ASSIGNMENT 5

FAROOQUE AKHTAR | NUID- 001403864

Part 1 – 10 pts) As described in class load the 4 yearly files (1st tab only in a for each loop) into a staging

1) rename the files if you need to

2) edit the file to remove the extra items at the top of the tab

3) leave the data formated as is

4) leave it as an xls file or you can modify so it is a tab delimited

5) you should have on emapping from source file to staging area.

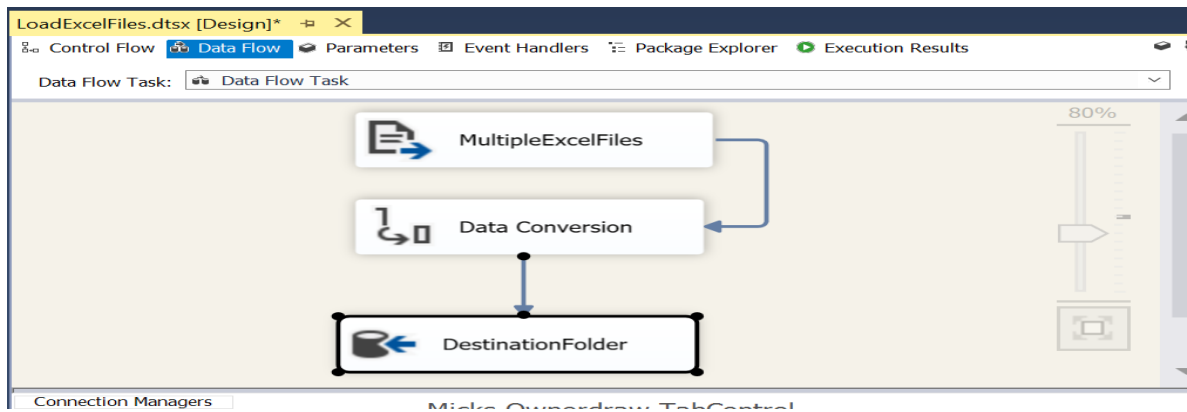STEP 1 – File >> New >> Project >> Integration Service Project >> Data Flow Task (Drag)

STEP 2 - Create a variable with the name InputFolder and give the path(folder path) in value

STEP 3 – Drag For Each loop in the Control Flow Tab. My loop will run the same number as the number of files in my set source path.

STEP 4 – Select DataConversion and select the appropriate data types.
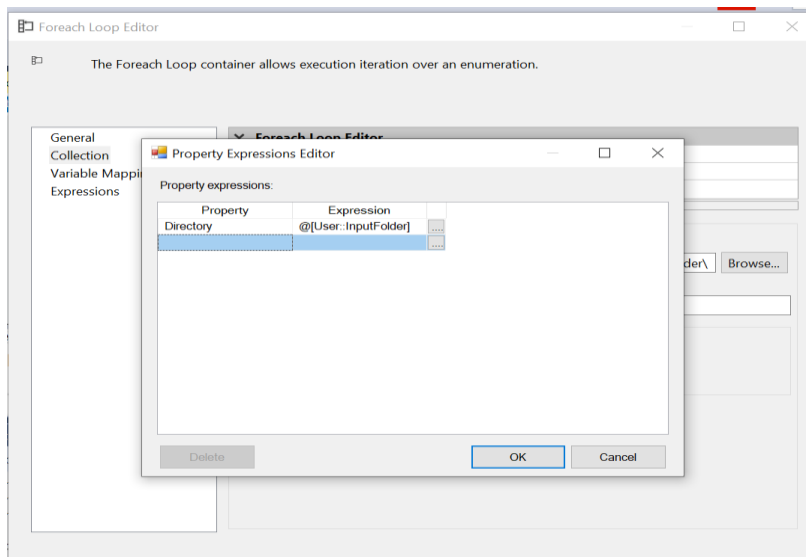
STEP 5 – Select OLE DB destination and select the destination table, and do the mapping.

STEP 6 – After Running the above task, we will have 204 rows in our staging table.

Part 2- 10 pts) move the staging data into a destination table
   1) add a column for the year
   2) add another for the load date. All rows for a file should have the same date and time for a load date.

We will select source files and converting the data type as per the domain understanding.

After loading the files and data conversion, we will create an OLE DB Destination where I am creating a new table for staging the data where my data from all the source files are being combined and stored.

After this, Destination Table is configured and also an Truncate SQL statement is added before "Storing Data into Staging Area" flow task out of the Forloop so that duplicate rows are not added into the Destination Table

```
SELECT * FROM Destination_Table
```

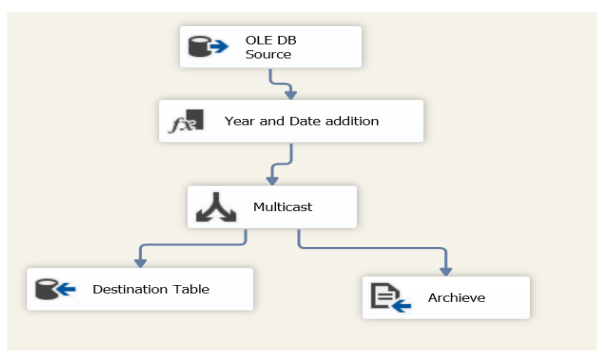|	State Name	State Abbreviation	State FIPS	Part D Prescribers	Part D Opioid Prescribers	Opioid Claims	Extended Release C
1	Alabama	AL	1	12820	7422	2260284	118523
2	Alaska	AK	2	2275	1099	86517	8602
3	Arizona	AZ	4	20542	10510	1545138	151086
4	Arkansas	AR	5	7909	4700	1128356	57519
5	California	CA	6	109571	49671	7228782	474288
6	Colorado	CO	8	17270	8814	1084129	89225
7	Connecticut	CT	9	16017	6306	671688	36924
8	Delaware	DE	10	3205	1568	233078	20044
9	District of Columbia	DC	11	4324	1299	81482	3371

Query executed successfully.   DESKTOP-P50DT8L\SQLDEV16 (1...   DESKTOP-P50DT8L\HP (53)   staging   00:00:00   204 rows

- After executing the package, data from each file is stored in the table namely "StagingCombinedNew Table"

Derived Column Name	Derived Column	Expression	Data Type	Le
YEAR	<add as new column>	(DT_STR,12,1252)SUBSTRING(@[User::FileName],1,4)	string [DT_STR]	1.
DATE	<add as new column>	GETDATE()	database timestamp [DT_...	

Part 3 - 5 pts) After completing the load you should move the data to an archive area with the additional column

- Added a MultiCast in 'Moving Data from Staging to Destination Table' data flow after adding columns
- Data is moved to both Archive File and Destination Table after Running the Package.
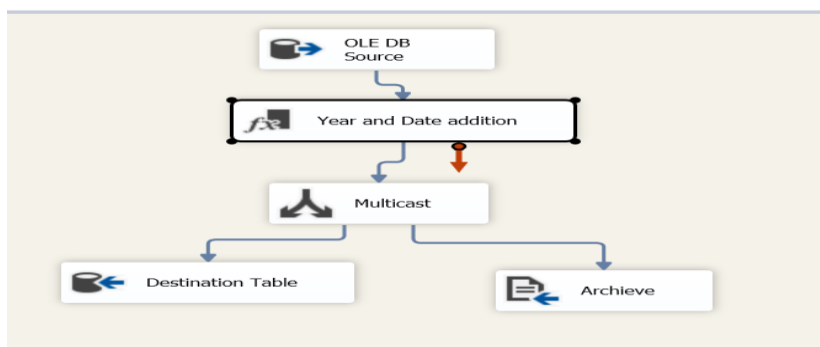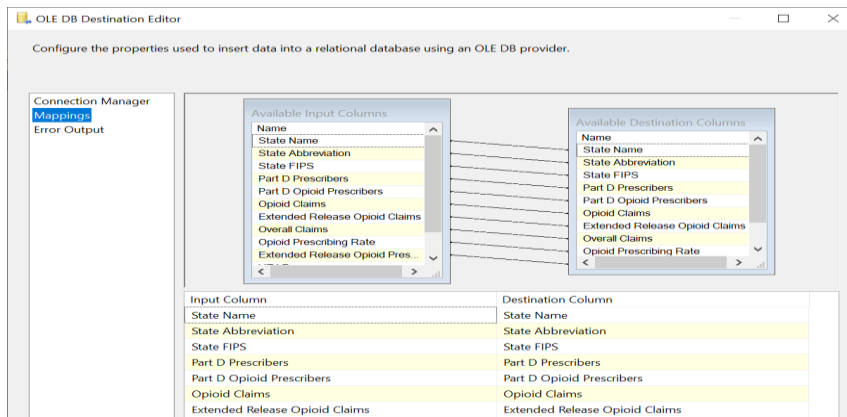
Part 4 - 10 pts) national totals row as a way to validate the load of each file

1) do not load the national total into the destination table

- Firstly Created an Data Flow Task namely "National Total Rows" to store the sum in Table namely "Aggregate"
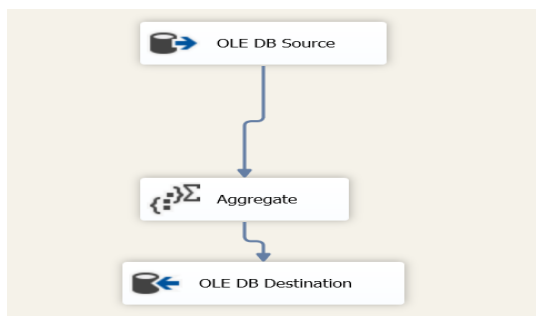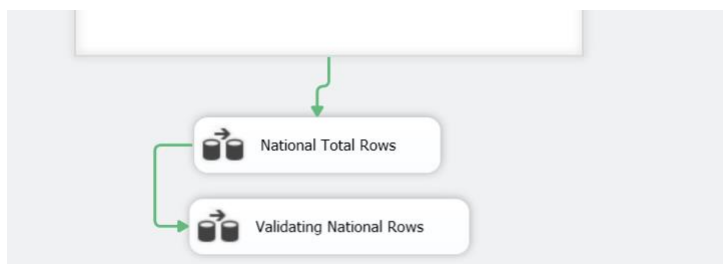

```
SELECT * FROM Aggregates

SELECT * FROM NationalRows
```

	Part D Prescribers	Part D Opioid Prescribers	Extended Release Opioid Claims	Opioid Claims	Overall Claims	Opioid Prescribing Rate	Ex
1	4151080	1966000	19093144	312182732	5368386620	6.07647059945499	6.

- **After this did the validation using data flow task using Lookup and Matched data is stored in Match table while unmatched data in Unmatch Table.**
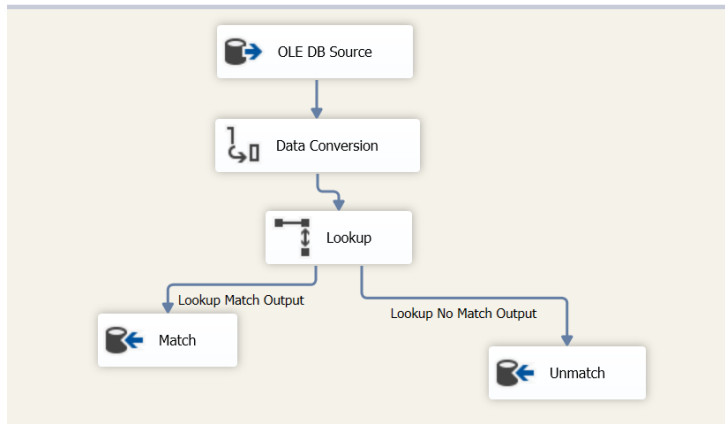
Part 5 - 5 pts) create an audit table that stores the results of each file loaded
   1) store the data and time of the load along with a validation

- Firstly, created table namely "PackageLoadInformation" to store results of each file loaded

```
Create table dbo.PackageLoadInformation(
PkgId int Identity(1,1),
PackageName VARCHAR(100),
InsertRecordCount int,
LoadDateTime datetime
default getdate())

SELECT * FROM dbo.PackageLoadInformation
```

- Secondly, created an data flow task namely "Audit" to make connection to database and read the records

- Then created a Row Count Transformation to save count of total variables that just passed and configured it by creating a new variable namely "RowCount"

- After this created an Execute SQL task namely "Creating Audit Table" and configured it to insert data into "PackageLoadInformation" table
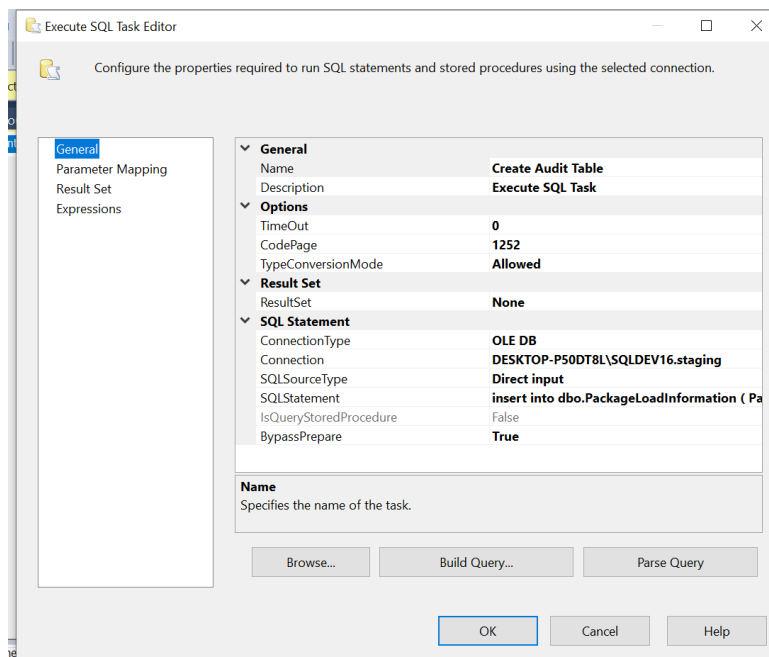
- After running the package, the data is stored in the table


```
SELECT * FROM dbo.PackageLoadInformation

insert into dbo.PackageLoadInformation ( PackageName, InsertRecordCount ) values (?,?)
```

	PkgId	PackageName	InsertRecordCount	LoadDateTime
1	1	LoadExcelFiles	51	2019-02-13 01:26:23.123
2	2	LoadExcelFiles	102	2019-02-13 01:26:23.927
3	3	LoadExcelFiles	153	2019-02-13 01:26:24.500
4	4	LoadExcelFiles	204	2019-02-13 01:26:25.177

Part 6 - submit a Toad data model of your solution

[1,1] **[2,1]**

2013.txt — ...tionship1

2015.txt — Relationship7

2014.txt — ...onship6

2016.txt

StagingCombinedNew — ...ip8

Column	Type
State Name	Varchar(max)
State Abbreviation	Varbinary(max)
State FIPS	Int
Part D Prescribers	Bigint
Part D Opioid Prescribers	Bigint
Opioid Claims	Bigint
Extended Release Opioid Claims	Bigint
Overall Claims	Bigint
Opioid Prescribing Rate	Float
Extended Release Opioid Prescribing Rate	Float
IX_Relationship6 (IX1)	
IX_Relationship7 (IX2)	
IX_Relationship8 (IX3)	

Relationship9

Destination_Table

Column	Type
State Name	Varchar(max)
State Abbreviation	Varbinary(max)
State FIPS	Int
Part D Prescribers	Bigint
Part D Opioid Prescribers	Bigint
Opioid Claims	Bigint
Extended Release Opioid Claims	Bigint
Overall Claims	Bigint
Opioid Prescribing Rate	Float
Extended Release Opioid Prescribing Rate	Float
Date	Date

Relationship10 Relationship11

PackageLoadInformation

Column	Type
PkgId	Int
PackageName	Varchar(max)
InsertRecordCount	Int
LoadDateTime	Datetime
IX_Relationship10 (IX1)	

Aggregates

Column	Type
Part D Prescribers	Bigint
Part D Opioid Prescribers	Bigint
Opioid Claims	Bigint
Extended Release Opioid Claims	Bigint
Overall Claims	Bigint
Opioid Prescribing Rate	Float
Extended Release Opioid Prescribing Rate	Float
IX_Relationship9 (IX1)	

Relationship12

Relationship13

Match

Column	Type
Part D Prescribers	Bigint
Part D Opioid Prescribers	Bigint
Opioid Claims	Bigint
Extended Release Opioid Claims	Bigint
Overall Claims	Bigint
Opioid Prescribing Rate	Float
Extended Release Opioid Prescribing Rate	Float
IX_Relationship9 (IX1)	

[1,2]

Unmatch

Column	Type
Part D Prescribers	Bigint
Part D Opioid Prescribers	Bigint
Opioid Claims	Bigint
Extended Release Opioid Claims	Bigint
Overall Claims	Bigint
Opioid Prescribing Rate	Float
Extended Release Opioid Prescribing Rate	Float
IX_Relationship9 (IX1)	

[2,2]