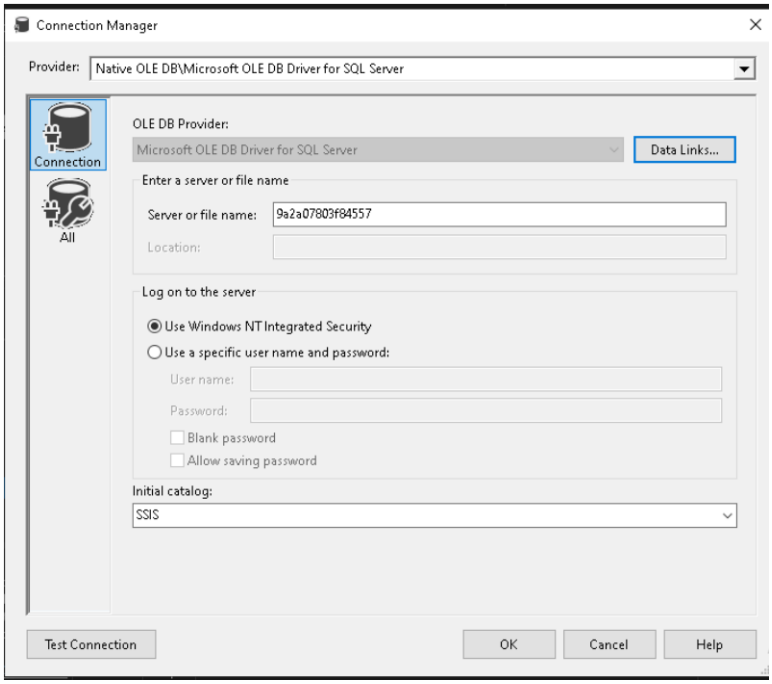
ASSIGNMENT 9

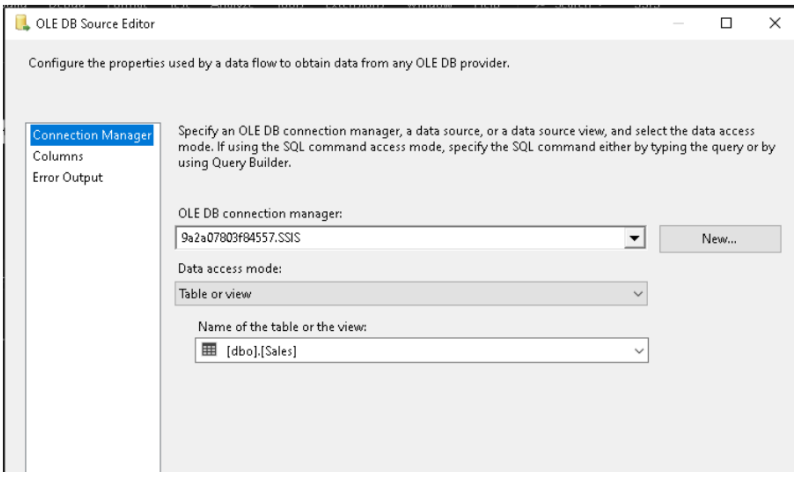
SSIS

Task 1: Integration with ETL Data Warehouse (DWH) Scenario: Your company has a data warehouse designed to consolidate data from various sources for analytical purposes. You need to create an SSIS package that extracts data from a transactional database and loads it into the data warehouse.

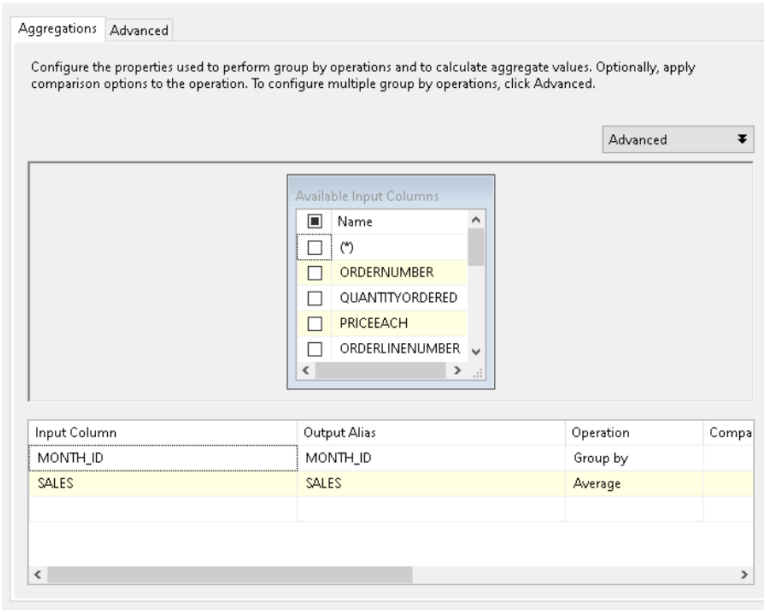
1. Create a Connection Manager



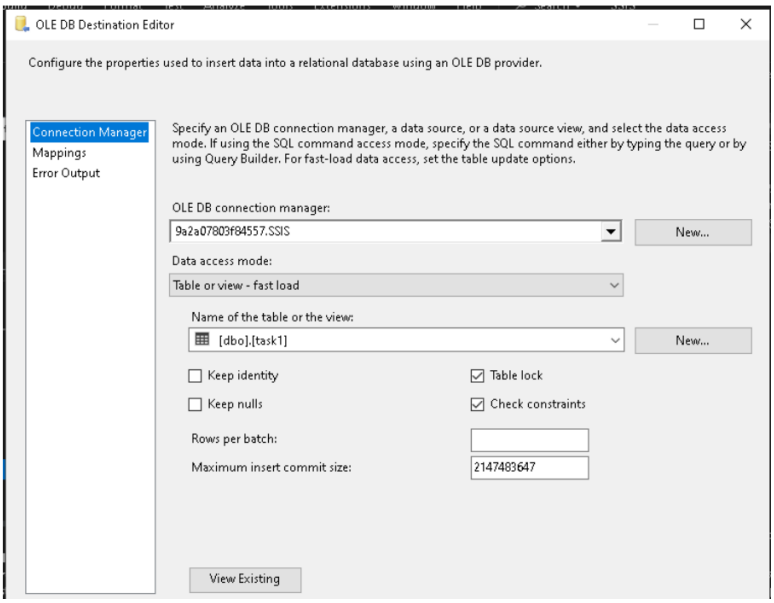
2. Extract Data from a transactional TABLE using an OLE DB Source

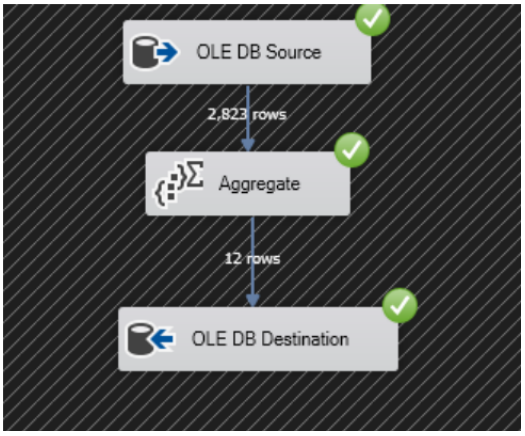


3. Apply necessary transformations such as data type conversions, data cleansing, and calculation3.



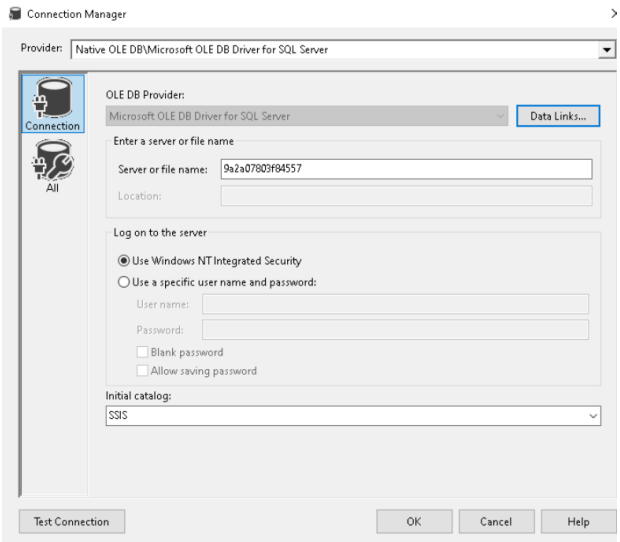
4. Load Data into the data warehous

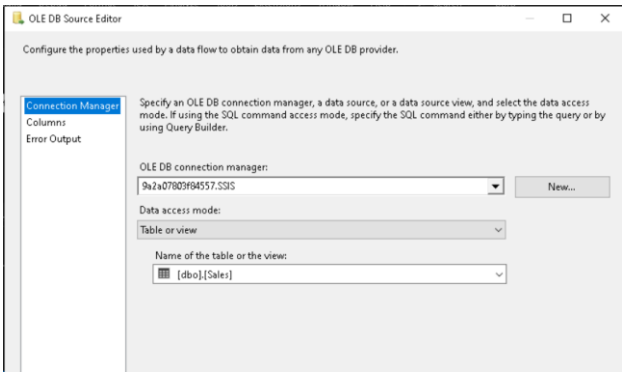




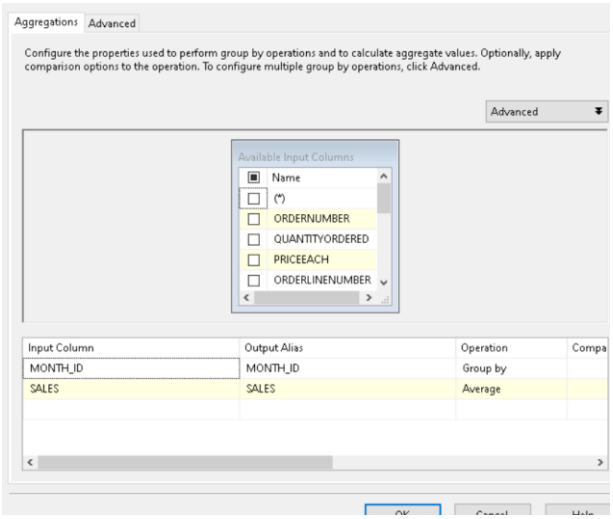
Task 2: Data Warehouse Migrations Scenario: Your organization is migrating its data warehouse from one server to another. You need to create an SSIS package that facilitates this migration.

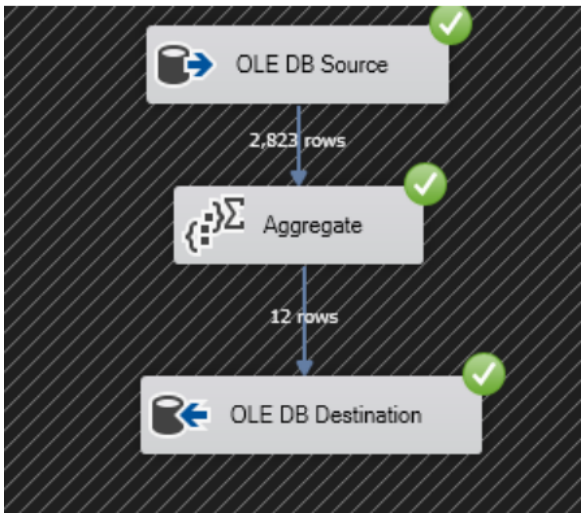
1. Create Connection Managers



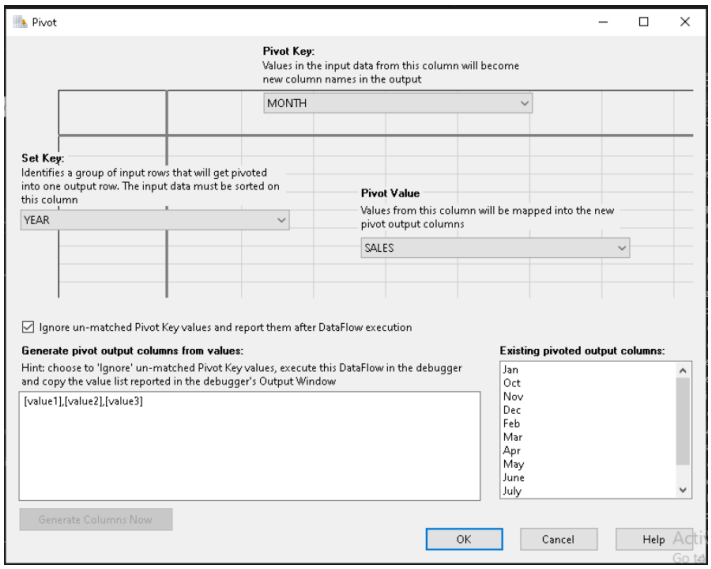


2. **Do aggregations**

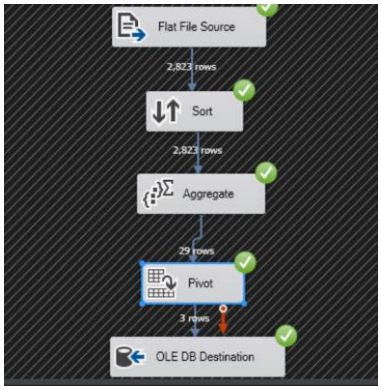




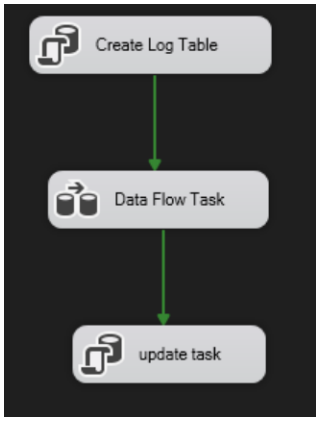
Task 3: Implementing a Pivot Transformation Scenario: You have data in a normalized format and need to pivot it for reporting purposes. Requirements: 1. Extract Data from the source table using an OLE DB Source. 2. Apply a Pivot Transformation to transform the normalized data into a pivoted format

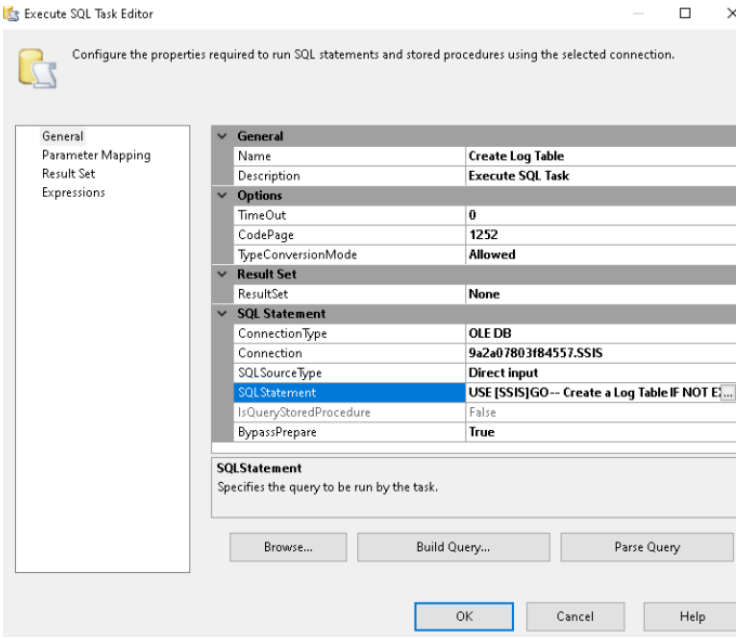


Load data

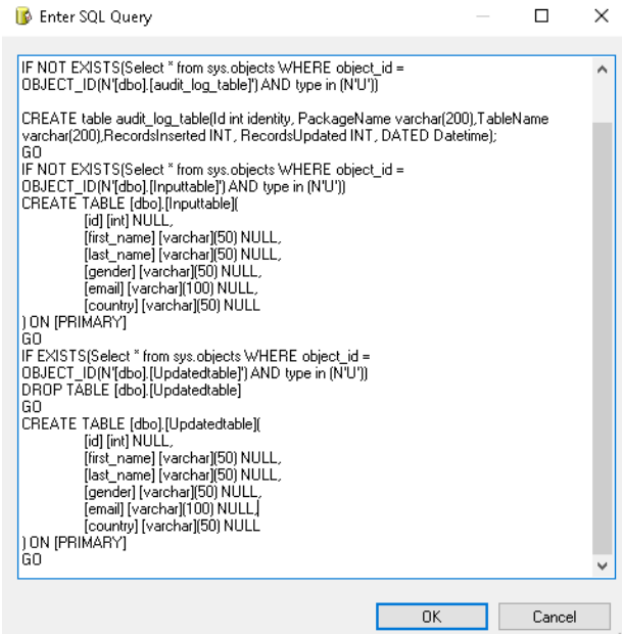


Task 4: Incremental Load Scenario: To optimize ETL processes, you need to implement an incremental load to update only the changed data in the data warehouse. Requirements: 1. Identify Changed Data: Use methods such as timestamps, change data capture using lookup, or checksums. 2. Extract Only the Changed Data from the source. 3. Update the Data Warehouse with the new and changed data only.

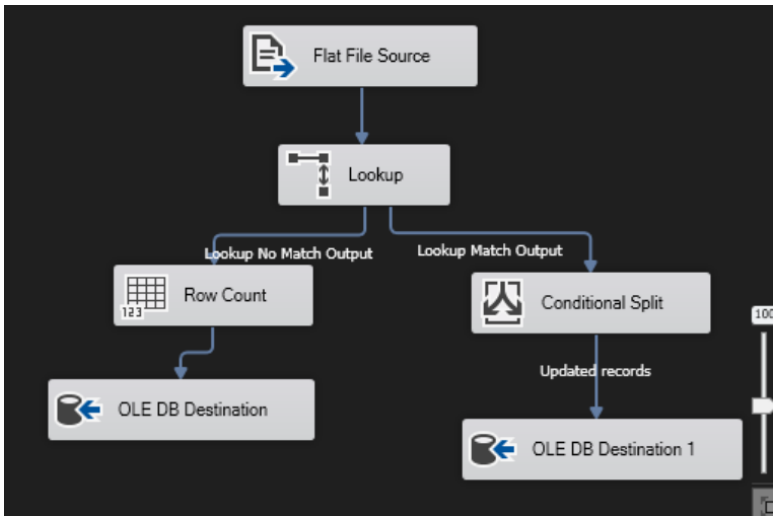




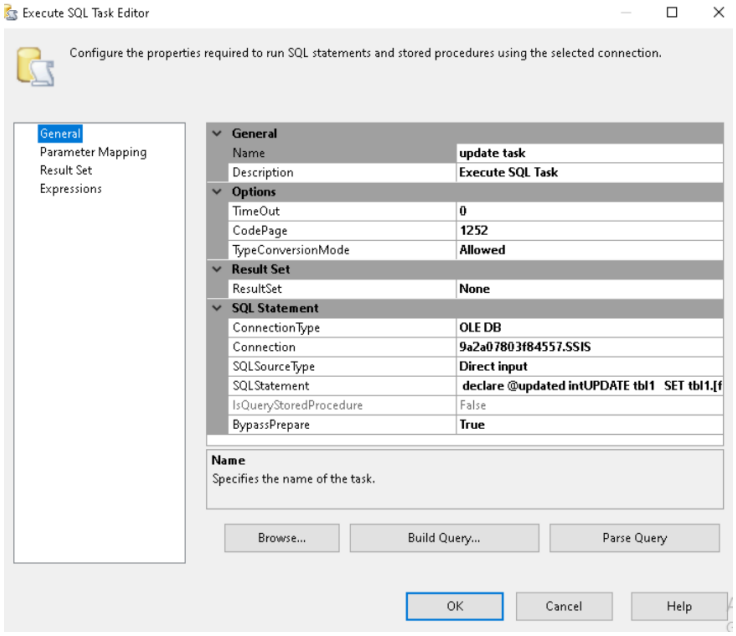
Write the SQL Query



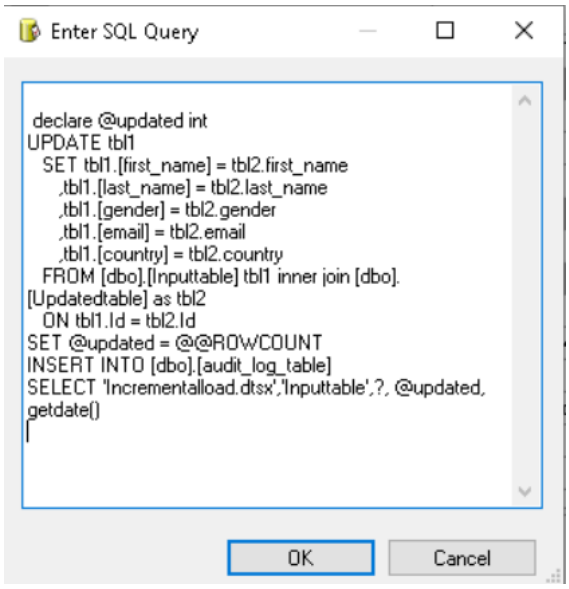
Create lookup no match and match outputs

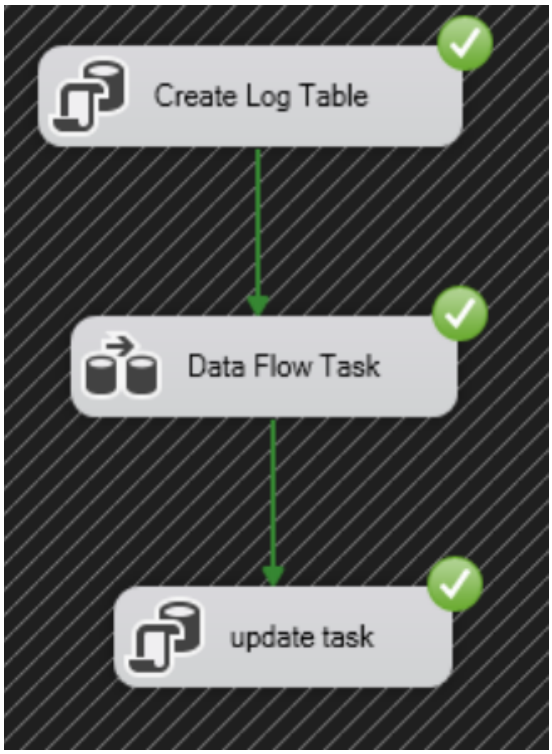


Set up the SQL settings



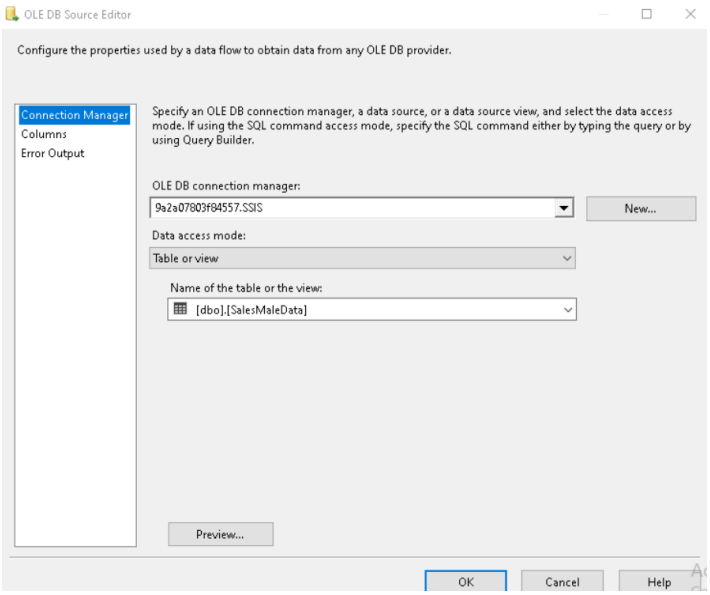
SQL Query to update table



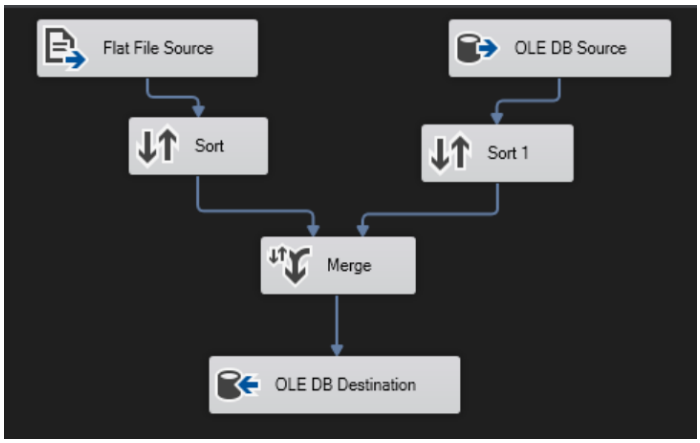


Task 5: Transformations Scenario: Your company needs to transform raw data into a format suitable for reporting. You need to perform multiple transformations within an SSIS package.

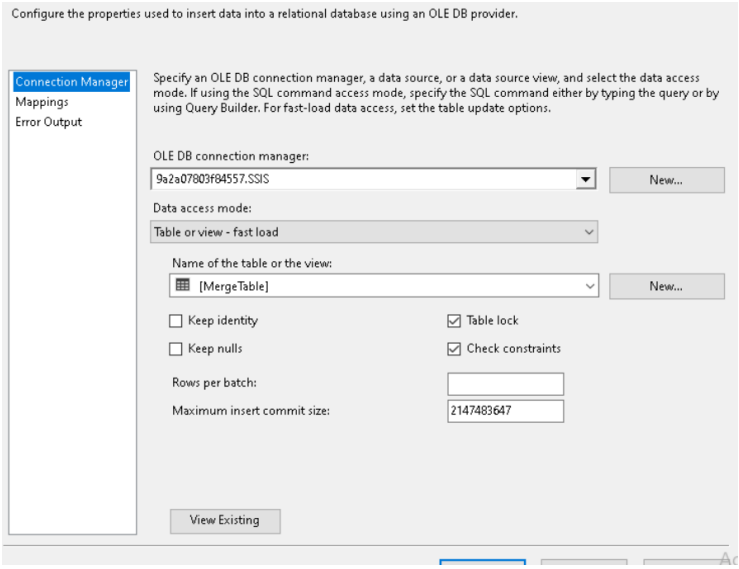
1. Extract Data from a source table using an OLE DB Source



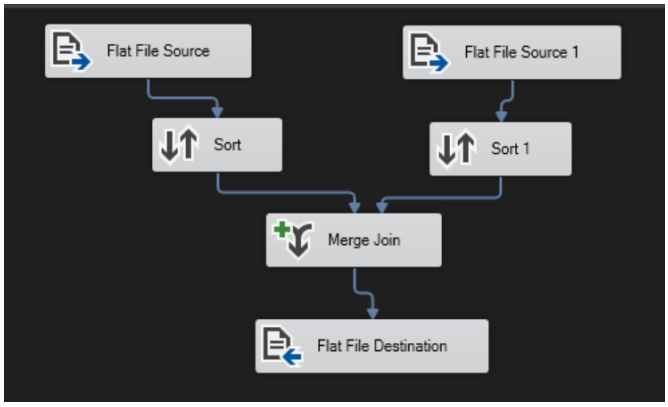
2. **Apply Sort and merge**



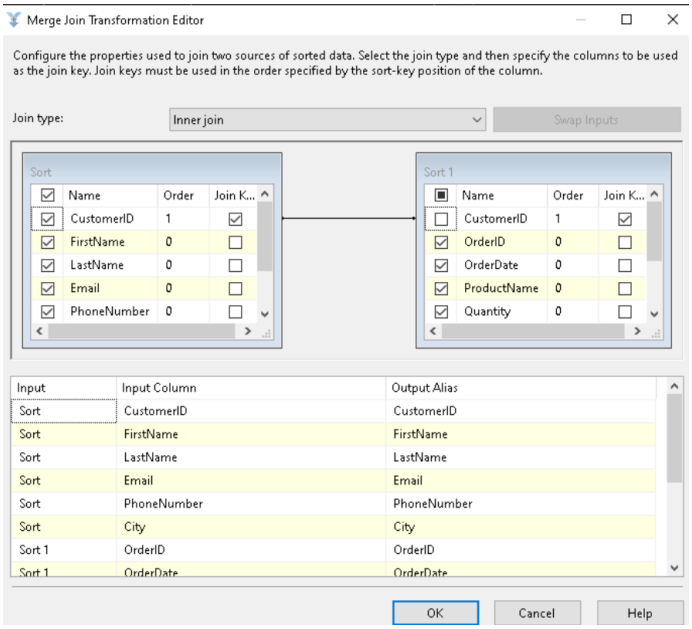
3. **Connection manager for destination**



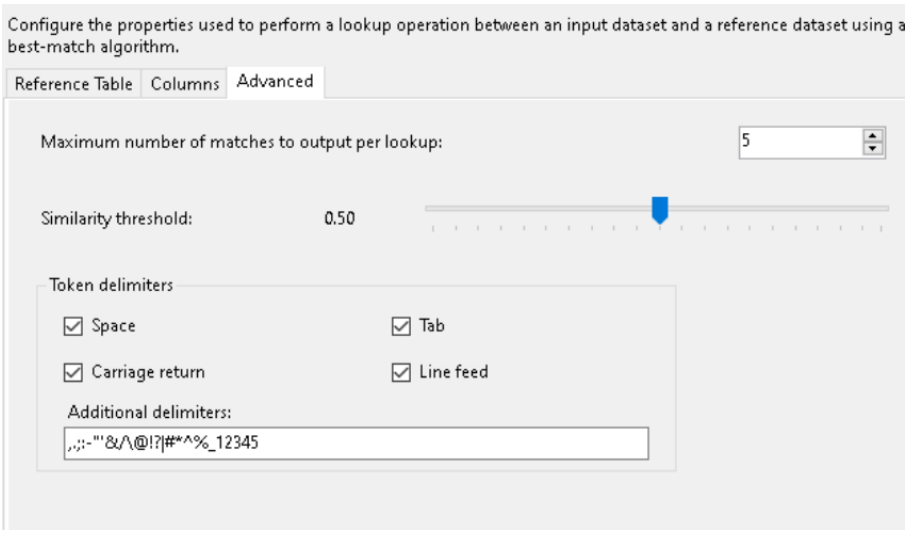
Task 6: MERGE & FUZZY LOOKUP Scenario: You need to merge two datasets and use fuzzy matching to handle potential duplicates. Requirements: 1. Extract Data from two source tables using OLE DB Sources. 2. Apply a Merge Join to combine the datasets based on a common key. 3. Use Fuzzy Lookup to identify and resolve duplicates in the merged data. 4. Load the Cleaned Data into a destination table



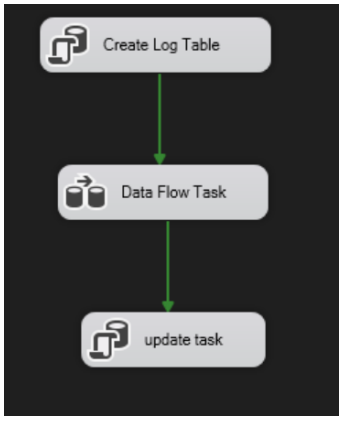
Apply inner join



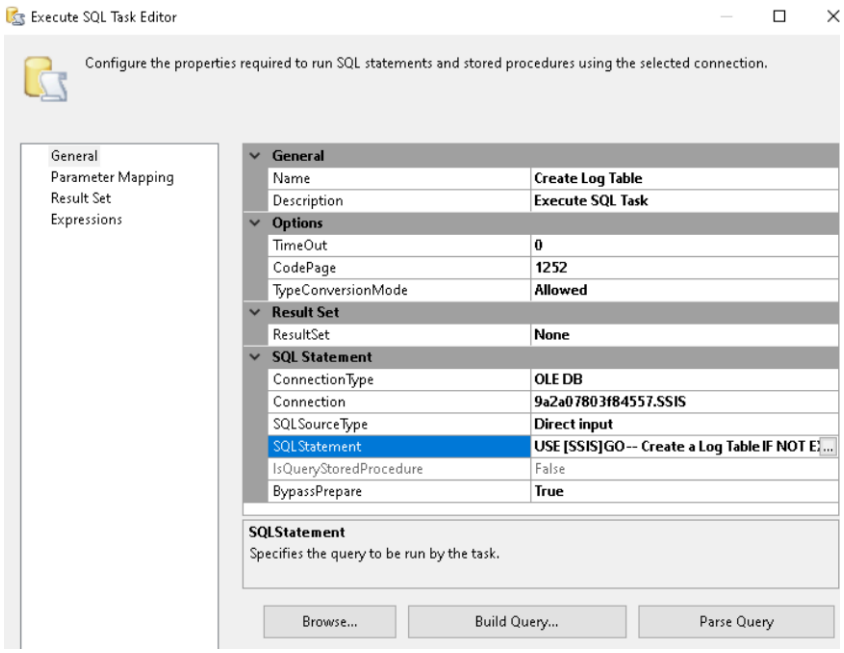
Set up fuzzy lookup



Task 7: Using Script Task Scenario: You need to perform a complex data transformation that is not supported by the standard SSIS components. A Script Task can be used to achieve this. Requirements: 1. Add a Script Task to the Control Flow. 2. Write a Script: that performs the required transformation. e.g. Reading data from a file, processing it, and writing the results to a database table. 3. Execute the Script Task within an SSIS package



SQL Execute query task



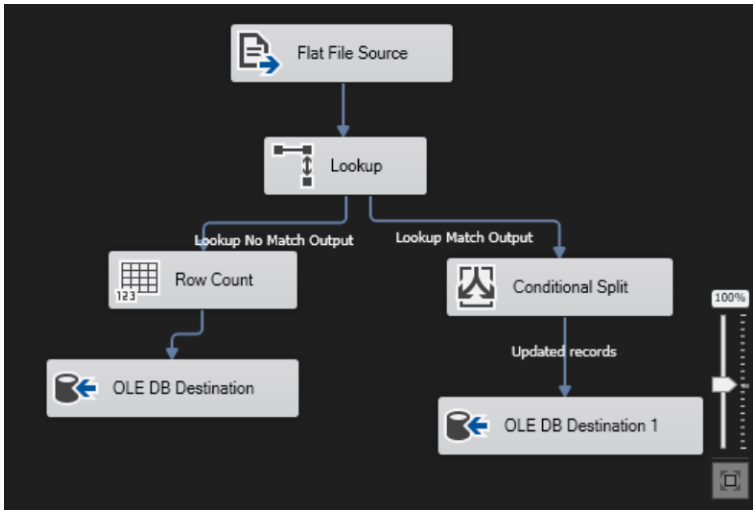
Enter SQL Query

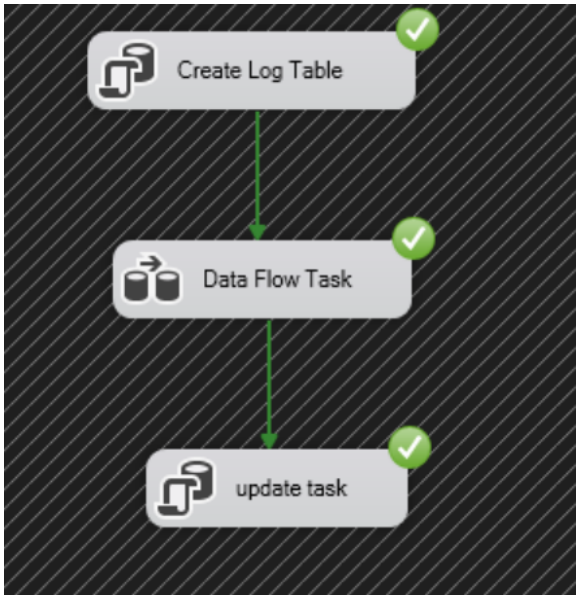
```
IF NOT EXISTS(Select * from sys.objects WHERE object_id =
OBJECT_ID(N'[dbo].[audit_log_table]') AND type in (N'U'))

CREATE table audit_log_table(Id int identity, PackageName varchar(200),TableName
varchar(200),RecordsInserted INT, RecordsUpdated INT, DATED Datetime);
GO
IF NOT EXISTS(Select * from sys.objects WHERE object_id =
OBJECT_ID(N'[dbo].[Inputtable]') AND type in (N'U'))
CREATE TABLE [dbo].[Inputtable](
        [id] [int] NULL,
        [first_name] [varchar](50) NULL,
        [last_name] [varchar](50) NULL,
        [gender] [varchar](50) NULL,
        [email] [varchar](100) NULL,
        [country] [varchar](50) NULL
) ON [PRIMARY]
GO
IF EXISTS(Select * from sys.objects WHERE object_id =
OBJECT_ID(N'[dbo].[Updatedtable]') AND type in (N'U'))
DROP TABLE [dbo].[Updatedtable]
GO
CREATE TABLE [dbo].[Updatedtable](
        [id] [int] NULL,
        [first_name] [varchar](50) NULL,
        [last_name] [varchar](50) NULL,
        [gender] [varchar](50) NULL,
        [email] [varchar](100) NULL,
        [country] [varchar](50) NULL
) ON [PRIMARY]
GO
```

OK Cancel

Final data flow




Create Log Table
Data Flow Task
update task