Faculty of Computers and artificial intelligenceCairo University
2024 – Second Semester

Project

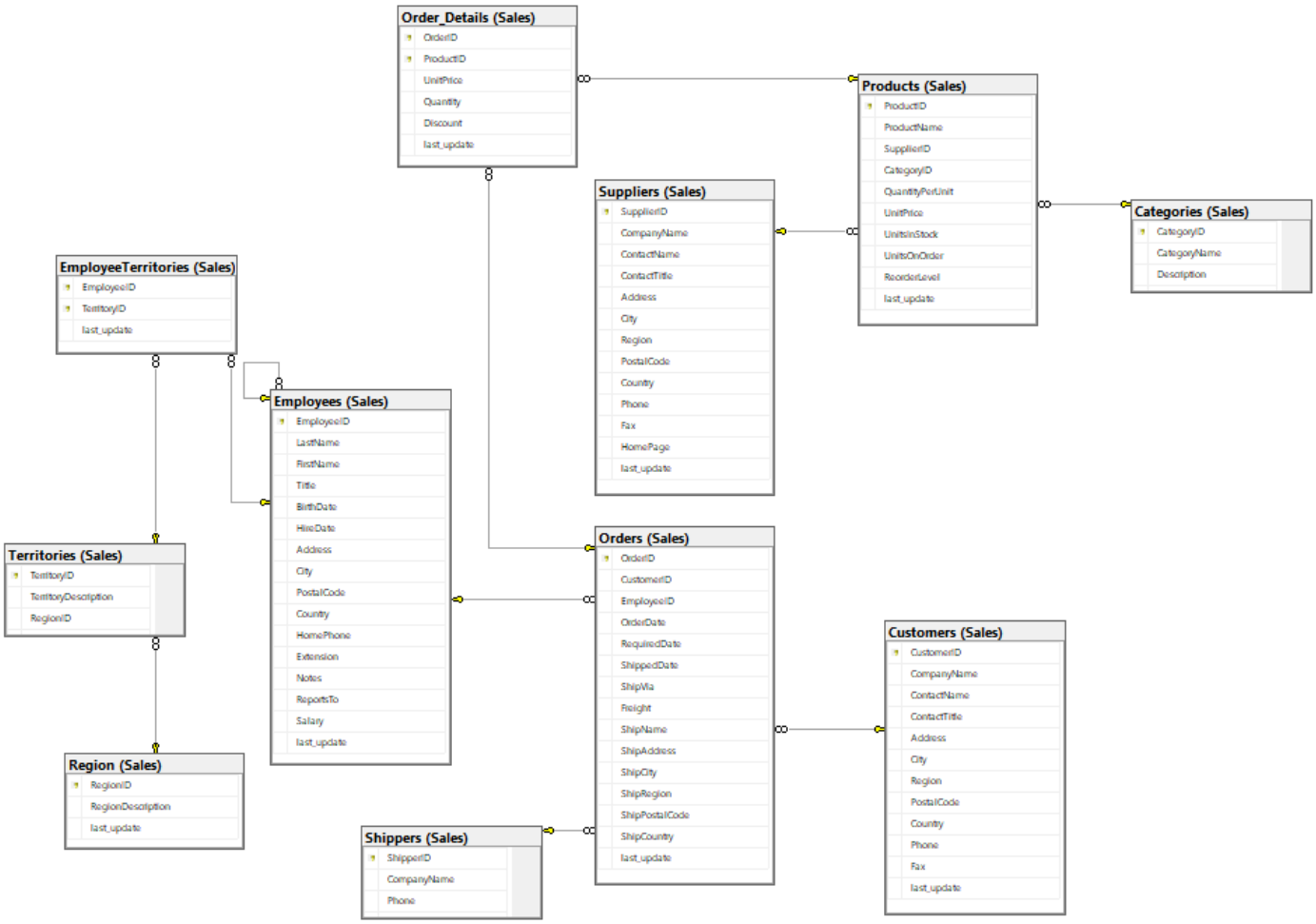
Data warehouse

Under supervision:

Eng.Ahmed Mohamed Galal

Name	ID
Ahmed Saad	20210020
Seif Ibrahim	20210164
Maryam Osama	20211090
Shahd Mostafa	20211054
Shahd Osama	20211052

2. physical Model



3.a. Business Process for the model: To ensure efficient product sales and customer satisfaction, while effectively managing employee information, tracking employee performance, and optimizing product sales through the management of product information.

KPIs

products sold for product by Customer in time period

Product

Customer

Time (day)

Product sold (additive)

Total Sales Revenue Generated by Each Employee

time (month)

Employee

Total Revenue (sum of the total sales amount (unitPrice -(Unitprice*Discount))*Quantity for each order) (additive)

Total number of orders shipped by Each shipper

time (day)

shipper

number of order(additive)

Number of employees worked in each region in period of time

Time(year)

Region

Number of employees

---

6 Dim:

Time

Customer

Employee

Shipper

Region

Product

4 fact completed : one for each KPI

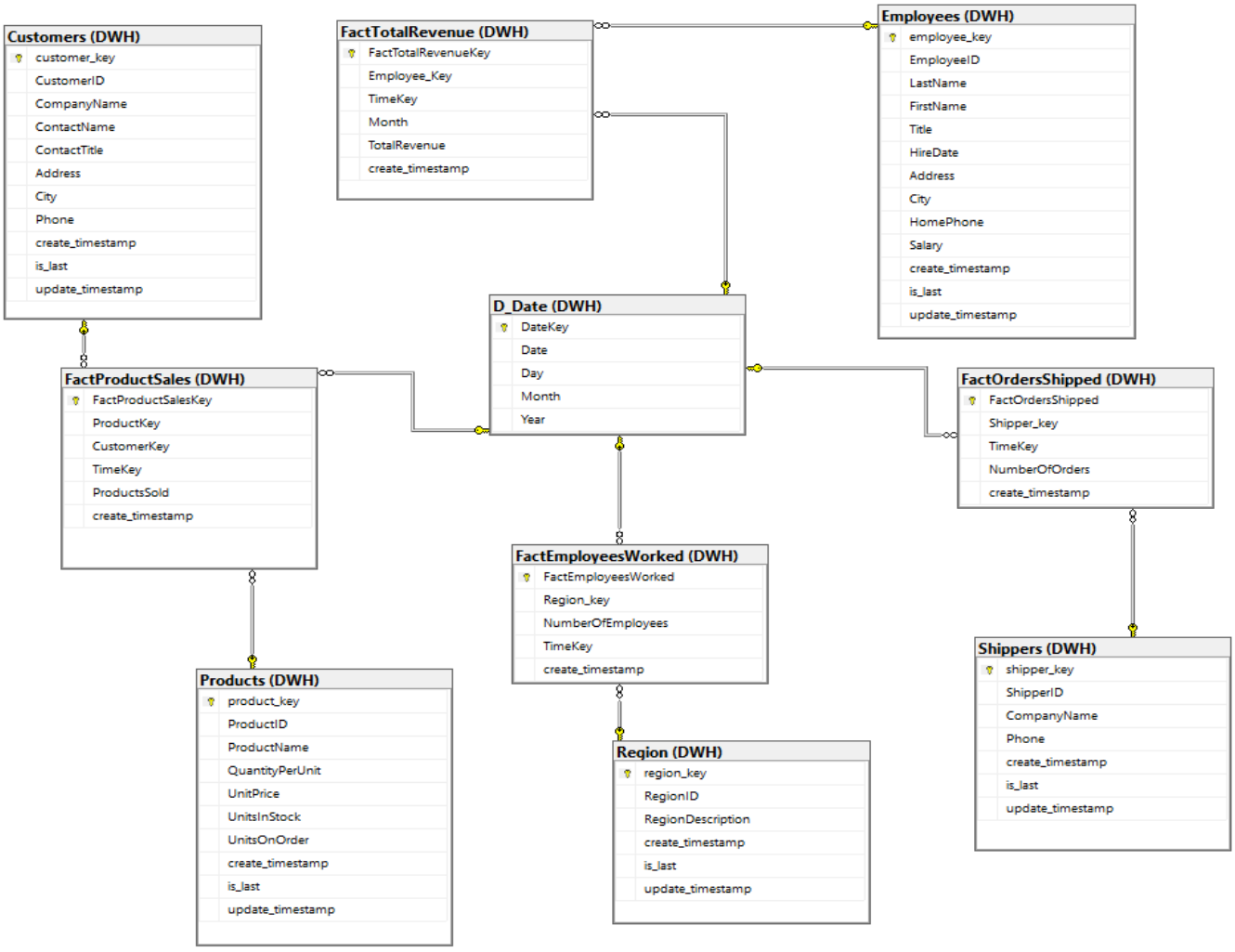
3.b.c.e

Fact Table	Representation	Type	Measures
Products_Fact	This fact table records the number of products bought by each customer on specific days. It contains data about products number, date, and other specific attributes with each product	Transaction Fact Table	Total sold products per customer at specific day (**additive**)
TotalRevenue_Fact	This fact table records the total sales revenue generated by each employee at specific months. It includes the date of sales, moth, and the total revenue that generated by each employee. The Total Revenue is calculated as the sum of (UnitPrice * Quantity) minus the discount for each order.	Transaction Fact Table	Total Revenue per Employee at specific day (**additive**)
OrdersShipped_Fact	This fact table records the total number of orders shipped by each shipper on specific days. It includes the date of shipment, the shipper responsible for shipping the orders, and the total number of orders shipped by that shipper on that particular day	Transaction Fact Table	Total number of orders per shipper at specific day (**additive**)
Employees_Fact	This fact table records the number of employees who worked in each region during specific years. It includes the year, the region, and the total number of employees who worked in that region during the specified year.	Transaction Fact Table	Total number of worked employees per region at specific year(**additive**)

3.d

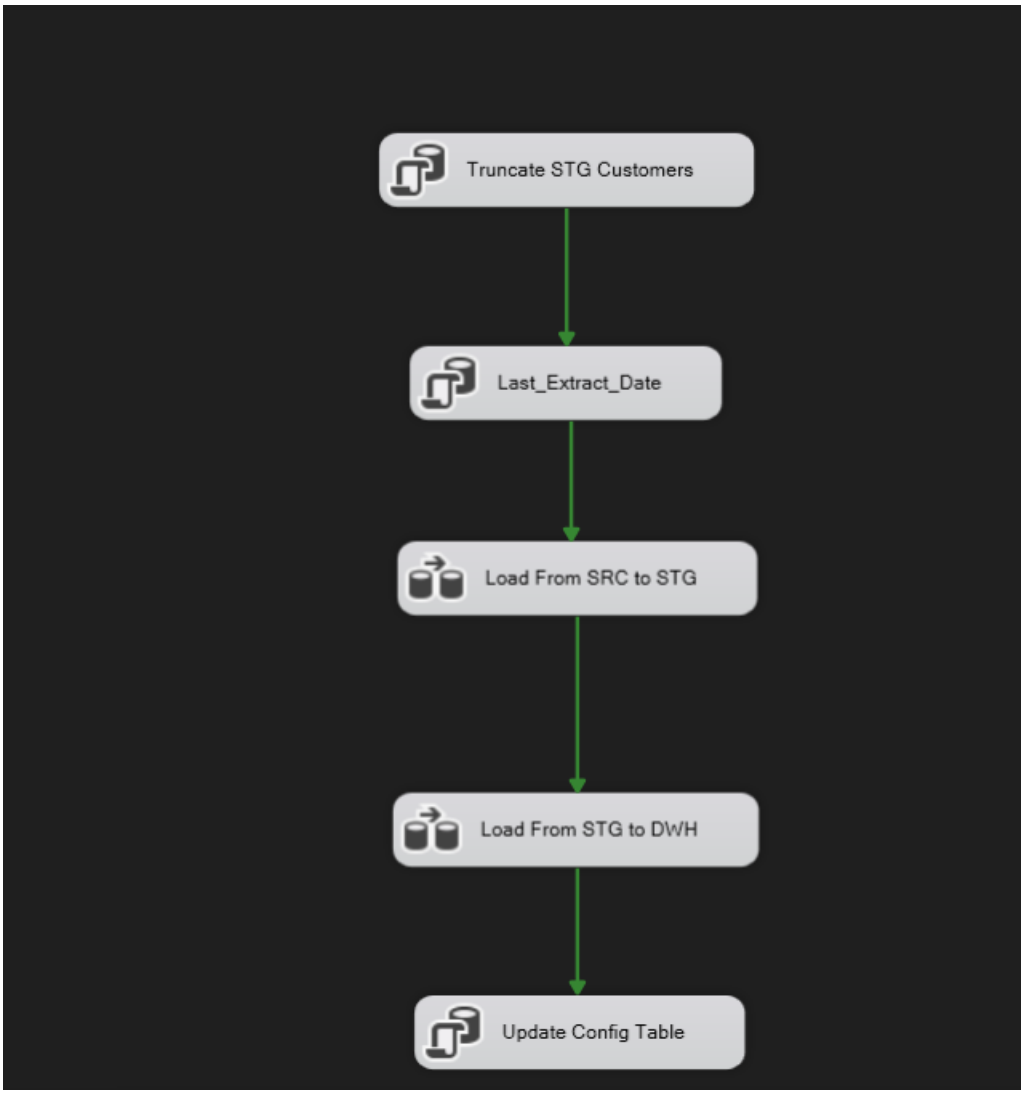
Dimension Table	Representation	Type
Time_Dim	This dimension is used to track various time-related information, such as date, day, month, quarter, year, etc. It can be used to filter and analyze data based on different time periods.	**Conformed** Dim, **Static** Dim,
Customers_Dim	This dimension table provides detailed information about each customer, including their contact information, address, and timestamps for record creation and last update. It can be used to filter and analyze data based on different customer attributes, enabling detailed customer analysis and reporting	**SCD2**
Employees_Dim	This dimension table provides detailed information about each employee, including their contact information (Name, Address, etc.), employment details (Hire date, salary, etc.), and timestamps for record creation and last update. It can be used to filter and analyze data based on different employee attributes, enabling detailed employee analysis and reporting	**SCD2**
Shippers_Dim	This dimension table provides detailed information about each shipper, including their contact information (Company name, phone) and timestamps for record creation and last update. It can be used to filter and analyze data based on different shipper attributes, enabling detailed shipper analysis and reporting	**SCD2**
Region_Dim	This dimension table provides information about different regions and their identifiers. It can be used to filter and analyze data based on different region attributes, enabling detailed region analysis and reporting	**SCD2**
Products_Dim	This dimension table provides detailed information about each product, including its name, pricing, Quantity, and stock levels. It can be used to filter and analyze data based on different product attributes, enabling detailed product analysis and reporting.	**SCD2**

3.f Galaxy Schema:

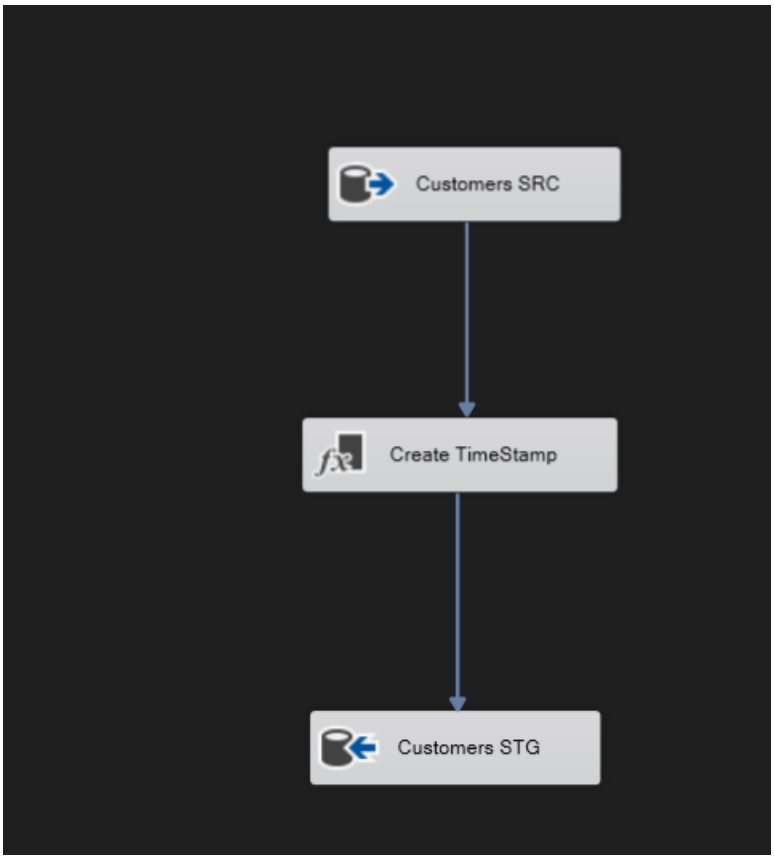


4. 1 Product Sold Fact Table:

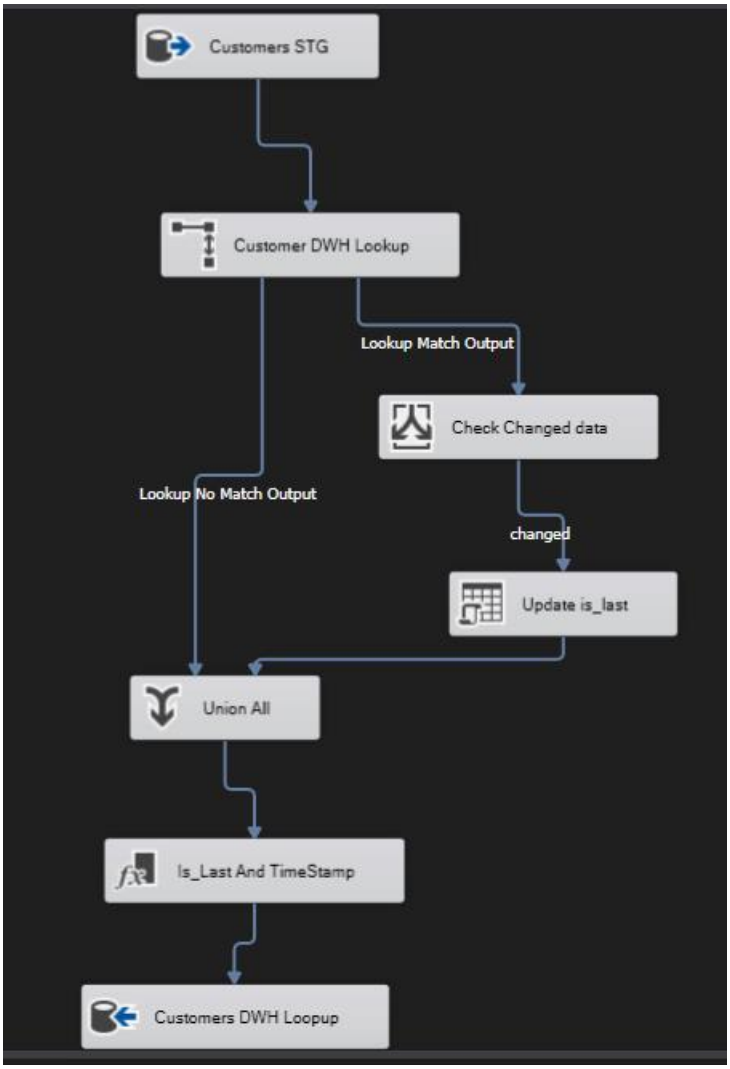
- **Customer Dim Control Flow**



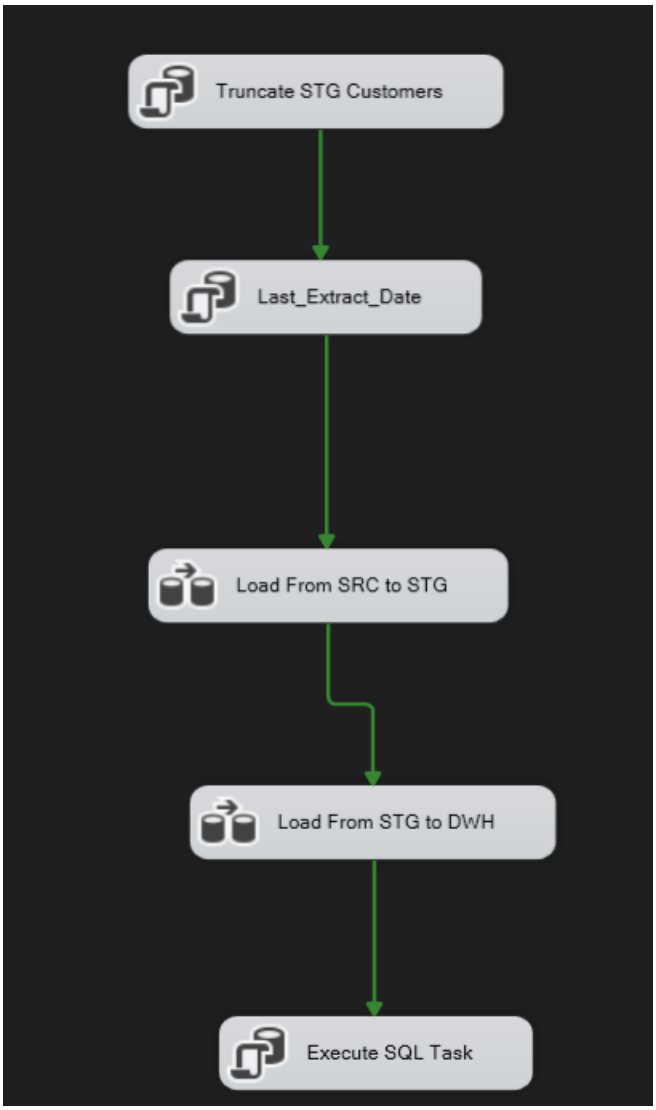
- **Data Flow from SRC to STG Customers**



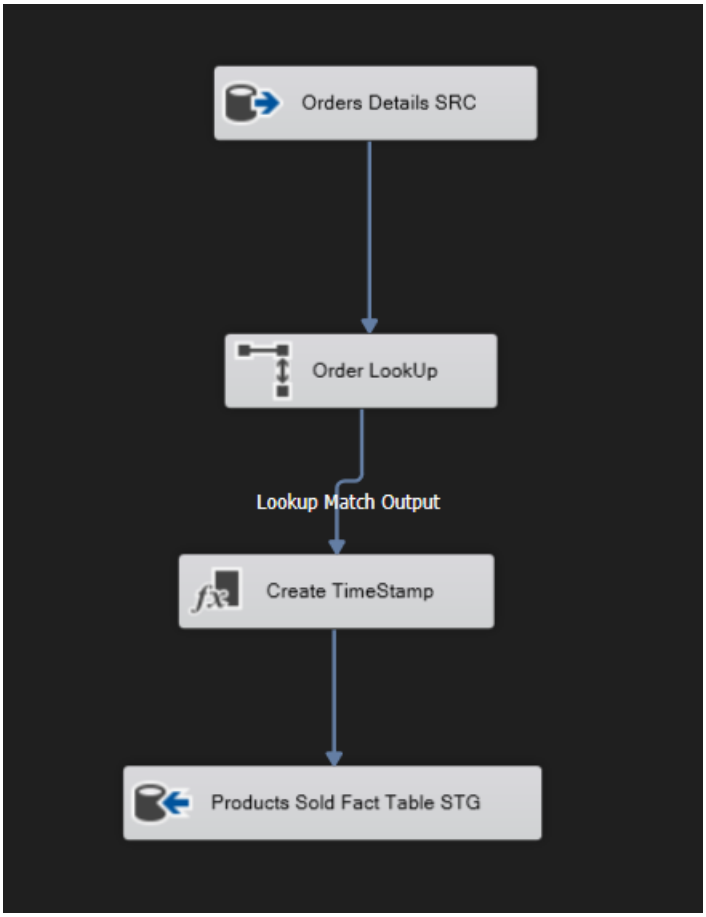
- **Data Flow from STG to DWH Customers**



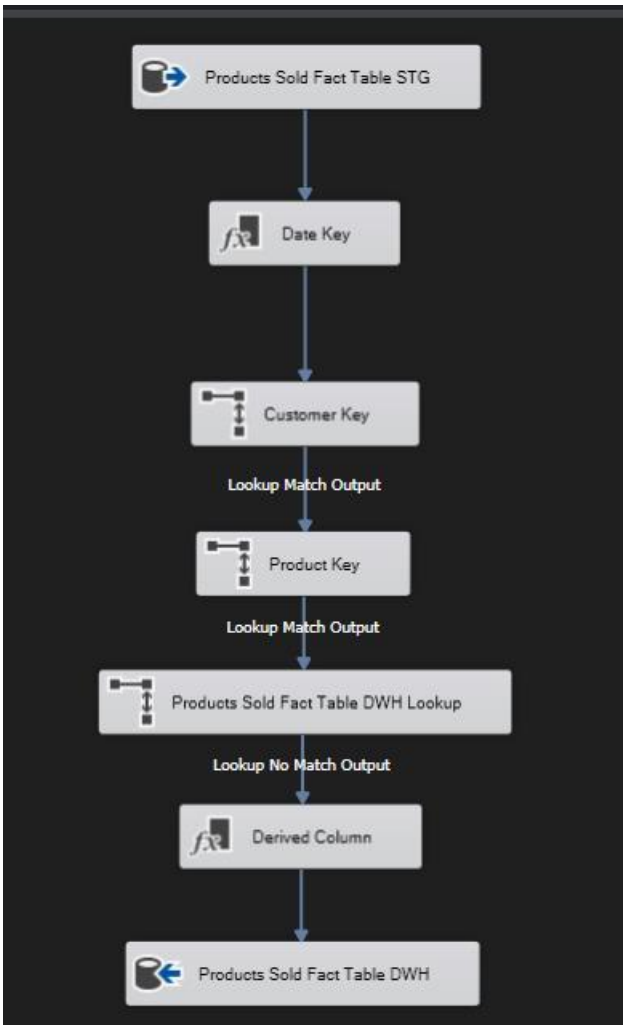
- **Control Flow of Product Sold Fact**



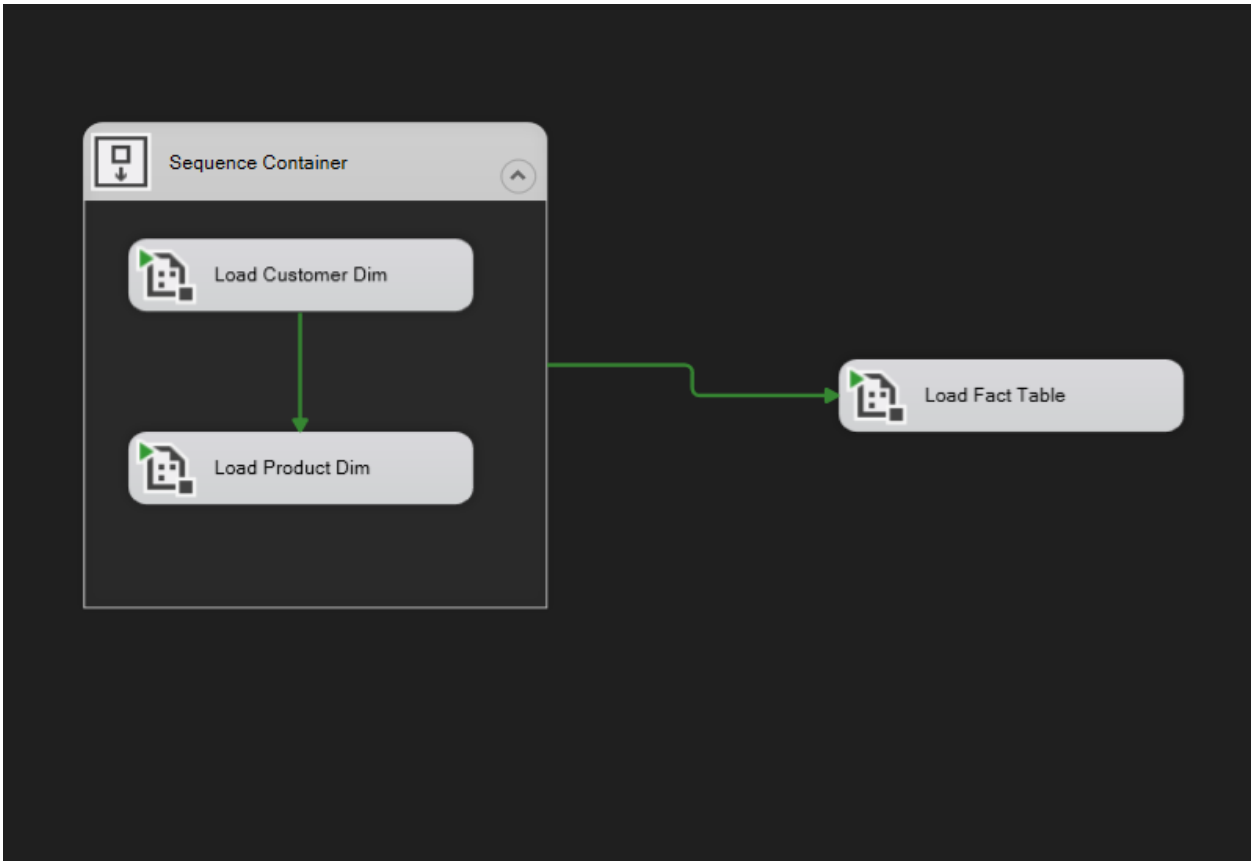
- **Data flow from src to stg product sold fact**



- **Data flow from STG to DWH**

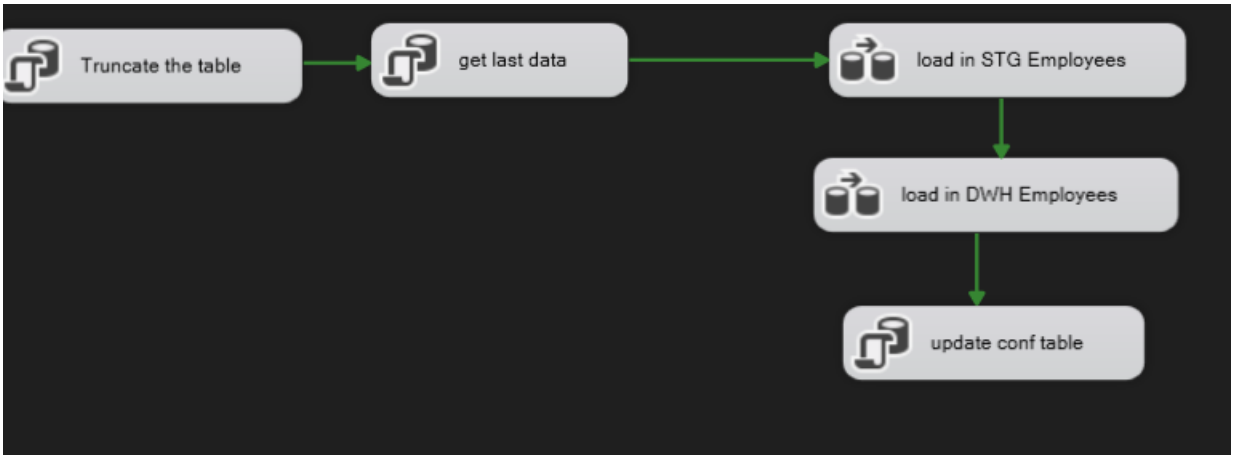


- **Batch to Run the DWH**

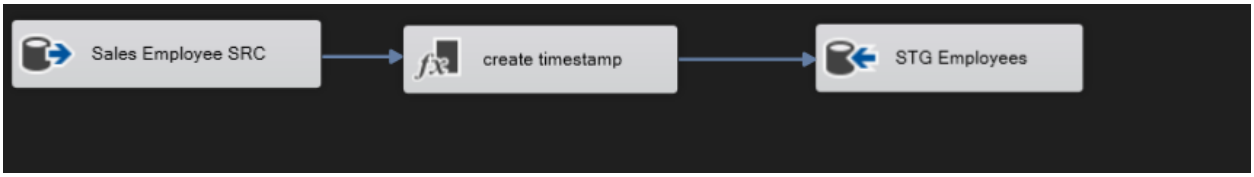


2- Emp Revenue Fact Table

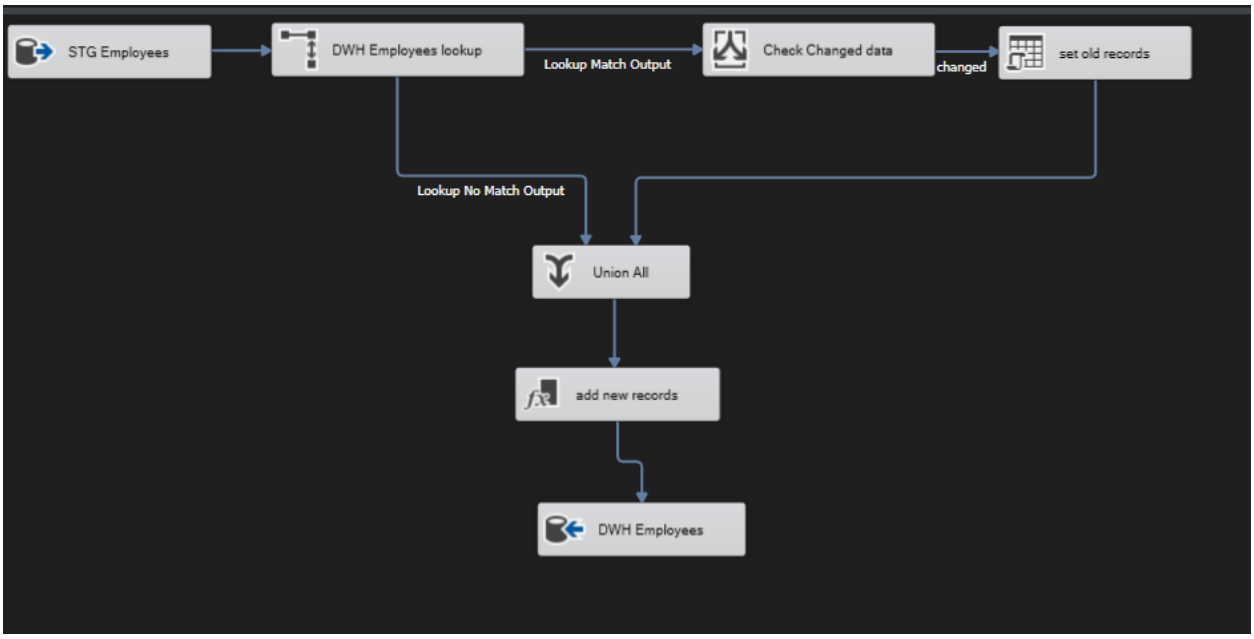
- **Control flow of Emp Dim**



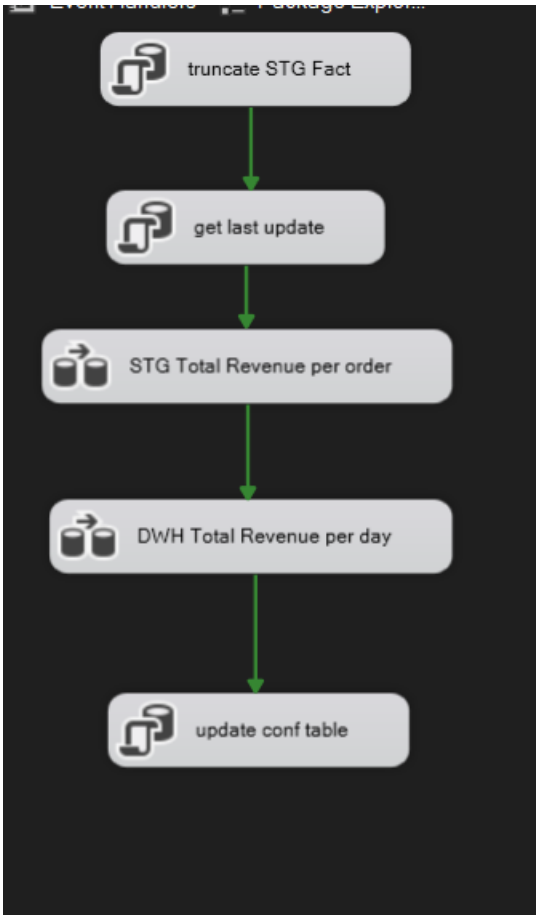
- **Data flow of Emp from src to stg**



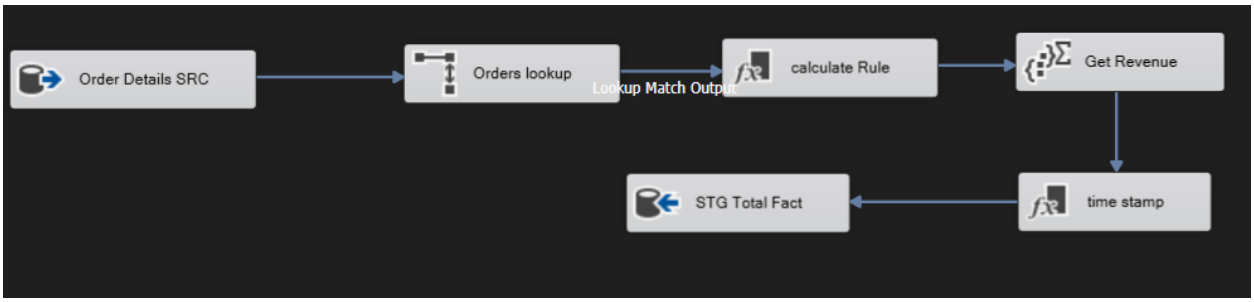
- **Data flow of Emp from stg to dwh**



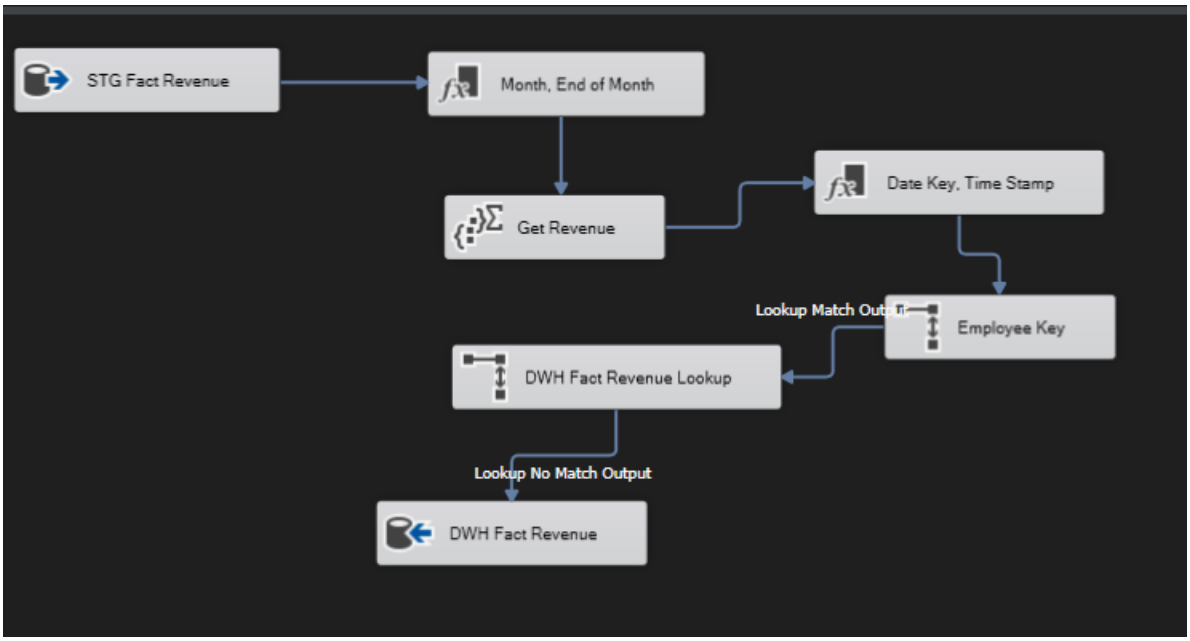
- **Control flow of Emp fact**



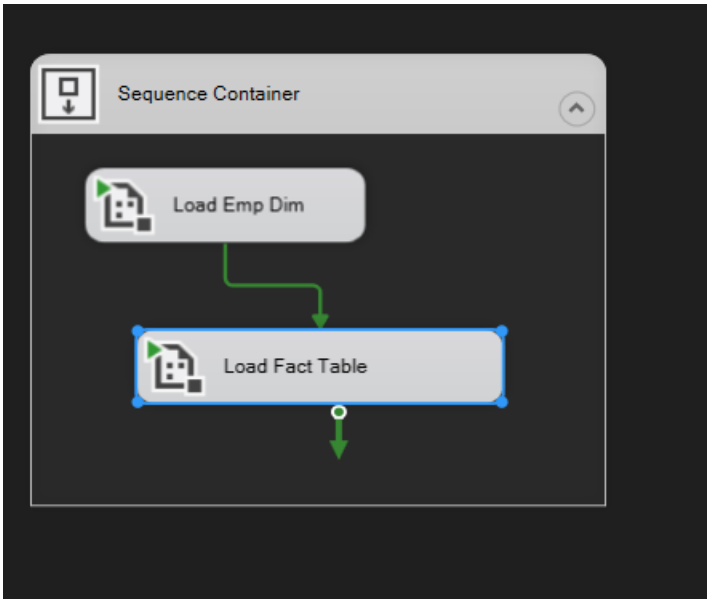
- **Data flow of fact from src to stg**



- **Data flow of fact from stg to dwh**

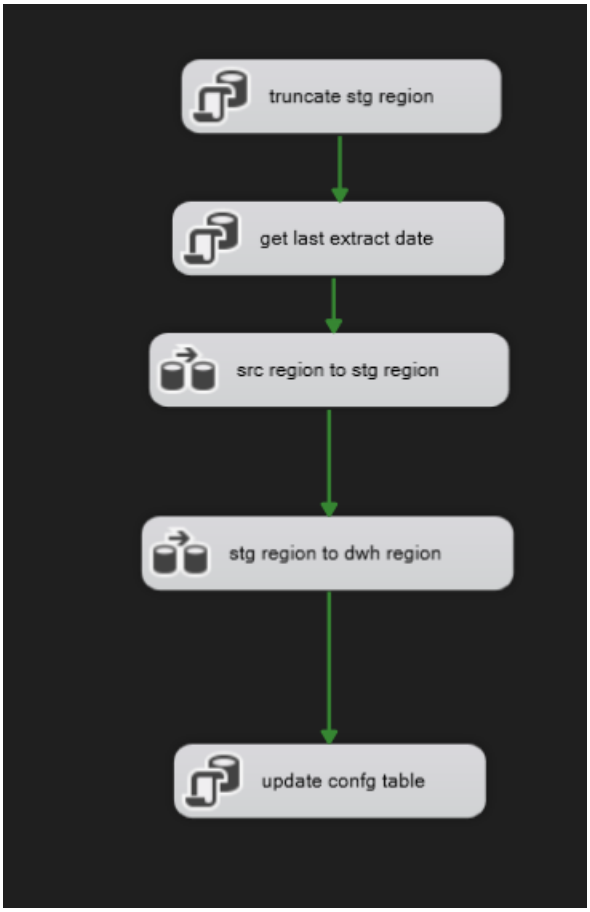

STG Fact Revenue
Month, End of Month
Get Revenue
Date Key, Time Stamp
Lookup Match Out
Employee Key
DWH Fact Revenue Lookup
Lookup No Match Output
DWH Fact Revenue

- **Run the DWH**

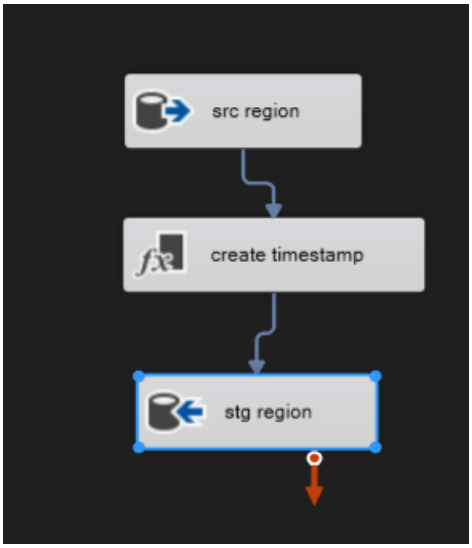

Sequence Container
Load Emp Dim
Load Fact Table

3 – Emp num Fact

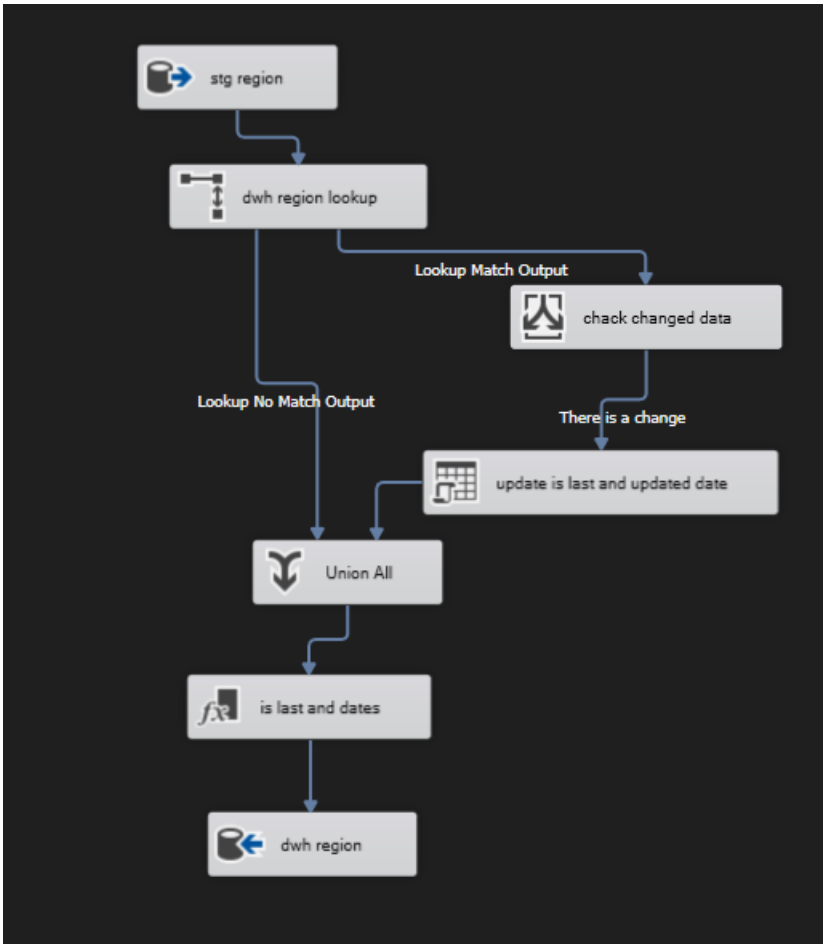
- **Control flow of the region Dim**



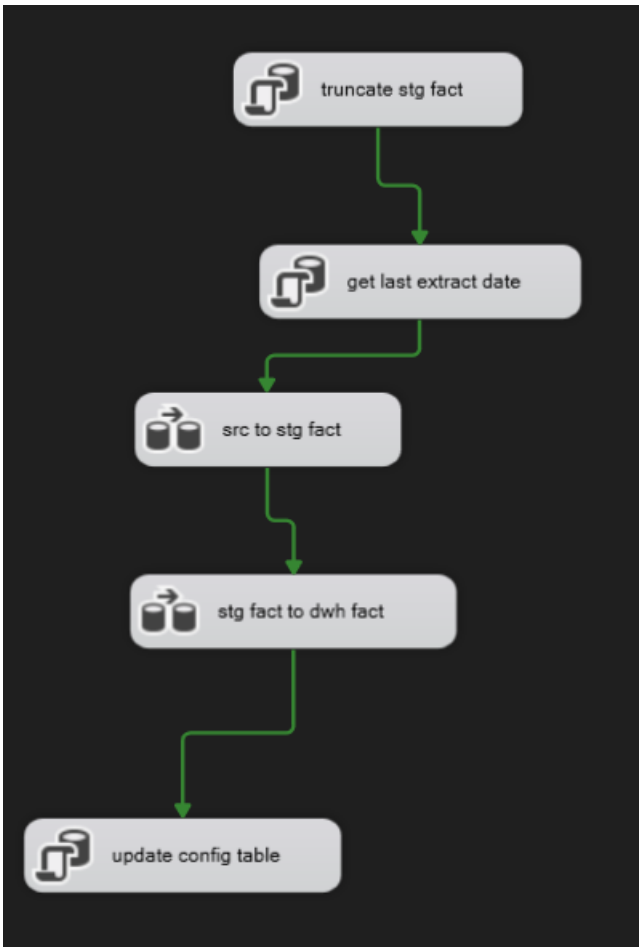
- **Data flow of region from src to stg**



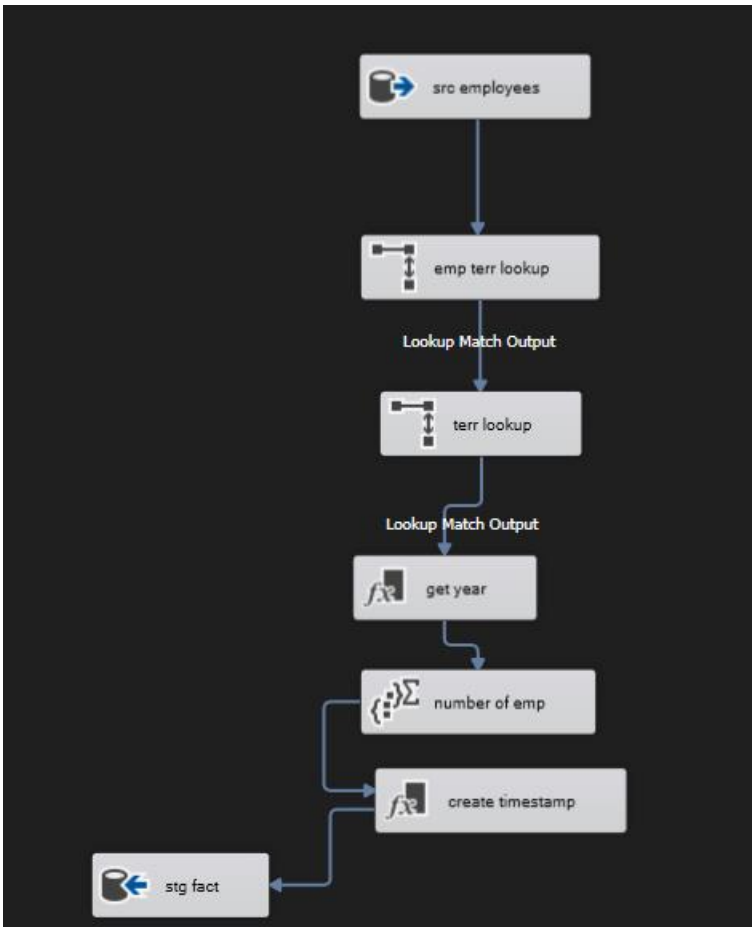
- **Data flow of region from stg to dwh**



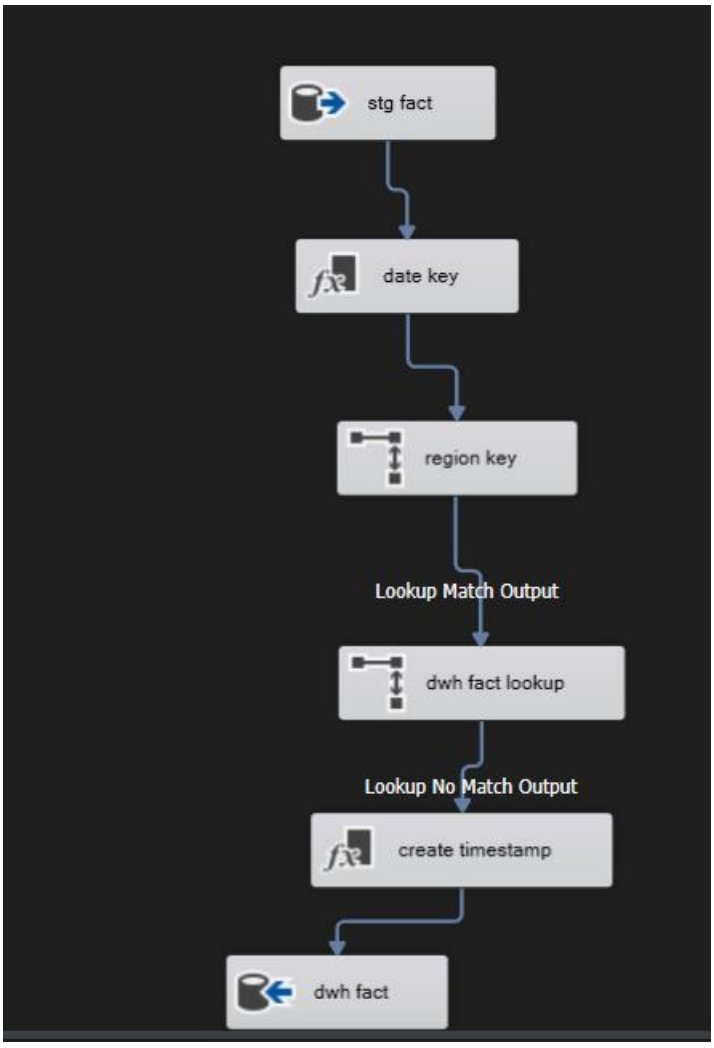
- **Control flow of fact**



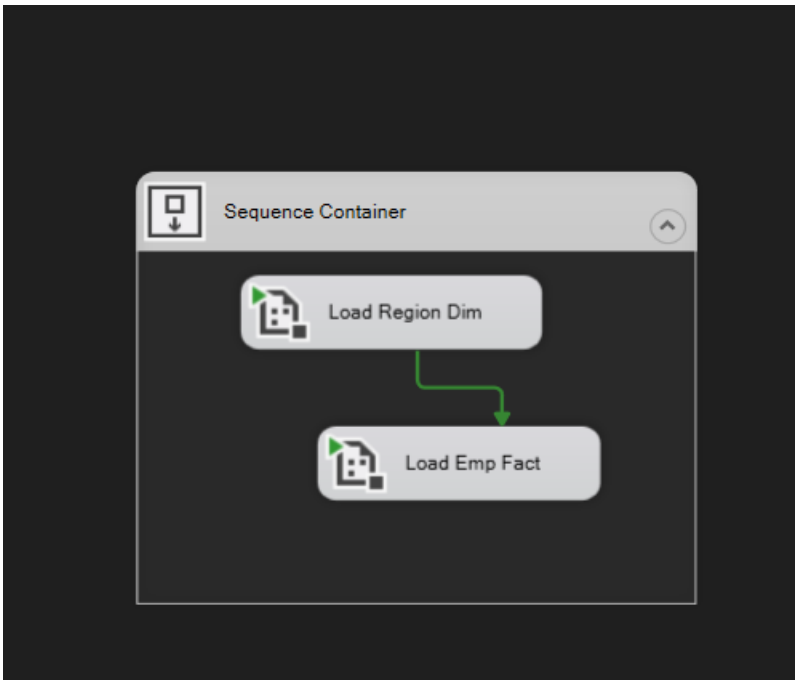
- **Data flow of fact from src to stg**



- **Data flow of fact from stg to dwh**

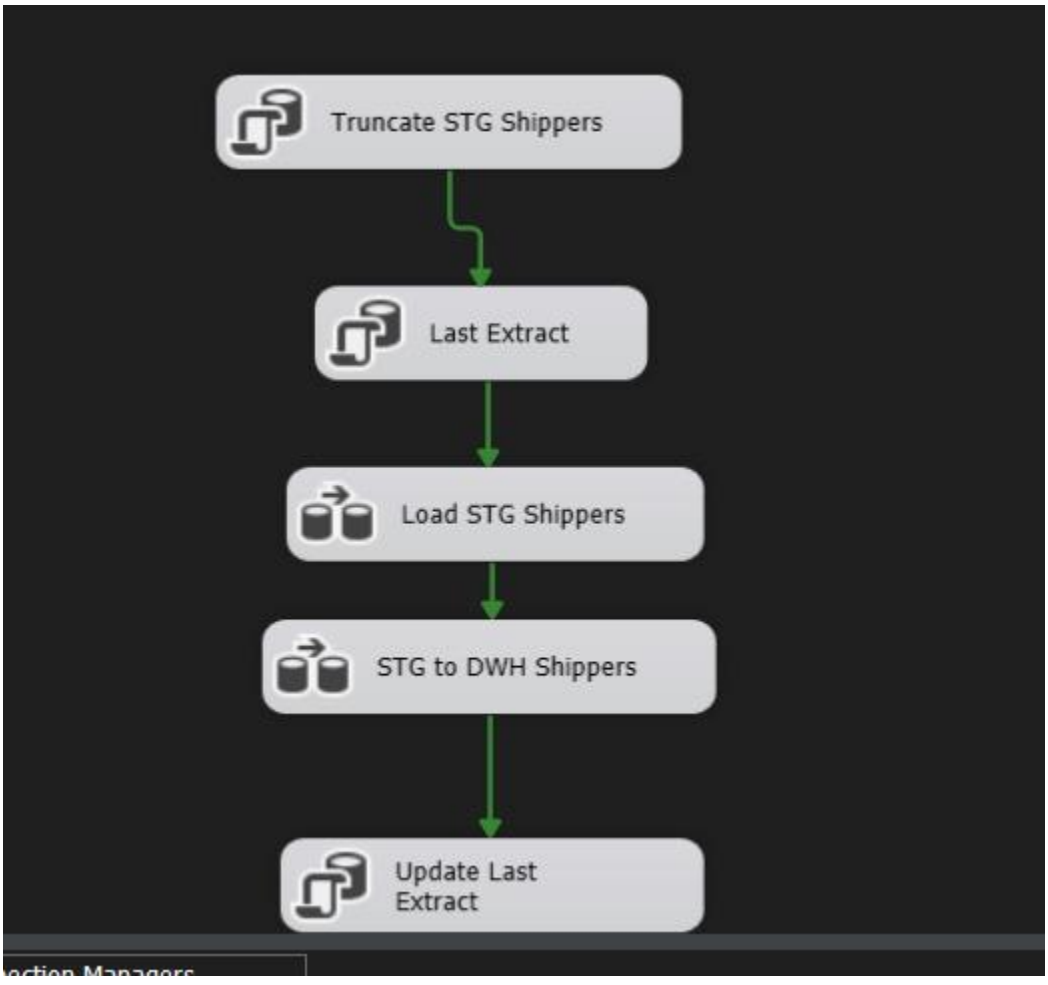


- **Run the DWH**

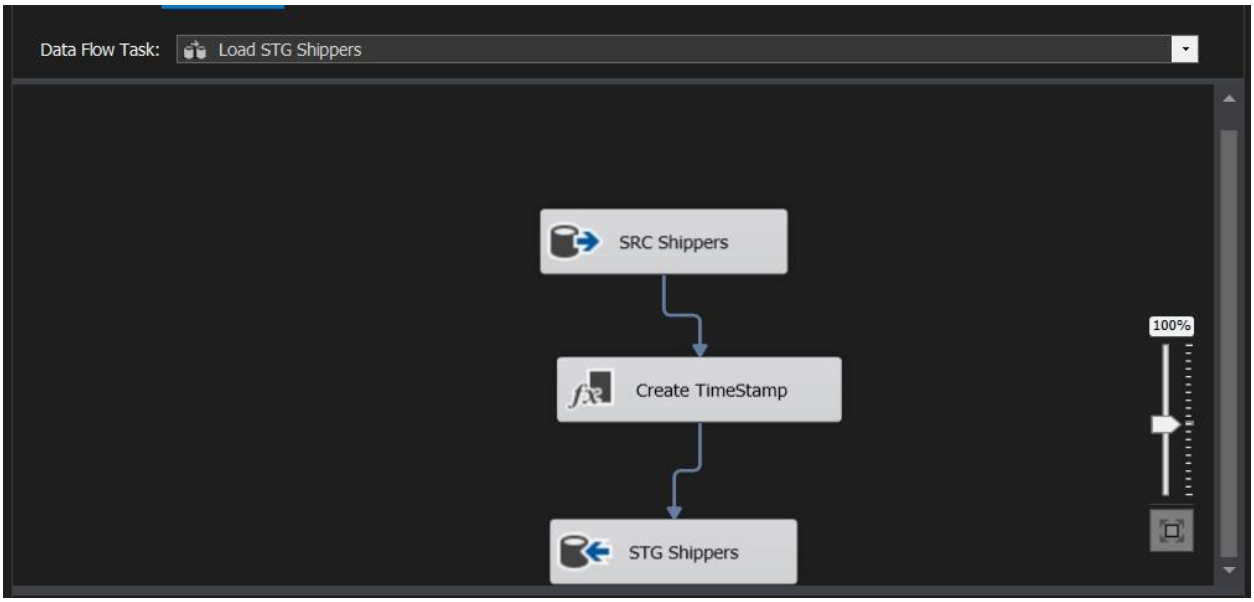


4- Orders num fact

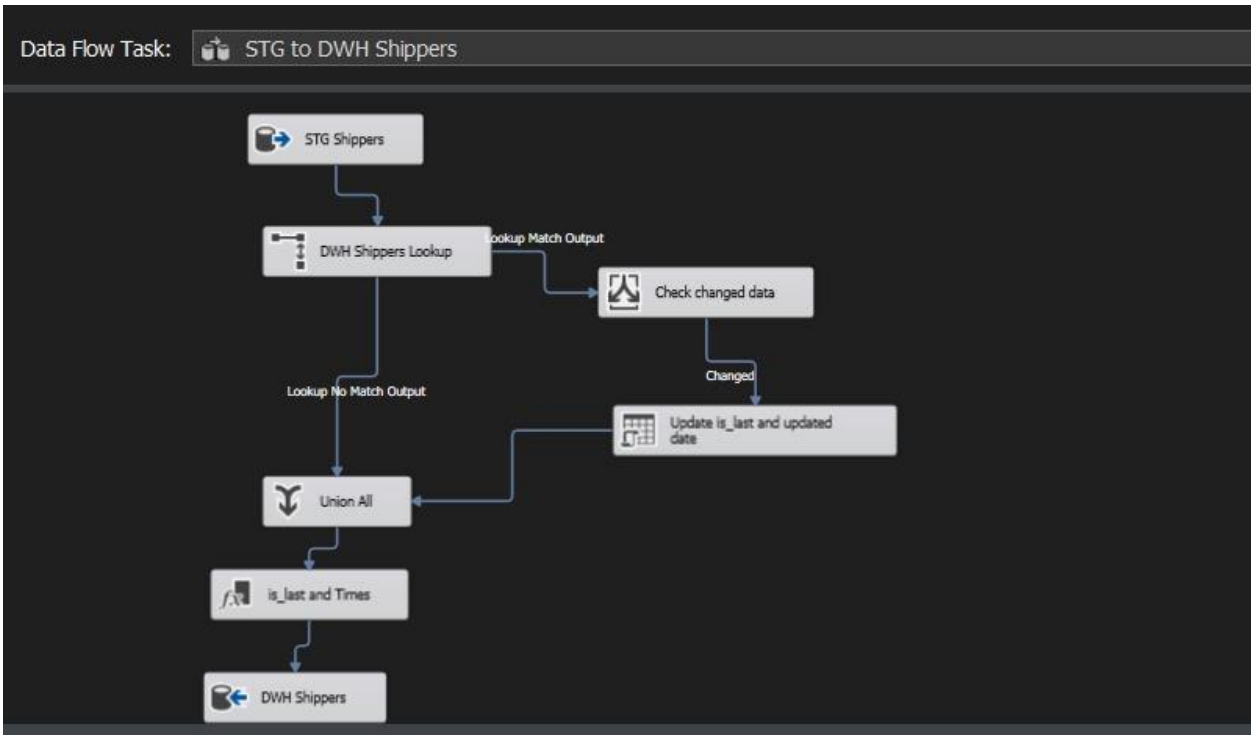
- **Control flow of the shipper DIM**



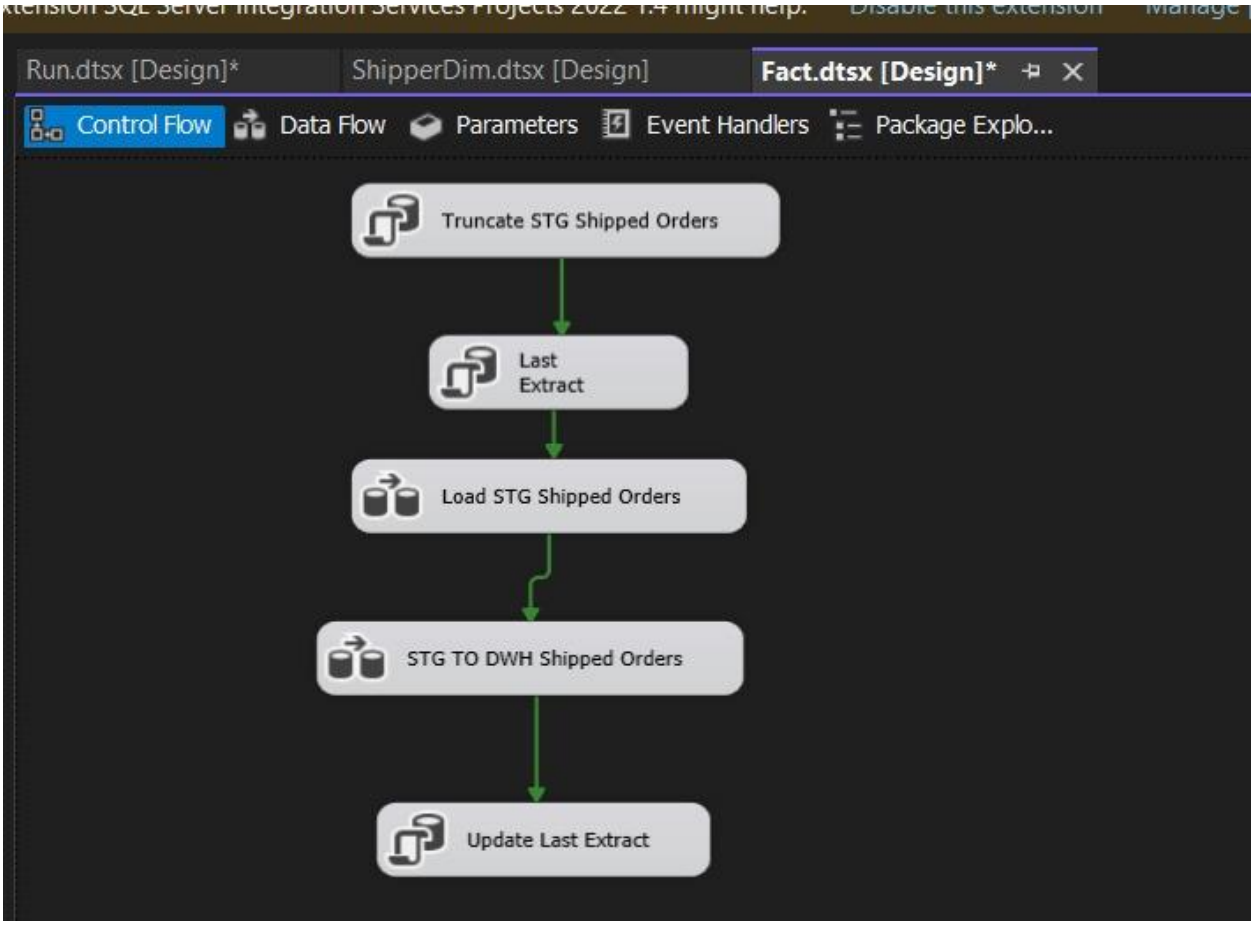
- **Data flow of shipper from src to stg**



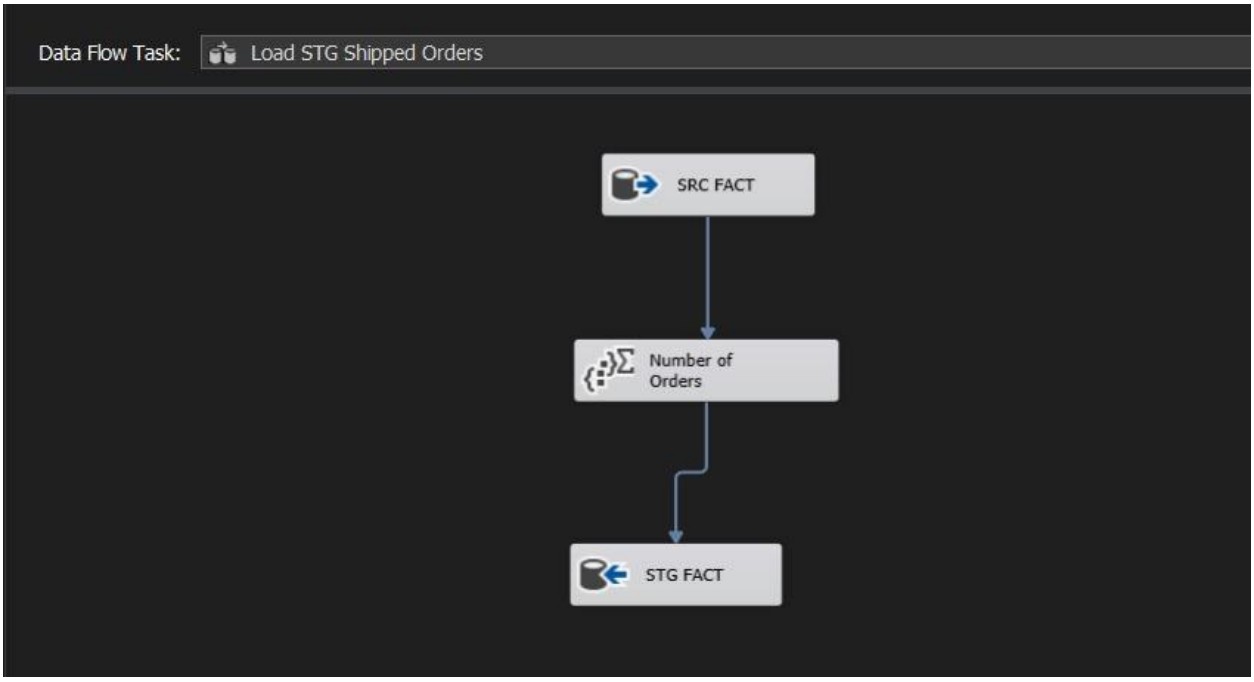
- **Data flow of shipper from stg to dwh**



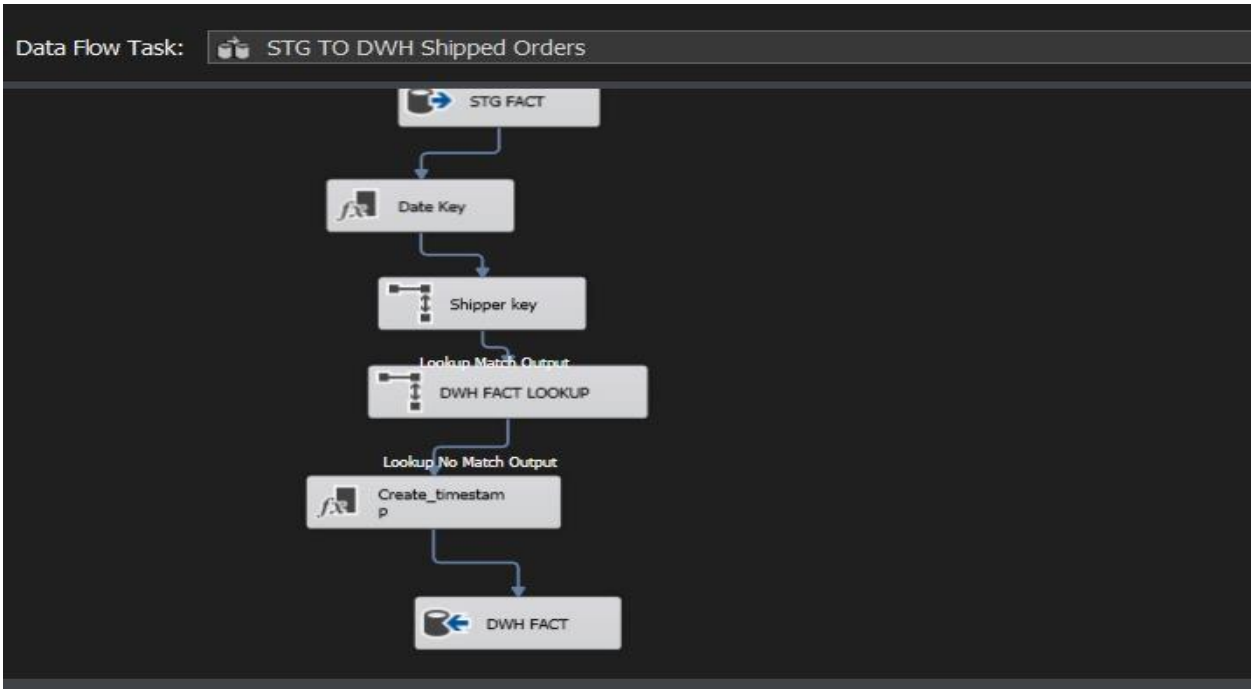
- **Control flow of fact**


Run.dtsx [Design]*
ShipperDim.dtsx [Design]
Fact.dtsx [Design]*
Control Flow
Data Flow
Parameters
Event Handlers
Package Explo...
Truncate STG Shipped Orders
Last Extract
Load STG Shipped Orders
STG TO DWH Shipped Orders
Update Last Extract

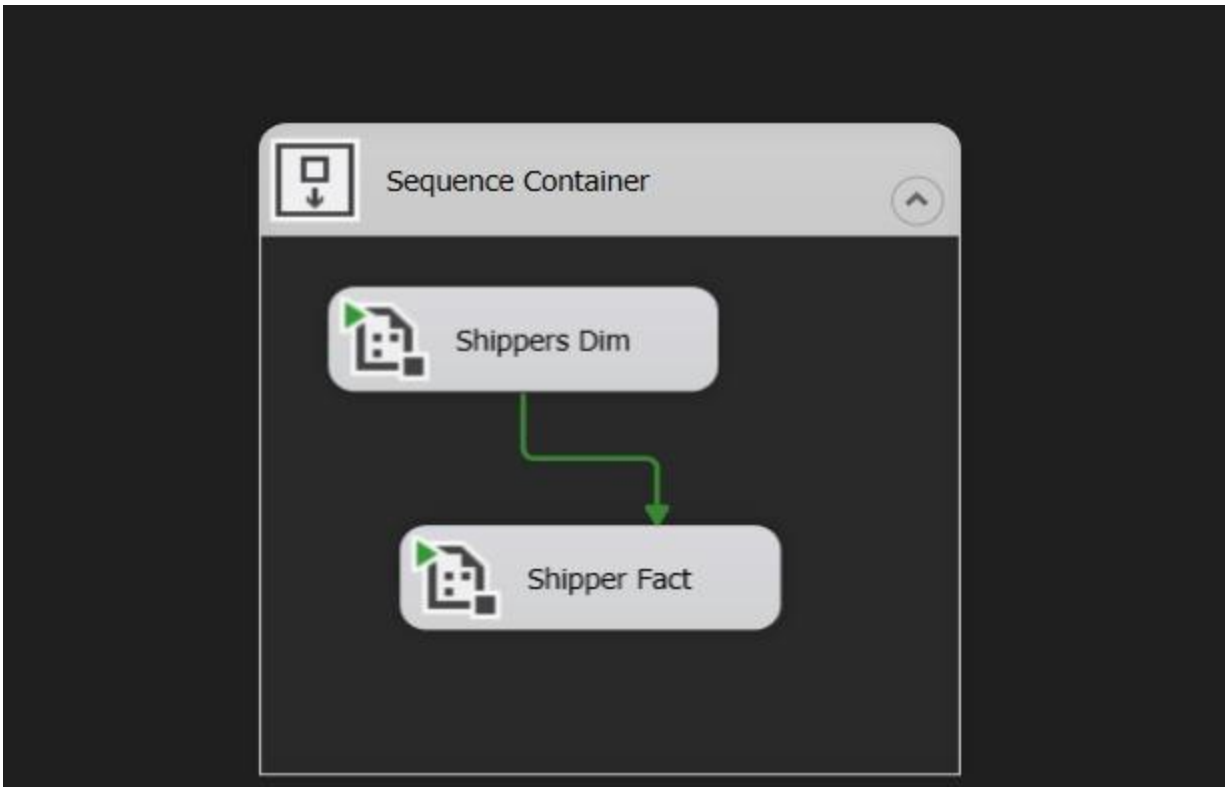
- **Data flow of fact from src to stg**



- **Data flow of fact from stg to dwh**



- **Run the DWH**


Sequence Container
Shippers Dim
Shipper Fact

5- 1- Product Sold Fact :

```sql
-- get the number of products sold along with their product names

SELECT
    p.ProductID,
    p.ProductName,
    SUM(f.ProductsSold) AS TotalProductsSold
FROM
    DWH.FactProductSales f
JOIN
    DWH.Products p ON f.ProductKey = p.product_key
GROUP BY
    p.ProductID, p.ProductName
ORDER BY
    TotalProductsSold DESC;
```

38 %

Results | Messages

	ProductID	ProductName	TotalProductsSold
1	85	Boston Crab Meat	6
2	17	Alice Mutton	4
3	3	Aniseed Syrup	4
4	40	Boston Crab Meat	4
5	60	Camembert Pierrot	4
6	100	Chartreuse verte	4
7	77	Cheddar	4
8	4	Chef Anton's Cajun Seasoning	4
9	48	Chocolade	4
10	52	Filo Mix	4
11	33	Geitost	4
12	37	Gravad lax	4
13	24	Guaraná Fantástica	4
14	69	Gudbrandsdalsost	4
15	97	Gudbrandsdalsost	4
16	44	Gula Malacca	4
17	56	Gnocchi di nonna Alice	4

```sql
-- calculates the total number of products sold to each customer


SELECT
    c.CustomerID,
    c.ContactName ,
    SUM(f.ProductsSold) AS TotalProductsSold
FROM
    DWH.FactProductSales f
JOIN
    DWH.Customers c ON f.CustomerKey = c.customer_key
GROUP BY
    c.CustomerID, c.ContactName
ORDER BY
    TotalProductsSold DESC;


-- number of product sold for each month in each year
```

esults | Messages

CustomerID	ContactName	TotalProductsSold
32	Sebastian Ward	21
31	Ella Turner	18
48	Avery Rivera	15
56	Aiden Cook	14
55	Lily Morris	14
64	Lily Morris	14
40	Ella Turner	13
63	Evelyn Hill	13
71	Lucas Reyes	11
8	William Rodriguez	11
16	Noah Hall	10
24	Henry Hall	10
39	Avery Rivera	10
23	Amelia Walker	10
47	Aiden Cook	10
33	Mila Bailey	9

```sql
-- number of product sold for each month in each year


SELECT
    d.Year,
    d.Month,
    SUM(f.ProductsSold) AS TotalProductsSold
FROM
    DWH.FactProductSales f
JOIN
    DWH.Products p ON f.ProductKey = p.product_key
JOIN
    DWH.D_Date d ON f.TimeKey = d.DateKey
GROUP BY
    d.Year,
    d.Month
ORDER BY
    d.Year,
    d.Month,
    TotalProductsSold DESC;
```

sults | Messages

Year	Month	TotalProductsSold
2021	1	18
2021	2	20
2021	3	26
2021	4	23
2021	5	15
2021	6	85
2021	7	78
2021	8	64
2021	9	20
2021	10	18
2021	11	22
2021	12	8

2- Orders num Fact :

```sql
-- Total number of orders shipped by each shipper
SELECT
    s.CompanyName,
    SUM(fos.NumberOfOrders) AS TotalOrders
FROM
    DWH.FactOrdersShipped fos
JOIN
    DWH.Shippers s ON fos.Shipper_key = s.shipper_key
GROUP BY
    s.CompanyName;


--Shippers who have shipped more than a certain number
```

150 %

Results | Messages

	CompanyName	TotalOrders
1	Federal Shipping	42
2	Speedy Express	40
3	United Package	43

```sql
-- Get Shippers Who Have Not Shipped Any Orders
SELECT
    s.CompanyName
FROM
    DWH.Shippers s
LEFT JOIN
    DWH.FactOrdersShipped fos ON s.shipper_key = fos.Shipper_key
WHERE
    fos.Shipper_key IS NULL;


--Get Orders Shipped by a Specific Shipper
```

150 %

Results | Messages

	CompanyName
1	Quickship
2	Express Cargo
3	Shippio
4	Fast Track
5	Swift Deliveries
6	Air Express
7	Delivery Masters
8	Quick Deliveries
9	Speedy Ship
10	Rapid Express
11	Swift Shipments
12	Express Deliveries
13	Fast Freight
14	Quick Transport
15	Air Speed
16	Speedy Delivery
17	Priority Shippers
18	Express Movers

Query executed successfully. LAPTOP-

```sql
--Get Orders Shipped by a Specific Shipper
SELECT
    fos.FactOrdersShipped,
    fos.NumberOfOrders,
    fos.TimeKey
FROM
    DWH.FactOrdersShipped fos
JOIN
    DWH.Shippers s ON fos.Shipper_key = s.shipper_key
WHERE
    s.CompanyName = 'Speedy Express';
```

150 %

Results | Messages

	FactOrdersShipped	NumberOfOrders	TimeKey
1	5	2	20210822
2	6	1	20211207
3	13	1	20210605
4	14	1	20210615
5	16	1	20210705
6	17	1	20210715
7	26	1	20210620
8	29	2	20210810
9	31	1	20210830
10	34	1	20211115
11	39	1	20210608
12	44	1	20210828
13	47	1	20240818
14	48	1	20210403
15	49	1	20210503
16	50	1	20210523
17	53	1	20210903
18	54	1	20211203

3- Emp Region Fact

```sql
--Number of employees worked in each region in period of time
SELECT
    r.RegionDescription,
    SUM(f.NumberOfEmployees) AS TotalEmployeesWorked
FROM
    DWH.FactEmployeesWorked f
JOIN
    DWH.Region r ON f.Region_key = r.region_key
WHERE
    CONVERT(datetime, LEFT(CONVERT(varchar, f.TimeKey), 8)) BETWEEN '2010-01-01' AND '2015-05-18'
GROUP BY
    r.RegionDescription;
```

50 %

Results | Messages

	RegionDescription	TotalEmployeesWorked
1	Appalachia	1
2	Central Coast	1
3	Central Valley	1
4	Chico Metropolitan Area	1
5	Deep South	1
6	Eastern	1
7	Gold Country	1
8	Hanford-Corcoran Metropolitan Area	2
9	Los Angeles-Long Beach-Glendale Metropolitan Area	1
10	Madera Metropolitan Area	1
11	Merced Metropolitan Area	1
12	Midwest	2
13	Modesto Metropolitan Area	1
14	Napa Metropolitan Area	1
15	Northeast	1
16	Northern	1
17	Northern Rockies	1
18	Oxnard Metropolitan Area	1

```sql
--Total number of employees worked per year
SELECT
    YEAR(CONVERT(datetime, CONVERT(varchar, f.TimeKey))) AS Year,
    SUM(f.NumberOfEmployees) AS TotalEmployeesWorked
FROM
    DWH.FactEmployeesWorked f
GROUP BY
    YEAR(CONVERT(datetime, CONVERT(varchar, f.TimeKey)));
```

50 %

Results | Messages

	Year	TotalEmployeesWorked
1	2006	1
2	2007	7
3	2008	9
4	2009	7
5	2010	8
6	2011	6
7	2012	8
8	2013	11
9	2014	5
10	2015	8
11	2016	10
12	2017	9
13	2018	5
14	2019	6

```sql
-- Count the number of employees worked by each region
SELECT
    r.RegionDescription,
    SUM(f.NumberOfEmployees) AS TotalEmployeesWorked
FROM
    DWH.FactEmployeesWorked f
JOIN
    DWH.Region r ON f.Region_key = r.region_key
GROUP BY
    r.RegionDescription;
```

0 %

Results | Messages

	RegionDescription	TotalEmployeesWorked
	Appalachia	1
	Bakersfield Metropolitan Area	2
	Bay Area	1
	Central	2
	Central Coast	1
	Central Valley	1
	Chico Metropolitan Area	1
	Deep South	1
	Down East	1
0	Eastern	2
1	El Centro Metropolitan Area	1
2	Fairfield Metropolitan Area	1
3	Four Corners	1
4	Fresno Metropolitan Area	2
5	Gold Country	1
6	Great Lakes	1
7	Gulf Coast	1
8	Hanford-Corcoran Metropolitan Area	2

4- Emp Revenue Fact:

```sql
-- Select the total revenue for each month across all employees
SELECT
    DD.Year,
    FR.Month,
    SUM(FR.TotalRevenue) AS TotalMonthlyRevenue
FROM
    DWH.FactTotalRevenue FR
JOIN
    DWH.D_Date DD ON FR.TimeKey = DD.DateKey
GROUP BY
    DD.Year,
    FR.Month
ORDER BY
    DD.Year,FR.Month ;
```

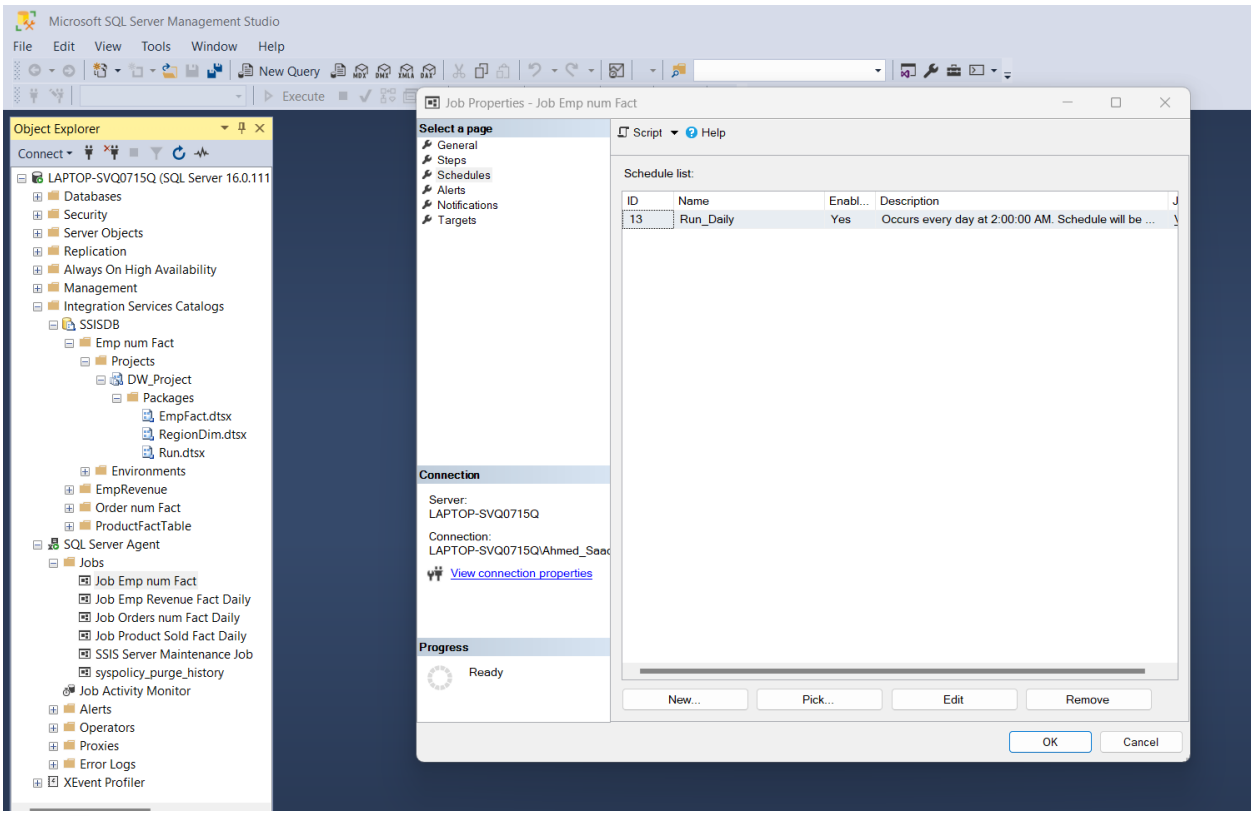
117 %

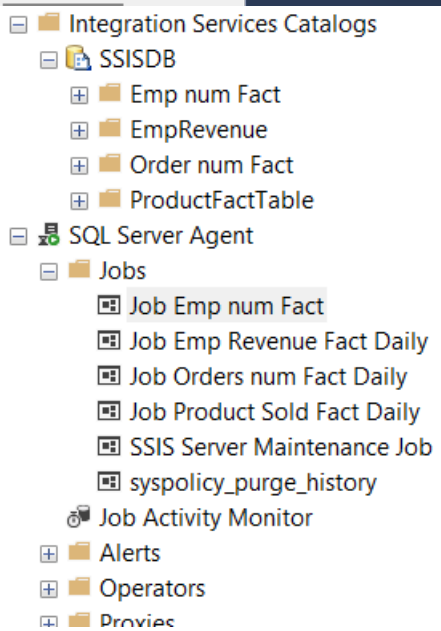
Results | Messages

	Year	Month	TotalMonthlyRevenue
1	2021	1	471
2	2021	2	437.7
3	2021	3	503.5
4	2021	4	449.675
5	2021	5	293
6	2021	6	1543.15
7	2021	7	1472.7
8	2021	8	1341.35
9	2021	9	446.85
10	2021	10	502.5
11	2021	11	544.77
12	2021	12	190.5

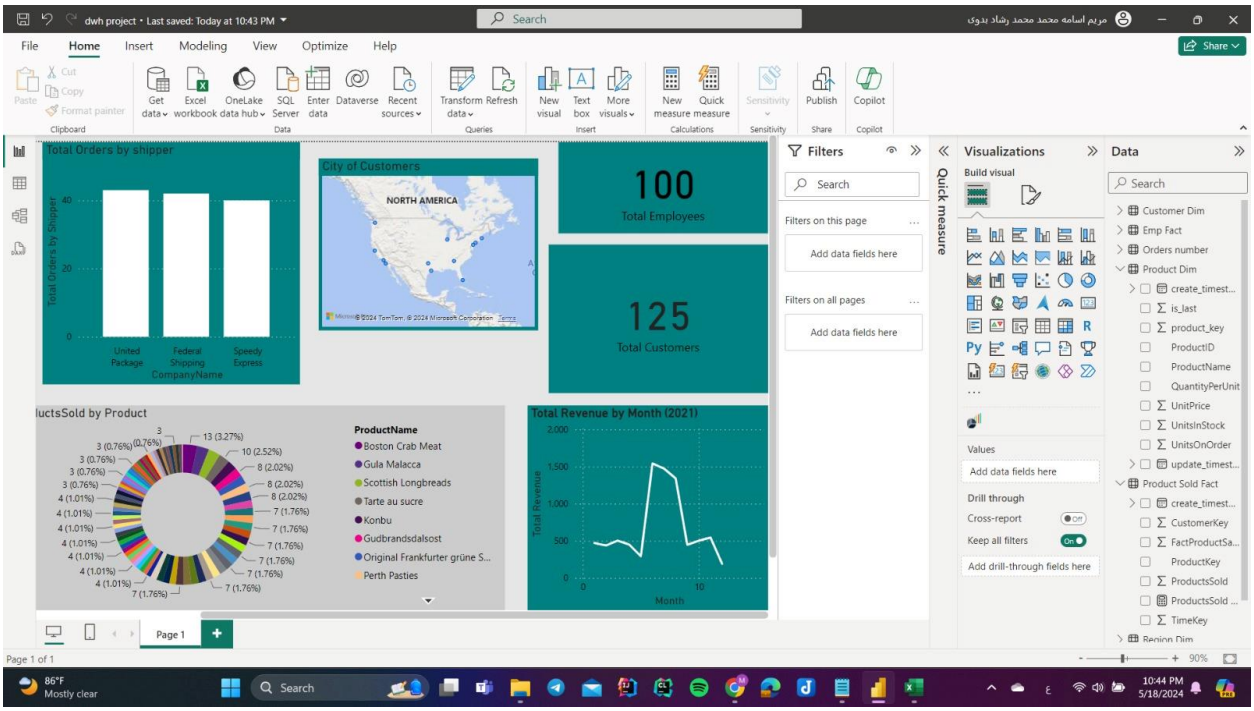
Query executed successfully.

6- Deploy and Schedule :


Microsoft SQL Server Management Studio
File Edit View Tools Window Help
New Query
Execute
Object Explorer
Connect
LAPTOP-SVQ0715Q (SQL Server 16.0.111
Databases
Security
Server Objects
Replication
Always On High Availability
Management
Integration Services Catalogs
SSISDB
Emp num Fact
Projects
DW_Project
Packages
EmpFact.dtsx
RegionDim.dtsx
Run.dtsx
Environments
EmpRevenue
Order num Fact
ProductFactTable
SQL Server Agent
Jobs
Job Emp num Fact
Job Emp Revenue Fact Daily
Job Orders num Fact Daily
Job Product Sold Fact Daily
SSIS Server Maintenance Job
syspolicy_purge_history
Job Activity Monitor
Alerts
Operators
Proxies
Error Logs
XEvent Profiler
Job Properties - Job Emp num Fact
Select a page
General
Steps
Schedules
Alerts
Notifications
Targets
Script
Help
Schedule list:
ID
Name
Enabl...
Description
13
Run_Daily
Yes
Occurs every day at 2:00:00 AM. Schedule will be ...
Connection
Server:
LAPTOP-SVQ0715Q
Connection:
View connection properties
Progress
Ready
New...
Pick...
Edit
Remove
OK
Cancel


Integration Services Catalogs
SSISDB
Emp num Fact
EmpRevenue
Order num Fact
ProductFactTable
SQL Server Agent
Jobs
Job Emp num Fact
Job Emp Revenue Fact Daily
Job Orders num Fact Daily
Job Product Sold Fact Daily
SSIS Server Maintenance Job
syspolicy_purge_history
Job Activity Monitor
Alerts
Operators
Proxies

7- Dash Board :


dwh project • Last saved: Today at 10:43 PM
Search
مريم اسامه محمد محمد رشاد بدوى
File
Home
Insert
Modeling
View
Optimize
Help
Share
Total Orders by shipper
Total Orders by Shipper
United Package
Federal Shipping
Speedy Express
CompanyName
City of Customers
NORTH AMERICA
100
Total Employees
125
Total Customers
luctsSold by Product
ProductName
Boston Crab Meat
Gula Malacca
Scottish Longbreads
Tarte au sucre
Konbu
Gudbrandsdalsost
Original Frankfurter grüne S...
Perth Pasties
Total Revenue by Month (2021)
Total Revenue
Month
Filters
Search
Filters on this page
Add data fields here
Filters on all pages
Add data fields here
Visualizations
Build visual
Values
Add data fields here
Drill through
Cross-report
Keep all filters
Add drill-through fields here
Data
Search
Customer Dim
Emp Fact
Orders number
Product Dim
create_timest...
is_last
product_key
ProductID
ProductName
QuantityPerUnit
UnitPrice
UnitsInStock
UnitsOnOrder
update_timest...
Product Sold Fact
create_timest...
CustomerKey
FactProductSa...
ProductKey
ProductsSold
ProductsSold ...
TimeKey
Region Dim
Quick measure
Page 1
Page 1 of 1
90%
86°F
Mostly clear
Search
10:44 PM
5/18/2024