Name: Girum Obse
Date: 02/01/2025
Course: BIDD 320 A /Data Migration Techniques (ETL Processing)
Assignment: A03_G.Obse

SSIS ETL Process - Technical Manual

Table of Contents

---

1. Introduction

The purpose of this document is to provide a comprehensive guide to the SSIS (SQL Server Integration Services) ETL (Extract, Transform, Load) process. The focus is on the ETLProcess.dtsx package, which extracts data from the AdventureWorks_Basics database, transforms it according to the required business logic, and loads the data into the DWAdventureWorks_Basics data warehouse. This guide is designed for new hires and technicians who need to understand and troubleshoot the ETL process.

Overview

The ETL process follows these steps:

- **Extract**: Data is extracted from operational systems (source tables in the AdventureWorks_Basics database).
- **Transform**: Data is cleansed, transformed, and enriched based on business rules.
- **Load**: Transformed data is loaded into the DWAdventureWorks_Basics data warehouse.

The ETLProcess.dtsx package is designed to handle these operations efficiently, ensuring that data is accurately moved and transformed for reporting and analytics purposes.

---

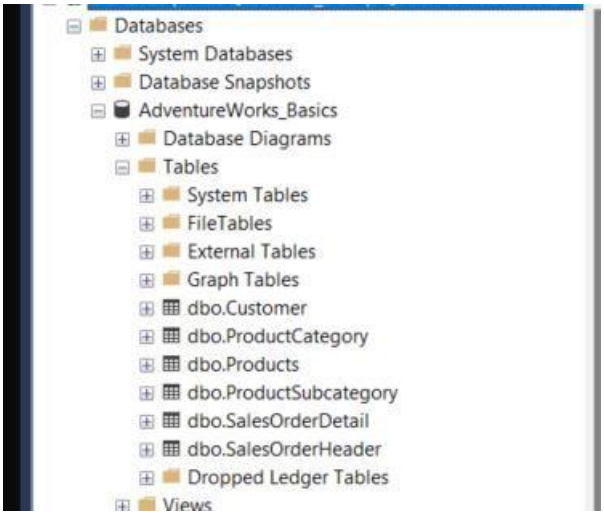
2. Source Tables in SQL server

The source tables reside in the AdventureWorks_Basics database within SQL Server Management Studio (SSMS). These tables contain transactional and operational data, which serves as input for the ETL process.

Source Tables Overview

The following tables are the primary source tables:

- **DimProduct**: Contains details about products sold in the business.
- **DimCustomer**: Stores information about customers.
- **DimDate**: Holds date-related data for time-based analysis.
- **FactSalesOrder**: Stores transactional data related to sales orders.



Screenshot of source databse

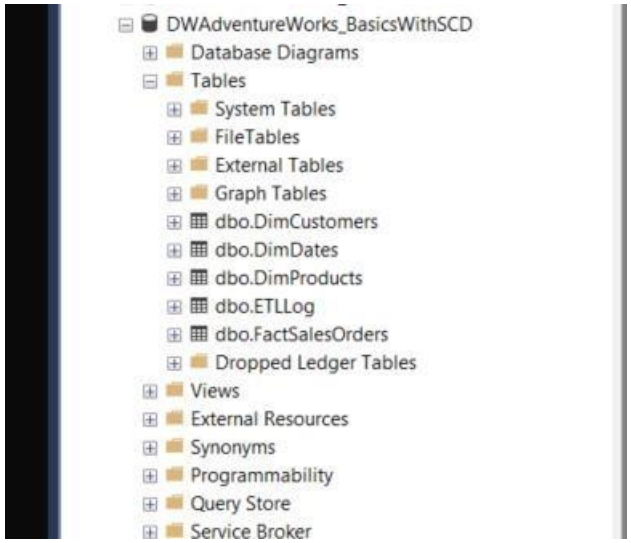
Steps to Access and Validate Source Tables:

1. **Restore Database**: The AdventureWorks_Basics database should be restored using the script 1-Restore AdventureWorks_Basics Script-AsNeeded.sql. This script ensures that the database is set up with the required data and schema.
2. **Validate Table Structures and Relationships**:
    - Open the AdventureWorks_Basics database in SSMS. ○ Examine table structures, ensuring that all foreign key relationships and indexes are properly defined.
    - Verify data integrity by checking that tables are populated and consistent.
3. **Data Validation**:
    - Before the ETL process begins, validate that the data is clean and ready for transformation.
    - Ensure that no missing or corrupted data exists in the source tables.

---

3. Data Warehouse in SQL server

The data warehouse, DWAdventureWorks_Basics, is where transformed data is loaded and stored for reporting and analytics.

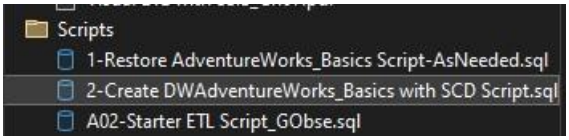


Screnshot showing datawarehouse in SQL server

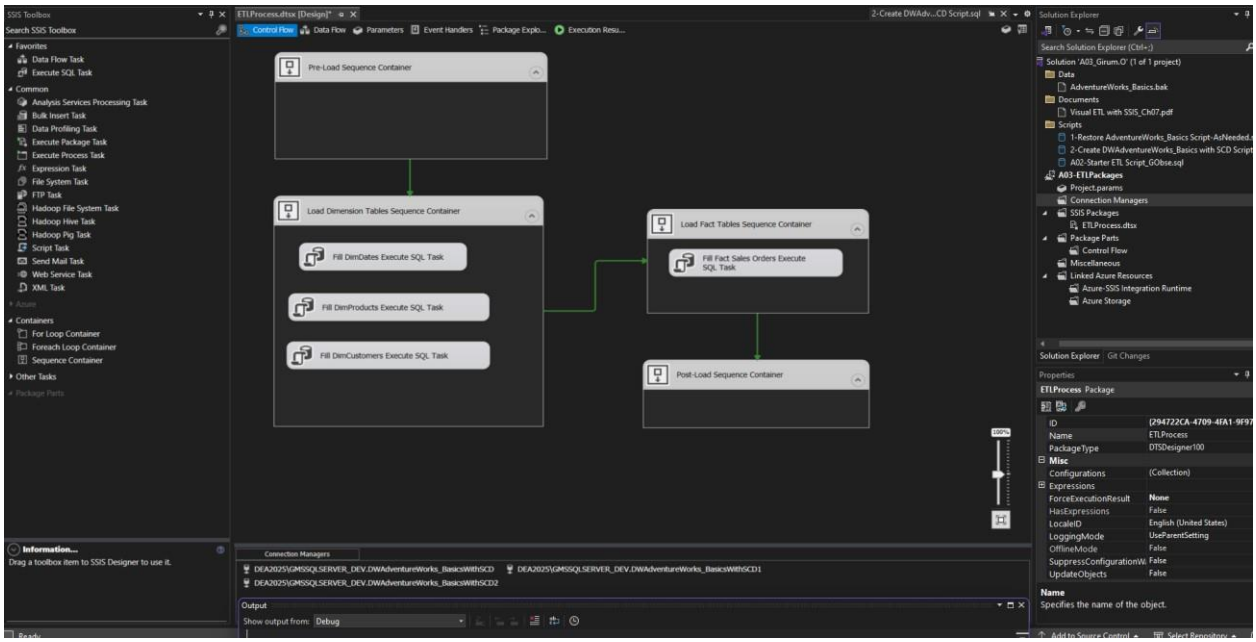
Steps to Create and Set Up the Data Warehouse:

1. **Create Data Warehouse**: Execute the script 2-Create DWAdventureWorks_Basics with SCD.sql to create the data warehouse. This script sets up the database with necessary tables and slowly changing dimensions (SCDs) to support historical tracking.

Screenshot showing the scripts that can be executed in visual studio.



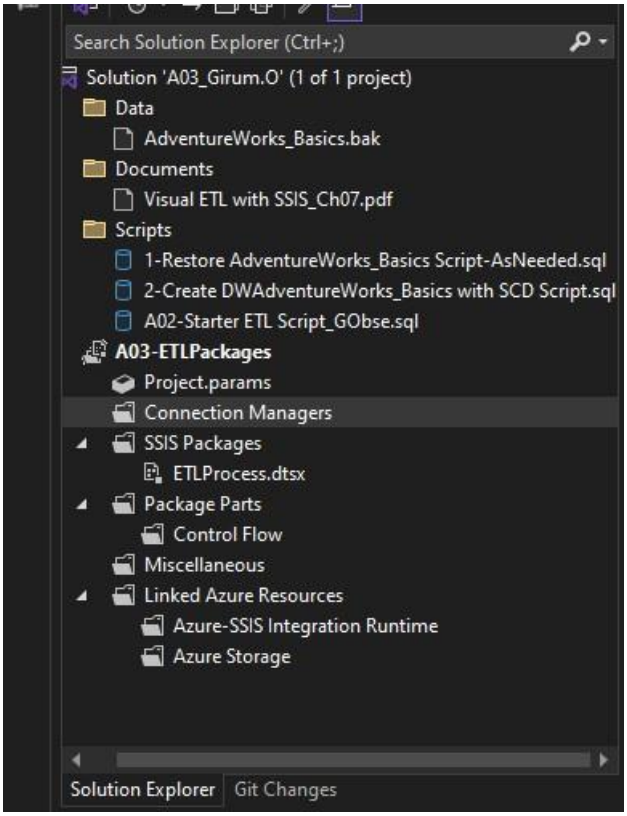
4. Project Setup in Visual Studio



Here's a short summary of the screenshots in Visual Studio:

1. **Package Overview: Displays the ETLProcess.dtsx package structure, including all tasks and flows.**
2. **Control Flow: Shows tasks such as Execute SQL Tasks, Data Flow Tasks, and Sequence Containers.**
3. **Sequence Containers: Illustrates the four sequence containers (Pre-Load, Load Dimension Tables, Load Fact Tables, Post-Load).**
4. **Connection Managers: Shows the configured OLE DB connections for both source (AdventureWorks_Basics) and target (DWAdventureWorks_Basics) databases.**

These screenshots provide a concise view of the SSIS package setup and execution process.



This screenshot shows all components in the Solution Explorer, including the SSIS project, package files, and connection managers.

To begin working with SSIS, the project must be set up in Visual Studio. Follow the steps below to create and configure the SSIS project.

Steps:

1. **Create a Blank Solution**: Launch Visual Studio and create a new blank solution. Name it A03_YourNameHere.
2. **Add an SSIS Project**: In the solution, add a new SSIS project titled A03_ETLPackages.
3. **Rename the Default Package**: By default, Visual Studio creates a package named Package.dtsx. Rename it to ETLProcess.dtsx to reflect the purpose of the package.
4. **Add SSIS Components**: From the SSIS toolbox, drag the necessary components to create the ETL workflow. Components include Data Flow Tasks, Execute SQL Tasks, Sequence Containers, and Connection Managers.

5.Control Flow

The Control Flow in SSIS orchestrates the ETL operations, ensuring that tasks are executed in the correct order and with proper dependencies.

Key Components of Control Flow:

5. **Execute SQL Tasks**: These tasks are used to execute stored procedures or SQL queries to manage the data extraction and transformation process.
6. **Sequence Containers**: Sequence containers help to organize tasks logically and can be used to group related operations together.
7. **Precedence Constraints**: These constraints define the execution order of tasks, ensuring that tasks are executed based on the success or failure of previous tasks.

Example:

- **Execute SQL Task** for extracting data from the source.
- **Data Flow Task** for transforming and loading the data.
- **Execute SQL Task** for post-load validation or cleanup.

---

1. **OLE DB Source**: This component is used to extract data from the source database (AdventureWorks_Basics). The connection is configured through a Connection Manager.
2. **Transformations**:
    - **Derived Column Transformation**: Used to create new columns or modify existing ones based on expressions.
    - **Lookup Transformation**: Used to find matching records from other data sources or tables for enrichment.
    - **Data Conversion**: Converts data types as needed during the transformation.
3. **OLE DB Destination**: This component is used to load the transformed data into the target data warehouse (DWAdventureWorks_Basics).

---

7. Sequence Containers

Sequence containers group tasks together for better organization and error handling. Four sequence containers are defined in the ETLProcess.dtsx package.

Containers:

1. **Pre-Load Sequence Container**: This container ensures that any necessary environment setup, such as truncating destination tables or preparing staging areas, is done before the data load begins.
2. **Load Dimension Tables Sequence Container**: This container handles the loading of dimension tables (e.g., DimProduct, DimCustomer, DimDate).
3. **Load Fact Tables Sequence Container**: This container is responsible for loading fact tables, such as FactSalesOrder.
4. **Post-Load Sequence Container**: After data loading is complete, this container handles finalization tasks, such as logging and cleanup.

---

8. Connection Managers

Connection Managers are used to manage connections between SSIS and the SQL Server databases.

Steps to Configure Connection Managers:

1. **AdventureWorks_Basics Connection**:
    - Define an OLE DB Connection Manager to connect to the source database AdventureWorks_Basics.
    - Configure the server name, authentication method, and database name.
2. **DWAdventureWorks_Basics Connection**:
    - Define another OLE DB Connection Manager for the target data warehouse DWAdventureWorks_Basics. ○ Ensure that the connection is properly configured and tested before running the package.
3. **Test Connections**: Always test the connections after configuration to ensure that SSIS can connect to both source and destination databases.

---

9. Execution and Troubleshooting

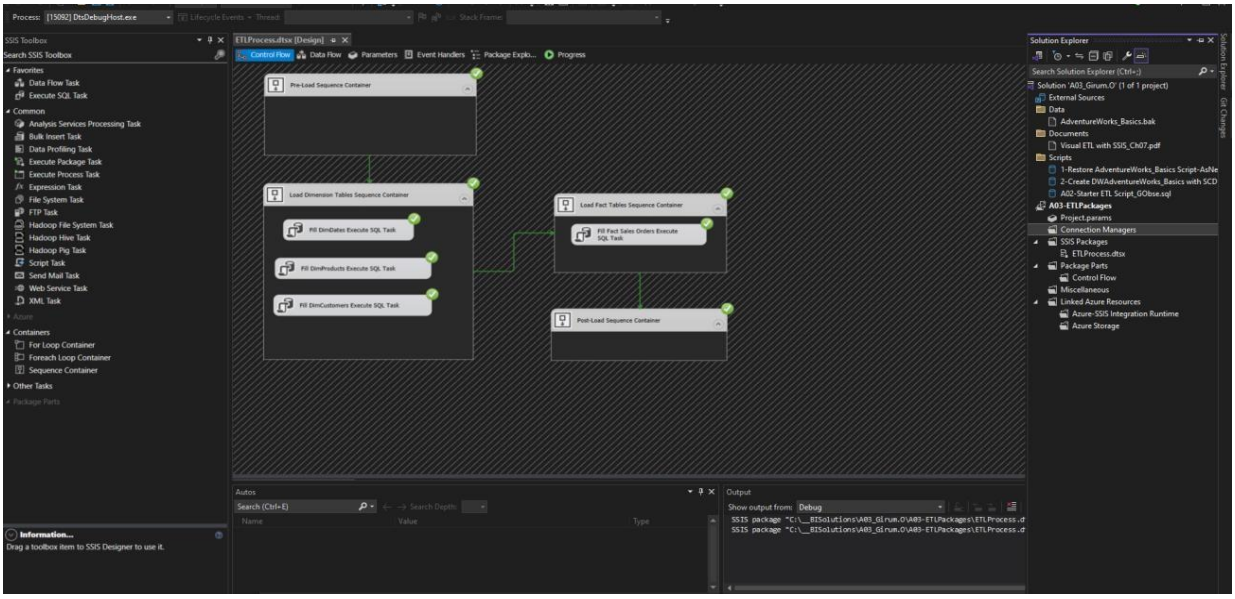
Executing the SSIS package involves running it in Visual Studio and monitoring the progress.

Execution Steps:

1. **Open Package**: Open the ETLProcess.dtsx package in Visual Studio.
2. **Execute the Package**: Click on the Start Debugging button to execute the SSIS package.
3. **Monitor Progress**: Monitor the progress in the Execution Results window. If the package completes successfully, verify that data is loaded into DWAdventureWorks_Basics.

Common Issues and Solutions:

Issue	Possible Cause	Solution
Connection Failure	Incorrect credentials or server issues	Verify credentials and server settings in the Connection Manager.
Data Mismatch	Errors in transformation logic	Review transformation steps and mappings in the Data Flow.
ETL Process Failure	Missing stored procedures or permissions	Check the existence and permissions of necessary stored procedures.



This screenshot shows all green indicators, indicating the SSIS package executed successfully with no errors.

Summery

This manual provides a detailed guide to the SSIS (SQL Server Integration Services) ETL (Extract, Transform, Load) process using the ETLProcess.dtsx package. It explains how data is extracted from the AdventureWorks_Basics database, transformed based on business rules, and loaded into the DWAdventureWorks_Basics data warehouse for reporting and analytics. The guide is aimed at new hires and technicians who need to understand and troubleshoot the ETL process. This technical manual serves as a comprehensive resource for understanding, implementing, and troubleshooting the SSIS ETL process for the DWAdventureWorks_Basics data warehouse, ensuring data is accurately moved and transformed for reporting and analysis.