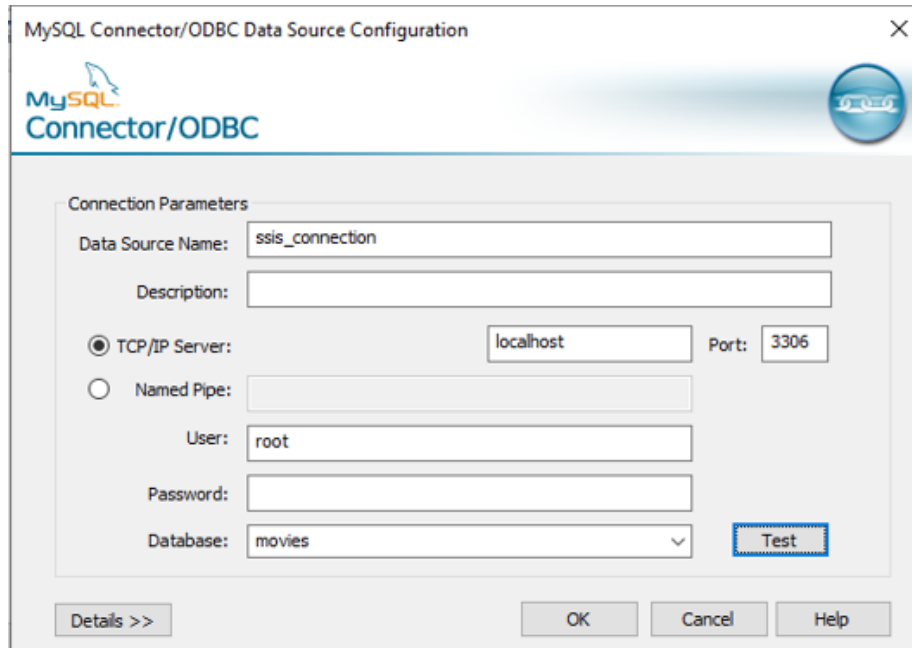


ETL:

Step 1: Creating Operational Database Connection:



MySQL Connector/ODBC Data Source Configuration

MySQL Connector/ODBC

Connection Parameters

Data Source Name:

Description:

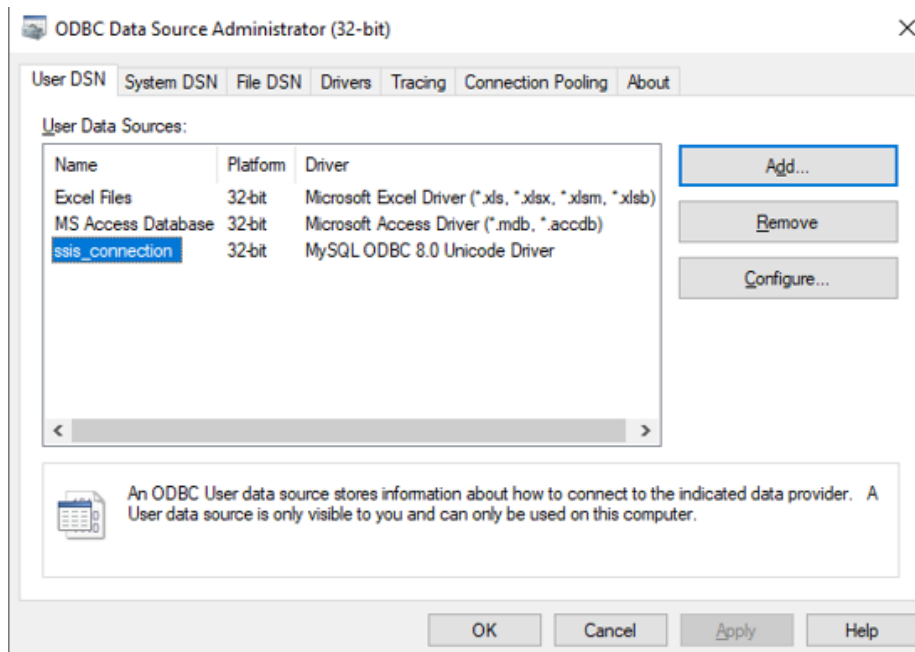
☒ TCP/IP Server: Port:

☐ Named Pipe:

User:

Password:

Database:



ODBC Data Source Administrator (32-bit)

User DSN System DSN File DSN Drivers Tracing Connection Pooling About

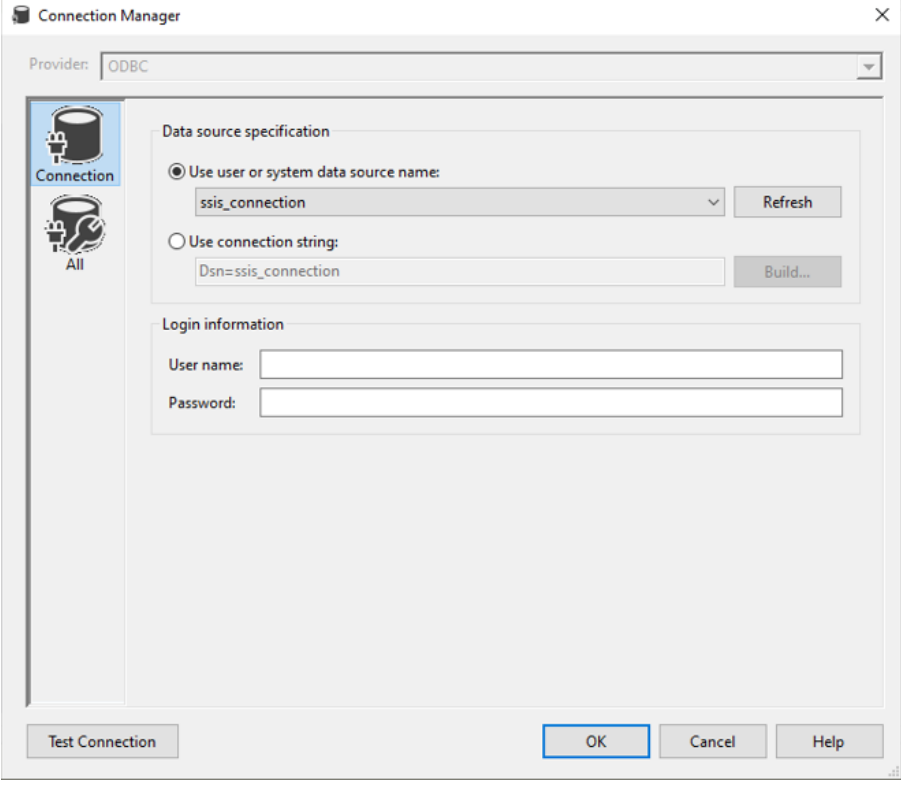
User Data Sources:

Name	Platform	Driver
Excel Files	32-bit	Microsoft Excel Driver (*.xls, *.xlsx, *.xsm, *.xlsb)
MS Access Database	32-bit	Microsoft Access Driver (*.mdb, *.accdb)
ssis_connection	32-bit	MySQL ODBC 8.0 Unicode Driver

An ODBC User data source stores information about how to connect to the indicated data provider. A User data source is only visible to you and can only be used on this computer.

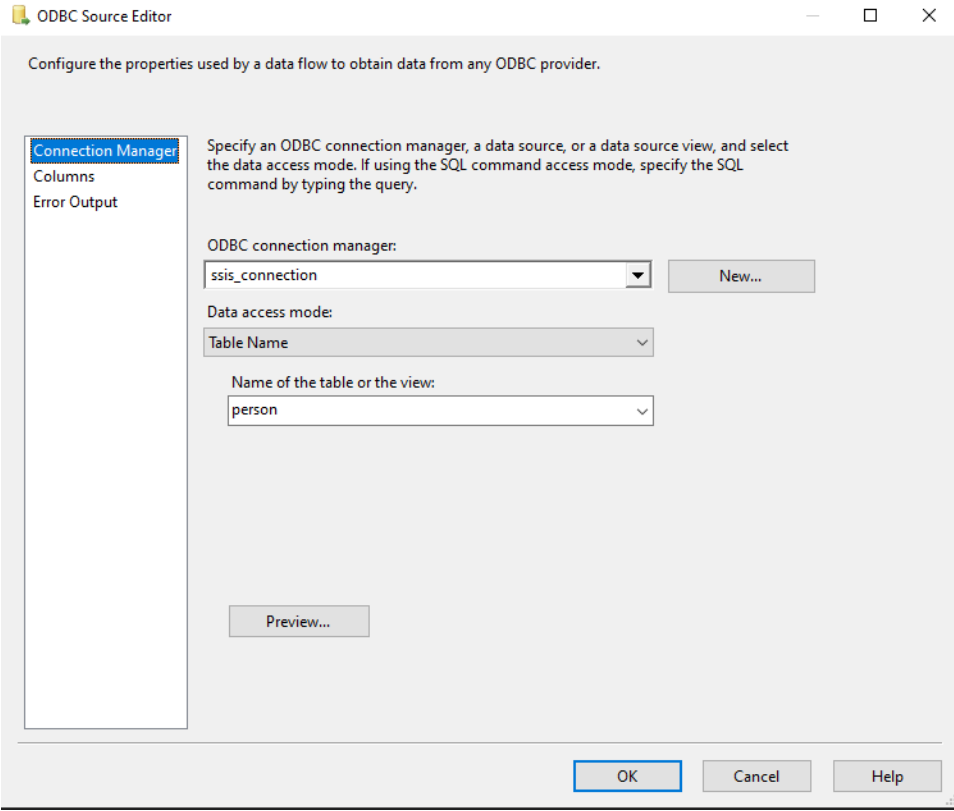
Step 2: Extraction:

1. Create Connections:



The screenshot shows the 'Connection Manager' dialog box. The 'Provider' dropdown is set to 'ODBC'. On the left, there are two icons: 'Connection' (a cylinder) and 'All' (a cylinder with a plug). The 'Data source specification' section has two radio buttons: 'Use user or system data source name:' (selected) and 'Use connection string:'. The 'Use user or system data source name:' option has a dropdown menu showing 'ssis_connection' and a 'Refresh' button. The 'Use connection string:' option has a text box containing 'Dsn=ssis_connection' and a 'Build...' button. The 'Login information' section has two text boxes: 'User name:' and 'Password:'. At the bottom, there are four buttons: 'Test Connection', 'OK', 'Cancel', and 'Help'.

2. Extracting Table from Operation Database:



The screenshot shows the 'ODBC Source Editor' dialog box. The title bar says 'ODBC Source Editor'. The main text says 'Configure the properties used by a data flow to obtain data from any ODBC provider.' On the left, there are three tabs: 'Connection Manager' (selected), 'Columns', and 'Error Output'. The 'Connection Manager' tab has a text box for 'ODBC connection manager:' with a dropdown menu showing 'ssis_connection' and a 'New...' button. Below that is a 'Data access mode:' dropdown menu showing 'Table Name'. Below that is a 'Name of the table or the view:' dropdown menu showing 'person'. At the bottom left, there is a 'Preview...' button. At the bottom right, there are three buttons: 'OK', 'Cancel', and 'Help'.

ODBC Source Editor

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

Connection Manager
Columns
Error Output

Specify an ODBC 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 by typing the query.

ODBC connection manager:

Data access mode:

Name of the table or the view:

Step 3: Transformations:

- Choose a Transformation to be performed on the Extracted Data such as:
 - Aggregate (Group By)**

Aggregate Transformation Editor

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.

Available Input Columns

- ☐ Name
- ☐ (*)
- ☒ person_id
- ☒ movie_id
- ☐ department_id

Input Column	Output Alias	Operation	Comparison
person_id	person_id	Group by	
movie_id	movie_count	Count	

- Sort

Sort Transformation Editor

Specify the columns to sort, and set their sort type and their sort order. All nonselected columns are copied unchanged.

Available Input Columns

<input type="checkbox"/>	Name	Pass Throu...
<input checked="" type="checkbox"/>	person_id	<input type="checkbox"/>
<input type="checkbox"/>	movie_count	<input checked="" type="checkbox"/>

Input Column	Output Alias	Sort Type	Sort Order	Con
person_id	person_id	ascending	1	

☐ Remove rows with duplicate sort values

OK Cancel Help

- Merge 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 1

<input checked="" type="checkbox"/>	Name	Order	Join K...
<input checked="" type="checkbox"/>	person_id	1	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>	person_name	0	<input type="checkbox"/>

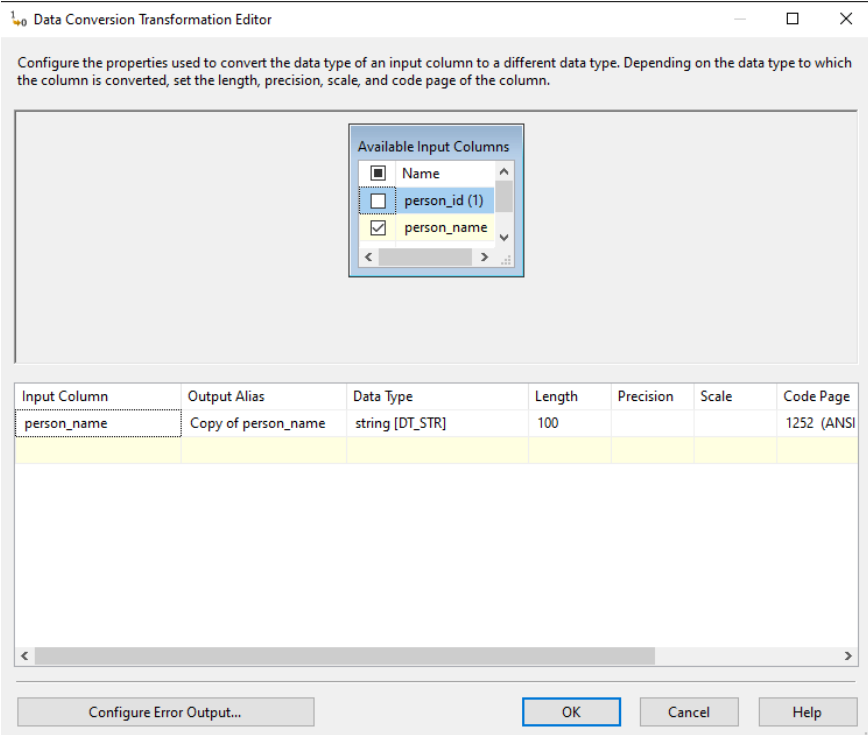
Sort

<input type="checkbox"/>	Name	Order	Join K...
<input type="checkbox"/>	person_id	1	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>	movie_count	0	<input type="checkbox"/>

Input	Input Column	Output Alias
Sort 1	person_id	person_id (1)
Sort 1	person_name	person_name
Sort	movie_count	movie_count

OK Cancel Help

- **Data Conversion**



Data Conversion Transformation Editor

Configure the properties used to convert the data type of an input column to a different data type. Depending on the data type to which the column is converted, set the length, precision, scale, and code page of the column.

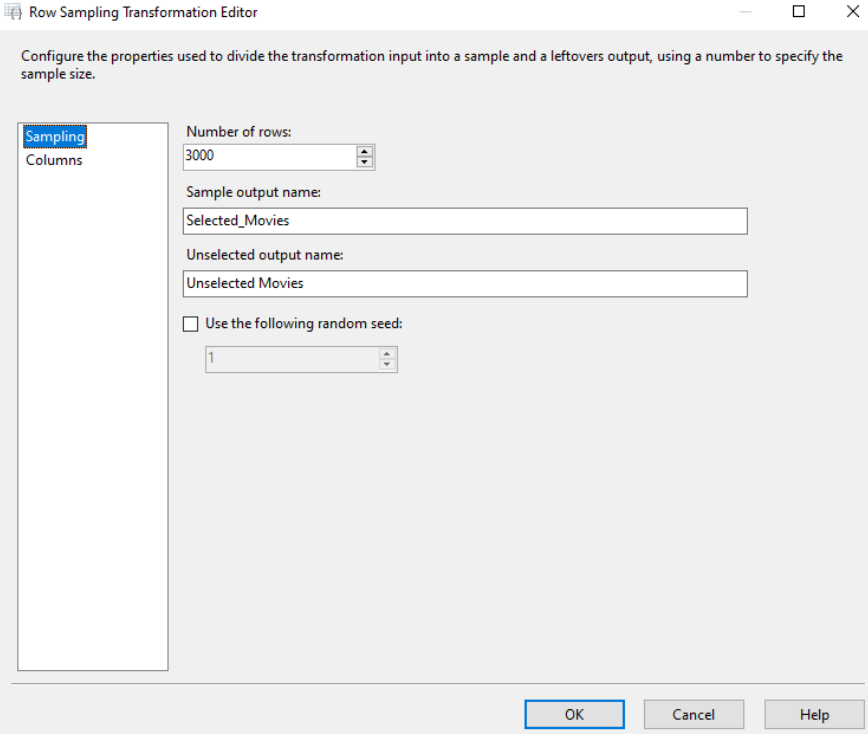
Available Input Columns

- ☐ Name
- ☐ person_id (1)
- ☒ person_name

Input Column	Output Alias	Data Type	Length	Precision	Scale	Code Page
person_name	Copy of person_name	string [DT_STR]	100			1252 (ANSI)

Configure Error Output... OK Cancel Help

- **Row Sampling**



Row Sampling Transformation Editor

Configure the properties used to divide the transformation input into a sample and a leftovers output, using a number to specify the sample size.

Sampling Columns

Number of rows: 3000

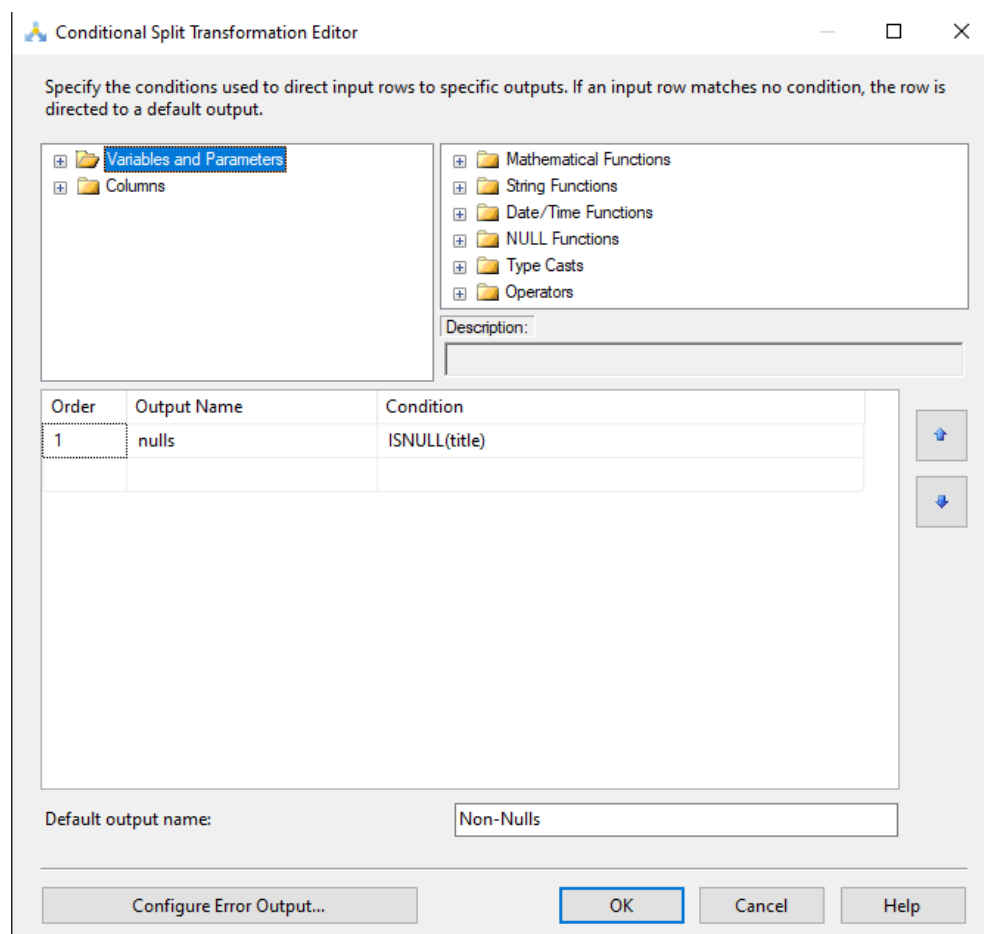
Sample output name: Selected_Movies

Unselected output name: Unselected Movies

☐ Use the following random seed: 1

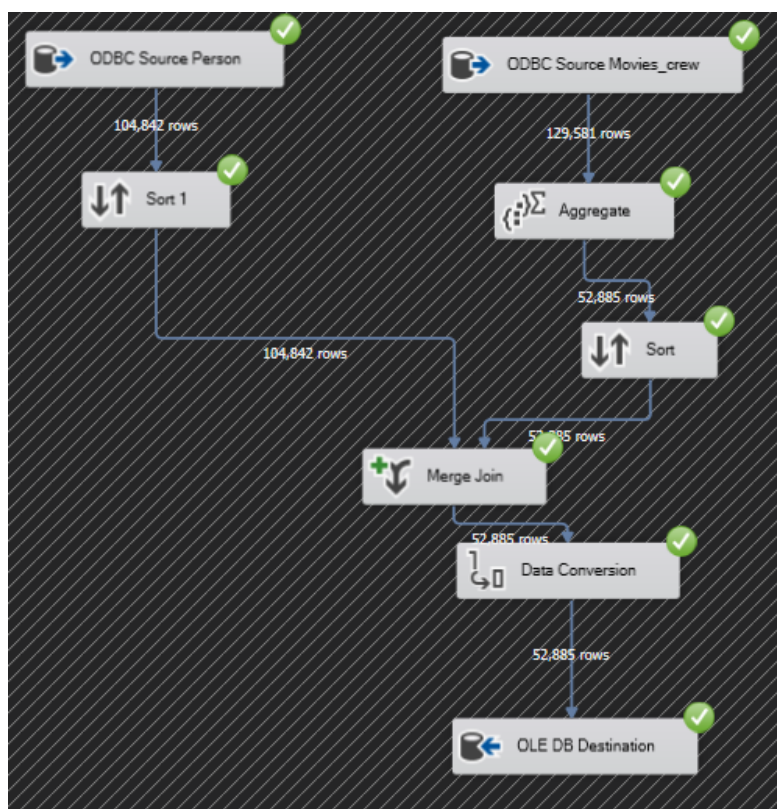
OK Cancel Help

- **Conditional Split:**

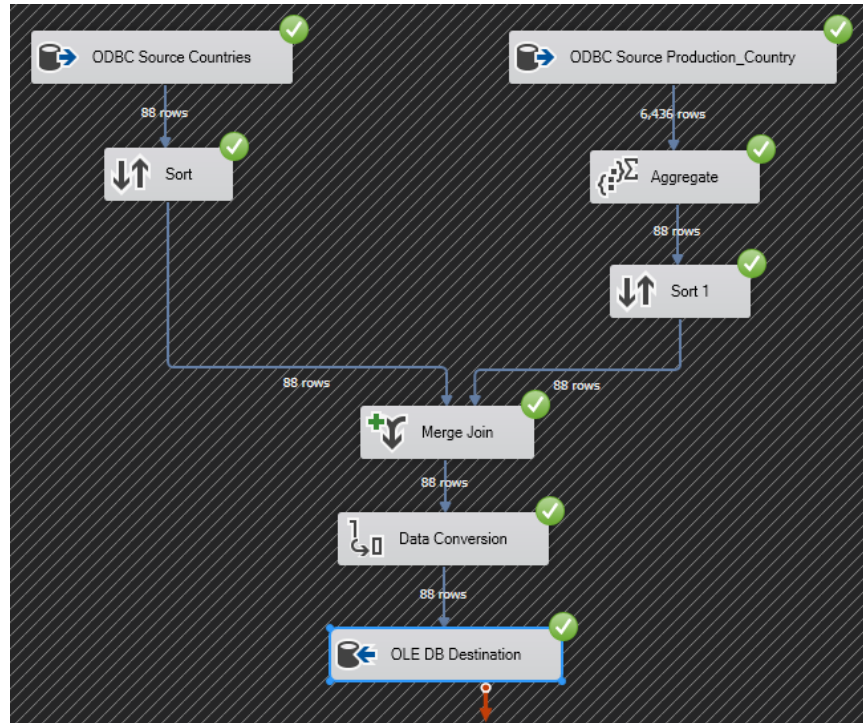


2. Perform the Transformations:

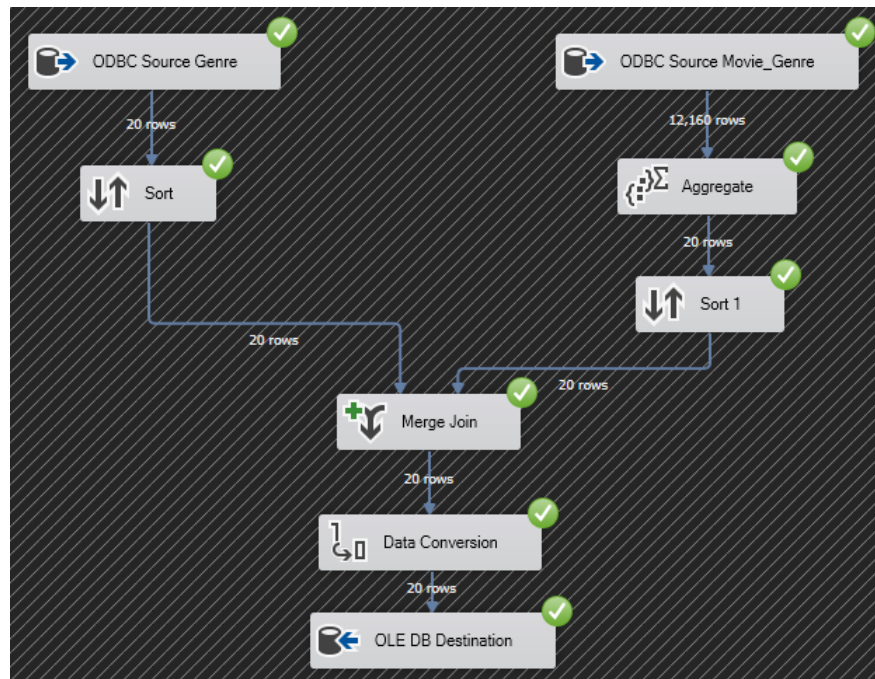
- **Person_Dim:**



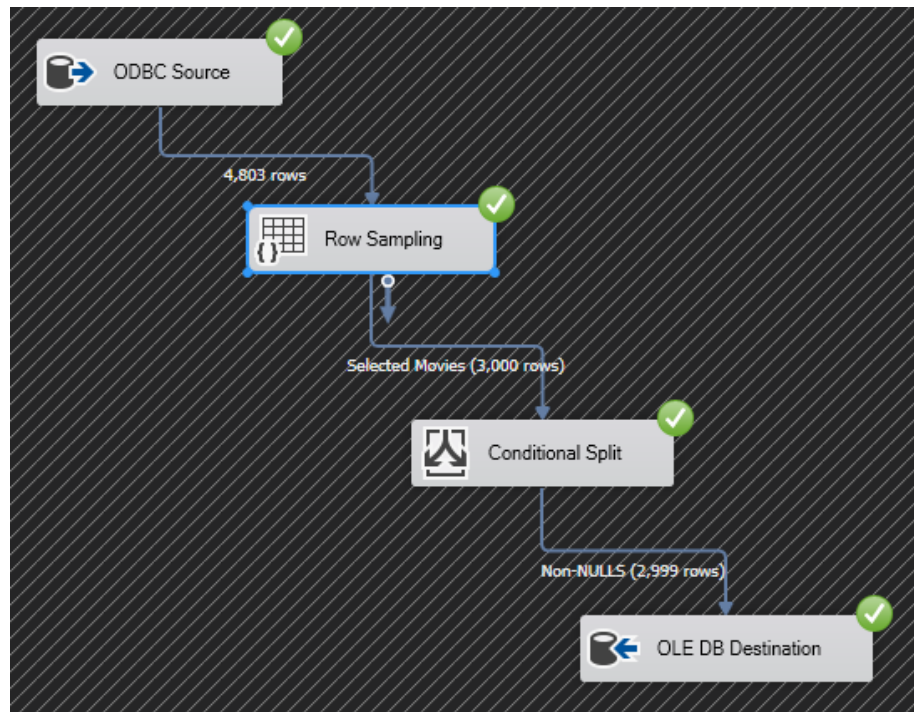
- **Countries_dim:**



- **Genre_dim**



- **Movies_dim**



- **Budget_dim**

