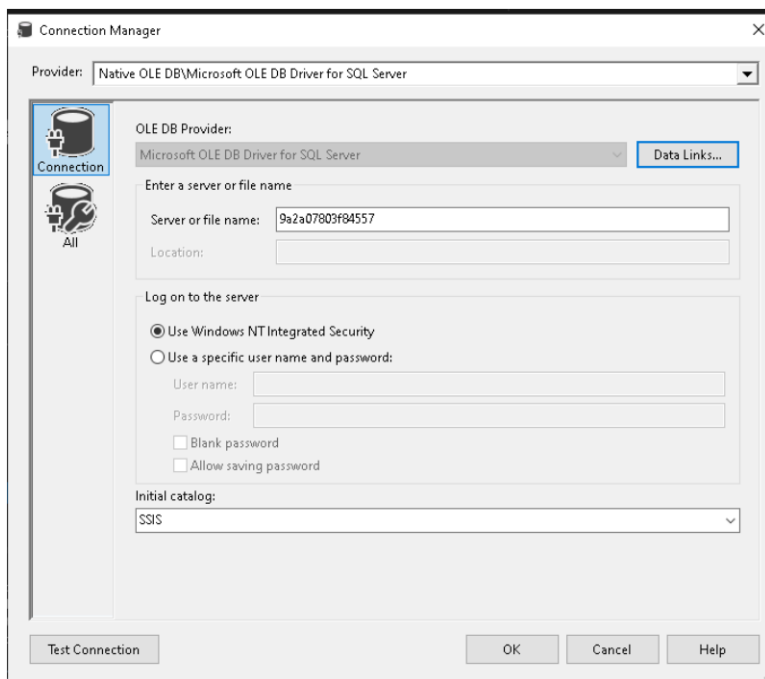


ASSIGNMENT 9

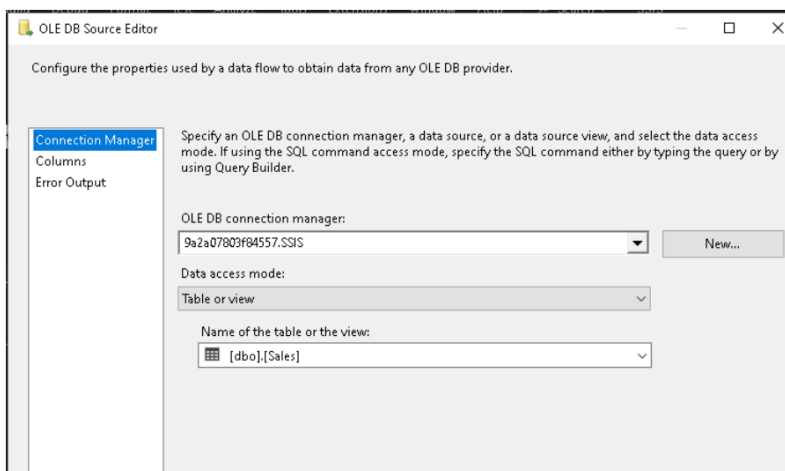
SSIS

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 calculation.

Aggregations Advanced

Configure the properties used to perform group by operations and to calculate aggregate values. Optionally, apply comparison options to the operation. To configure multiple group by operations, click Advanced.

Advanced

Available Input Columns

- ☒ Name
- ☐ (*)
- ☐ ORDERNUMBER
- ☐ QUANTITYORDERED
- ☐ PRICEEACH
- ☐ ORDERLINENUMBER

Input Column	Output Alias	Operation	Comparison
MONTH_ID	MONTH_ID	Group by	
SALES	SALES	Average	

4. Load Data into the data warehouse

OLE DB Destination Editor

Configure the properties used to insert data into a relational database using an OLE DB provider.

Connection Manager

Specify an OLE DB connection manager, a data source, or a data source view, and select the data access mode. If using the SQL command access mode, specify the SQL command either by typing the query or by using Query Builder. For fast-load data access, set the table update options.

OLE DB connection manager:

9a2a07803f84557.SSIS

New...

Data access mode:

Table or view - fast load

Name of the table or the view:

[dbo].[task1]

New...

☐ Keep identity ☒ Table lock

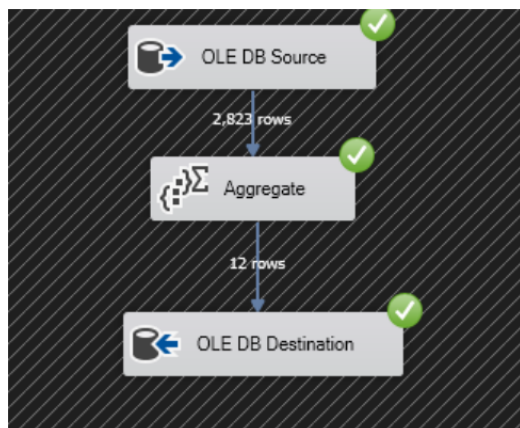
☐ Keep nulls ☒ Check constraints

Rows per batch:

Maximum insert commit size:

2147483647

View Existing



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

Connection Manager

Provider: Native OLE DB\Microsoft OLE DB Driver for SQL Server

OLE DB Provider: Microsoft OLE DB Driver for SQL Server [Data Links...](#)

Enter a server or file name

Server or file name: 9a2a07803f84557

Location:

Log on to the server

☒ Use Windows NT Integrated Security

☐ Use a specific user name and password:

User name:

Password:

☐ Blank password

☐ Allow saving password

Initial catalog: SSIS

Test Connection OK Cancel Help

OLE DB Source Editor

Configure the properties used by a data flow to obtain data from any OLE DB provider.

Connection Manager

Columns

Error Output

Specify an OLE DB connection manager, a data source, or a data source view, and select the data access mode. If using the SQL command access mode, specify the SQL command either by typing the query or by using Query Builder.

OLE DB connection manager: 9a2a07803f84557.SSIS [New...](#)

Data access mode:

Table or view:

Name of the table or the view: [dbo].[Sales]

2. Do aggregations

Aggregations Advanced

Configure the properties used to perform group by operations and to calculate aggregate values. Optionally, apply comparison options to the operation. To configure multiple group by operations, click Advanced.

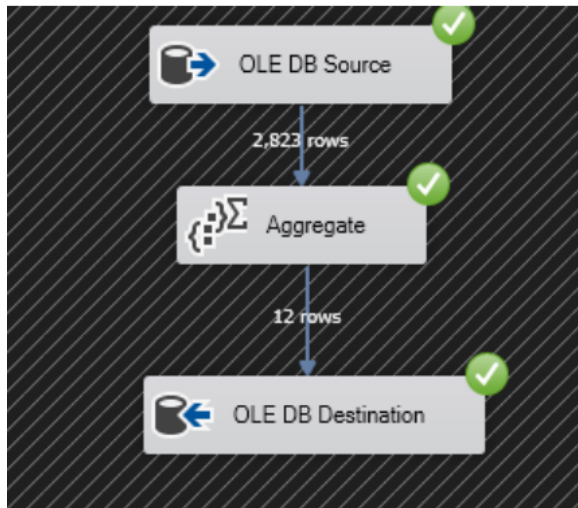
Advanced

Available Input Columns

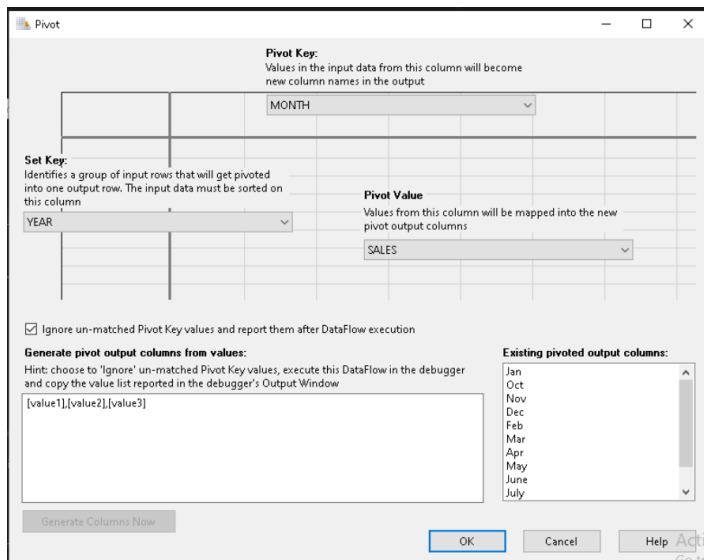
- ☒ Name
- ☐ (*)
- ☐ ORDERNUMBER
- ☐ QUANTITYORDERED
- ☐ PRICEEACH
- ☐ ORDERLINENUMBER

Input Column	Output Alias	Operation	Comparison
MONTH_ID	MONTH_ID	Group by	
SALES	SALES	Average	

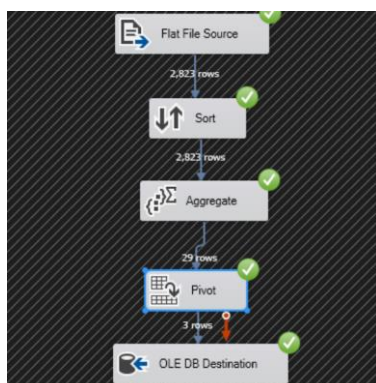
OK Cancel Help



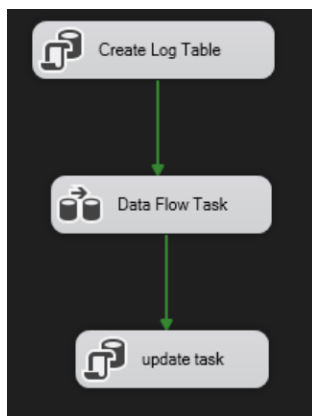
Task 3: Implementing a Pivot Transformation Scenario: You have data in a normalized format and need to pivot it for reporting purposes. Requirements: 1. Extract Data from the source table using an OLE DB Source. 2. Apply a Pivot Transformation to transform the normalized data into a pivoted format



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.



Execute SQL Task Editor

Configure the properties required to run SQL statements and stored procedures using the selected connection.

General	
Name	Create Log Table
Description	Execute SQL Task

Options	
Timeout	0
CodePage	1252
TypeConversionMode	Allowed

Result Set	
ResultSet	None

SQL Statement	
ConnectionType	OLE DB
Connection	9a2a07803f84557.SSIS
SQLSourceType	Direct input
SQLStatement	USE [SSIS]GO-- Create a Log Table IF NOT E...
IsQueryStoredProcedure	False
BypassPrepare	True

SQL Statement
Specifies the query to be run by the task.

Browse... Build Query... Parse Query

OK Cancel Help

Write the SQL Query

Enter SQL Query

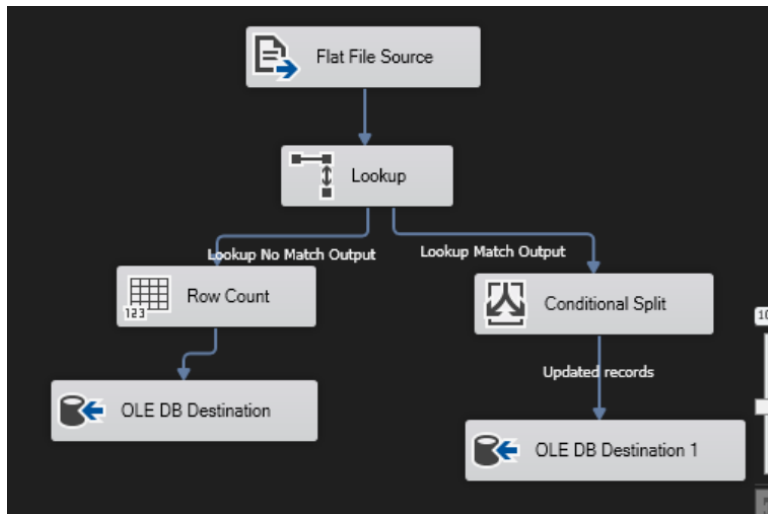
```

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

Create lookup no match and match outputs



Set up the SQL settings

Execute SQL Task Editor

Configure the properties required to run SQL statements and stored procedures using the selected connection.

General

Parameter Mapping
Result Set
Expressions

General	
Name	update task
Description	Execute SQL Task
Options	
TimeOut	0
CodePage	1252
TypeConversionMode	Allowed
Result Set	
ResultSet	None
SQL Statement	
ConnectionType	OLE DB
Connection	9a2a07803f84557.SSIS
SQLSourceType	Direct input
SQLStatement	declare @updated intUPDATE tbl1 SET tbl1.[f
IsQueryStoredProcedure	False
BypassPrepare	True

Name
Specifies the name of the task.

Browse... Build Query... Parse Query

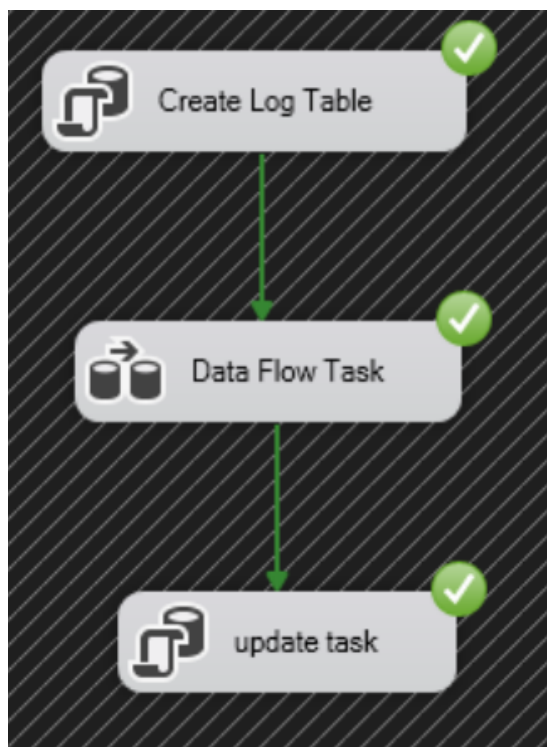
OK Cancel Help

SQL Query to update table

Enter SQL Query

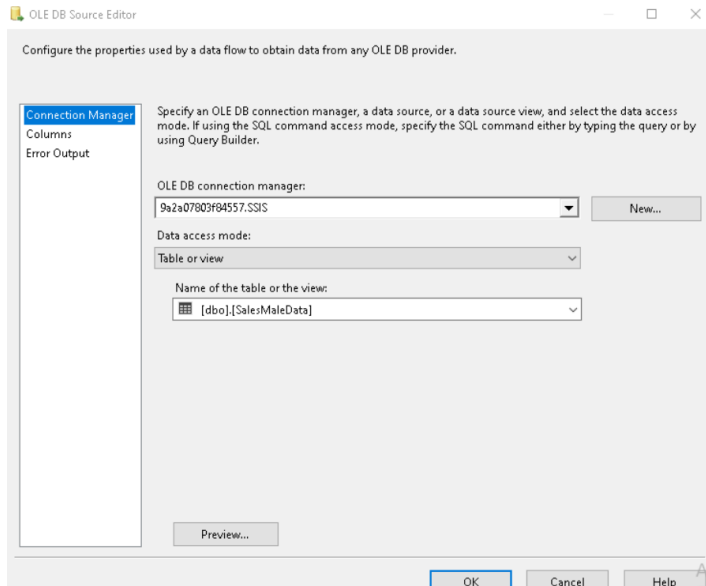
```
declare @updated int
UPDATE tbl1
SET tbl1.[first_name] = tbl2.first_name
  ,tbl1.[last_name] = tbl2.last_name
  ,tbl1.[gender] = tbl2.gender
  ,tbl1.[email] = tbl2.email
  ,tbl1.[country] = tbl2.country
FROM [dbo].[Inputtable] tbl1 inner join [dbo].[Updatedtable] as tbl2
ON tbl1.Id = tbl2.Id
SET @updated = @@ROWCOUNT
INSERT INTO [dbo].[audit_log_table]
SELECT 'Incrementalload.dtsx','Inputtable',?, @updated,
getdate()
```

OK Cancel

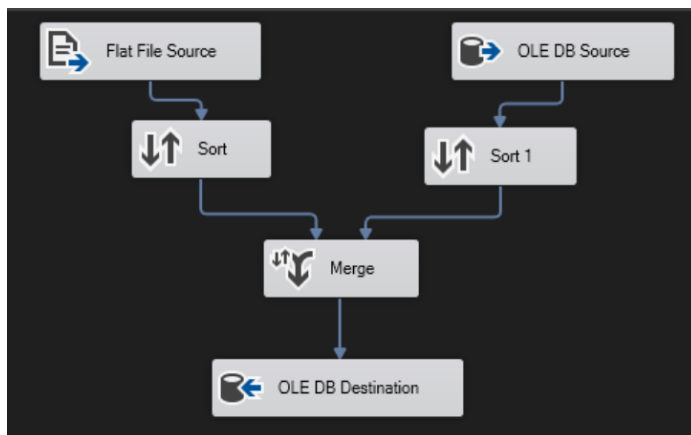


Task 5: Transformations Scenario: Your company needs to transform raw data into a format suitable for reporting. You need to perform multiple transformations within an SSIS package.

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



2. Apply Sort and merge



3. Connection manager for destination

Configure the properties used to insert data into a relational database using an OLE DB provider.

Connection Manager
Mappings
Error Output

Specify an OLE DB connection manager, a data source, or a data source view, and select the data access mode. If using the SQL command access mode, specify the SQL command either by typing the query or by using Query Builder. For fast-load data access, set the table update options.

OLE DB connection manager:
 9a2a07803f84557.SSIS New...

Data access mode:
 Table or view - fast load

Name of the table or the view:
 [MergeTable] New...

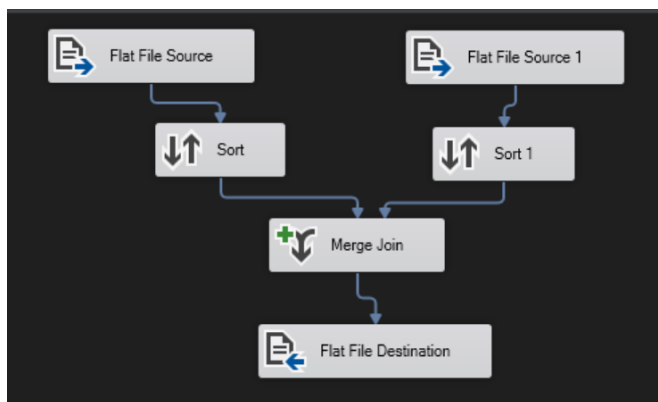
☐ Keep identity ☒ Table lock
☐ Keep nulls ☒ Check constraints

Rows per batch:

Maximum insert commit size:

View Existing

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

Merge Join Transformation Editor

Configure the properties used to join two sources of sorted data. Select the join type and then specify the columns to be used as the join key. Join keys must be used in the order specified by the sort-key position of the column.

Join type: Inner join Swap Inputs

Sort

<input checked="" type="checkbox"/>	Name	Order	Join K...
<input checked="" type="checkbox"/>	CustomerID	1	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>	FirstName	0	<input type="checkbox"/>
<input checked="" type="checkbox"/>	LastName	0	<input type="checkbox"/>
<input checked="" type="checkbox"/>	Email	0	<input type="checkbox"/>
<input checked="" type="checkbox"/>	PhoneNumber	0	<input type="checkbox"/>

Sort 1

<input type="checkbox"/>	Name	Order	Join K...
<input type="checkbox"/>	CustomerID	1	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>	OrderID	0	<input type="checkbox"/>
<input checked="" type="checkbox"/>	OrderDate	0	<input type="checkbox"/>
<input checked="" type="checkbox"/>	ProductName	0	<input type="checkbox"/>
<input checked="" type="checkbox"/>	Quantity	0	<input type="checkbox"/>

Input	Input Column	Output Alias
Sort	CustomerID	CustomerID
Sort	FirstName	FirstName
Sort	LastName	LastName
Sort	Email	Email
Sort	PhoneNumber	PhoneNumber
Sort	City	City
Sort 1	OrderID	OrderID
Sort 1	OrderDate	OrderDate

OK Cancel Help

Set up fuzzy lookup

Configure the properties used to perform a lookup operation between an input dataset and a reference dataset using a best-match algorithm.

Reference Table Columns Advanced

Maximum number of matches to output per lookup: 5

Similarity threshold: 0.50

Token delimiters

☒ Space

☒ Tab

☒ Carriage return

☒ Line feed

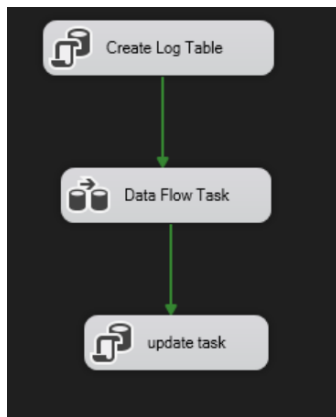
Additional delimiters:

.,:;'-"/&^@!~|`#**^%_12345

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

Execute SQL Task Editor

Configure the properties required to run SQL statements and stored procedures using the selected connection.

General
Parameter Mapping
Result Set
Expressions

General	
Name	Create Log Table
Description	Execute SQL Task
Options	
TimeOut	0
CodePage	1252
TypeConversionMode	Allowed
Result Set	
ResultSet	None
SQL Statement	
ConnectionType	OLE DB
Connection	9a2a07803f84557.SSIS
SQLSourceType	Direct input
SQL Statement	USE [SSIS]GO-- Create a Log Table IF NOT E...
IsQueryStoredProcedure	False
BypassPrepare	True

SQLStatement
Specifies the query to be run by the task.

Browse... Build Query... Parse Query

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

