for the transformation pipeline.</i>
Target	Destination Adapter	<i>Represents the place where data in the transformation pipeline will eventually reside.</i>
Transformation	Transformation	<i>A component that applies an operation to some data.</i>
Workflow	Control flow or package	<i>Groups the designed objects into a single, executable, unit of work.</i>
Worklet	Sequence Container	<i>Logically group related components into a single unit of work.</i>

Source – MS whitepaper - Jamie



Copyright © Capgemini 2015. All Rights Reserved 22

Microsoft Business Intelligence Platform Overview



Integration Services: An Introduction

- Integration Services is a platform for building high performance data integration and workflow solutions, including extraction, transformation, and loading (ETL) operations for data warehousing.
- Integration Services includes graphical tools and wizards for building and debugging packages; tasks for performing workflow functions such as FTP operations, SQL statement execution, and e-mail messaging; data sources and destinations for extracting and loading data; transformations for cleaning, aggregating, merging, and copying data; a management service, the Integration Services service, for administering Integration Services packages; and application programming interfaces (APIs) for programming the Integration Services object model.
- Integration Services=>
- Provides functionality commonly referred to as Extract, Transform, Load (ETL).
- Moves and transforms data between sources and destinations, regardless of format.
- Cleanses data and ensures data integrity.
- Integrates heterogeneous data sources.



Copyright © Capgemini 2015. All Rights Reserved 24

SSIS Background

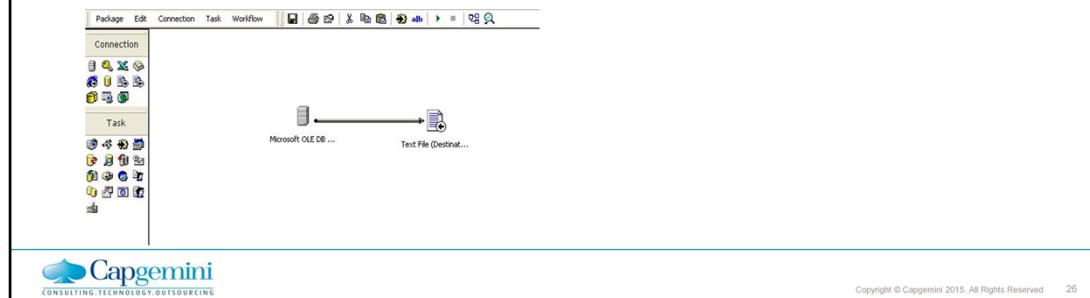
Lesson 2:

SQL Server 2000 DTS Limitations

Offered a limited set of tasks and transformations.

The graphical designer combined the workflow and data flow on a single design surface and offered limited control flow options.

- In DTS, the designer consists of a single pane: the Connections and the Workflow task
- SSIS Data Flow engine is significant faster than DTS Data Pump



DTS provided only a limited set of tasks and transformations. The graphical designer that was used to create packages combined the workflow and data flow on a single design surface and offered limited control flow options.

The DTS and SSIS are both used in SQL Server Transformations. The DTS was used for transformation purpose upto SQL Server 2000 and SSIS (Including SSAS and SSRS) are the advanced to DTS and are included in SQL Server 2005 onwards. The capability of SSIS tool in comparison with DTS is tremendous. This SSIS tool is in competition with other ETL Tools in the present. The SSIS has lot added features as per the current requirements.

Advantages of SSIS 2005, 2008, R2 and 2012,2014,2016 over DTS

The SSIS has lot added features as per the current requirements as compared to DTS

- SSIS, the designer is split into 5 design panes



- SSIS has new many built in tasks which can save significant amount of time compare to manual coding in DTS (e.g. Looping through files in a folder is inbuilt in SSIS - See Foreach Loop Container)
- Making package dynamic at runtime is easier using expression and configurations
- DataViewer support to debug DataFlow
- Event Handlers support with separate control flow for each event
- ScriptTask in SSIS uses VB.net language with rich .net framework support. Script Task code is compiled so run much faster than DTS ActiveX script.
- Expression based control flow along with Success, Failure and Complete

The Evolution of Microsoft Data Platform



Copyright © Capgemini 2015. All Rights Reserved 27

Added features of SSIS 2016

- Better deployment:
- Better debugging
- Better package management
- Expanded connectivity on premises
- Usability and productivity

SQL Server 2016 editions – what's new

SQL Server 2016 is the biggest leap forward in the Microsoft data platform history with real-time operational analytics, rich visualizations on mobile devices, built-in advanced analytics, new advanced security technology, and new hybrid cloud scenarios.

Enterprise	Standard	Express	Developer
<p>Enterprise SQL Server Enterprise delivers comprehensive data management, security, and performance capabilities for mission-critical workloads with minimal IT resources.</p>			
<p>Mission critical performance</p> <ul style="list-style-type: none"> ▪ Operating system max ▪ Enhanced in-memory OLTP performance ▪ Columnstore analytics ▪ Enhanced AlwaysOn with session join (WS 2016) ▪ QueryStore ▪ Temporal <p>Disk-based OLTP</p> <ul style="list-style-type: none"> ▪ 24 cores max and 128 GB ▪ 2-node single database (readable secondary) ▪ Transaction log compression ▪ Temporal <p>Temporal</p> <ul style="list-style-type: none"> ▪ 1 GB memory, max 10 GB memory ▪ Basic OLTP 	<p>Security</p> <ul style="list-style-type: none"> ▪ Always Encrypted ▪ Role-level security ▪ Dynamic data masking ▪ Enhanced separation of duties ▪ Enhanced SQL Server auditing ▪ Transparency data encryption ▪ Policy-based management <p>Data warehousing</p> <ul style="list-style-type: none"> ▪ Row-level security ▪ Dynamic data masking ▪ Role authentication ▪ Separation of duties ▪ Policy-based management <p>Business intelligence</p> <ul style="list-style-type: none"> ▪ PolyBase (compute node only) ▪ Support for JSON <p>Advanced analytics</p> <ul style="list-style-type: none"> ▪ End-to-end mobile BI on ▪ Enhanced in-memory OLTP ▪ Polybase in scale-out configuration (three or more compute nodes) ▪ Deployment rights for APS ▪ Distributed query processing ▪ Support for JSON <p>Hybrid cloud</p> <ul style="list-style-type: none"> ▪ In-database advanced analytics ▪ R integration with massive parallel processing for parallel computing ▪ Works with in-memory OLTP ▪ Run in database or standalone ▪ Pin reports to Power BI ▪ Enhanced multi-tier models ▪ Basic tabular (16GB) ▪ Modifying parallel query ▪ Modernized reports ▪ Pin reports to Power BI ▪ Enhanced multi-tier models ▪ Basic reporting and analytics ▪ Modernized reports 	<p>Express SQL Server Express is a free edition of SQL Server designed for development and test only, and not for production environments or use with production data.</p> <p>Developer SQL Server Developer is now a free edition of SQL Server designed for development and test only, and not for production environments or use with production data.</p>	
			<p>Microsoft</p>

This content was developed prior to the product or service's release and as such, we cannot guarantee that all details included herein will be exactly as what is found in the shipping product. Because Microsoft must support the product or service until its end-of-life date, any changes made to the product or service after publication may not be reflected in this document. The information represents the product or service at the time this document was shared and should be used for planning purposes only. Information subject to change at any time without prior notice.

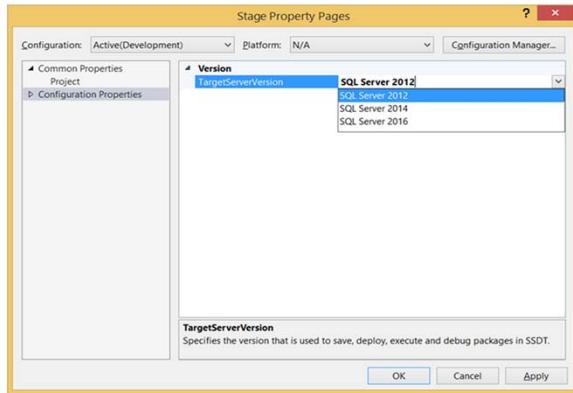


Copyright © Capgemini 2015. All Rights Reserved

28

Added features of SSIS 2016

- SSDT for VS2015 supports SSIS 2012, 2014 and 2016



*Not in current preview/screenshot photoshopped



Copyright © Capgemini 2015. All Rights Reserved 29

When you upgrade SSIS projects from previous versions to the current version, the project-level connection managers continue to work as expected and the package layout and annotations are retained.

Added features of SSIS 2016 - Error column

- Error column
- ErrorCode (integer)
-
- ErrorCode can be translated to a description with a Script Component
- ErrorColumnName?
 - Custom Transformation
 - Looping through copy of package
 - Now similar solution as for the Error description



Copyright © Capgemini 2015. All Rights Reserved 30

<http://microsoft-ssis.blogspot.com/2015/09/sql-server-2016-ctp-23-get-error.html>
// C# Code
public override void Input0_ProcessInputRow(Input0Buffer Row)
{
 Row.ErrorDescription =
 this.ComponentMetaData.GetErrorDescription(Row.ErrorCode);
 Row.ErrorColumnName =
 this.ComponentMetaData.GetIdentificationStringByLineageID(Row.ErrorColumn);
}

Added features of SSIS 2016- OData source and connection manager

- Open Data protocol initiated by Microsoft in 2007 (v1 to v3)
- Standardized in 2014 (v4)
- Available as separate download for SSIS 2012 and 2014
- Included in SSDT CTP 2.3
- Supports v4 protocol



Copyright © Capgemini 2015. All Rights Reserved 31

OData protocol: allows the creation and consumption of queryable and interoperable RESTful APIs in a simple and standard way

<http://microsoft-ssis.blogspot.com/2015/09/sql-server-2016-ctp-23-odata-v4-protocol.html>

Added features of SSIS 2016-New database roles for SSIS

- ssis_monitor - used for AlwaysOn, not designed to use directly
- ssis_logreader - Allows you to see all executions on catalog reports. Alternative for ssis_admin which gives too much rights



Copyright © Capgemini 2015. All Rights Reserved 32

<http://microsoft-ssis.blogspot.com/2015/09/sql-server-2016-ctp-23-new-database.html>

Added features of SSIS 2016-Logging Levels

- 2008 => developers decided what to log
- 2012 => introduction of 4 logging levels None, Performance, Basic and Verbose
- 2016 => Extending the 4 logging levels with your own customized logging level.



Copyright © Capgemini 2015. All Rights Reserved 33

<http://microsoft-ssis.blogspot.com/2015/08/sql-server-2016-ctp-23-custom-logging.html>

Added features of SSIS 2016 - Logging level RuntimeLineage

- Is intended to support a lineage library, which is not yet released.
- Lineage library will probably not included in SSIS, but in other tools.
Probably in Azure Data Factory.



Copyright © Capgemini 2015. All Rights Reserved 34

Azure Data Factory is a guess

Added features of SSIS 2016 –Deployment model

- 2008 => Package Deployment Model
- 2012 => Project Deployment Model
- 2016 => Introduction of Incremental Package Deployment for the Project Deployment Model

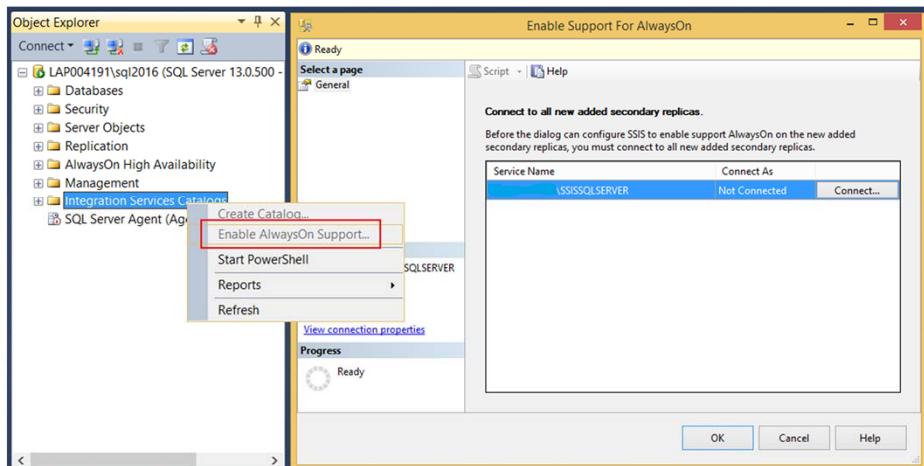


Copyright © Capgemini 2015. All Rights Reserved 35

<http://microsoft-ssis.blogspot.com/2015/06/ssis-2016-ctp2-incremental-deployment.html>

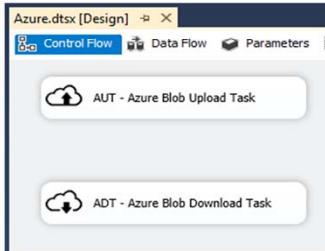
Note: can't deploy project connection managers or parameters

Added features of SSIS 2016 - Support for AlwaysOn



Added features of SSIS 2016 - Azure Pack

- Azure Blob Upload Task
- Azure Blob Download Task
- Azure File System Task (still missing)

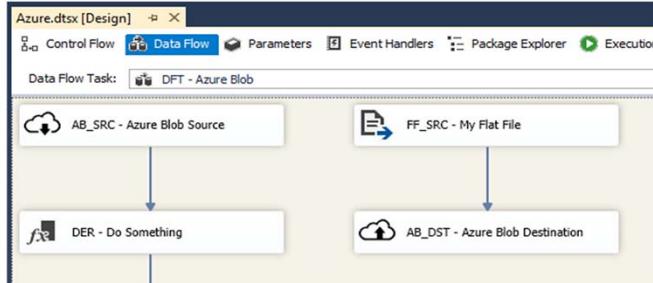


Copyright © Capgemini 2015. All Rights Reserved 37

<http://microsoft-ssis.blogspot.com/2015/06/azure-upload-and-download-tasks.html>

Added features of SSIS 2016 - Azure Pack

- Azure Blob Source
- Azure Blob Destination

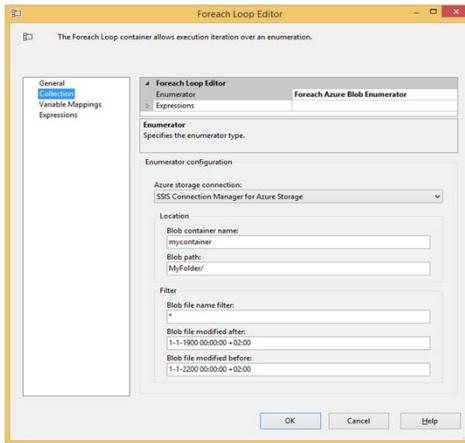


Copyright © Capgemini 2015. All Rights Reserved 38

<http://microsoft-ssis.blogspot.com/2015/06/azure-blob-source-and-destination.html>

Added features of SSIS 2016 - Azure Pack

- Foreach Azure Blob Enumerator

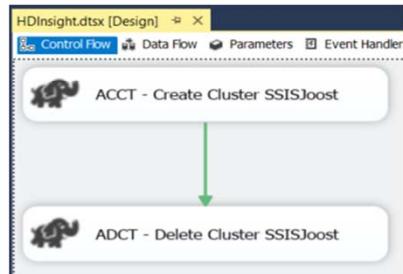


Copyright © Capgemini 2015. All Rights Reserved 39

<http://microsoft-ssis.blogspot.com/2015/06/azure-blob-enumerator.html>

Added features of SSIS 2016 - Azure Pack

- Azure HDInsight Create Cluster Tasks
- Azure HDInsight Delete Cluster Tasks

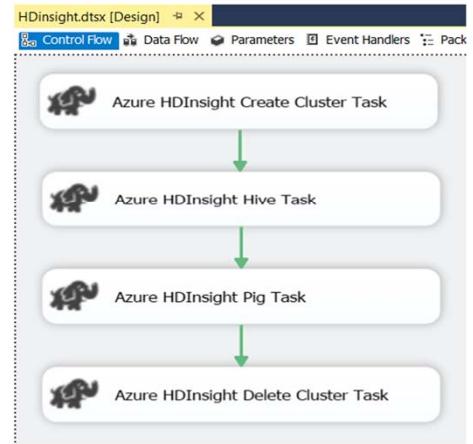


Copyright © Capgemini 2015. All Rights Reserved 40

<http://microsoft-ssis.blogspot.com/2015/06/azure-hdinsight-create-delete-cluster.html>

Added features of SSIS 2016 - Azure Pack

- Azure HDInsight Hive Tasks
- Azure HDInsight Pig Tasks



Copyright © Capgemini 2015. All Rights Reserved 41

Typical Uses of Integration Services

Lesson 3

Typical Uses

Merging Data from Heterogeneous Data Stores

Populating Data Warehouses and Data Marts

Cleaning and Standardizing Data

Building Business Intelligence into a Data Transformation Process

Automating Administrative Functions and Data Loading



Copyright © Capgemini 2015. All Rights Reserved 43

Merging Data from Heterogeneous Data Stores

Data is typically stored in many different data storage systems, and extracting data from all sources and merging the data into a single, consistent dataset is challenging. This situation can occur for a number of reasons. For example:

Many organizations archive information that is stored in legacy data storage systems. This data may not be important to daily operations, but it may be valuable for trend analysis that requires data collected over a long period of time.

Branches of an organization may use different data storage technologies to store the operational data. The package may need to extract data from spreadsheets as well as relational databases before it can merge the data.

Data may be stored in databases that use different schemas for the same data. The package may need to change the data type of a column or combine data from multiple columns into one column before it can merge the data.

Populating Data Warehouses and Data Marts

The data in data warehouses and data marts is usually updated frequently, and the data loads are typically very large.

Integration Services includes a task that bulk loads data directly from a flat file into SQL Server tables and views, and a destination component that bulk loads data into a SQL Server database as the last step in a data transformation process.

An SSIS package can be configured to be restartable. This means you can rerun the package from a predetermined checkpoint, either a task or container in the package. The ability to restart a package can save a lot of time, especially if the package processes data from a large number of sources.

Cleaning and Standardizing Data

Whether data is loaded into an online transaction processing (OLTP) or online analytic processing (OLAP) database, an Excel spreadsheet, or a file, it needs to be cleaned and standardized before it is loaded. Data may need to be updated for the following reasons:

Data is contributed from multiple branches of an organization, each using different conventions and standards. Before the data can be used, it may need to be formatted differently. For example, you may need to combine the first name and the last name into one column.

Data is rented or purchased. Before it can be used, the data may need to be standardized and cleaned to meet business standards. For example, an organization wants to verify that all the records use the same set of state abbreviations or the same set of product names.

Data is locale-specific. For example, the data may use varied date/time and numeric formats. If data from different locales is merged, it must be converted to one locale before it is loaded to avoid corruption of data.

Building Business Intelligence into a Data Transformation Process

A data transformation process requires built-in logic to respond dynamically to the data it accesses and processes.

The data may need to be summarized, converted, and distributed based on data values. The process may even need to reject data, based on an assessment of column values.

To address this requirement, the logic in the SSIS package may need to perform the following types of tasks:

Merging data from multiple data sources.

Evaluating data and applying data conversions.

Splitting a dataset into multiple datasets based on data values.

Applying different aggregations to different subsets of a dataset.

Loading subsets of the data into different or multiple destinations.

Integration Services provides containers, tasks, and transformations for building business intelligence into SSIS packages.

Automating Administrative Functions and Data Loading

Administrators frequently want to automate administrative functions such as backing up and restoring databases, copying SQL Server databases and the objects they contain, copying SQL Server objects, and loading data. Integration Services packages can perform these functions.

Integration Services includes tasks that are specifically designed to copy SQL Server database objects such as tables, views, and stored procedures; copy SQL Server objects such as databases, logins, and statistics; and add, change, and delete SQL Server objects and data by using Transact-SQL statements.

Administration of an OLTP or OLAP database environment frequently includes the loading of data. Integration Services includes several tasks that facilitate the bulk loading of data. You can use a task to load data from text files directly into SQL Server tables and views, or you can use a destination component to load data into SQL Server tables and views after applying transformations to the column data.

Tools

Lesson 4

Tools

- SQL Server Data Tools
- BIDS can perform the following tasks:

SQL Server Data Tools
creates packages that include complex control flow, data flow, event-driven logic, and logging.

Test and debug packages

Create configurations

Create a deployment utility that can install packages

Save copies of packages to the SQL Server msdb database, the SSIS Package Store,



Copyright © Capgemini 2015. All Rights Reserved 45

Microsoft SQL Server includes two studios for working with Integration Services:
SQL Server Data Tools, for developing the Integration Services packages that a business solution requires, and SQL Server Management Studio, for managing packages in a production environment.

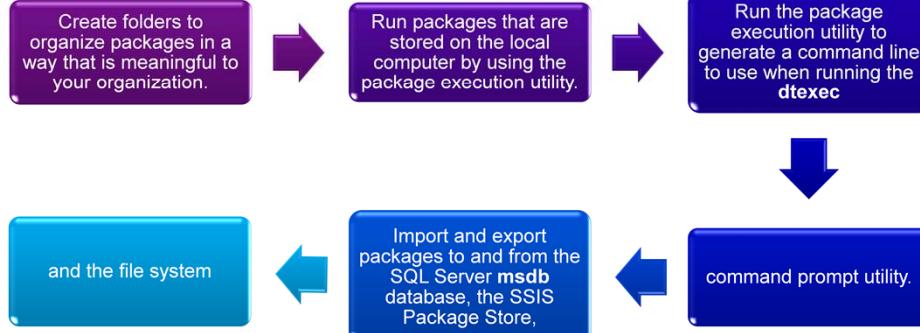
SQL Server Data Tools provides the Integration Services project in which you create packages, their data sources, and data source views.

Working in you can perform the following tasks:

- Run the SQL Server Import and Export Wizard to create basic packages that copy data for a source to a destination.
- Create packages that include complex control flow, data flow, event-driven logic, and logging.
- Test and debug packages by using the troubleshooting and monitoring features in SSIS Designer, and the debugging features in Business Intelligence Development Studio.
- Create configurations that update the properties of packages and package objects at run time.
- Create a deployment utility that can install packages and their dependencies on other computers.
- Save copies of packages to the SQL Server **msdb** database, the SSIS Package Store, and the file system.

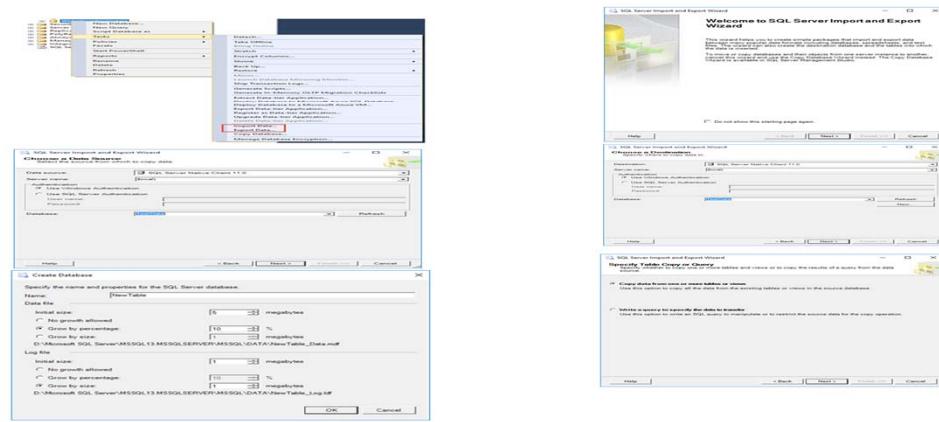
Tools...

- SQL Server Management Studio provides the Integration Services service to manage packages and monitor running packages.
- It performs following tasks



TOOLS.....

- Import and Export Wizard**
 - The SQL Server Import and Export Wizard provides the simplest method of copying data between data sources and of constructing basic packages



Capgemini
CONSULTING TECHNOLOGY OUTSOURCING

Copyright © Capgemini 2015. All Rights Reserved 47

Third tools is Import and Export Wizard . The SQL Server Import and Export Wizard provides the simplest method of copying data between data sources and of constructing basic packages.

Here we will see that how can we load data using Import export wizard.

To start the SQL Server Import and Export Wizard

On the Start menu, point to All Programs, point to Microsoft SQL Server 2008, and then click Import and Export Data.

—or—

In Business Intelligence Development Studio, right-click the SSIS Packages folder, and then click SSIS Import and Export Wizard.

—or—

In Business Intelligence Development Studio, on the Project menu, click SSIS Import and Export Wizard.

—or—

In SQL Server Management Studio, connect to the Database Engine server type, expand Databases, right-click a database, point to Tasks, and then click Import Data or Export data.

—or—

In a command prompt window, run DTWizd.exe, located in C:\Program Files\Microsoft SQL Server\100\DT\Binn.

To use the SQL Server Import and Export Wizard for importing and exporting data

Start the SQL Server Import and Export Wizard.

On the corresponding wizard pages, select a data source and a data destination.

In our example Browse and choose the Customers.xls which is located in the Sample Data Folder Verify that the Excel version box contains Microsoft Excel 97-2005 and the First Row has column names check box is selected

On the Choose Destination page, in the Destination list, Select SQL Native Client, and in the Server name box, type localhost.

In the Database list, select AdventureWorks.

On the Specify Table Copy or Query page, click Write a query to specify the data to transfer and click Next If you want to copy it from a table or views then click the other option.

On the Provide a Source Query page, in the SQL statement box,type the query.

Click Next

On the Select Source Tables and Views page, click [AdventureWorks].[dbo].[Query] in the Destination list, and change the table name, Query, to Prospective Customers.

To edit column metadata and table options, click Edit Mappings

On the Columns Mappings page, verify that the Create Destination table option is selected, select the Drop and re-create destination table check box, and modify the metadata of the destination columns

On the Complete the Wizard page, review information about the new package and click Finish

On the Performing Operations page, view the actions that the wizard performs. When finished, the Status column for each action should display Success

TOOLS.....

- Import and Export Wizard

The screenshot shows the SQL Server Import and Export Wizard interface with four main steps:

- Detect Source Tables and Views:** Shows a list of tables from the Adventureworks database.
- Map Data Types:** Shows data type mappings between source and destination columns.
- Convert Types Without Conversions:** Shows conversion details for specific columns.
- Complete the Wizard:** Summary page with options to run immediately or schedule the package.

Third tools is Import and Export Wizard . The SQL Server Import and Export Wizard provides the simplest method of copying data between data sources and of constructing basic packages.

Here we will see that how can we load data using Import export wizard.

To start the SQL Server Import and Export Wizard

On the Start menu, point to All Programs, point to Microsoft SQL Server 2008, and then click Import and Export Data.

--or--

In Business Intelligence Development Studio, right-click the SSIS Packages folder, and then click SSIS Import and Export Wizard.

--or--

In Business Intelligence Development Studio, on the Project menu, click SSIS Import and Export Wizard.

--or--

In SQL Server Management Studio, connect to the Database Engine server type, expand Databases, right-click a database, point to Tasks, and then click Import Data or Export data.

--or--

In a command prompt window, run DTSPWIZD.exe, located in C:\Program Files\Microsoft SQL Server\100\DTSP\Binn.

To use the SQL Server Import and Export Wizard for importing and exporting data

Start the SQL Server Import and Export Wizard.

On the corresponding wizard pages, select a data source and a data destination.

In our example Browse and choose the Customers.xls which is located in the Sample Data Folder

Verify that the Excel version box contains Microsoft Excel 97-2005 and the First Row has column names check box is selected

On the Choose Destination page, in the Destination list, Select SQL Native Client, and in the Server name box, type localhost.

In the Database list, select AdventureWorks.

On the Specify Table Copy or Query page, click Write a query to specify the data to transfer and click Next If you want to copy it from a table or views then click the other option.

On the Provide a Source Query page, in the SQL statement box,type the query.

Click Next

On the Select Source Tables and Views page, click [AdventureWorks].[dbo].[Query] in the Destination list, and change the table name, Query, to Prospective Customers.

To edit column metadata and table options, click Edit Mappings

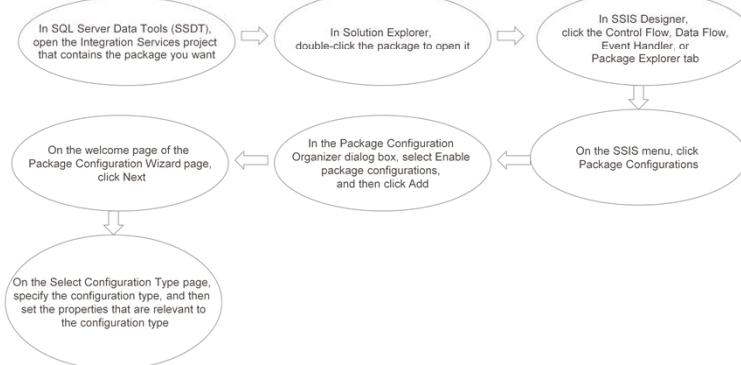
On the Columns Mappings page, verify that the Create Destination table option is selected, select the Drop and re-create destination table check box, and modify the metadata of the destination columns

On the Complete the Wizard page, review information about the new package and click Finish

On the Performing Operations page, view the actions that the wizard performs. When finished, the Status column for each action should display Success

Package Configuration Wizard

- Use the Package Configuration Wizard to create configurations that update the properties of an Integration Services package and its objects at run time. This wizard runs when you add a new configuration or modify an existing one in the Package Configurations Organizer dialog box. To open the Package Configurations Organizer dialog box, select Package Configurations on the SSIS menu in SQL Server Data Tools (SSDT).



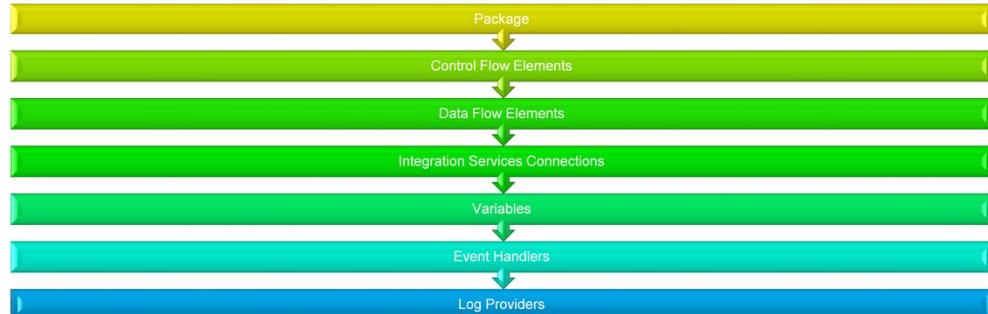
Copyright © Capgemini 2015. All Rights Reserved 49

SSIS Objects

Lesson 5:

Objects and Concepts

Topic	Description
Integration Services Packages	Describes packages.
Control Flow Elements	Describes containers, tasks, and precedence constraints.
Data Flow Elements	Describes sources, transformations, destinations, and paths.
Integration Services Connections	Describes connection managers.
Integration Services Variables	Describes user-defined variables and the system variables that Integration Services provides.
Integration Services Event Handlers	Describes the run-time events and the event handlers you can build for the events.
Integration Services Log Providers	Describes the log providers that Integration Services includes.



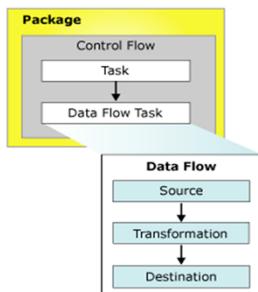
Copyright © Capgemini 2015. All Rights Reserved 51

To start SSIS in details like how to create SSIS project, first you should be familiar with some Integration Services concepts to help you create Integration Services packages successfully from the very beginning. These objects and concepts are relevant to the use of the Integration Services tools. Those concepts include the following:

- Package
- Control Flow Elements
- Data Flow Elements
- Integration Services Connections
- Variables
- Event Handlers
- Log Providers

Package..

- A package is an organized collection of connections, control flow elements, data flow elements, event handlers, variables, and configurations.,
- The package is the unit of work that is retrieved, executed, and saved.
- The following diagram shows a simple package that contains a control flow with a Data Flow task, which in turn contains a data flow.



Copyright © Capgemini 2015. All Rights Reserved 52

A package is an organized collection of connections, control flow elements, data flow elements, event handlers, variables, and configurations, that you assemble using either the graphical design tools that SQL Server Integration Services provides, or build programmatically. We then save the completed package to SQL Server, the SSIS Package Store, or the file system.

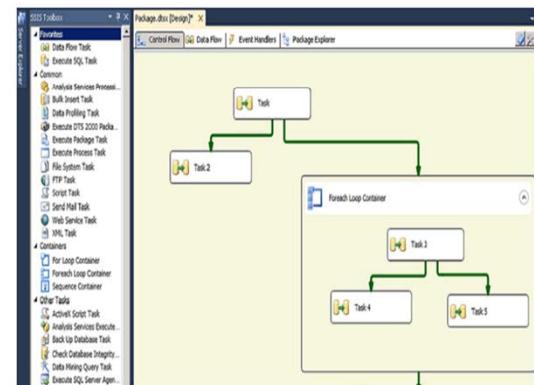
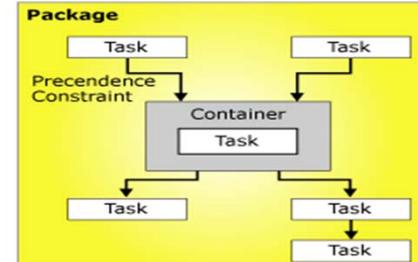
When we first create a package, it is an empty object that does nothing. To add functionality to a package, we add a control flow and, optionally, one or more data flows to the package.

AS diagram shows a simple package with a control flow with a Data Flow task, which in turn contains a data flow.

In next slides we will go through with control flow and data flow.

Control Flow Elements

- Control flow consists of below components:
 - Create the control flow in a package by using the Control Flow tab in SSIS Designer.
 - When the Control Flow tab is active, the Toolbox lists the tasks and containers that are available for control flow.

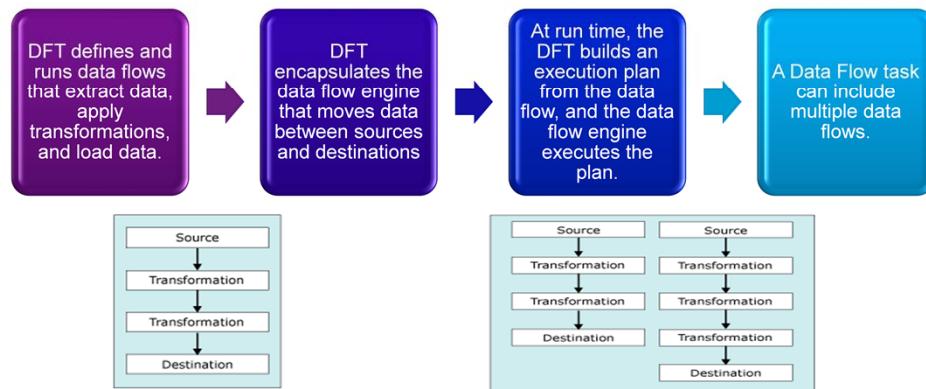


➤A **[control flow](#)** consists of the tasks ,containers and **precedence constraints** . **containers** that provide structures in packages.

The tasks perform specific types of work such as executing SQL statements or sending email messages, and the containers define repeating subsets of the control flow or group subsets of the control flow to make the package easier to manage. The tasks and containers are usually connected by precedence constraints that specify the sequence in which tasks and containers are executed and the conditions that must be satisfied to run the next task or container in the control flow.

Tasks

- Task: Tasks do the work in packages. Integration Services includes tasks for performing a variety of functions.
- The Data Flow task



Tasks do the work in packages. Integration Services includes tasks for performing a variety of functions.

The Data Flow task encapsulates the data flow engine that moves data between sources and destinations, and lets the user transform, clean, and modify data as it is moved. Addition of a Data Flow task to a package control flow makes it possible for the package to extract, transform, and load data.

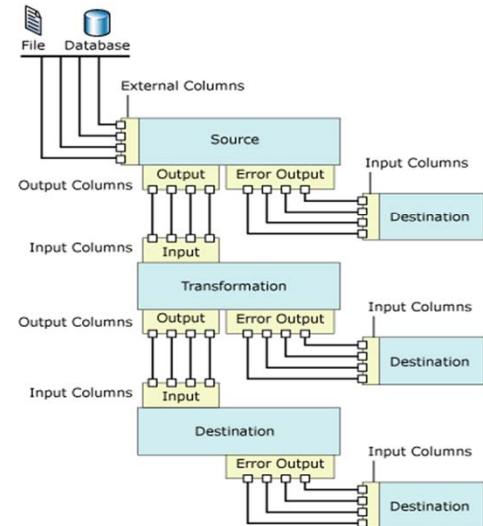
A data flow consists of at least one data flow component, but it is typically a set of connected data flow components: sources that extract data; transformations that modify, route, or summarize data; and destinations that load data. Components are connected in the data flow by paths. Each path specifies the two components that are the start and the end of the path.

A Data Flow task can include multiple data flows. If a task copies several sets of data, and if the order in which the data is copied is not significant, it can be more convenient to include multiple data flows in the Data Flow task. For example, you might create five data flows, each copying data from a flat file into a different dimension table in a data warehouse star schema.

However, the data flow engine determines order of execution when there are multiple data flows within one data flow task. Therefore, when order is important, the package should use multiple Data Flow tasks, each task containing one data flow. You can then apply precedence constraints to control the execution order of the tasks.

Data Flow Elements

- Data flow components



Copyright © Capgemini 2015. All Rights Reserved 55

Definition of Connection Manager

- “A Connection Manager is a logical representation of a connection”. At design time, one sets its properties to describe the physical connection that SSIS creates when the package runs. For instance, a connection manager includes the Connection String property that you set at design time; at run time, a physical connection is created using the value in the connection string property.
- NOTE: A package can use multiple instances of a connection manager type, and you can set the properties on each instance. At run time, each instance of a connection manager type creates a connection that has different attributes.



Copyright © Capgemini 2015. All Rights Reserved 56

Connection Managers

- The following table(s) show the various connections that can be created using Connection Manager:

Type	Description
ADO	Connects to ActiveX Data Objects (ADO) objects.
ADO.NET	Connects to a data source by using a .NET provider.
EXCEL	Connects to an Excel workbook file.
FILE	Connects to a file or a folder.
FLATFILE	Connect to data in a single flat file.
FTP	Connect to an FTP server.
HTTP	Connects to a web server.
MSMQ	Connects to a message queue.
MSOLAP90	Connects to an instance of SQL Server 2008 Analysis Services (SSAS) or an Analysis Services project.

Type	Description
MULTIFILE	Connects to multiple files and folders.
MULTIPLATFILE	Connects to multiple data files and folders.
OLEDB	Connects to a data source by using an OLE DB provider.
ODBC	Connects to a data source by using ODBC.
SMO Server	Connects to a SQL Management Objects (SMO) server.
SMTP	Connects to an SMTP mail server.
SQLMOBILE	Connects to a SQL Server Mobile database.
WMI	Connects to a server and specifies the scope of Windows Management Instrumentation (WMI) management on the server.



Copyright © Capgemini 2015. All Rights Reserved 57

Connection Managers

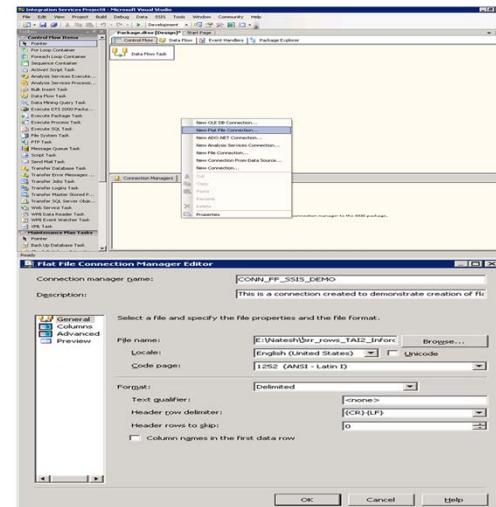
- The following sections will explain at length, the method(s) to create the below mentioned types of connections:
- FLAT FILE
- EXCEL
- OLEDB
- ODBC



Copyright © Capgemini 2015. All Rights Reserved 58

Flat File Connection

- Right-click in the “Connection Managers” pane and select “New Flat File Connection...”
- Completion of step 1 opens the “Flat File Connection Manager Editor”. Select the “General” section if it’s not the default.
- Specify the ---
- Name of this Connection Manager,
- Description,
- File Name



Copyright © Capgemini 2015. All Rights Reserved 59

Flat File Source

- Support for varying number of columns

OrderID,CustomerID,EmployeeID,ShipVia
10248,VINET,5,3
10249,TOMSP,6
10250,HANAR,4,2

OrderID	CustomerID	EmployeeID	ShipVia
10248	VINET	5	3
10249	TOMSP	6	
10250	HANAR	4	2

- Embedded qualifiers

ID,Title
1148,'Can''t Buy a Thrill'
1149,'Echoes, Silence, Patience & Grace'

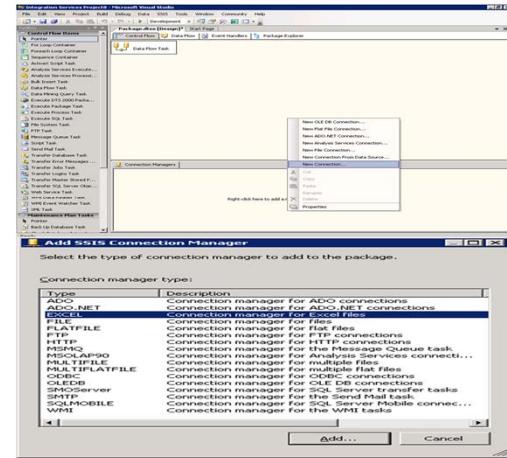
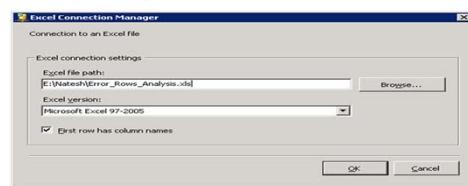
ID	Title
1148	Can't buy a thrill
1149	Echoes, Silence, Patience & Grace



Copyright © Capgemini 2015. All Rights Reserved 60

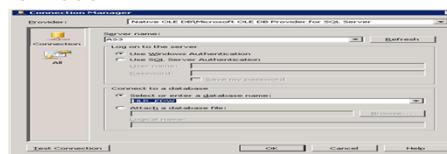
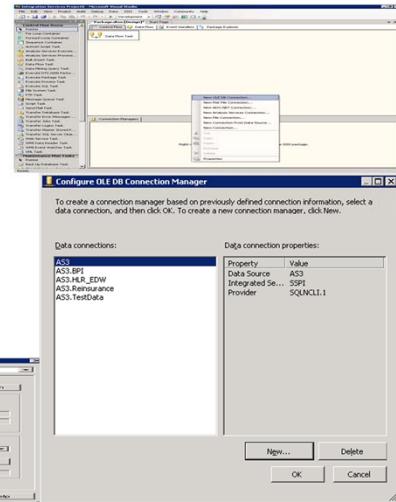
Excel Connection

- Right-click in the “Connection Managers” pane and select “New Connection...”
- Completion of STEP 1 opens the “Add SSIS Connection Manager” dialog box. Select the type as “EXCEL”
- Specify the Excel File Path



OLE DB Connection

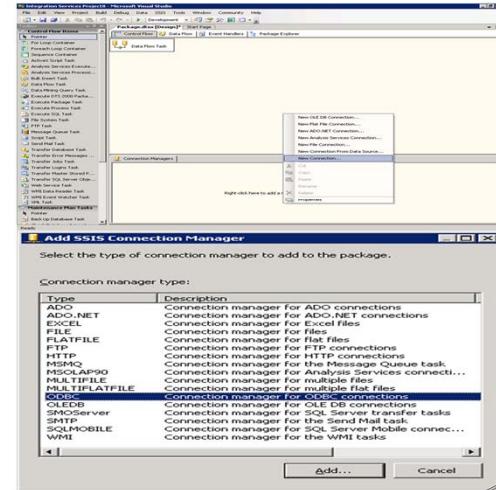
- Right-click in the “Connection Managers” pane and select “New OLE DB Connection...”
- Completion of STEP 1 opens the “Configure OLE DB Connection Manager” dialog box. Select New if you would like to have something different from the existing ones (if any)
- When one selects New, a new dialog box will be opened for Selecting a provider & its respective information
- Select ‘OK’



Copyright © Capgemini 2015. All Rights Reserved 62

ODBC Connection

- Right-click in the “Connection Managers” pane and select “New Connection...”



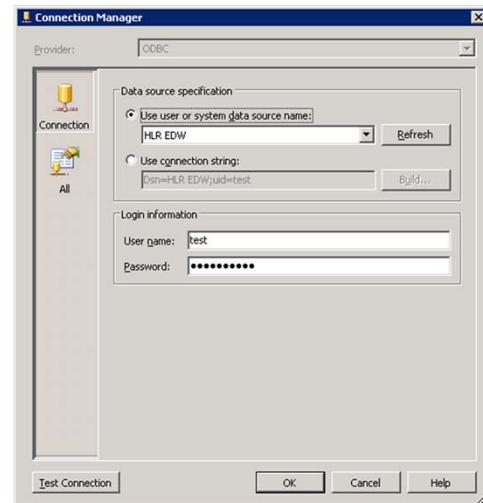
- Completion of STEP 1 opens the “Add SSIS Connection Manager” dialog box. Select the type as “ODBC”. Then Click on “Add..”



Copyright © Capgemini 2015. All Rights Reserved 63

ODBC Connection

- Completion of STEP 2 opens the “Configure ODBC Connection Manager” dialog box. Select New if you would like to have something different from the existing ones (if any).
- When one selects New, a new dialog box will be opened for specifying User or System DSN & Login Information
- Select ‘OK’ to successfully complete creating the connection



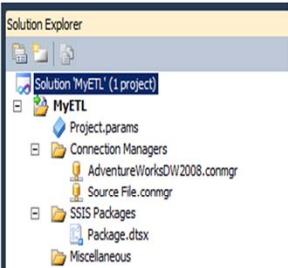
Copyright © Capgemini 2015. All Rights Reserved 64

ODBC Connection

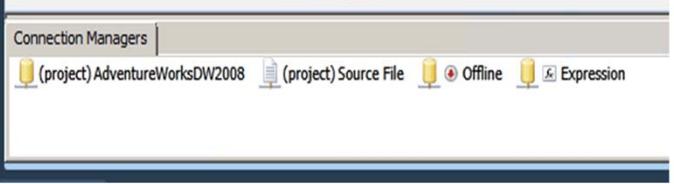
- Explicit addition of ODBC Source and Destination task in data flow tool box



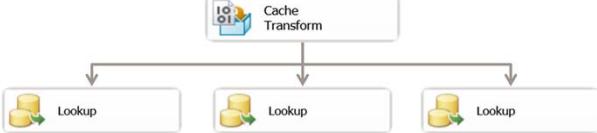
Connection Managers



Offline Validation & Expression Markers



Share In-Memory Cache Across Packages



Shared Connection Managers

 Capgemini
CONSULTING TECHNOLOGY OUTSOURCING

Copyright © Capgemini 2015. All Rights Reserved 66