

Azure Datawarehouse Pipeline

CS/2019/028

M.W.C.J.Wijethunga

Azure services



Resources

Recent Favorite

Name	Type	Last Viewed
sourceDB	Azure SQL Database Hyperscale	7 hours ago
targetDW	Azure SQL Database Hyperscale	8 hours ago
jsonstorageassignment	Storage account	16 hours ago
chiraserver	SQL server	17 hours ago
chiradatafactory	Data factory (V2)	21 hours ago
rg2	Resource group	2 days ago
Azure for Students	Subscription	2 days ago

See all

SECTION A: DIMENSIONAL MODELING (I)

Measures- • Sales

- Quantity Sold
- Marketing Expenditure

Dimensions-

Product (product key, product_name, price, category, ProductID)

Customer (customer key, customer_name, contact_no, CustomerID)

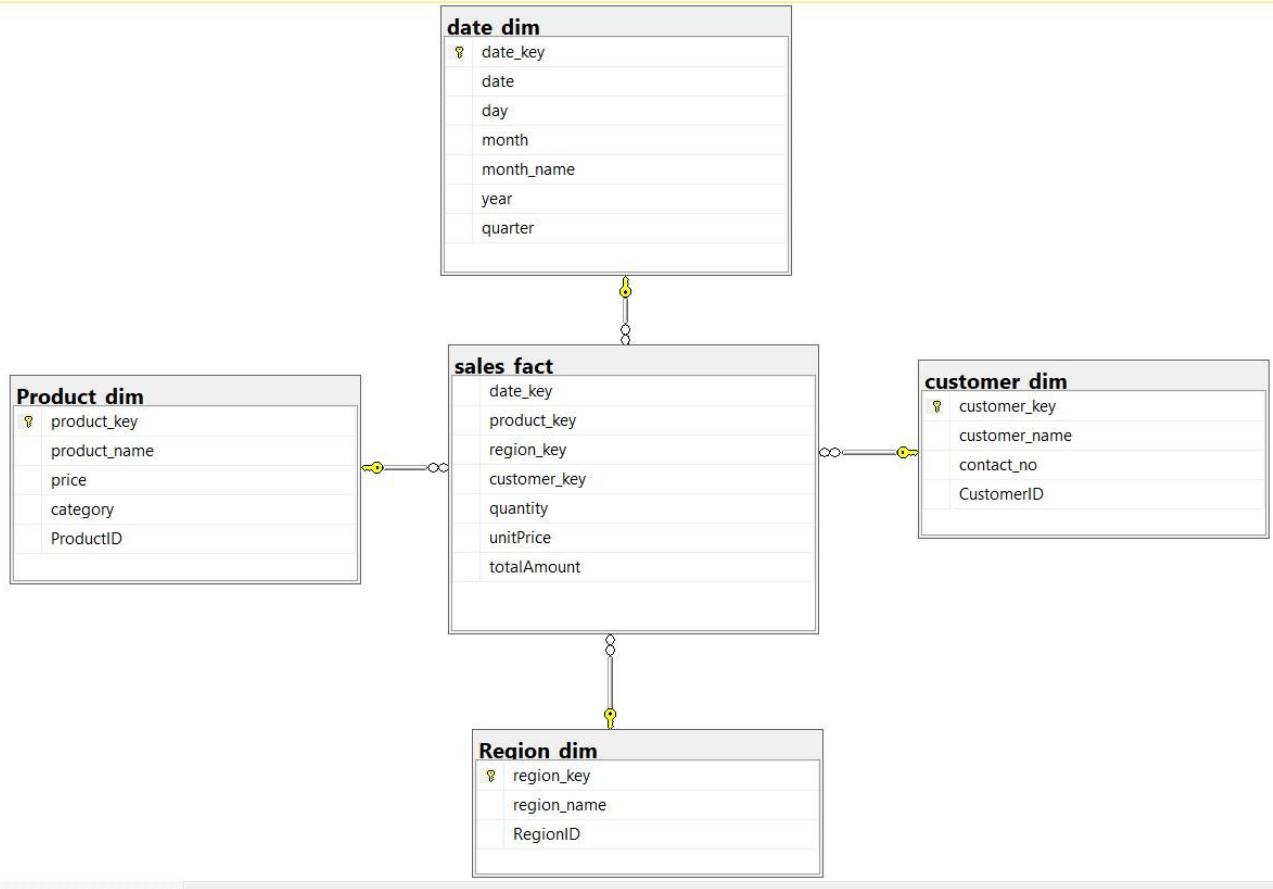
Region (region key, region_name, RegionID)

Date (date key, date, day, month, month_name, year, quarter)

Bus Matrix-

	Product	Customer	Region	Date
Sales	X	X	X	X
Quantity Sold	X	X	X	X
Marketing Expenditure			X	X

(II)



SECTION B: PHYSICAL DW DESIGN AND ETL USING ADF

(III)

Created a source database using Azure SQL database.

1) Created SQL server named “chiraserver” in Azure.

- Resource Group- rg2
- Authentication- SQL Authentication

Home >

chiraserver SQL server

Search Create database New elastic pool New dedicated SQL pool (formerly SQL DW) Import database Reset password Move Delete Feedback

Overview

Essentials

Resource group (move)	: rg2	Server admin	: chirantha
Status	: Available	Networking	: Show networking settings
Location	: Central India	Microsoft Entra admin	: Not configured
Subscription (move)	: Azure for Students	Server name	: chiraserver.database.windows.net
Subscription ID	: 5b64cd6e-b500-4ee8-a66e-073302845fb5		
Tags (edit)	: Add tags		

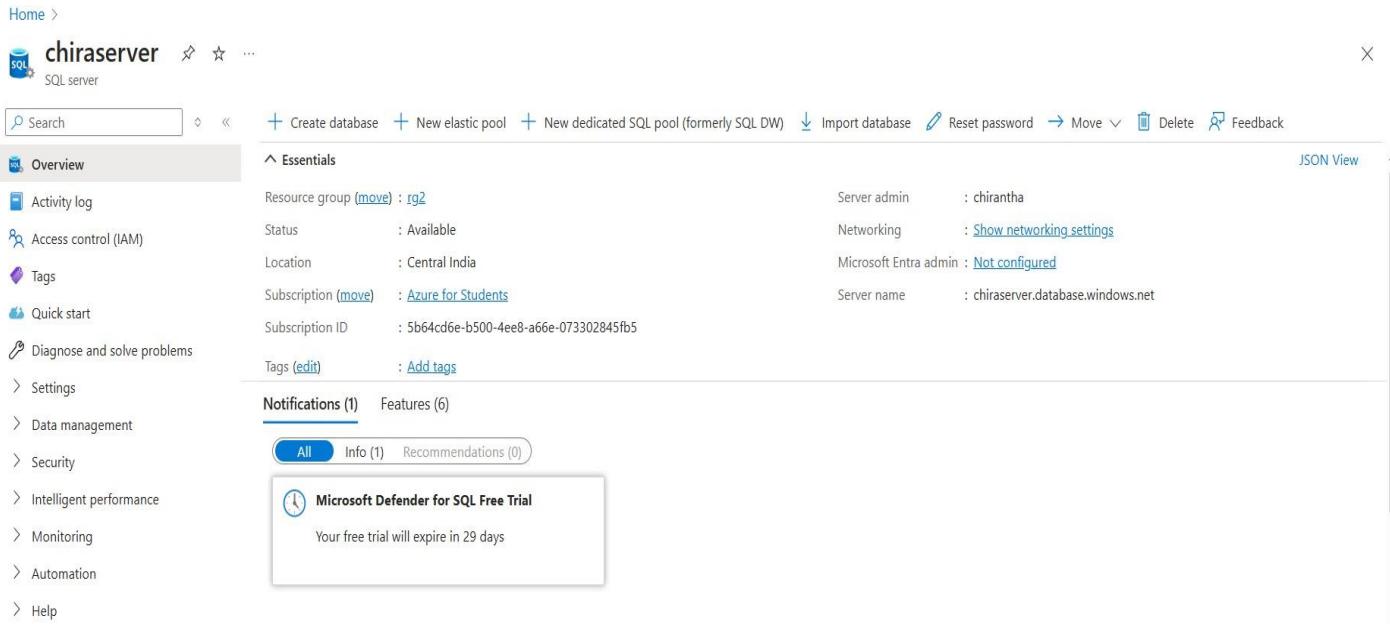
Activity log Access control (IAM) Tags Quick start Diagnose and solve problems Settings Data management Security Intelligent performance Monitoring Automation Help

Notifications (1) Features (6)

All Info (1) Recommendations (0)

Microsoft Defender for SQL Free Trial
Your free trial will expire in 29 days

JSON View



After this went to chiraserver ➔ Security ➔ Networking And then set public network IP address.

2) Created SQL database for Source Database named “sourceDB” in Azure.

Home >

sourceDB (chiraserver/sourceDB) Azure SQL Database Hyperscale

Search Copy Restore Export Set server firewall Delete Connect with... Feedback

Overview

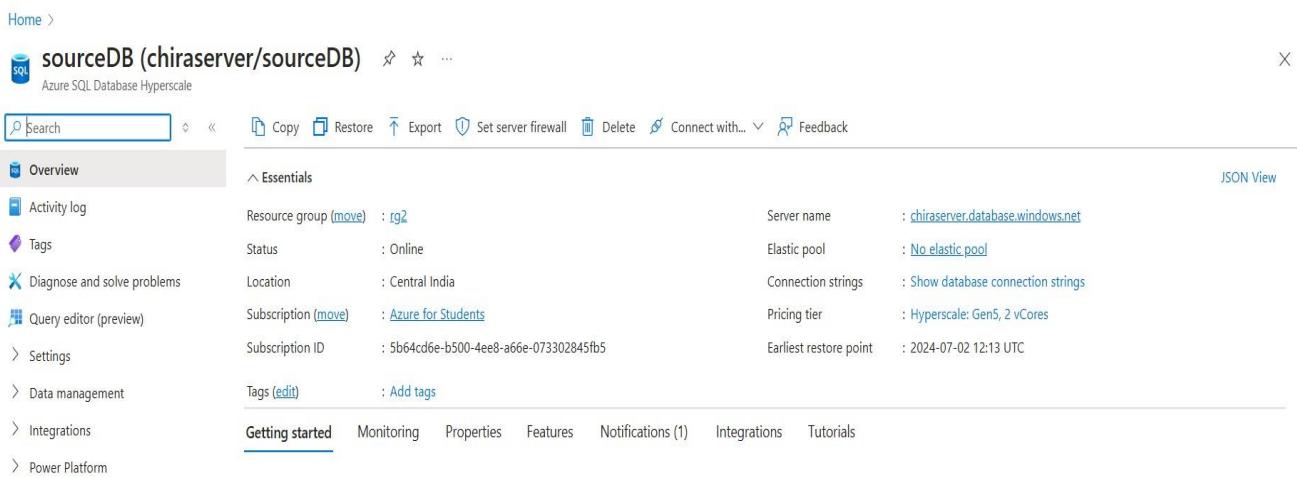
Essentials

Resource group (move)	: rg2	Server name	: chiraserver.database.windows.net
Status	: Online	Elastic pool	: No elastic pool
Location	: Central India	Connection strings	: Show database connection strings
Subscription (move)	: Azure for Students	Pricing tier	: Hyperscale: Gen5, 2 vCores
Subscription ID	: 5b64cd6e-b500-4ee8-a66e-073302845fb5	Earliest restore point	: 2024-07-02 12:13 UTC
Tags (edit)	: Add tags		

Activity log Tags Diagnose and solve problems Query editor (preview) Settings Data management Integrations Power Platform

Getting started Monitoring Properties Features Notifications (1) Integrations Tutorials

JSON View



Then created tables inside this “sourceDB” and inserted data into the 4 tables provided.

Queries to create tables in sourceDB-

```
CREATE TABLE Products (
    ProductID VARCHAR(10) PRIMARY KEY,
    ProductName VARCHAR(100),
    Price DECIMAL(10, 2),
    Category VARCHAR(50)
);
```

```
CREATE TABLE Regions (
    RegionID INT PRIMARY KEY,
    RegionName VARCHAR(100)
);
```

```
CREATE TABLE Customers (
    CustomerID VARCHAR(10) PRIMARY KEY,
    CustomerName VARCHAR(100),
    ContactNumber VARCHAR(15),
    RegionID INT,
    FOREIGN KEY (RegionID) REFERENCES Regions(RegionID)
);
```

```
CREATE TABLE Sales (
    SaleID VARCHAR(10) PRIMARY KEY,
    ProductID VARCHAR(10),
    CustomerID VARCHAR(10),
```

```
Date DATE,  
Quantity INT,  
TotalAmount DECIMAL(10, 2),  
FOREIGN KEY (ProductID) REFERENCES Products(ProductID),  
FOREIGN KEY (CustomerID) REFERENCES  
Customers(CustomerID)  
);
```

Queries to insert data into tables in sourceDB-

```
INSERT INTO Regions (RegionID, RegionName) VALUES  
(1, 'North'),  
(2, 'South'),  
(3, 'East');
```

```
INSERT INTO Products (ProductID, ProductName, Price, Category)  
VALUES  
('P100', 'Widget A', 120, 'Electronics'),  
('P101', 'Widget B', 150, 'Electronics'),  
('P102', 'Gadget C', 90, 'Home'),  
('P103', 'Gadget D', 110, 'Home');
```

```
INSERT INTO Customers (CustomerID, CustomerName,  
ContactNumber, RegionID) VALUES  
('C100', 'John Doe', '1234567890', 1),  
('C101', 'Jane Smith', '1234567891', 2),  
('C102', 'Alice Johnson', '1234567892', 3),
```

```
('C103', 'Bob Brown', '1234567893', 1);
```

```
INSERT INTO Sales (SaleID, ProductID, CustomerID, Date,
Quantity, TotalAmount) VALUES
('S011', 'P100', 'C100', '2024-01-01', 2, 240),
('S012', 'P101', 'C101', '2024-01-02', 1, 150),
('S013', 'P102', 'C102', '2024-01-03', 3, 270),
('S014', 'P103', 'C103', '2024-01-04', 1, 110);
```

select*from Regions; -- this is for a example

Then got below table lists for SourceDB.

The screenshot shows the Azure SQL Database Hyperscale Query editor interface. On the left, the 'Tables' sidebar lists three tables: dbo.Customers, dbo.Products, and dbo.Regions. The 'dbo.Sales' table is currently selected, showing its schema with columns: SaleID (PK, varchar, not null), ProductID (varchar, null), CustomerID (varchar, null), Date (date, null), Quantity (int, null), and TotalAmount (decimal(10, 2)). A query editor window titled 'Query 1' contains the SQL code for creating the Sales table, which includes a FOREIGN KEY constraint referencing the ProductID column in the Products table. Below the query editor is a results grid showing the data from the dbo.Regions table, with three rows: RegionID 1 (RegionName North), RegionID 2 (RegionName South), and RegionID 3 (RegionName East). At the bottom of the interface, a message bar indicates 'Query succeeded | 0s'.

customer

sourceDB (chiraserver/sourceDB) | Query editor (preview)

Azure SQL Database Hyperscale

chirantha

sourceDB (chirantha)

Showing limited object explorer here. For full capability please click here to open Azure Data Studio.

Tables

- dbo.Customers
- dbo.Products
- dbo.Regions
- dbo.Sales
- Views
- Stored Procedures

Query 1

Run Cancel query Save query Export data as Show only Editor

```
1 select * from customers;
```

Results Messages

CustomerID	CustomerName	ContactNumber	RegionID
C100	John Doe	1234567890	1
C101	Jane Smith	1234567891	2
C102	Alice Johnson	1234567892	3
C103	Bob Brown	1234567893	1

Query succeeded | 0s

Products

Home > sourceDB (chiraserver/sourceDB)

sourceDB (chiraserver/sourceDB) | Query editor (preview)

Azure SQL Database Hyperscale

chirantha

sourceDB (chirantha)

Showing limited object explorer here. For full capability please click here to open Azure Data Studio.

Tables

- dbo.Customers
- dbo.Products
- dbo.Regions
- dbo.Sales
- Views
- Stored Procedures

Query 1

Run Cancel query Save query Export data as Show only Editor

```
1 select * from Products;
```

Results Messages

ProductID	ProductName	Price	Category
P100	Widget A	120.00	Electronics
P101	Widget B	150.00	Electronics
P102	Gadget C	90.00	Home
P103	Gadget D	110.00	Home

Query succeeded | 0s

Region

The screenshot shows the Azure SQL Database Hyperscale Query editor (preview) interface. The left sidebar displays the database schema for 'sourceDB (chiraserver/sourceDB)', specifically the 'Tables' section which includes 'dbo.Customers', 'dbo.Products', 'dbo.Regions', 'dbo.Sales', 'Views', and 'Stored Procedures'. The main area is titled 'Query 1' and contains the SQL command: 'select * from Regions;'. The results pane shows the following data:

RegionID	RegionName
1	North
2	South
3	East

A message at the bottom indicates 'Query succeeded | 0s'.

Sales

The screenshot shows the Azure SQL Database Hyperscale Query editor (preview) interface. The left sidebar displays the database schema for 'sourceDB (chiraserver/sourceDB)', specifically the 'Tables' section which includes 'dbo.Customers', 'dbo.Products', 'dbo.Regions', 'dbo.Sales', 'Views', and 'Stored Procedures'. The main area is titled 'Query 1' and contains the SQL command: 'select * from sales;'. The results pane shows the following data:

SaleID	ProductID	CustomerID	Date	Quantity	TotalAmount
S011	P100	C100	2024-01-01	2	240.00
S012	P101	C101	2024-01-02	1	150.00
S013	P102	C102	2024-01-03	3	270.00
S014	P103	C103	2024-01-04	1	110.00

A message at the bottom indicates 'Query succeeded | 0s'.

(IV) Created target Database named “targetDW” in Azure.

Home >

targetDW (chiraserver/targetDW) ⚡ ☆ ...

Azure SQL Database Hyperscale

Search Copy Restore Export Set server firewall Delete Connect with... Feedback

Overview ▲ Essentials JSON View

Activity log Resource group (move) : rg2 Server name : [chiraserver.database.windows.net](#)
Tags Status : Online Elastic pool : [No elastic pool](#)
Diagnose and solve problems Location : Central India Connection strings : Show database connection strings
Query editor (preview) Subscription (move) : [Azure for Students](#) Pricing tier : Hyperscale: Gen5, 2 vCores
Settings Subscription ID : 5b64cd6e-b500-4ee8-a66e-073302845fb5 Earliest restore point : 2024-07-02 12:47 UTC
Data management Tags (edit) : Add tags
Integrations Getting started Monitoring Properties Features Notifications (1) Integrations Tutorials
Power Platform

Then created tables inside this “targetDW” .

Queries to create tables in targetDW -

```
CREATE TABLE Product_dim ( product_key INT  
IDENTITY(1,1) PRIMARY KEY, product_name  
VARCHAR(100), price DECIMAL(10, 2), category  
VARCHAR(50),  
ProductID VARCHAR(10)  
);
```

```
CREATE TABLE Region_dim ( region_key INT  
IDENTITY(1,1) PRIMARY KEY, region_name  
VARCHAR(100),  
RegionID VARCHAR(50)  
);
```

```
CREATE TABLE customer_dim (  customer_key INT  
IDENTITY(1,1) PRIMARY KEY,  customer_name  
VARCHAR(100),  contact_no VARCHAR(15),  
CustomerID VARCHAR(10)  
);
```

```
CREATE TABLE date_dim (  date_key INT  
IDENTITY(1,1) PRIMARY KEY,  date DATE,  day  
INT,  month INT,  month_nam VARCHAR(15),  
year INT,  quarter INT  
);
```

```
CREATE TABLE sales_fact(  
  
date_key int FOREIGN KEY REFERENCES date_dim(date_key),  product_key int  
FOREIGN KEY REFERENCES Product_dim(product_key),  region_key int FOREIGN KEY  
REFERENCES Region_dim(region_key),  customer_key int FOREIGN KEY REFERENCES  
customer_dim(customer_key),      quantity INT,  
unitPrice DECIMAL(10, 2),  totalAmount  
DECIMAL(10, 2),  
);
```

Then got below table list for targetDW.

In below figure we can see 4 Dimension tables and 1 fact table.

Dimension tables - customer_dim,date_dim,Product_dim,Region_dim. Fact table- sales_fact

The screenshot shows the Azure SQL Database Hyperscale Query editor interface. The left sidebar displays the database structure with a 'Tables' section containing 'customer_dim', 'date_dim', 'Product_dim', and 'Region_dim'. The main area is titled 'Query 1' and contains the following T-SQL code:

```
10 CREATE TABLE Region_dim (
11     region_key INT IDENTITY(1,1) PRIMARY KEY,
12     region_name VARCHAR(100),
13     RegionID VARCHAR(50)
14 );
15
16
17 CREATE TABLE customer_dim (
18     customer_key INT IDENTITY(1,1) PRIMARY KEY,
```

Before consider the ETL part, created an Azure storage account named “jsonstorageassignment” to store Real_Time_Sales_Data.json file.

Home > Storage accounts >

Create a storage account

storage accounts 

Project details

Select the subscription in which to create the new storage account. Choose a new or existing resource group to organize and manage your storage account together with other resources.

Subscription *	Azure for Students
Resource group *	rg2 Create new

Instance details

Storage account name * 	assignment
Region * 	(Asia Pacific) Central India Deploy to an Azure Extended Zone
Performance * 	<input checked="" type="radio"/> Standard: Recommended for most scenarios (general-purpose v2 account) <input type="radio"/> Premium: Recommended for scenarios that require low latency.
Redundancy * 	Geo-redundant storage (GRS) <input checked="" type="checkbox"/> Make read access to data available in the event of regional unavailability.

[Previous](#)

[Next](#)

Review + create

Then uploaded the json file to a created named “chirantha container” container. (In upcoming slides we’ll see for where should we use this json file) Next we have to consider the ETL part. (V)

Use Azure Data Factory to perform ETL from 3 data sources to the target data warehouse. **1)**

First created Azure data factory named “**chiradataproxy**”

- Resource Group- rg2

Home > chiradataproject Data factory (V2)

Overview

Essentials

Resource group (move) : rg2
Status : Succeeded
Location : Central India
Subscription (move) : Azure for Students
Subscription ID : 5b64cd6e-b500-4ee8-a66e-073302845fb5

Type : Data factory (V2)
Getting started : Quick start

Activity log
Access control (IAM)
Tags
Diagnose and solve problems
Settings
Getting started
Monitoring
Automation
CLI / PS
Tasks (preview)
Help

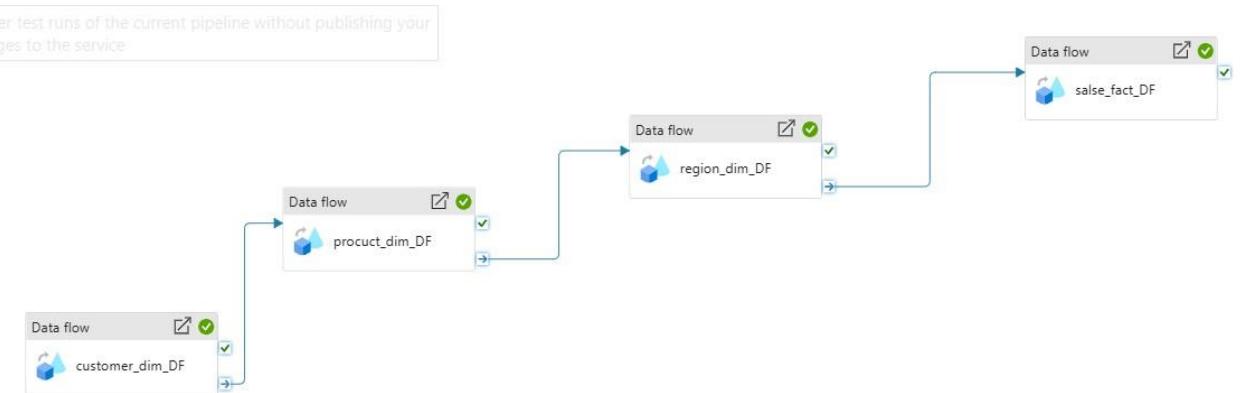
Azure Data Factory Studio

Launch studio

Then clicked on ‘Launch studio’ button.

2) Then created new pipeline named “ETL pipe”.

Next created separate data flows named, “region_dim_DF”, “customer_dim_DF” and “product_dim_DF”, “sales_fact_DF” Like below.



To understand this let’s consider creating “region_dim_DF”.

- First double clicked on created ETL pipe and dragged and dropped a Data Flow. Then rename it as “region_dim_DF” and Debug turned on.
-

But still this “region_dim_DF” is not display under Data flows. To display it,

Settings clicked new and then renamed it as “region_dim_DF”.

Then clicked Add Source.

iii) Go to created “ETL pipe”. Then,

Settings Under Data flow, select “region_dim_DF”.

iv) Go to Source setting set Output stream name as “regionSource”. Then under Data set click +New and select Azure SQL database.

Under Linked service click +New and create link service then hit test.

New linked service

Azure SQL Database [Learn more](#)

Version
 Recommended Legacy

[Connection string](#) [Azure Key Vault](#)

Account selection method
 From Azure subscription Enter manually

Azure subscription
Azure for Students (5b64cd6e-b500-4ee8-a66e-073302845fb5)

Server name *
chiraserver

Database name *
sourceDB

Authentication type *
SQL authentication

User name *
chirantha

[Password](#) [Azure Key Vault](#)

Password *

Connection successful Test connection

[Create](#) [Cancel](#)

After click on create, the Linked service is created.

Then in ‘Set properties’ used the above created new link and select region table under table name.

V) Then again test connection. After that we can see like below.

The screenshot shows the Microsoft Power BI Data Flow interface. At the top, there are tabs for 'ETL pipe', 'customer_dim_DF', 'product_dim_DF', and 'region_dim_DF'. Below the tabs, there are buttons for 'Validate' (with a checkmark), 'Data flow debug' (with a blue toggle switch), and 'Debug Settings'. A search icon and a '+' icon are also present.

The main area displays a source named 'regionresource' with a preview icon. It shows 'Columns: 2 total'. Below this, the 'Source settings' tab is selected, followed by 'Source options', 'Projection', 'Optimize', 'Inspect', and 'Data preview' (with a green dot).

Under 'Source settings', the 'Output stream name' is set to 'regionresource' with a 'Learn more' link. The 'Description' field contains 'Import data from Regionsourcetable' with a 'Reset' button. The 'Source type' section has two options: 'Dataset' (selected) and 'Inline'. The 'Dataset' dropdown is set to 'Regionsourcetable'.

Below the source type, there is a 'Connection successful' message with a checkmark, a 'Test connection' button, and 'Open' and 'New' buttons. Under 'Options', there are three checkboxes: 'Allow schema drift' (checked), 'Infer drifted column types' (unchecked), and 'Validate schema' (unchecked). Under 'Sampling', there are 'Enable' and 'Disable' radio buttons.

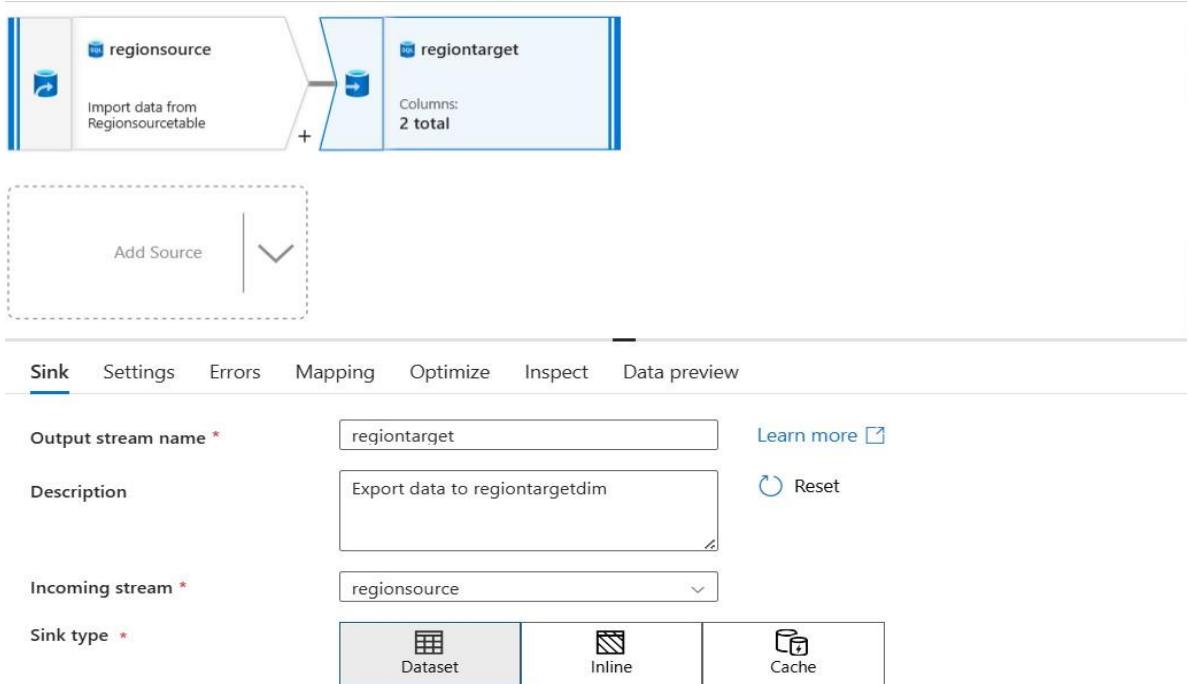
If it display 'connection successful', Go to the Data preview and refresh.

Then we can see the data in region table like below.

The screenshot shows the Azure Data Studio interface for an ETL pipe. At the top, there are tabs for 'ETL pipe', 'customer_dim_DF', 'product_dim_DF', and 'region_dim_DF'. Below the tabs, there are buttons for 'Validate', 'Data flow debug' (which is turned on), and 'Debug Settings'. On the left, a source dataset named 'regionsource' is selected, showing '2 total' columns. A dashed box labeled 'Add Source' is visible. To the right, there are icons for search, add, remove, and edit. Below the source selection, a 'Data preview' tab is active, showing the following data:

RegionID	RegionName
1	North
2	South
3	East

- v) Then we have to configure Sink. Using created ‘regionsource’ add a Destination sink call “regiontarget”. Added new dataset named “regiontargetdim”. (It will show in upcoming slides.)

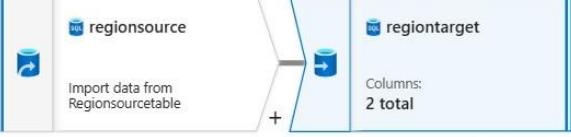


Now again we have to create new link service. (Followed same way which we used to create previous link service.)

Like below select “regiontarget” and Test connection.

ETL pipe ● customer_dim_DF ● procut_dim_DF ● region_dim_DF

✓ Validate Data flow debug   Debug Settings



Sink Settings Errors Mapping Optimize Inspect Data preview 

Output stream name * regiontarget  

Description Export data to regiontargetdim 

Incoming stream * regionresource 

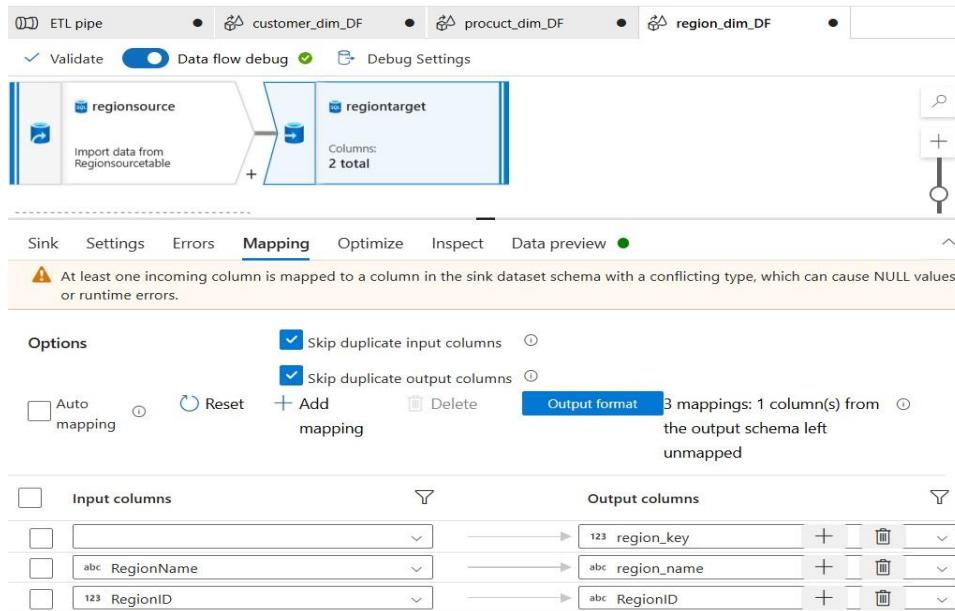
Sink type *  Dataset  Inline  Cache

 Connection successful

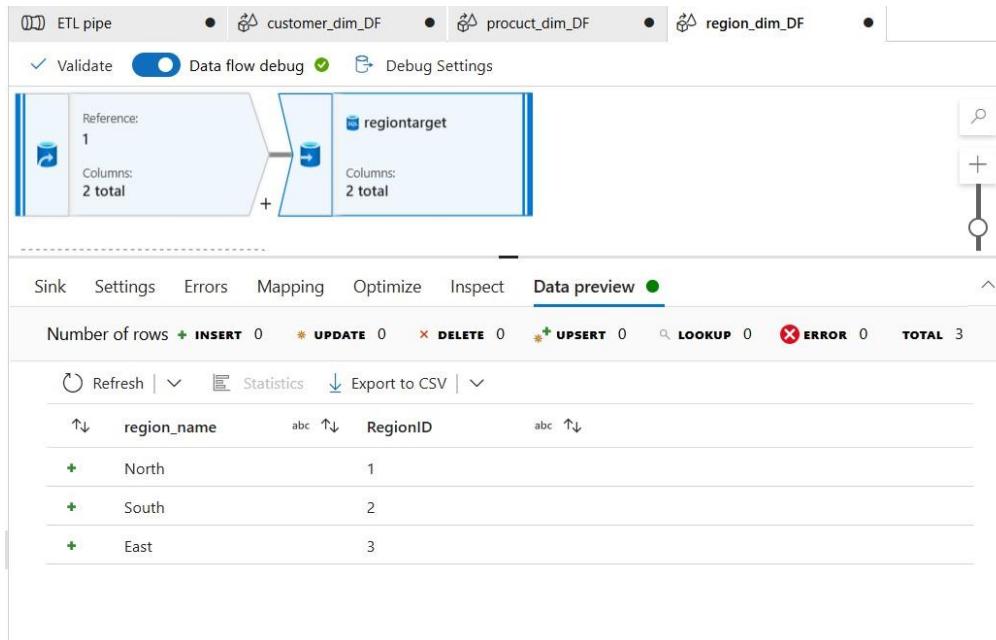
Dataset * regiontargetdim   Open  New

Options Allow schema drift  Validate schema 

Next we have to do the mapping manually.



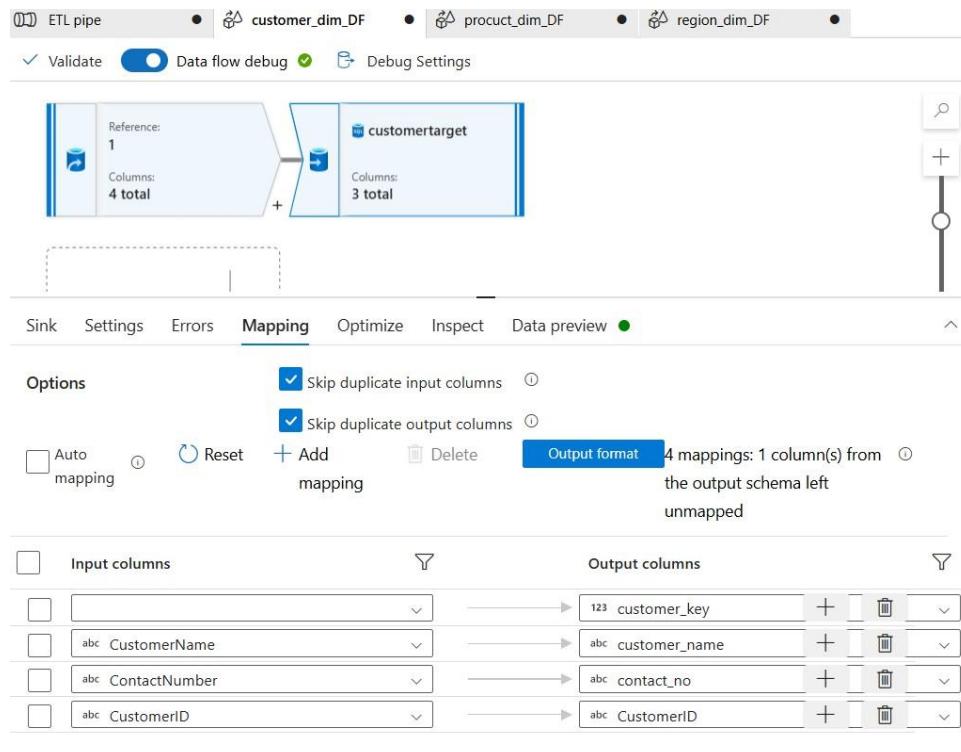
Then go to data preview and check whether mapping is correct. If it is successfully done, it will show like below.



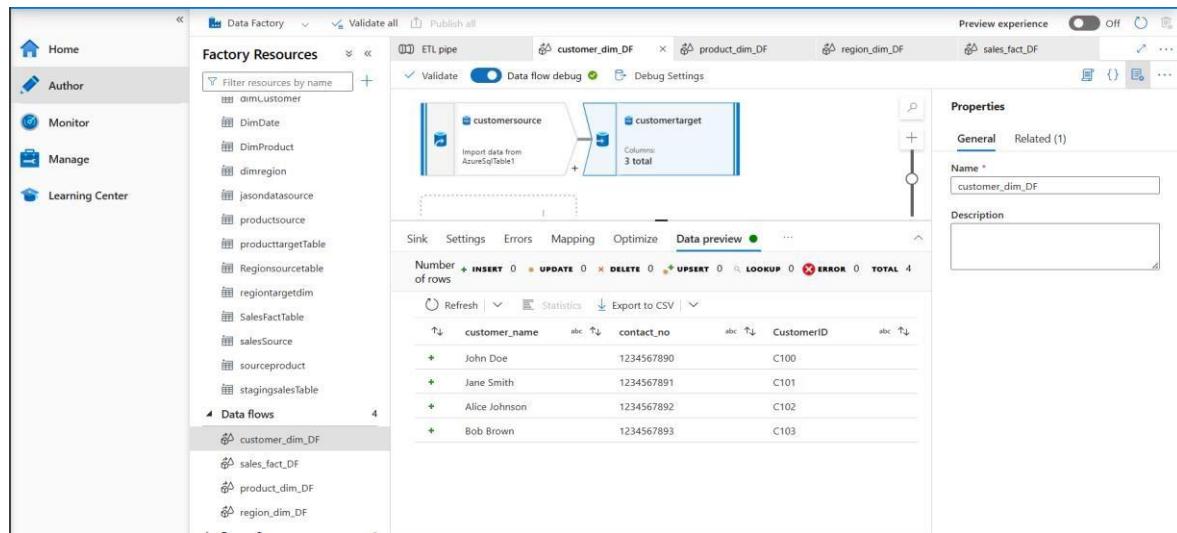
We can follow same procedure to create data flows for “customer_dim_DF” and “product_dim_DF” like above. (I did not create a data flow for date_dim.

Populated date dimension manually using SQL query like below. We'll discuss it in (VI))

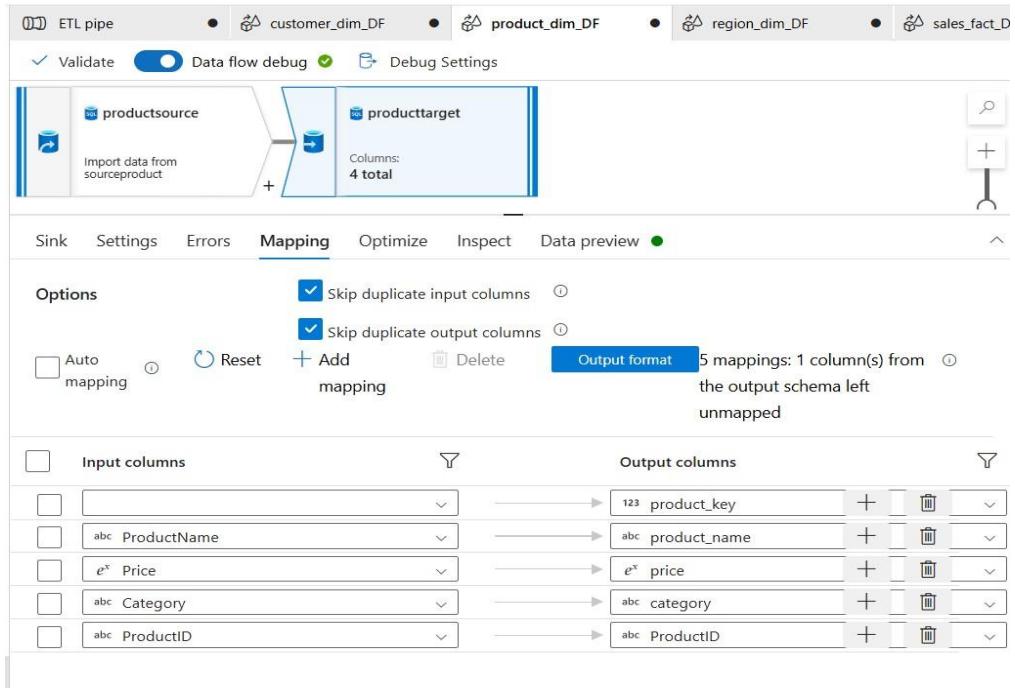
customer_dim_DF mapping-



Preview after the Link service -



product_dim_DF mapping-



Preview after the Link service –

The screenshot shows the 'Author' blade of the Data Factory interface. The 'product_dim_DF' data flow is selected. The 'Data preview' tab is active, displaying the following data:

product_n...	abc	price	e ^x	category	abc	ProductID	abc
Widget A		120.00		Electronics		P100	
Widget B		150.00		Electronics		P101	
Gadqet C		90.00		Home		P102	
Gadqet D		110.00		Home		P103	

Properties

- Name:** product_dim_DF
- Description:** this product dim

We have to get sales data from 2 sources.(From source database and json file)

To get sales data from sourceDB, used same procedure which we followed above.

The screenshot shows the 'Source settings' tab for a 'sales_fact' source in an ETL pipe. The top navigation bar includes tabs for 'ETL pipe', 'customer_dim_DF', 'product_dim_DF', 'region_dim_DF', and 'sales_fact'. Below the tabs, there are buttons for 'Validate' (green checkmark), 'Data flow debug' (blue toggle switch), and 'Debug Settings'.

The main area displays a preview of the 'salesfactsource' dataset, showing 6 total columns. The 'Source settings' tab is selected, followed by 'Source options', 'Projection', 'Optimize', 'Inspect', and 'Data preview'.

Configuration details:

- Output stream name ***: salesfactsource
- Description**: Import data from salesSource
- Source type ***: Dataset (selected)
- Dataset**: salesSource
- Options**:
 - Connection successful
 - Test connection
 - Open
 - New
- Samplina ***:
 - Enable
 - Disable

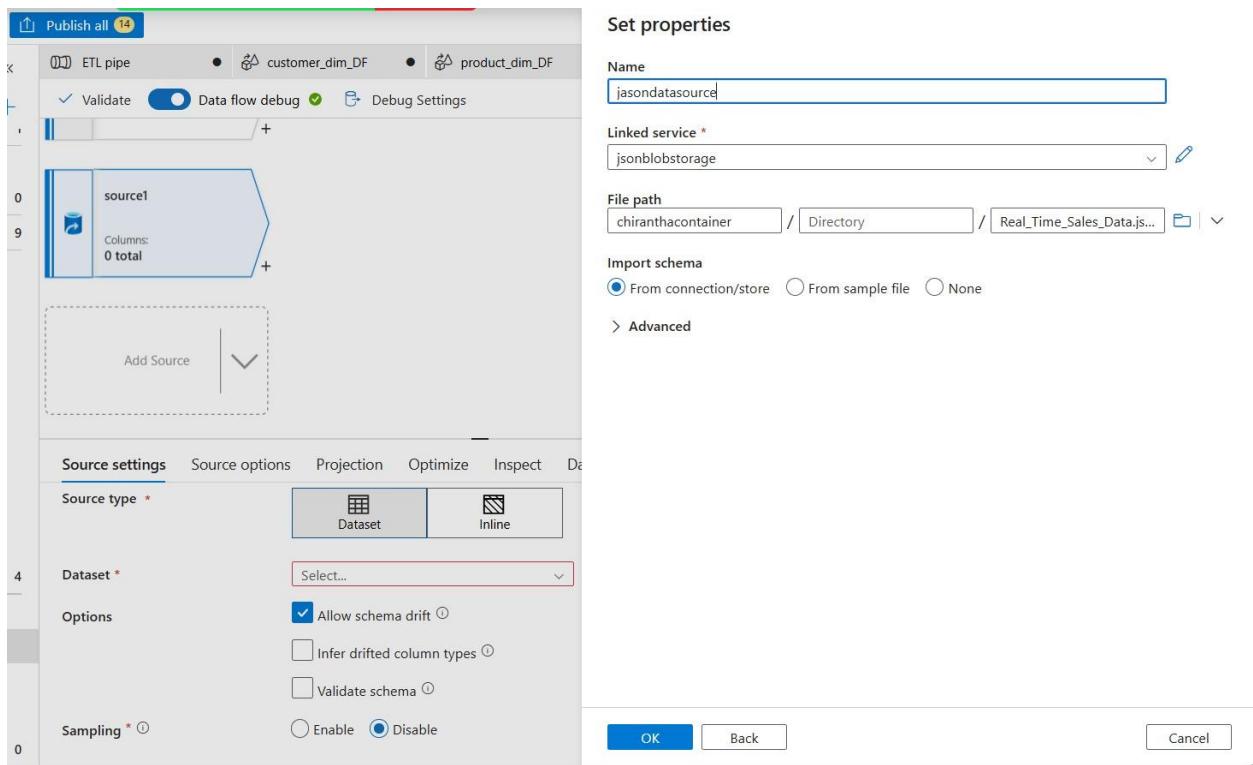
We have to get sales data from json file also. For that , first created Azure storage account named “jsonstorageassignment” and inside it created a container named

“chiranthaccontainer”. After creation of blob storage uploaded the json file to it. Then we can access it through the Azure Data Factory.

Then we should create another link service to access above blob storage. Lets follow below steps for that.

The screenshot shows the Azure Data Factory interface with the following details:

- Factory Resources:** A sidebar listing various resources including Datasets (AzureSqlTable1, AzureSqlTable2, AzureSqlTable3, productsource, producttargetsource, Regionsource, regiontargetdim, salesSource, sourceproduct), Data flows (customer_dim_DF, sales_fact_DF, product_dim_DF, region_dim_DF), and Power Query.
- ETL pipe:** The main workspace displays an ETL pipe named "ETL pipe" with a single stage named "source1". The "Source settings" tab is selected, showing "Dataset" as the source type and "Select..." as the dataset. Other options include "Allow schema drift" (checked), "Infer drifted column types", "Validate schema", and sampling settings.
- Select format:** A modal window titled "Select format" allows choosing the data format. Options include Avro, CSV, Excel, JSON, ORC, Parquet, XML, and Binary. "JSON" is selected.
- New linked service:** A modal window titled "New linked service" is open, specifically for "Azure Blob Storage". It includes fields for "Authentication type" (Account key), "Connection string" (selected), "Account selection method" (From Azure subscription selected), "Azure subscription" (Azure for Students selected), "Storage account name" (jsonstorageassignment selected), and "Additional connection properties". A "Create" button at the bottom right is highlighted.
- Status:** A message at the bottom right indicates "Connection successful" with a green checkmark icon.



After created the JSON source, the data look like below in Data Preview.
Column name- SalesData

The screenshot shows the 'Data preview' tab for the SalesData source. It indicates 1 inserted row. The preview table has one row with a single column labeled 'SalesData' containing '[...]'.

(It's like complex data type .To convert this nested JSON data into a flat, tabular format, we can use the Flatten transformation. In upcoming slide we'll discuss it.)

Added a new Column named “salesSourceUnitPrice” using derived column option.

Used expression ⑦ UnitPrice=TotalAmout/Quantity

The screenshot shows the Microsoft Power BI Data Flow interface. At the top, there are five tabs: ETL pipe, customer_dim_DF, product_dim_DF, region_dim_DF, and sales_fact_DF. Below these are three buttons: Validate (green checkmark), Data flow debug (blue switch), and Debug Settings. The main area displays a data flow diagram with a source named "salesSource" (represented by a blue cylinder icon) connected to a derived column step. The derived column step has a green hexagonal icon and is labeled "salesSourceUnitPrice". A tooltip indicates "Columns: 7 total". Below the diagram, there are several tabs: "Derived column's settings" (which is selected), Optimize, Inspect, and Data preview (with a green dot). Under "Derived column's settings", there are fields for "Output stream name" (set to "salesSourceUnitPrice"), "Description" (containing the text "Creating/updating the columns 'SaleID, ProductID, CustomerID, Date, Quantity, TotalAmount, UnitPrice'"), and "Incoming stream" (set to "salesSource"). There are also buttons for "Add", "Clone", "Delete", and "Open expression builder". In the "Columns" section, there is a table:

Column	Expression
UnitPrice	TotalAmount/Quantity

At the bottom right of the table, there is a link "Open expression builder".

Here used 'Flatten' transformation is used to transform hierarchical data structures, such as JSON, into a flat, tabular format. Set the "Unroll by" property to the SalesData array like below. Unroll by- [] SalesData

ETL pipe customer_dim_DF product_dim_DF region_dim_DF sales_fact_DF

Validate Data flow debug Debug Settings

salesjsonsource 6 Columns

Flatten settings Optimize Inspect Data preview

Output stream name *: salesjsonflatten Help Learn more

Description: Unrolling arrays from SalesData to with columns 'SaleID, ProductID, CustomerID, Date, Quantity, UnitPrice'

Incoming stream *: salesjsonsource

Unroll by * ⓘ: [] SalesData

Unroll root ⓘ: {}

Options:

- Skip duplicate input columns ⓘ
- Skip duplicate output columns ⓘ

Then Data Preview was like below.

The screenshot shows the Alteryx Data Designer interface with the following components:

- ETL pipe:** A horizontal bar at the top with five items: "ETL pipe" (selected), "customer_dim_DF", "product_dim_DF", "region_dim_DF", and "sales_fact_DF".
- Data flow debug:** A toggle switch labeled "Validate" and "Data flow debug" with a green checkmark.
- Debug Settings:** A button to open settings for debugging.
- Flow Diagram:** A visual representation of the data pipeline:
 - A "salesjsonsource" component (represented by a blue cylinder icon) is connected to a "salesjsonflatten" component (represented by a blue hexagon icon).
 - The "salesjsonsource" component has a tooltip: "Import data from jsondatasource".
 - The "salesjsonflatten" component has a tooltip: "Columns: 1 total".
- Data preview:** A tabbed section showing the preview of the data. The "Data preview" tab is selected, indicated by a green dot.
- Statistics:** A summary of the data:
 - Number of rows: 10 (with buttons for INSERT, UPDATE, DELETE, UPSERT, LOOKUP, and ERROR).
 - TOTAL: 10
- Action Buttons:** A row of buttons for data manipulation:
 - Refresh
 - Typecast
 - Modify
 - Map drifted
 - Statistics
 - Remove
 - Export to CSV
- Table:** A detailed preview of the data in a table format. The columns are:
 - SaleID
 - Prod...
 - Custo...
 - Date
 - Quan...
 - UnitP...
- Data Rows:** Eight rows of data are shown:

SaleID	Prod...	Custo...	Date	Quan...	UnitP...
S001	P100	C500	2024-...	2	120
S002	P101	C501	2024-...	1	150
S003	P102	C502	2024-...	3	90
S004	P103	C503	2024-...	1	110
S005	P104	C504	2024-...	5	95
S006	P105	C505	2024-...	4	100
S007	P106	C506	2024-...	2	110
S008	P107	C507	2024-...	3	115

After that added derived column transformation for derived TotalAmount and convert data type of Date.

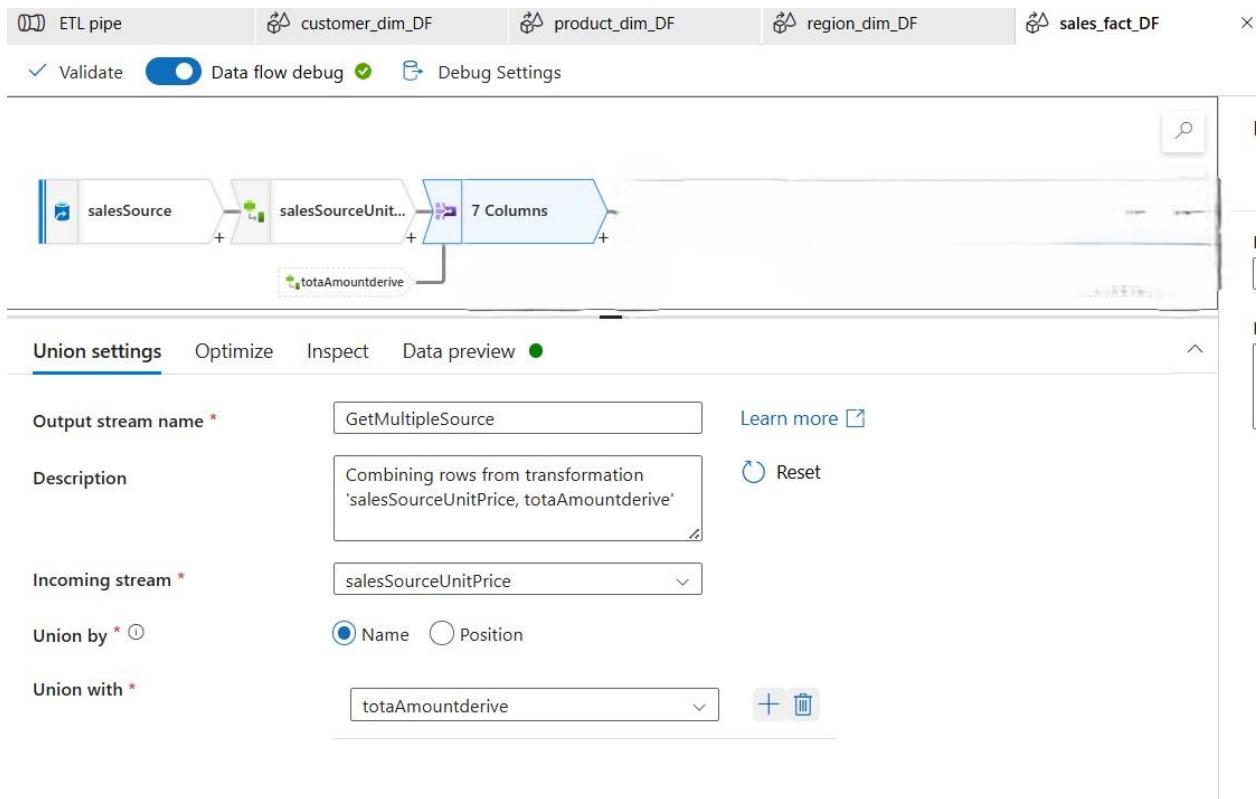
The screenshot shows the Azure Data Factory ETL pipe interface. At the top, there are five tabs: ETL pipe, customer_dim_DF, product_dim_DF, region_dim_DF, and sales_fact_DF. Below these are three buttons: Validate (checked), Data flow debug (on), and Debug Settings. The main area displays a data flow diagram with three stages: salesjsonsource, salesjsonflatten, and 7 Columns. The 7 Columns stage is currently selected. Below the diagram, there are four tabs: Derived column's settings (selected), Optimize, Inspect, and Data preview. The Derived column's settings tab contains the following fields:

- Output stream name ***: totaAmountderive
- Description**: Creating/updating the columns 'SaleID, ProductID, CustomerID, Date, Quantity, UnitPrice, TotalAmount'
- Incoming stream ***: salesjsonflatten

Below these fields are buttons for Add, Clone, Delete, and Open expression builder. The **Columns *** section lists three columns with their corresponding expressions:

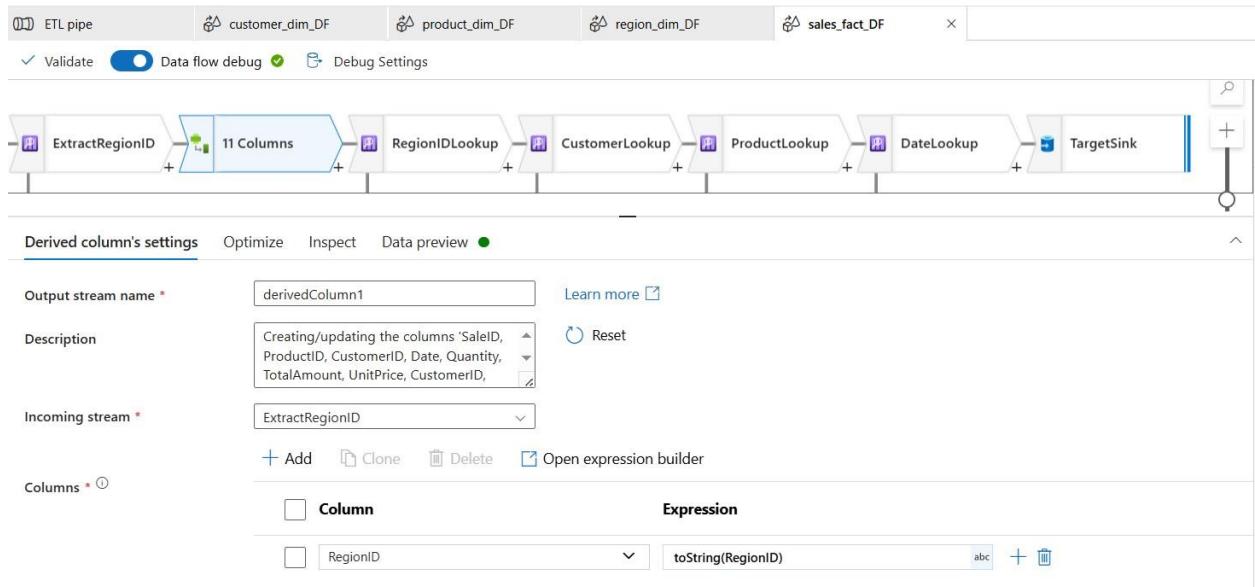
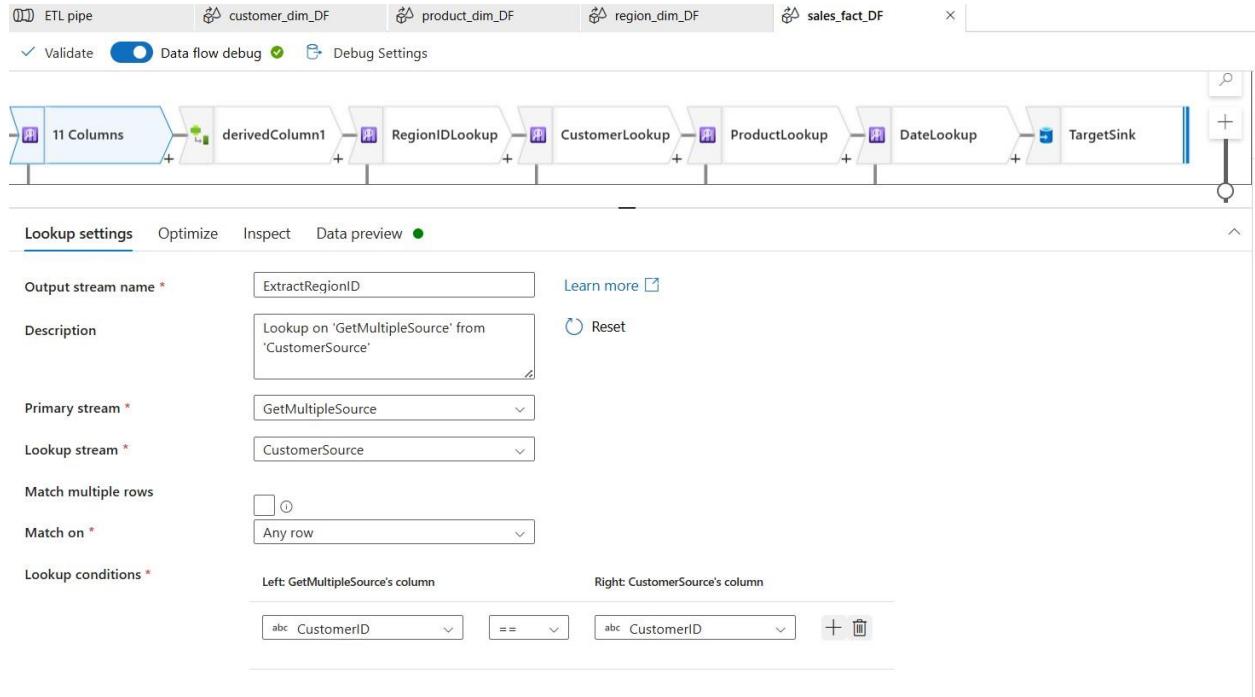
Column	Expression
TotalAmount	toDecimal(Quantity*UnitPrice)
Date	toDate(Date)
UnitPrice	toDecimal(UnitPrice,10,2)

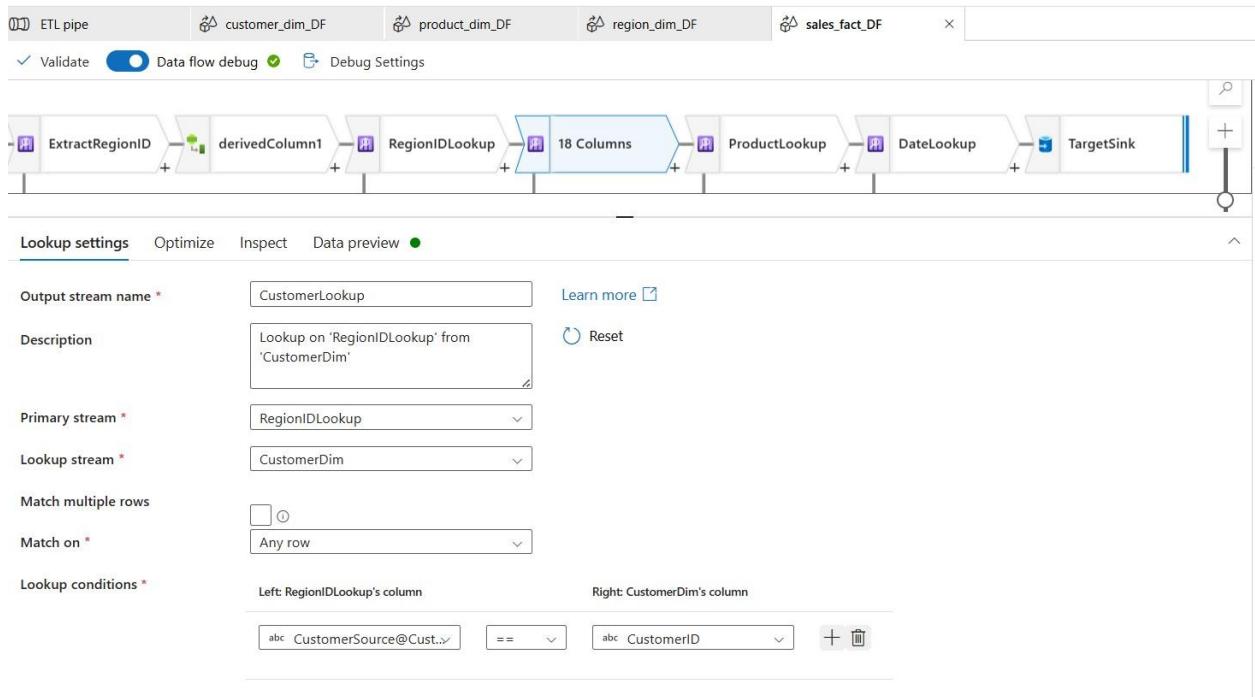
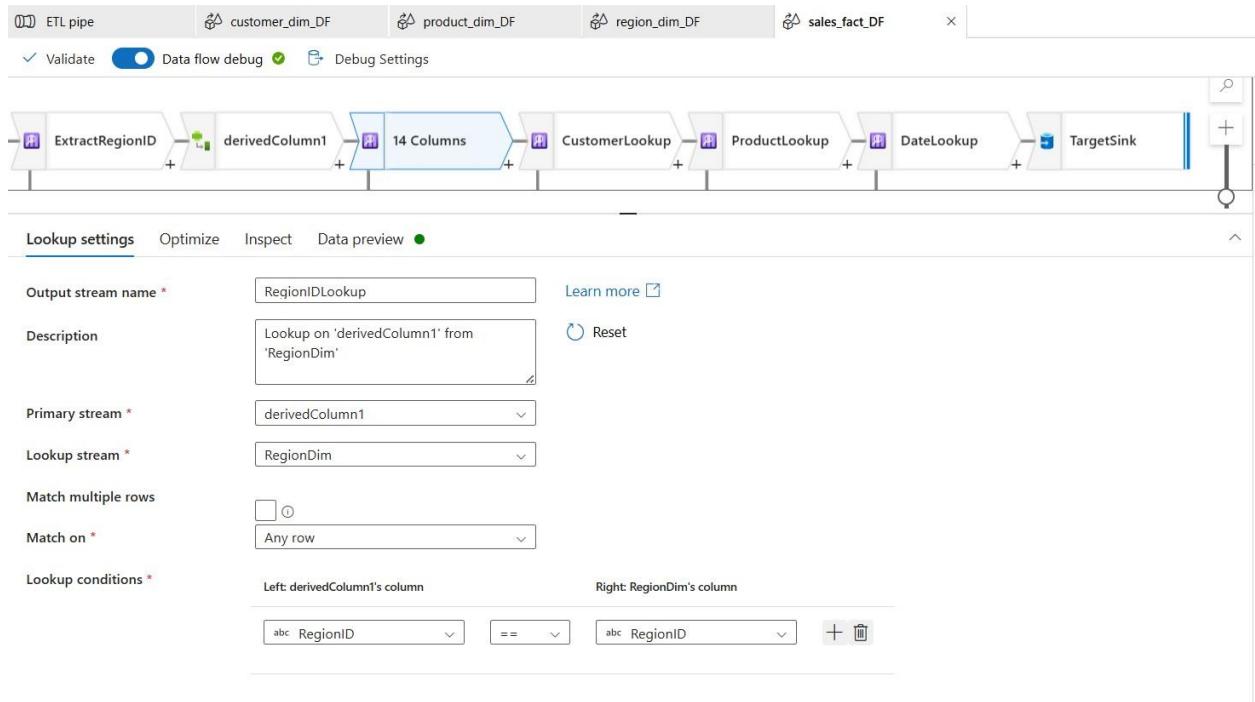
Then added the Union option to join Data from SalesSource and SalesSourceJSON

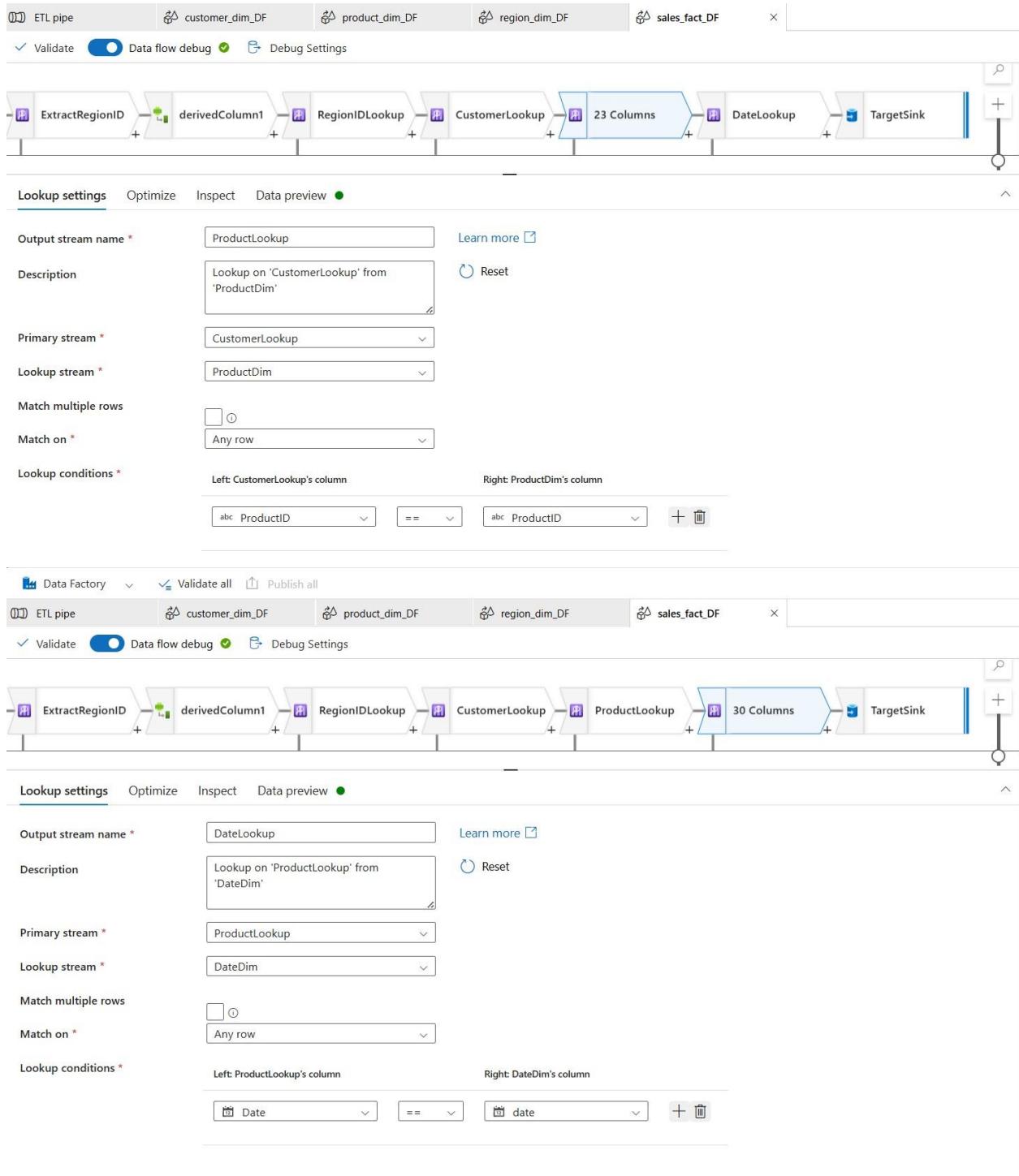


Next added several lookups to add Dimension Keys and derived columns like below. Lookup transformation is used to perform lookups on data from various sources.

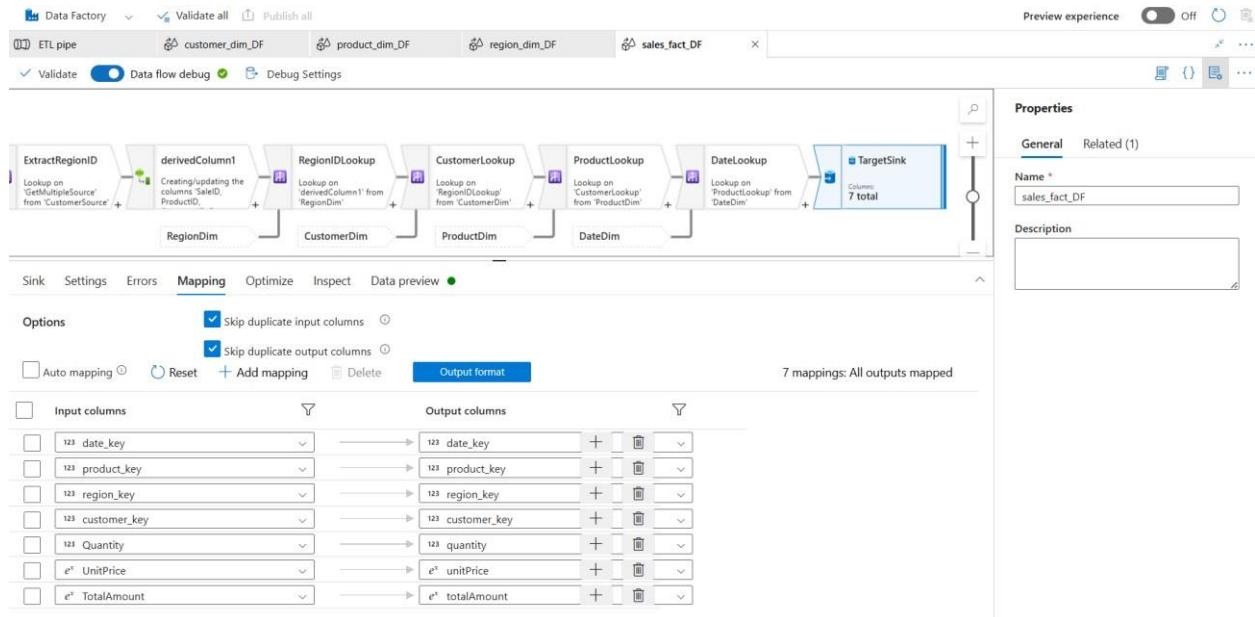
(Addition to lookups added a derived column for a purpose of data type conversion in RegionID)



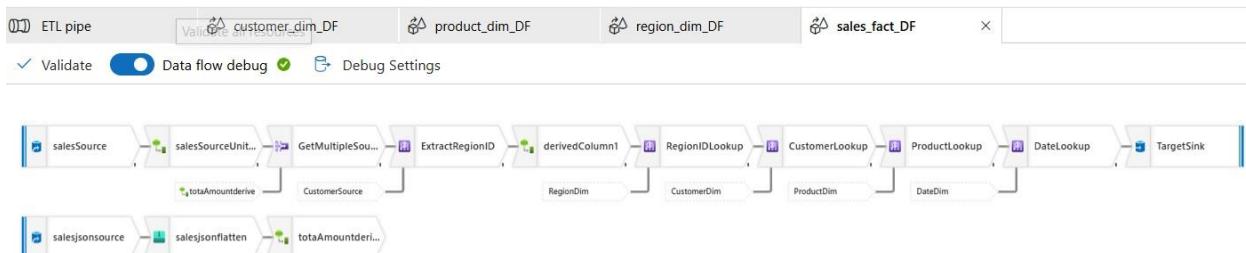




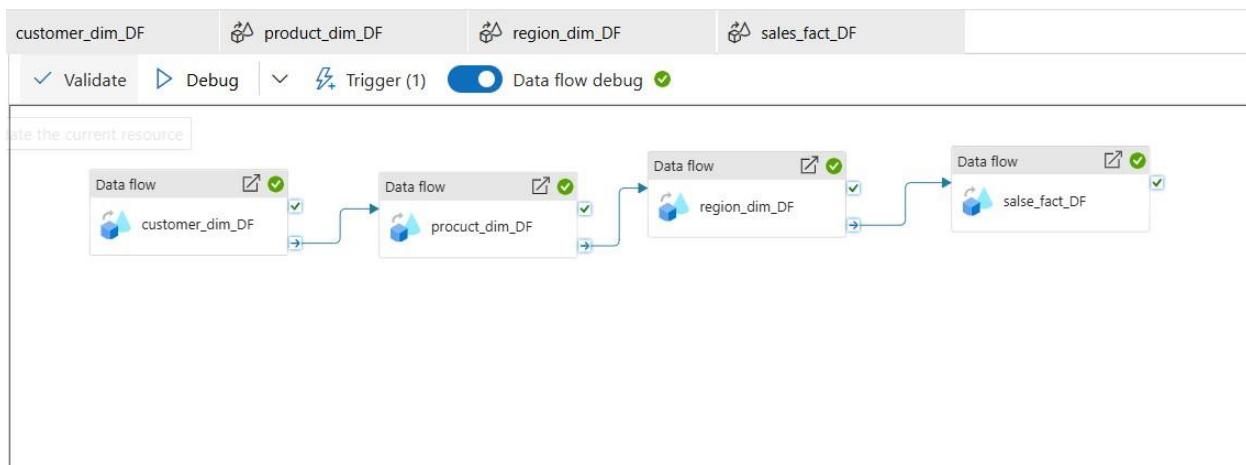
Finally added a Sink below.



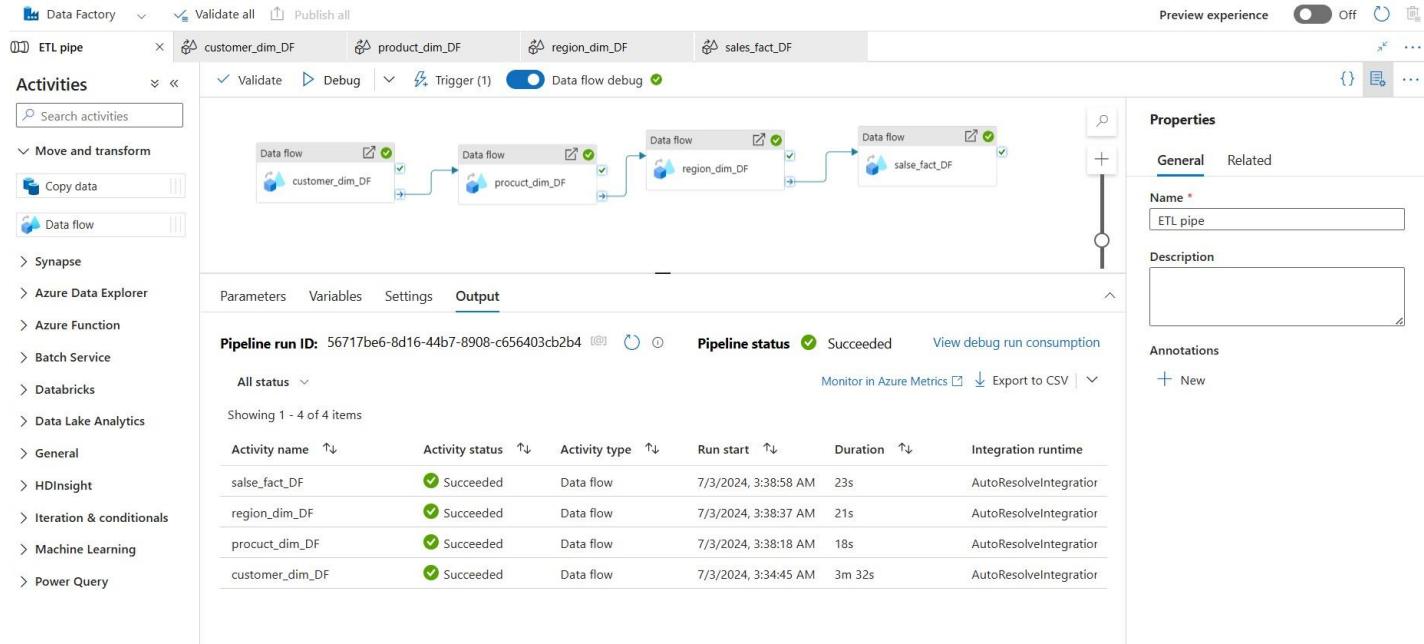
final Sales Data Flow structure



Final Pipeline structure-



After debu & run-



Everything was good.

VI. Date dim-

Assuming analysis needs to be done in the period 2024/1/1 – 2024/5/24, populated date dimension manually using SQL query like below.

targetDW (chiraserver/targetDW) | Query editor (preview)

Azure SQL Database Hyperscale

chirantha

Login New Query Open query Feedback Getting started

targetDW (chirantha)

Showing limited object explorer here. For full capability please click here to open Azure Data Studio.

Tables

- > dbo.customer_dim ...
- > dbo.date_dim ...
 - date_key (PK, int, not null)
 - date (date, null)
 - day (int, null)
 - month (int, null)
 - month_name (varchar, null)
 - year (int, null)
 - quarter (int, null)
- > dbo.Product_dim ...
- > dbo.Region_dim ...
- > dbo.sales_fact ...
- > Views ...
- > Stored Procedures ...

Query 1 × Query 2 ×

Run Cancel query Save query Export data as Show only Editor

```

1 2 DECLARE @startDate DATE = '2024-01-01';
3 3 DECLARE @endDate DATE = '2024-05-24';
4
5 WHILE @startDate <= @endDate
6 BEGIN
7   INSERT INTO date_dim (date, day, month, month_name, year, quarter)
8   VALUES (
9     @startDate,
10    DAY(@startDate),

```

Results Messages

Query succeeded: Affected rows: 145

Query succeeded | 0s

Home > targetDW (chiraserver/targetDW)

targetDW (chiraserver/targetDW) | Query editor (preview)

Azure SQL Database Hyperscale

chirantha

Login New Query Open query Feedback Getting started

targetDW (chirantha)

Showing limited object explorer here. For full capability please click here to open Azure Data Studio.

Tables

- > dbo.customer_dim ...
- > dbo.date_dim ...
 - date_key (PK, int, not null)
 - date (date, null)
 - day (int, null)
 - month (int, null)
 - month_name (varchar, null)
 - year (int, null)
 - quarter (int, null)
- > dbo.Product_dim ...
- > dbo.Region_dim ...
- > dbo.sales_fact ...
- > Views ...
- > Stored Procedures ...

Query 1 × Query 2 ×

Run Cancel query Save query Export data as Show only Editor

```

1 SELECT TOP (1000) * FROM [dbo].[date_dim]

```

Results Messages

Search to filter items...

date_key	date	day	month	month_name	year
1	2024-01-01	1	1	January	2024
2	2024-01-02	2	1	January	2024
3	2024-01-03	3	1	January	2024
4	2024-01-04	4	1	January	2024
5	2024-01-05	5	1	January	2024

Query succeeded | 1s

Lets go warehouse and Run SELECT queries and see output.

Home > targetDW (chiraserver/targetDW)

targetDW (chiraserver/targetDW) | Query editor (preview) ☆ ...

Azure SQL Database Hyperscale

chirantha

Login New Query Open query Feedback Getting started

targetDW (chirantha)

Showing limited object explorer here. For full capability please click here to open Azure Data Studio.

Tables

- > dbo.customer_dim
- > dbo.date_dim
- > dbo.Product_dim
- > dbo.Region_dim
- > dbo.sales_fact
- > dbo.sales_fact_staging

Views

Stored Procedures

Query 1

Run Cancel query Save query Export data as Show only Editor

```
1 select*from customer_dim;
```

Results Messages

Search to filter items...

customer_key	customer_name	contact_no	CustomerID
1	John Doe	1234567890	C100
2	Jane Smith	1234567891	C101
3	Alice Johnson	1234567892	C102
4	Bob Brown	1234567893	C103

Query succeeded | 0s

Home > targetDW (chiraserver/targetDW)

targetDW (chiraserver/targetDW) | Query editor (preview) ☆ ...

Azure SQL Database Hyperscale

chirantha

Login New Query Open query Feedback Getting started

targetDW (chirantha)

Showing limited object explorer here. For full capability please click here to open Azure Data Studio.

Tables

- > dbo.customer_dim
- > dbo.date_dim
- > dbo.Product_dim
- > dbo.Region_dim
- > dbo.sales_fact
- > dbo.sales_fact_staging

Views

Stored Procedures

Query 1

Run Cancel query Save query Export data as Show only Editor

```
1 select*from Product_dim;
```

Results Messages

Search to filter items...

product_key	product_name	price	category	ProductID
1	Widget A	120.00	Electronics	P100
2	Widget B	150.00	Electronics	P101
3	Gadget C	90.00	Home	P102
4	Gadget D	110.00	Home	P103

Query succeeded | 0s

Home > targetDW (chiraserver/targetDW)

targetDW (chiraserver/targetDW) | Query editor (preview) ☆ ...

Azure SQL Database Hyperscale

chirantha

Login New Query Open query Feedback Getting started

targetDW (chirantha)

Showing limited object explorer here. For full capability please click here to open Azure Data Studio.

Tables

- > dbo.customer_dim
- > dbo.date_dim
- > dbo.Product_dim
- > dbo.Region_dim
- > dbo.sales_fact
- > dbo.sales_fact_staging

Views

Stored Procedures

Query 1

Run Cancel query Save query Export data as Show only Editor

```
1 select*from Region_dim;
```

Results Messages

Search to filter items...

region_key	region_name	RegionID
1	North	1
2	South	2
3	East	3

Query succeeded | 0s

Home > targetDW (chiraserver/targetDW)

targetDW (chiraserver/targetDW) | Query editor (preview) ☆ ...

Azure SQL Database Hyperscale

chirantha

Login New Query Open query Feedback Getting started

Query 1

Run Cancel query Save query Export data as Show only Editor

```
1 select*from date_dim;
```

Results Messages

Search to filter items...

date_key	date	day	month	month_name	year	quarter
1	2024-01-01	1	1	January	2024	1
2	2024-01-02	2	1	January	2024	1
3	2024-01-03	3	1	January	2024	1

Query succeeded | 0s

The screenshot shows the Azure SQL Database Hyperscale Query editor interface. At the top, it displays the connection name "targetDW (chiraserver/targetDW)" and the page title "Query editor (preview)". Below the header, there's a toolbar with options like "Run", "Cancel query", "Save query", "Export data as", and "Show only Editor". A query window titled "Query 1" contains the SQL command "select * from sales_fact;". Below the query window, the "Results" tab is selected, showing a table with the following data:

date_key	product_key	region_key	customer_key	quantity	unitPrice	totalAmount
1	1	1	1	2	120.00	240.00
2	2	2	2	1	150.00	150.00
3	3	3	3	3	90.00	270.00

A message at the bottom of the results pane says "Query succeeded | 0s".

VII. trigger-

New trigger

Name *

Description

Type *

Start date * ⓘ

Time zone * ⓘ

Recurrence *

Every Minute(s)

Specify an end date

End On * ⓘ

Annotations

+ New

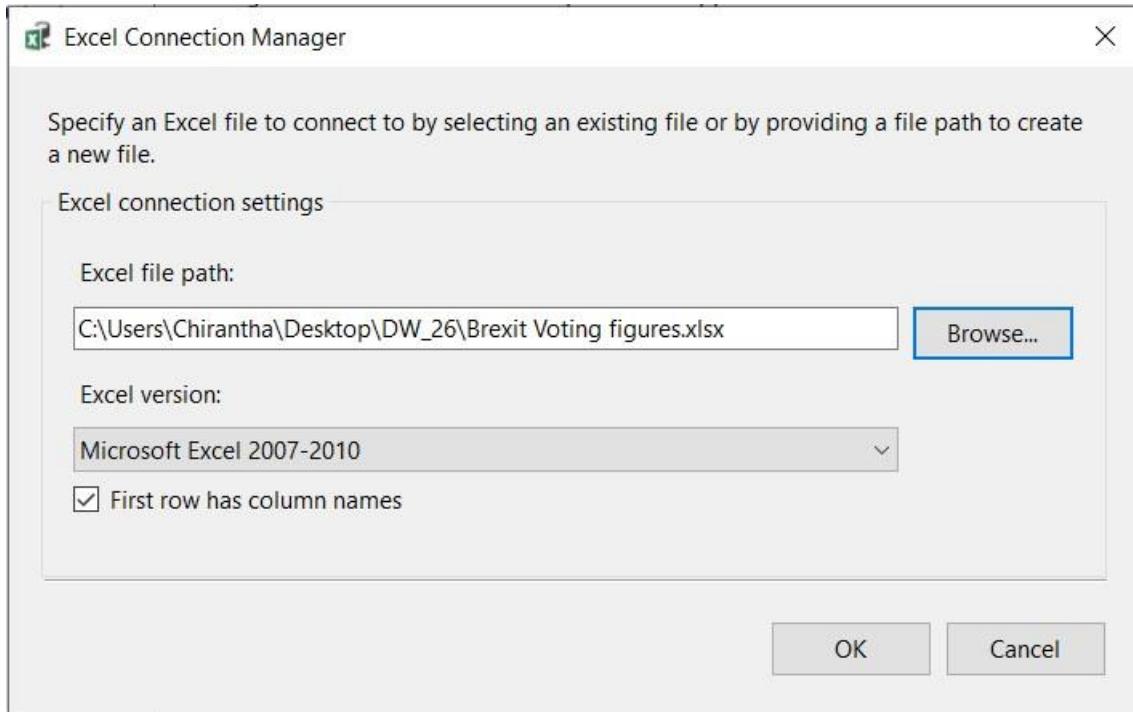
Start trigger ⓘ

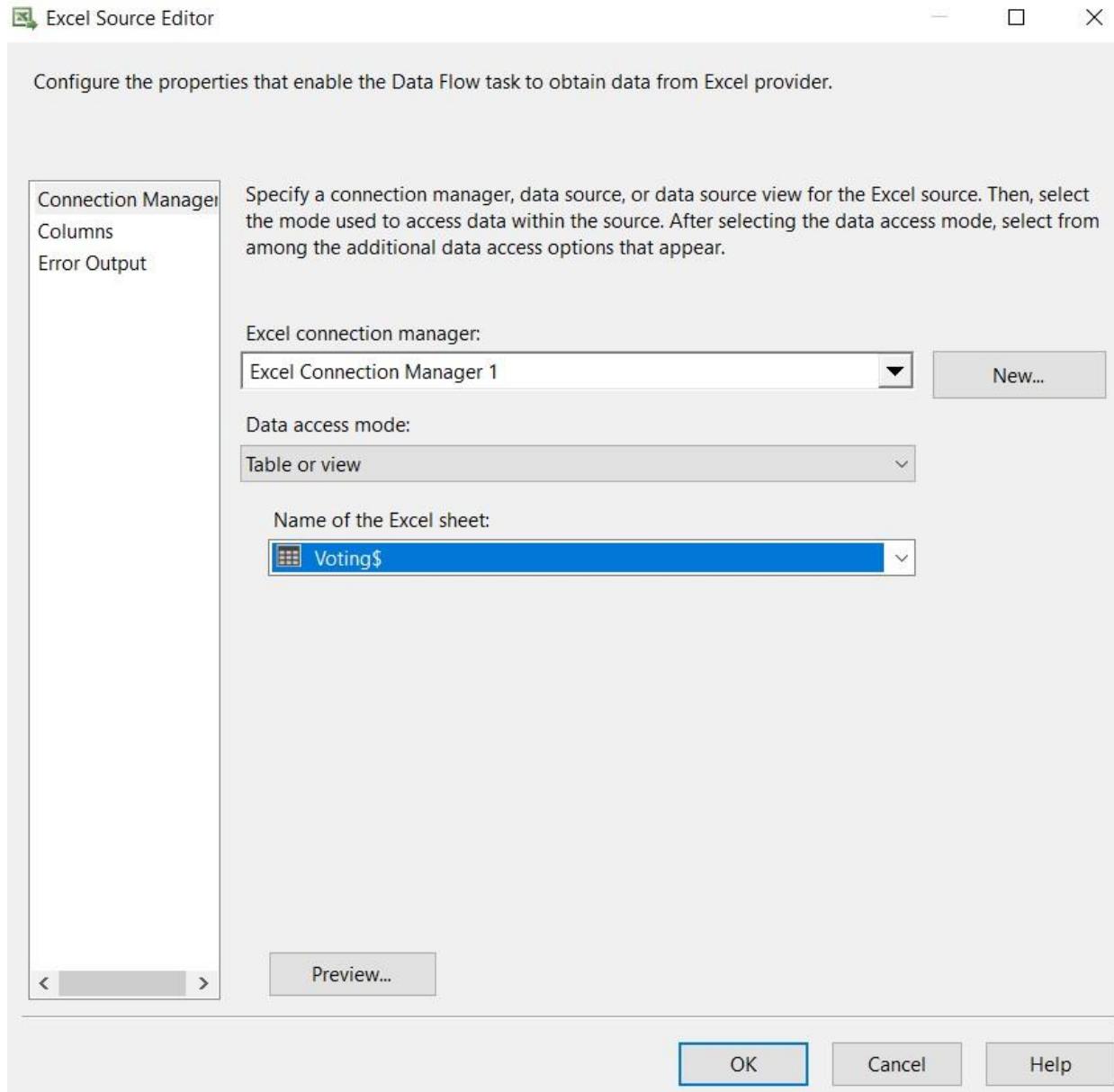
SECTION C: SSIS

(IX)

Step 1: Create a New SSIS Project and Add Data Flow Task

Step 2: Add Excel Source





Then checked the columns.

Excel Source Editor

Configure the properties that enable the Data Flow task to obtain data from Excel provider.

Connection Manager

Columns

Error Output

Available External ...	
<input checked="" type="checkbox"/>	Name
<input checked="" type="checkbox"/>	Region
<input checked="" type="checkbox"/>	Area
<input checked="" type="checkbox"/>	Electorate
<input checked="" type="checkbox"/>	ValidVotes
<input checked="" type="checkbox"/>	Remain

External Column	Output Column
Region	Region
Area	Area
Electorate	Electorate
ValidVotes	ValidVotes
Remain	Remain
Leave	Leave

< >

OK

Cancel

Help

Step 3: Add Derived Column Transformation

Derived Column Transformation Editor

Specify the expressions used to create new column values, and indicate whether the values update existing columns or populate new columns.

Variables and Parameters Columns

Mathematical Functions
String Functions
Date/Time Functions
NULL Functions
Type Casts
Operators

Description:

Derived Column Name	Derived Column	Expression	Data Type
RemainPercentage%	<add as new column>	(Remain * 100.0 / ValidVotes)	double-precision flo...
LeavePercentage%	<add as new column>	(Leave * 100.0 / ValidVotes)	double-precision flo...

Configure Error Output... OK Cancel Help

Step 4: Add Conditional Split Transformation

Conditional Split Transformation Editor

Specify the conditions used to direct input rows to specific outputs. If an input row matches no condition, the row is directed to a default output.

Order	Output Name	Condition
1	Case 1	Electorate >= ValidVotes

Variables and Parameters

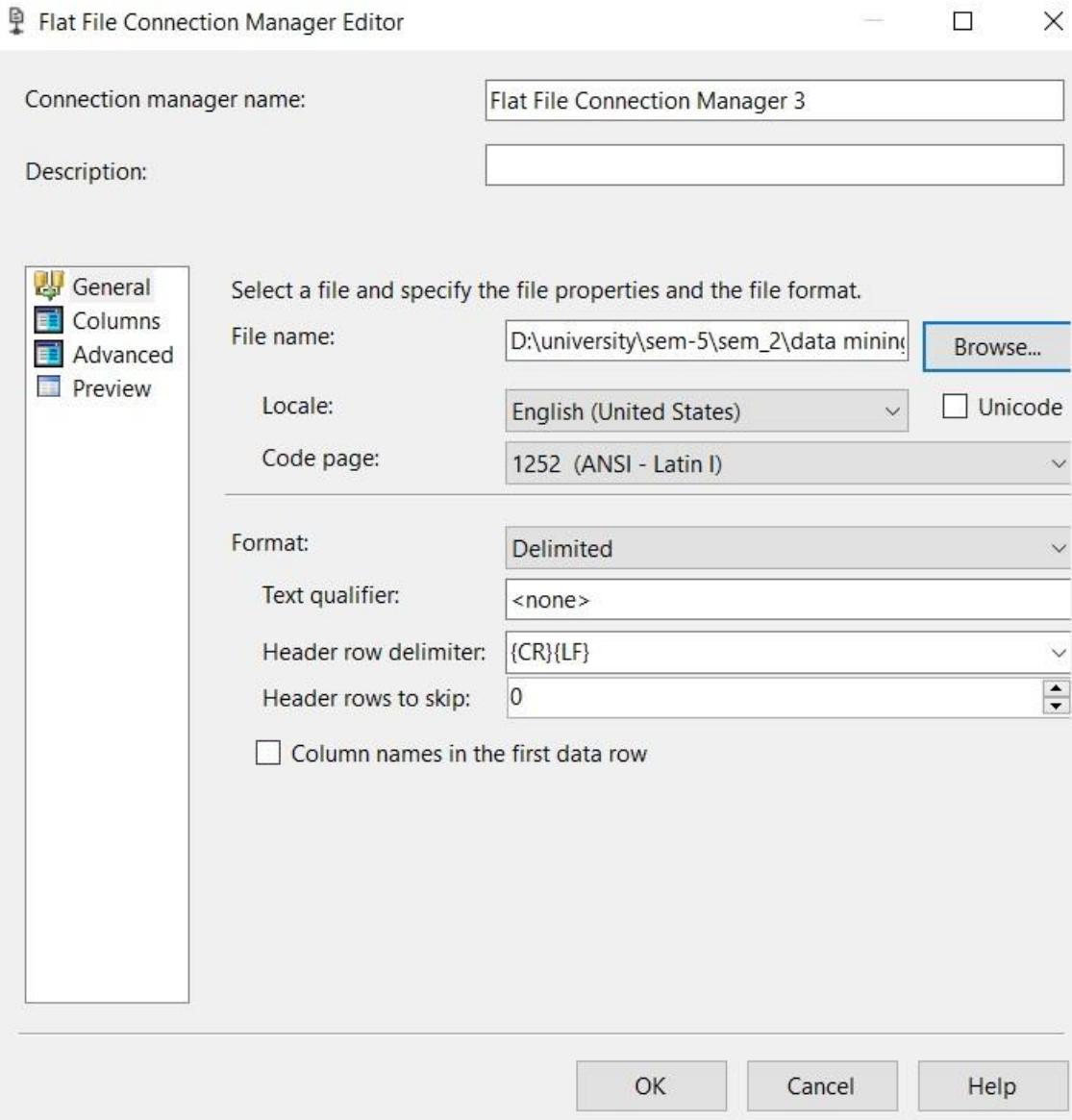
- Mathematical Functions
- String Functions
- Date/Time Functions
- NULL Functions
- Type Casts
- Operators

Description:

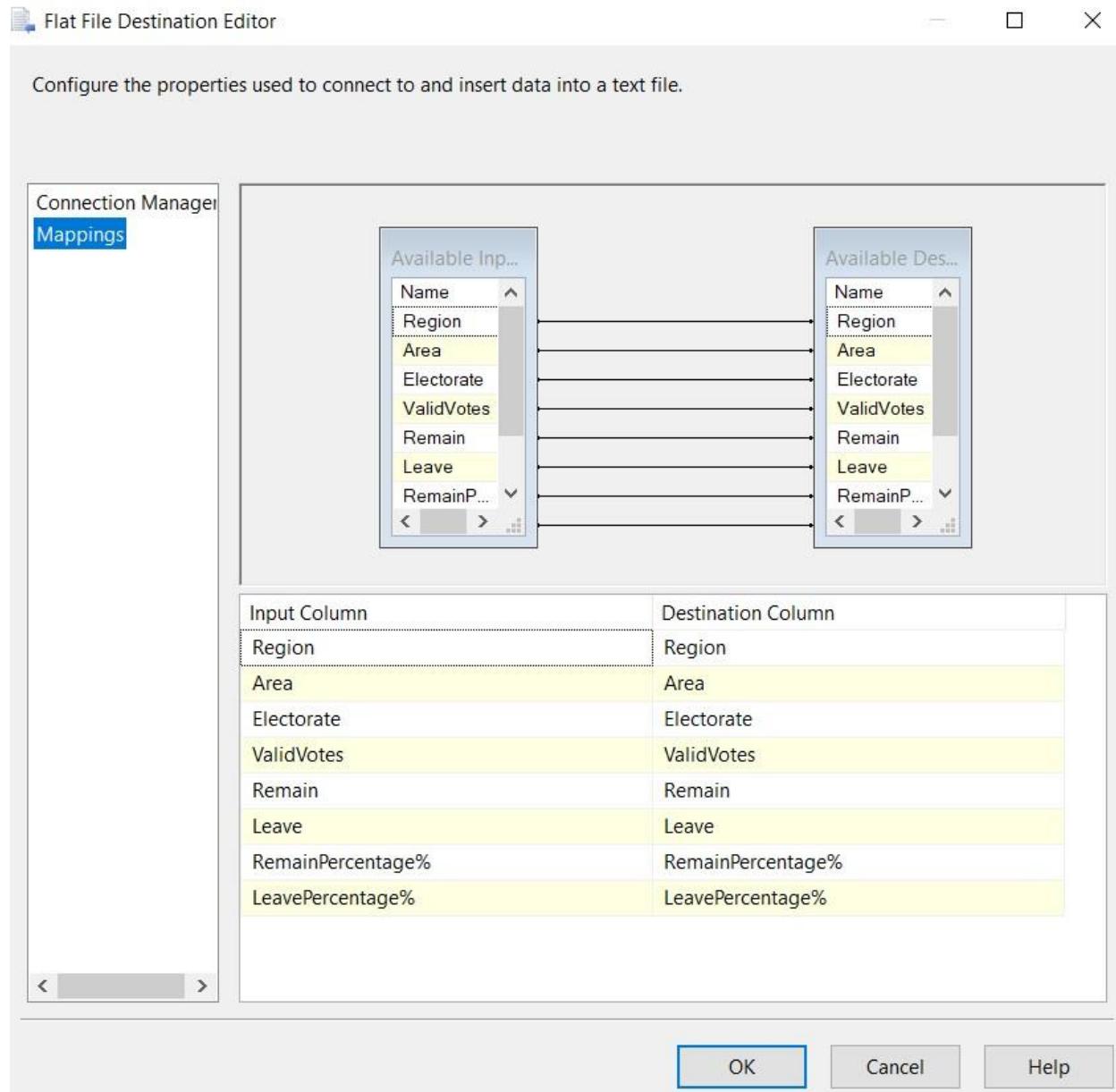
Default output name: Conditional Split Default Output

Configure Error Output... OK Cancel Help

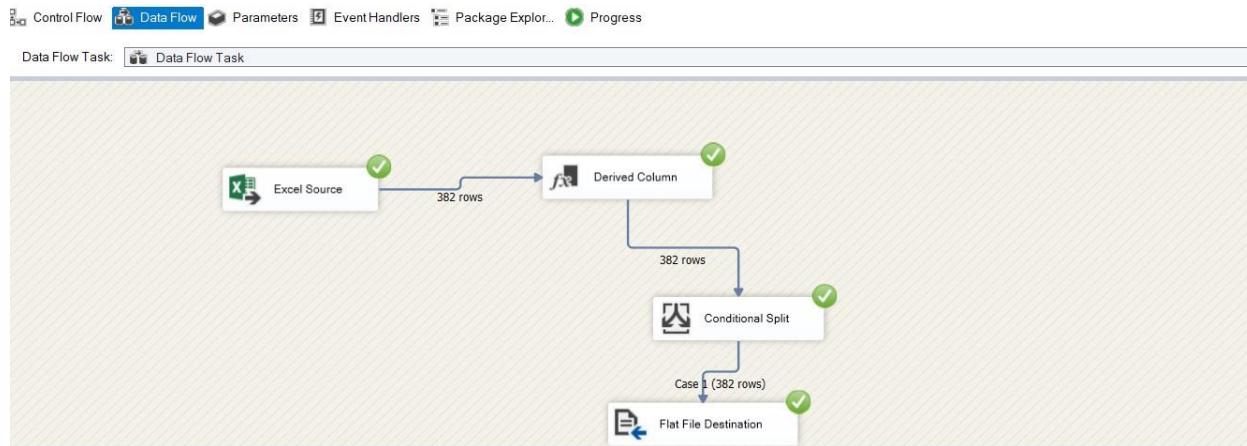
Step 5: Add Flat File Destination



Did mappings.



Step 6 : Execute the Package



Output-

percentage.txt - Notepad

```

File Edit Format View Help
Region,Area,Electorate,ValidVotes,Remain,Leave,RemainPercentage%,LeavePercentage%
East,Peterborough,120892,87392,34176,53216,39.106554375686564,60.893445624313436
East,Luton,127612,84481,36708,47773,43.451190208449241,56.548809791550759
East,Southend-on-Sea,128856,93870,39348,54522,41.917545541706616,58.082454458293384
East,Thurrock,109897,79916,22151,57765,27.717853746433754,72.282146253566239
East,Bedford,119530,86066,41497,44569,48.21532312411405,51.78467687588595
East,Central Bedfordshire,204004,158804,69670,89134,43.871690889398252,56.128309110601748
East,Cambridge,80108,57799,42682,15117,73.845568262426681,26.154431737573315
East,East Cambridgeshire,62435,48086,23599,24487,49.076654327662936,50.923345672337064
East,Fenland,71447,52626,15055,37571,28.607532398434234,71.39246760156577
East,Huntingdonshire,128486,99927,45729,54198,45.762406556786452,54.237593443213548
East,South Cambridgeshire,114830,93189,56128,37061,60.230284690253143,39.769715309746857
East,Basildon,132771,97999,30748,67251,31.375830365615975,68.624169634384018
East,Braintree,112562,86236,33523,52713,38.873556287397378,61.126443712602622
East,Brentwood,58777,46704,19077,27627,40.846608427543678,59.153391572456322
East,Castle Point,68860,51845,14154,37691,27.300607580287394,72.699392419712609
East,Chelmsford,129971,100794,47545,53249,47.170466496021589,52.829533503978411
East,Colchester,127520,95719,44414,51305,46.40040117427052,53.59959882572948
East,Epping Forest,100016,76852,28676,48176,37.313277468380782,62.686722531619218
East,Harlow,59124,43469,13867,29602,31.900894890611699,68.099105109388304
East,Maldon,49073,38831,14529,24302,37.415982076176249,62.584017923823751
East,Rochford,66589,52447,17510,34937,33.386085000095335,66.613914999904665
East,Tendring,111167,82657,25210,57447,30.499534219727305,69.500465780272691
East,Uttesford,64735,51943,25619,26324,49.321371503378707,50.678628496621293
East,Broxbourne,68997,50872,17166,33706,33.743513130995439,66.256486869004561
East,Dacorum,108965,86244,42542,43702,49.32748944854135,50.67251055145865
East,Hertsmere,73295,56125,27593,28532,49.163474387527842,50.836525612472158
East,North Hertfordshire,99316,77672,42234,35438,54.374806880214237,45.625193119785763
East,Three Rivers,67380,52848,25751,27097,48.726536481986074,51.273463518013926
East,Watford,65060,46586,23167,23419,49.72953247756837,50.27046752243163
  
```

(X)

Step 1: Create a New SSIS Project and Add Data Flow Task

Step 2: Add Excel Source- We can follow in (IX) step: 2 steps here also.

Step 3: Add Aggregate Transformation

Σ 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.

Advanced

Available Input Col...

Name
(*)
<input checked="" type="checkbox"/> Region
<input type="checkbox"/> Area
<input checked="" type="checkbox"/> Electorate
<input type="checkbox"/> ValidVotes
<input checked="" type="checkbox"/> Remain

Input Column	Output Alias	Operation	Compa
Region	Region	Group by	
Remain	Remain	Sum	
Leave	Leave	Sum	
Electorate	Electorate	Sum	

< >

OK Cancel Help

Input Column	Output Alias	Operation	Compa
Region	Region	Group by	
Remain	Remain	Sum	
Leave	Leave	Sum	
Electorate	Electorate	Sum	

Step 5: Add Derived Column Transformation

Derived Column Transformation Editor

Specify the expressions used to create new column values, and indicate whether the values update existing columns or populate new columns.

Variables and Parameters

Columns

Mathematical Functions
String Functions
Date/Time Functions
NULL Functions
Type Casts
Operators

Description:

Derived Column Name	Derived Column	Expression	Data Type	Length
WinningSide	<add as new column>	(Remain >= Leave ? "Remain" : "Leave")	Unicode string [DT_WSTR]	6

Step 6: Add Flat File Destination

The screenshot shows the 'Derived Column Transformation Editor' window. On the left, there's a navigation pane with 'Variables and Parameters' selected. Below it is a tree view with 'Columns' expanded, showing categories like Mathematical Functions, String Functions, etc. A 'Description:' text box is present. At the bottom, a table lists a single derived column named 'WinningSide' with the expression '(Remain >= Leave ? "Remain" : "Leave")'. The table has columns for 'Derived Column Name', 'Derived Column', 'Expression', 'Data Type', and 'Length'. The 'OK' button at the bottom is highlighted.

Flat File Connection Manager Editor

Connection manager name:

Description:

General

Select a file and specify the file properties and the file format.

File name:

Locale:

Code page:

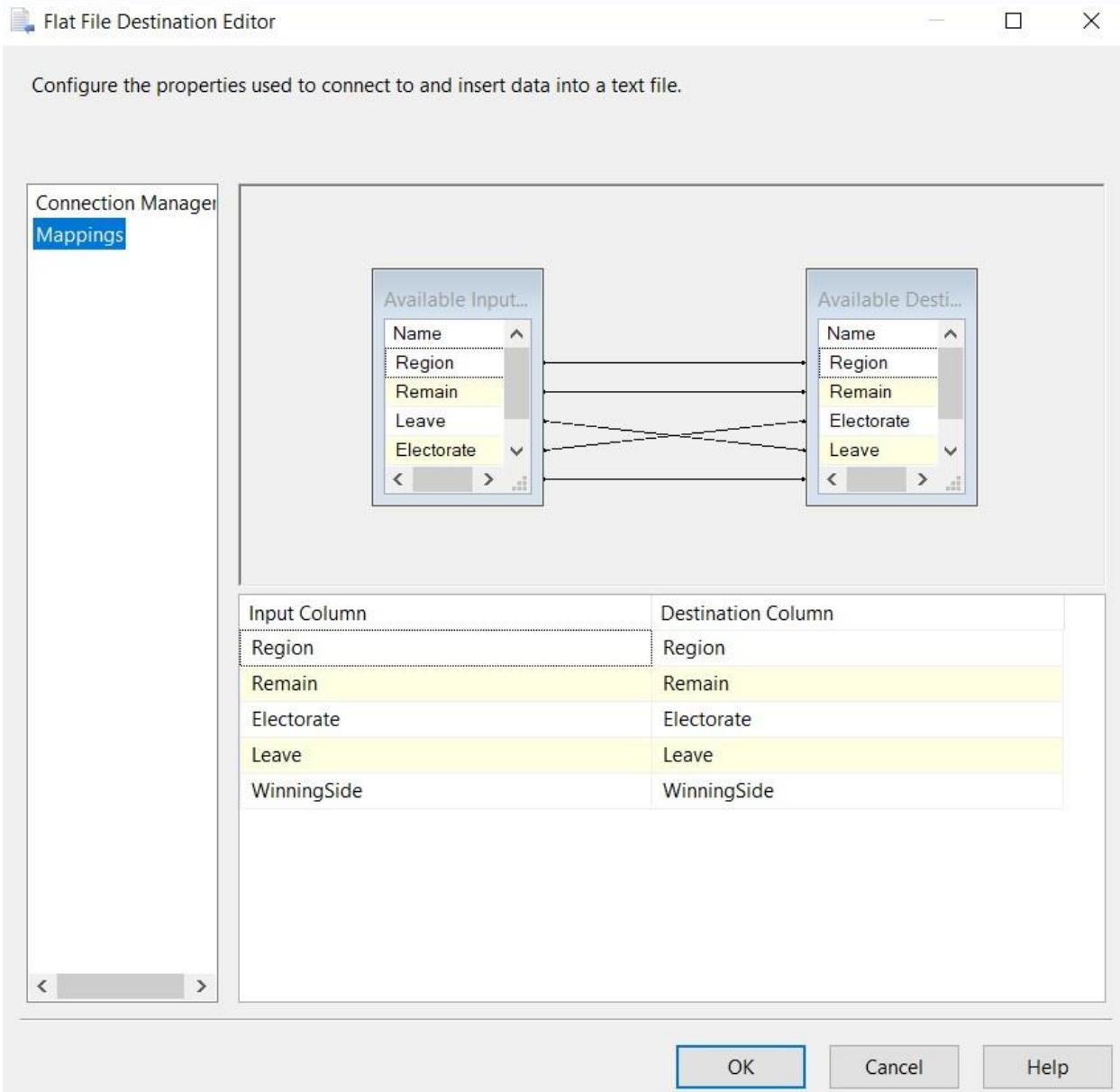
Format:

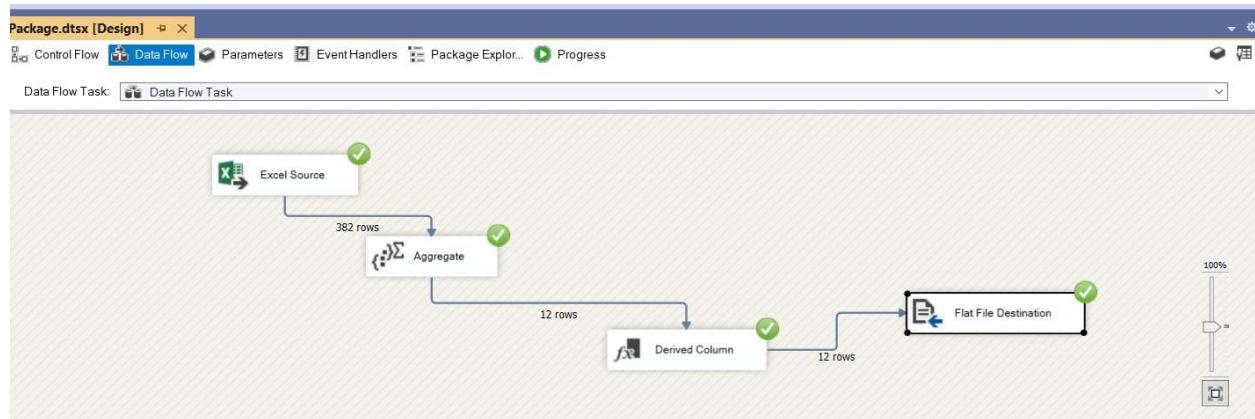
Text qualifier:

Header row delimiter:

Header rows to skip:

Column names in the first data row





Output-

```
winner.txt - Notepad
File Edit Format View Help
Region,Remain,Leave,Electorate,WinningSide
Scotland,1661191,1018322,3987112,Remain
Northern Ireland,440707,349442,1260955,Remain
Yorkshire and The Humber,1158298,1580937,3877780,Leave
South East,2391718,2567965,6465404,Leave
London,2263519,1513232,5424768,Remain
Wales,772347,854572,2270272,Leave
North West,1699020,1966925,5241568,Leave
West Midlands,1207175,1755687,4116572,Leave
South West,1503019,1669711,4138134,Leave
East Midlands,1033036,1475479,3384299,Leave
East,1448616,1880367,4398796,Leave
North East,562595,778103,1934341,Leave
```