Slowly changing Dimension Type 2

Introduction :

What is a Slowly Changing Dimension?

A Slowly Changing Dimension (SCD) is a dimension that stores and manages both current and historical data over time in a data warehouse. It is considered and implemented as one of the most critical ETL tasks in tracking the history of dimension records.

Type 2 SCDs - Creating another dimension record

A Type 2 SCD retains the full history of values. When the value of a chosen attribute changes, the current record is closed. A new record is created with the changed data values and this new record becomes the current record. Each record contains the effective time and expiration time to identify the time period between which the record was active.

Objective:

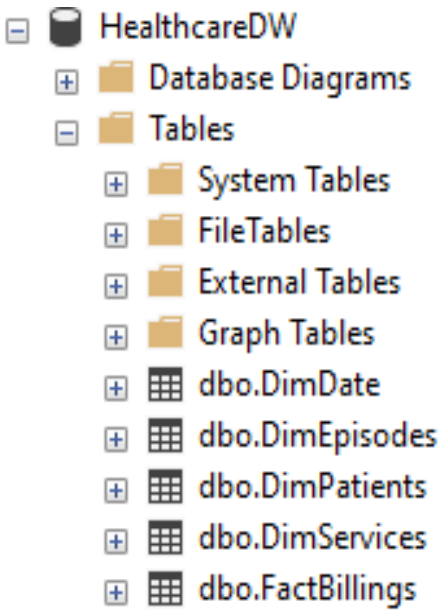
The Target of the project is to handle changes that happens in the Total Amount Billing and changes in the patient, Services or Episode Data .

The project lifecycle:

1- Create HealthCare Data Warehouse With Fact and Dimensions.
    a. Create Date Dimension.
    b. Create Episode Dimension.
    c. Create Patient Dimension.
    d. Create Services Dimension.
    e. Create Billing Fact.

2- Create Data Flow Task For Dimensions.
3- Make Slowly Changing Dimension For Each Dimension.
4- Show Results .
5- Testing Update and Insert.

1- Create HealthCare Data Warehouse With Fact and Dimensions tables



- **Patient Dimension**

```
SELECT TOP (1000) [patient_id_SK]
      ,[patient_id_BK]
      ,[Gov_Code]
      ,[Governance]
      ,[Gender_Key]
      ,[gender]
      ,[Nationality]
      ,[Nationality_Code]
      ,[Arabic_Name]
      ,[English_Name]
      ,[PatType]
      ,[Patient_Type]
      ,[FileID]
      ,[mobile_phone]
      ,[district_key]
      ,[District_Name]
      ,[pat_nrst_relv]
      ,[pat_addr]
      ,[_ValidFrom]
      ,[_ValidTo]
  FROM [HealthcareDW].[dbo].[DimPatients]
```

- **Episode Dimension**

```
SELECT TOP (1000) [Episode_Key_SK]
      ,[Episode_Key_BK]
      ,[Start_Date]
      ,[End_Date]
      ,[Episode_Type_Key]
      ,[Episode_Type]
      ,[Bed_Key]
      ,[Bed_Clinic]
      ,[_ValidFrom]
      ,[_ValidTo]
  FROM [HealthcareDW].[dbo].[DimEpisodes]
```

- **Date Dimension**

```
SELECT TOP (1000) [DateSK]
      ,[Date]
      ,[Day]
      ,[DaySuffix]
      ,[DayOfWeek]
      ,[DOWInMonth]
      ,[DayOfYear]
      ,[WeekOfYear]
      ,[WeekOfMonth]
      ,[Month]
      ,[MonthName]
      ,[Quarter]
      ,[QuarterName]
      ,[Year]
      ,[StandardDate]
      ,[HolidayText]
  FROM [HealthcareDW].[dbo].[DimDate]
```

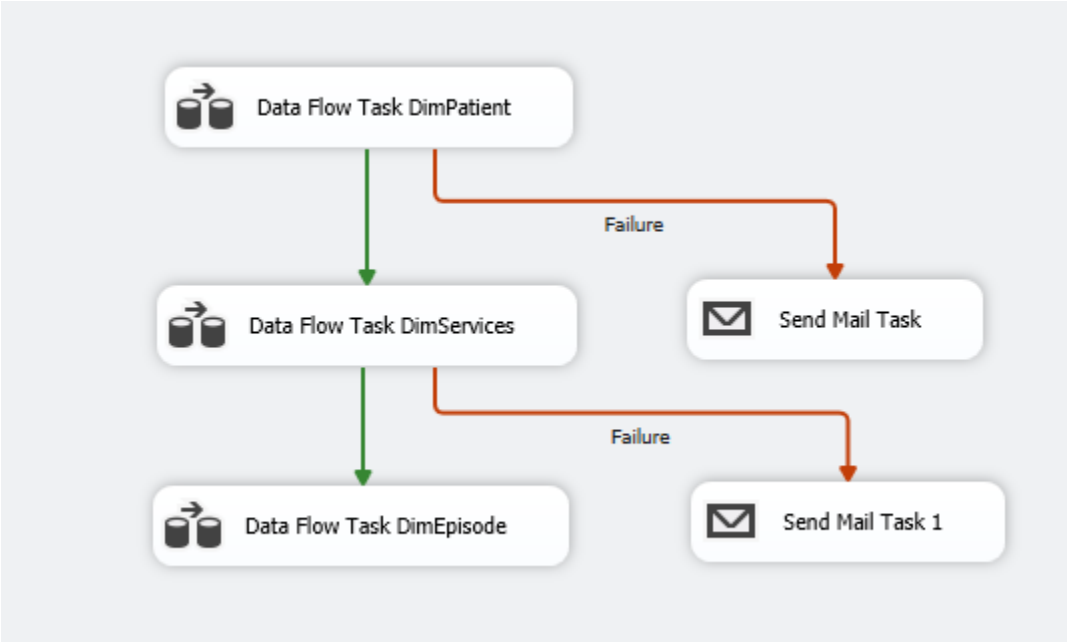
- **Fact Billing**

```
SELECT TOP (1000) [Service_SK]
      ,[patient_id_SK]
      ,[Episode_Key_SK]
      ,[DateSK]
      ,[Cash_Amount]
      ,[Credit_Amount]
      ,[Total_Amount]
  FROM [HealthcareDW].[dbo].[FactBillings]
```

- **Create Service Dimension**

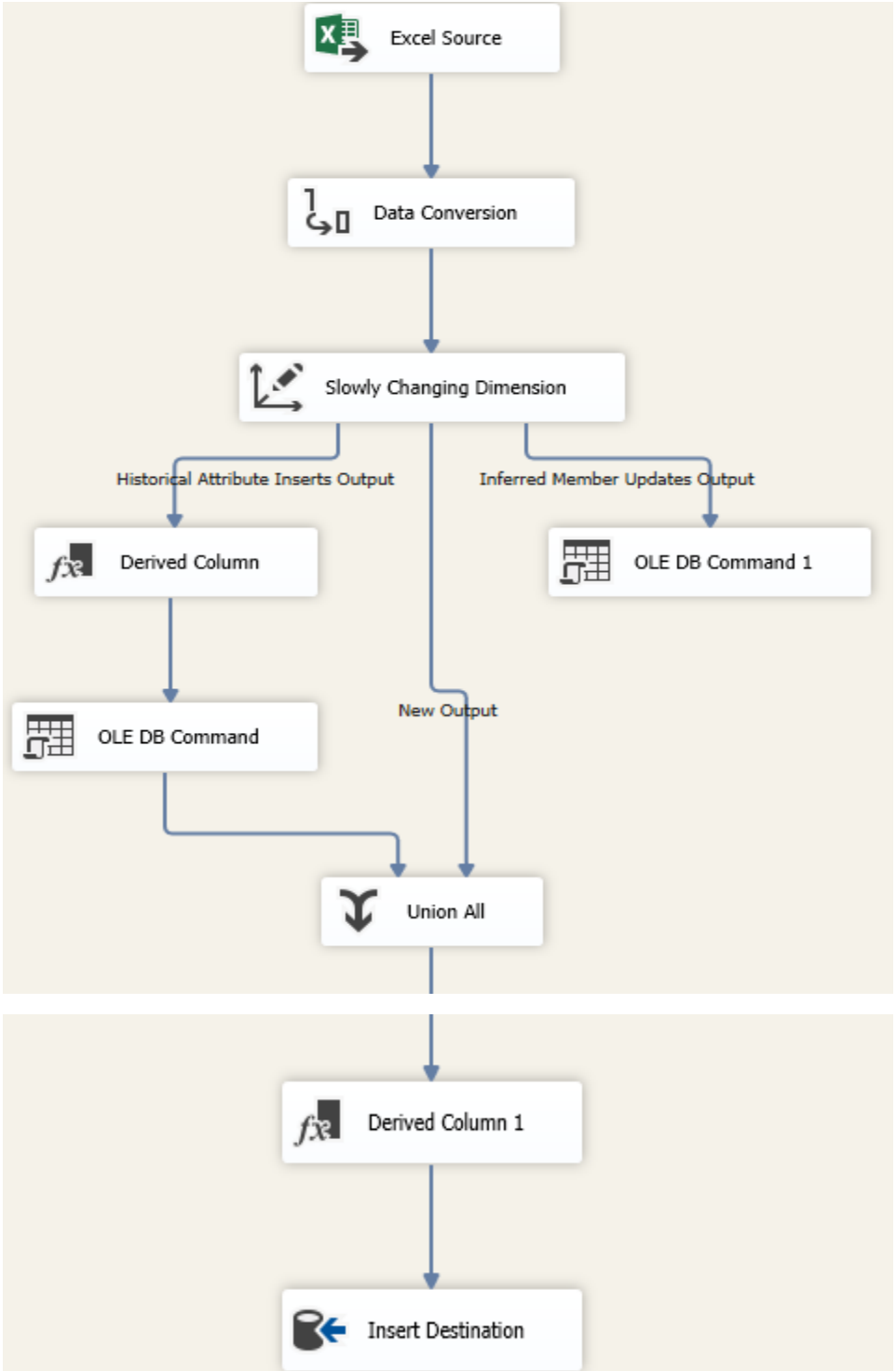
```
SELECT TOP (1000) [Service_SK]
      ,[Category_Key_BK]
      ,[Category_Name]
      ,[Service_Key_BK]
      ,[Service_Name]
      ,[_ValidFrom]
      ,[_ValidTo]
  FROM [HealthcareDW].[dbo].[DimServices]
```

2- Create Data Flow Task For Dimensions.

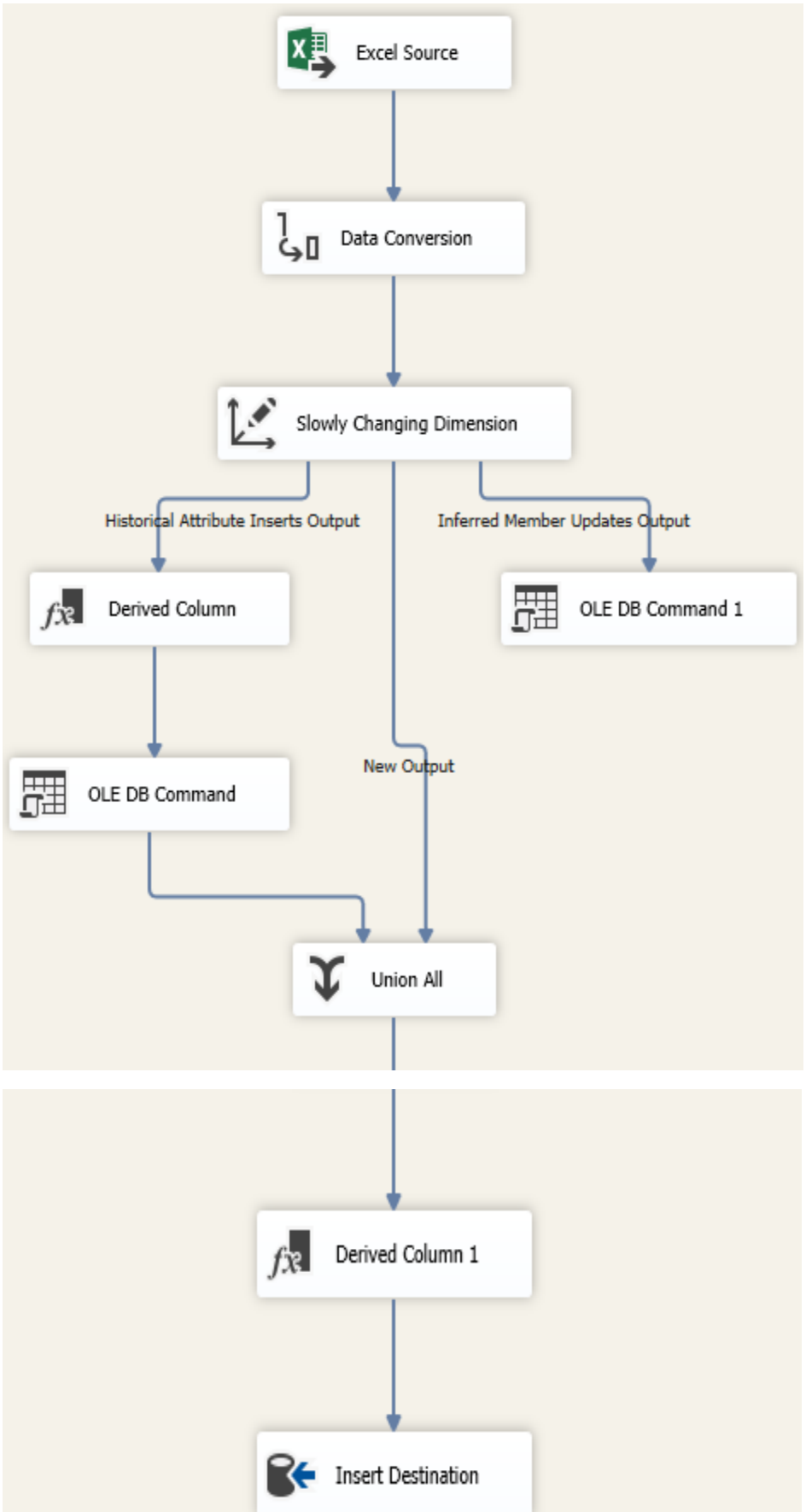


3- Make Slowly Changing Dimension For Each Dimensions

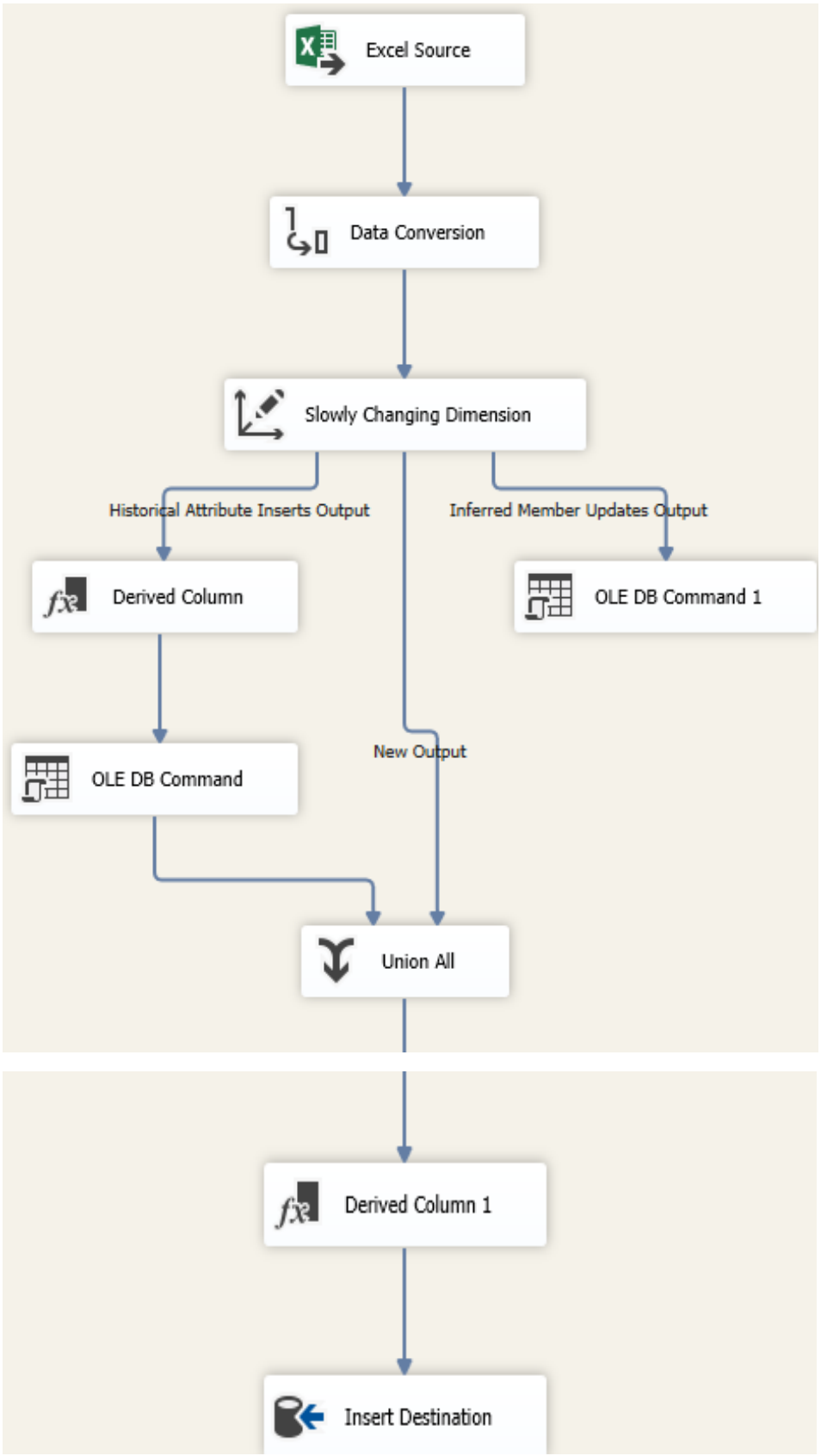
a. Create Data Flow Task For Patient Dimension



b. Create Data Flow Task For Services Dimension

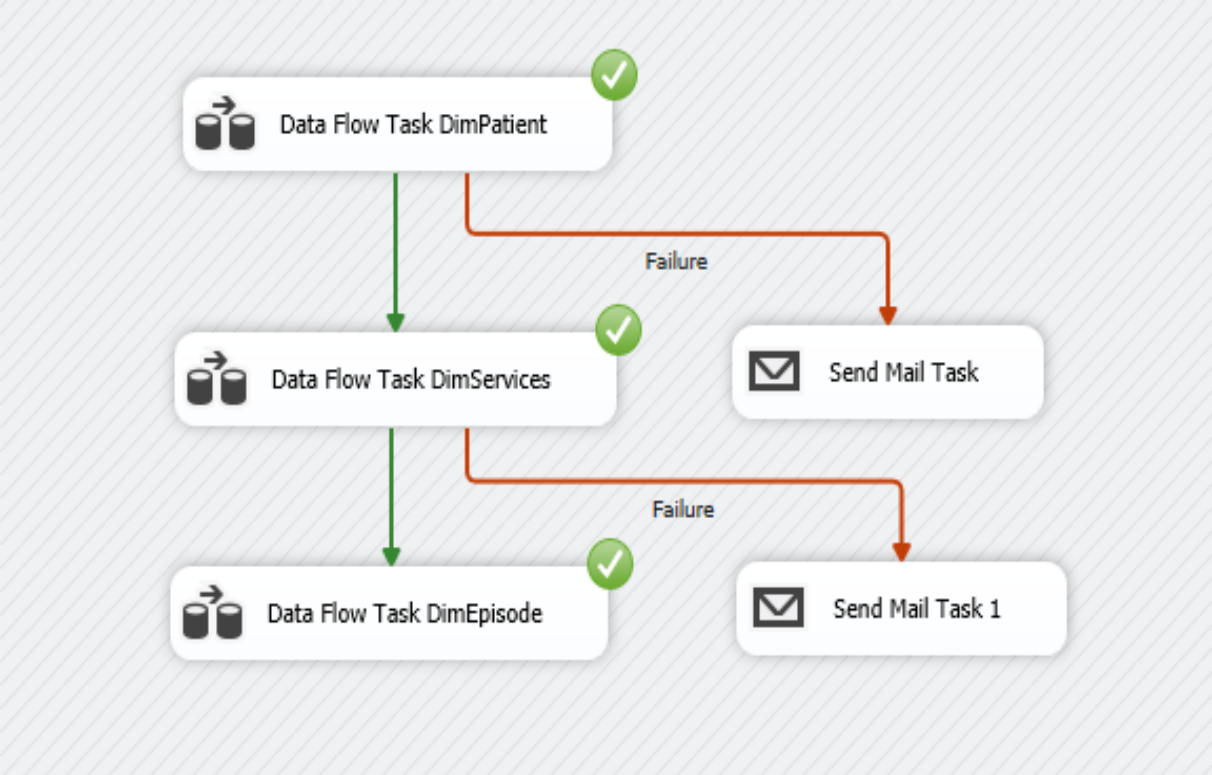


c. Create Data Flow Task For Episode Dimension



4- Show Results .

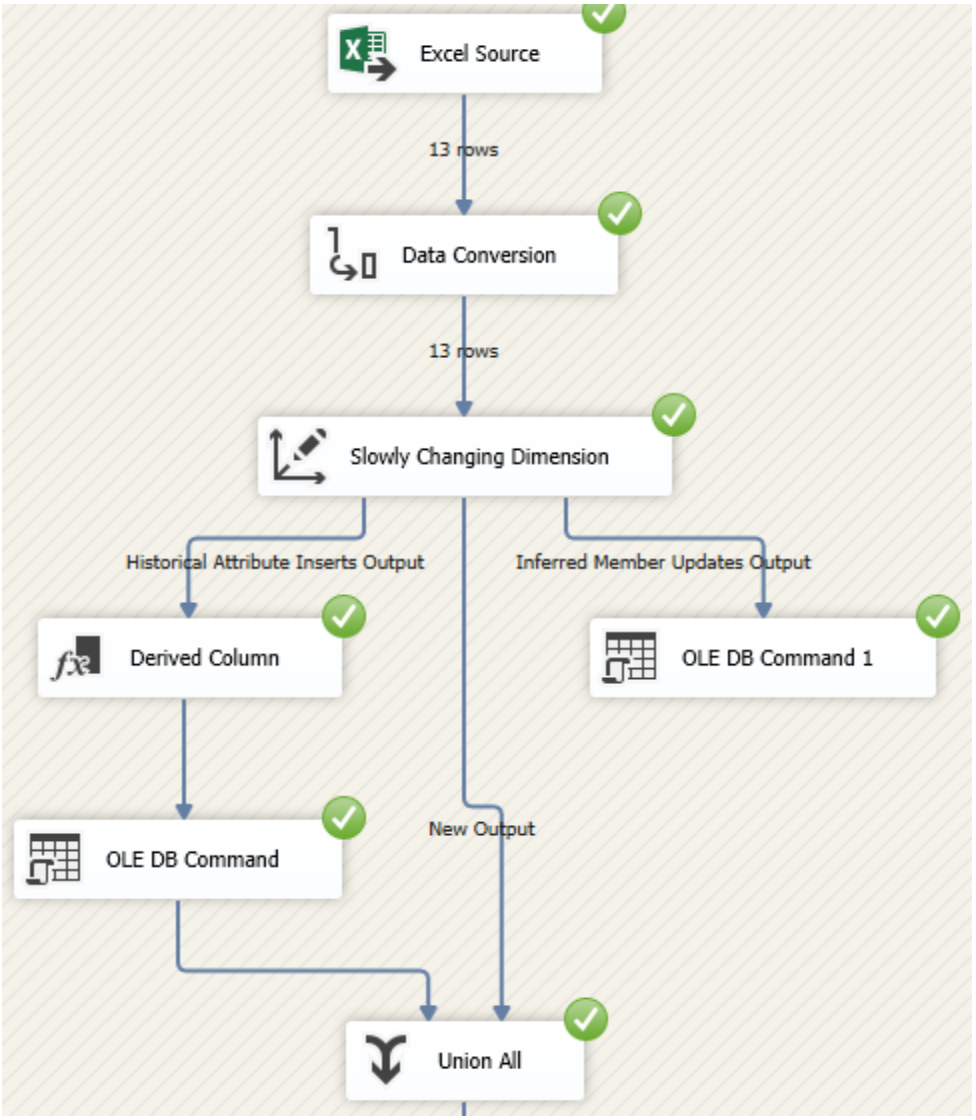
When Execute the control flow Tasks

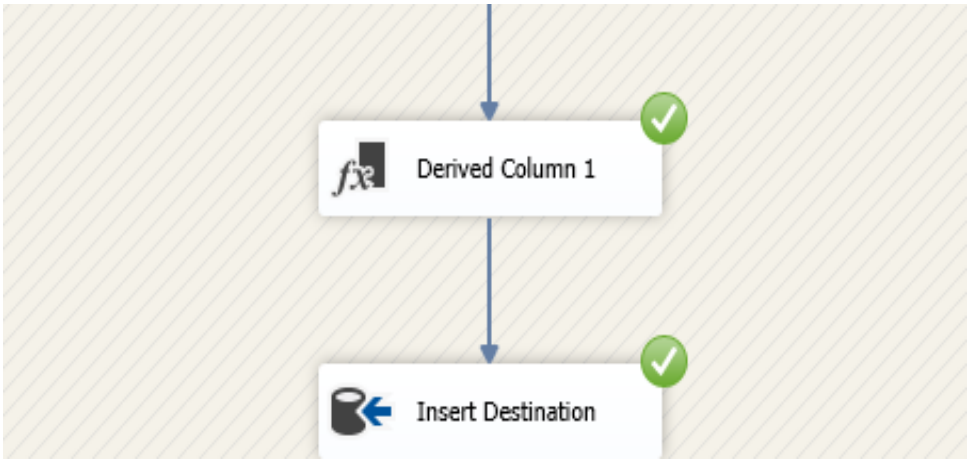


The first step is to get data from excel file patient and match columns with the columns in the patient dimension in HealthCare Data Warehouse, the second step is to matching the columns in slowly changing dimension but we found that there is a problem some columns must be matched with them with data types, data types was different so I needed to convert them first and then make slowly changing dimension so in the step we get the data from excel file and used data conversion to convert the data types and then make slowly changing dimension.

When Execute the Data flows tasks for each Dimension, the Patient Dimension

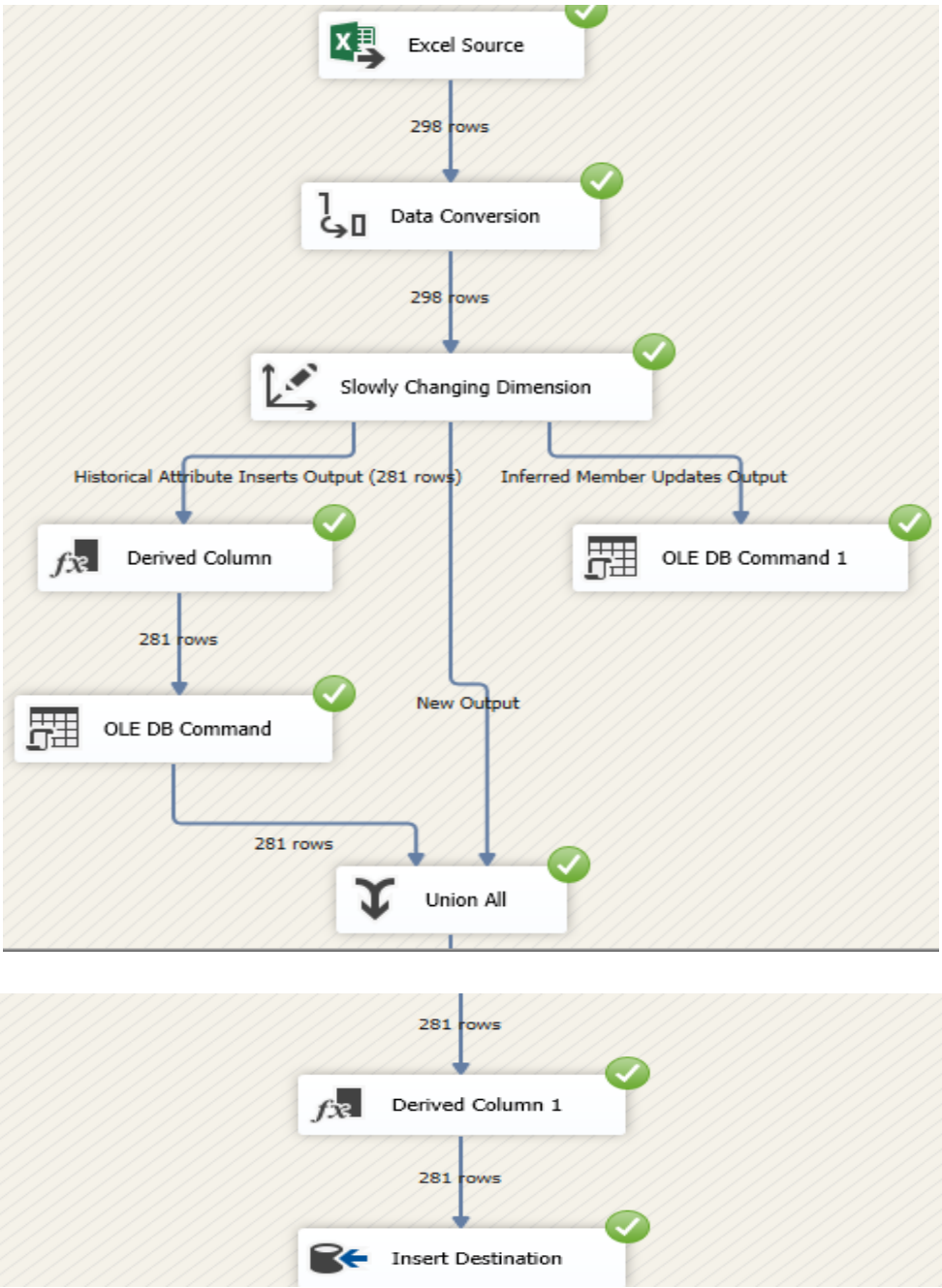
The data flow run successfully with any error after converting data types.


Excel Source
13 rows
Data Conversion
13 rows
Slowly Changing Dimension
Historical Attribute Inserts Output
Inferred Member Updates Output
Derived Column
OLE DB Command 1
OLE DB Command
New Output
Union All

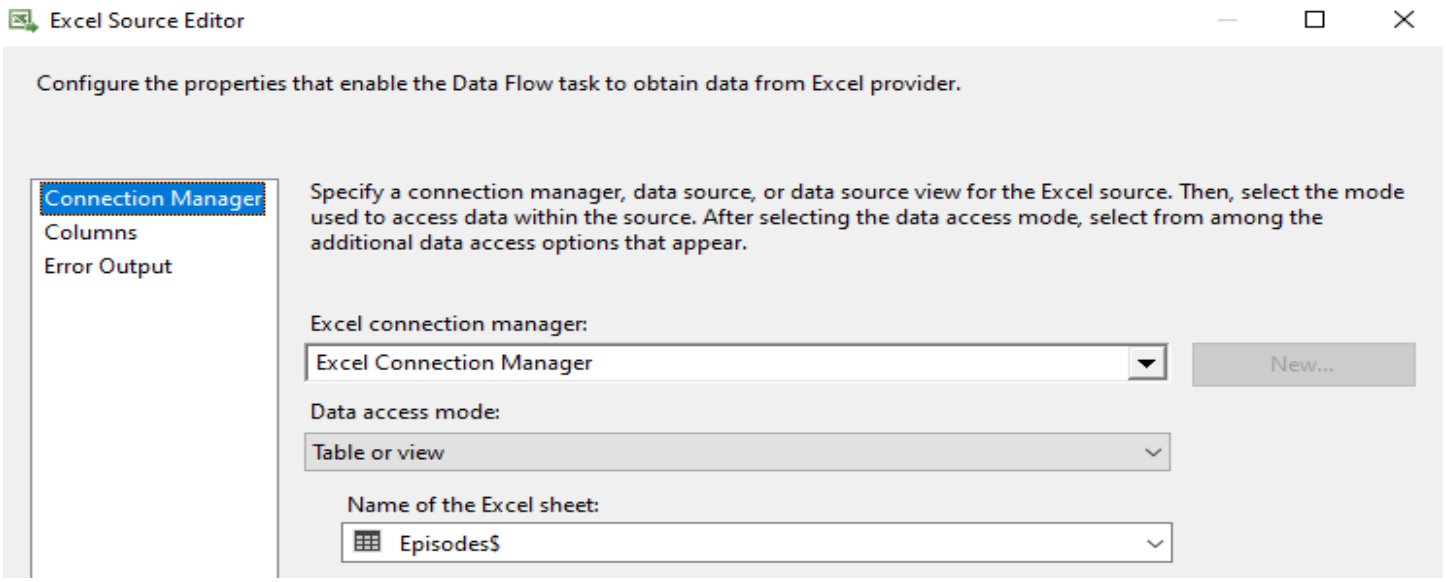

Derived Column 1
Insert Destination

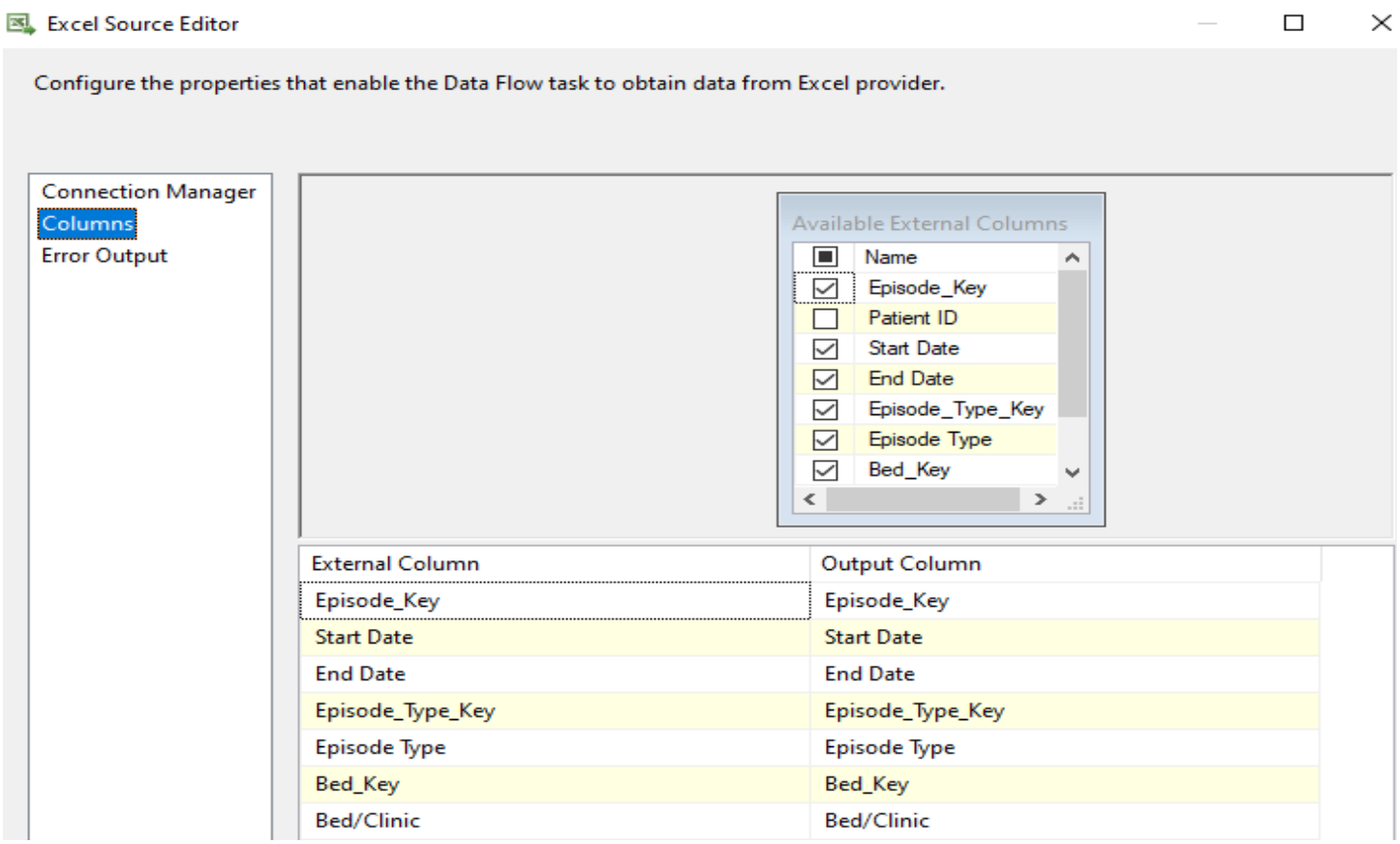
The Second step is to get data from excel file Services and match columns with the columns in the patient dimension in HealthCare Data Warehouse, the second step is to matching the columns in slowly changing dimension but we found that there is a problem some columns must be matched with them with data types, data types was different so I needed to convert them first and then make slowly changing dimension so in the step we get the data from excel file and used data conversion to convert the data types and then make slowly changing dimension.

When Execute the Data flows tasks for each Dimension, the Services Dimension



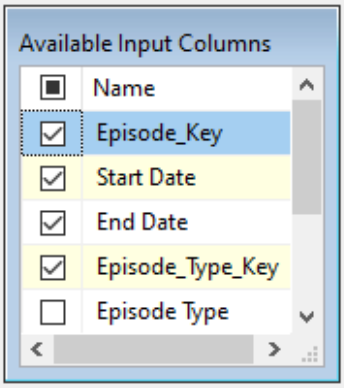
The Third step is to get data from excel file Episode and match columns with the columns in the patient dimension in HealthCare Data Warehouse, the second step is to matching the columns in slowly changing dimension but we found that there is a problem some columns must be matched with them with data types, data types was different so I needed to convert them first and then make slowly changing dimension so in the step we get the data from excel file and used data conversion to convert the data types and then make slowly changing dimension.





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.



Input Column	Output Alias	Data Type	Length	Precision	Scale	Code Page
Episode_Key	Con_Episode_Key	four-byte signed integer ...				
Start Date	Con_Start Date	database timestamp [DT_...				
End Date	Con_End Date	database timestamp [DT_...				
Episode_Type_Key	Con_Episode_Type_Key	four-byte signed integer ...				
Bed_Key	Con_Bed_Key	four-byte signed integer ...				

Slowly Changing Dimension Wizard

Select a Dimension Table and Keys

Select a dimension table to load and map columns in the transformation input to columns in the dimension table.

Connection manager:

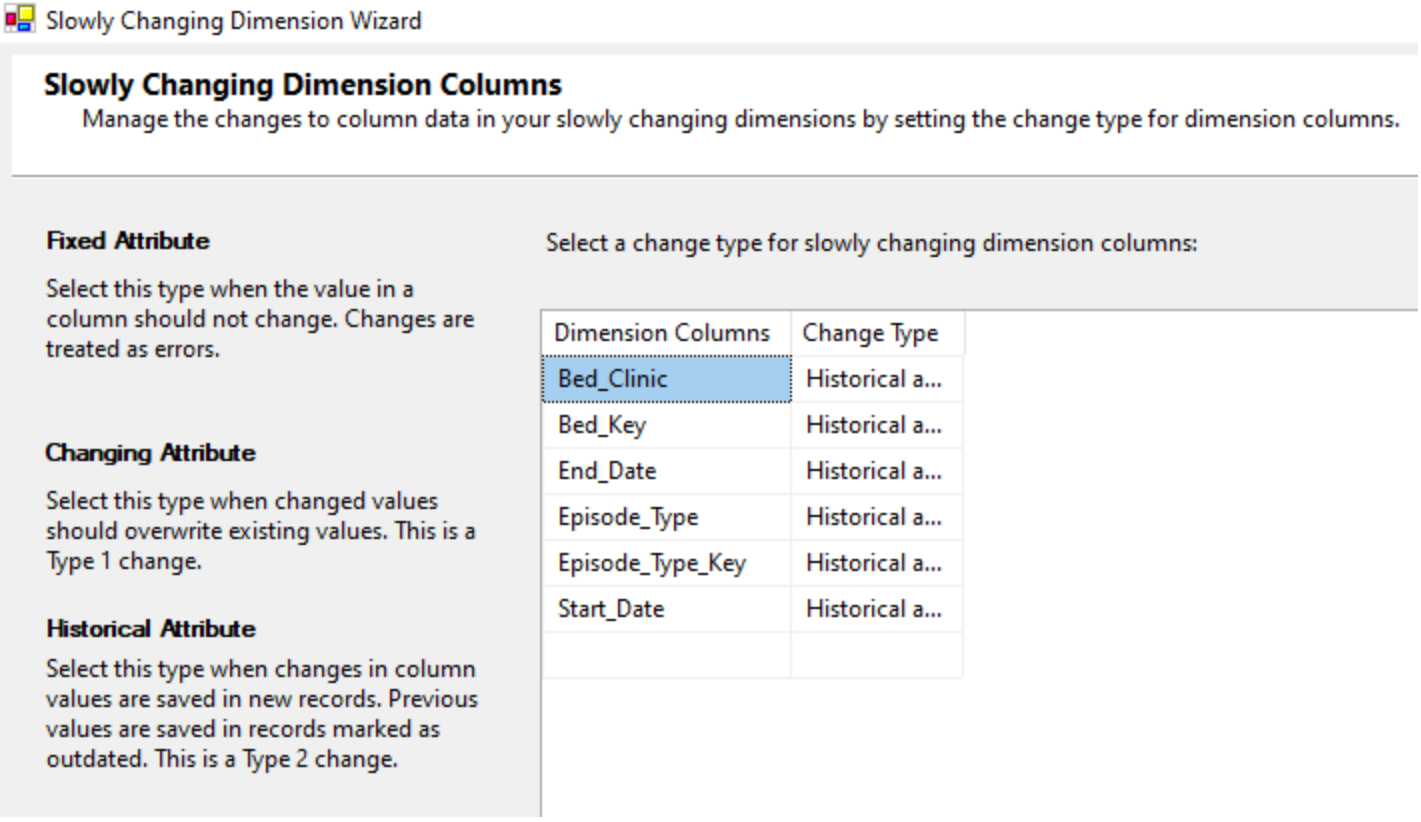
DESKTOP-6HREA33.HealthcareDW

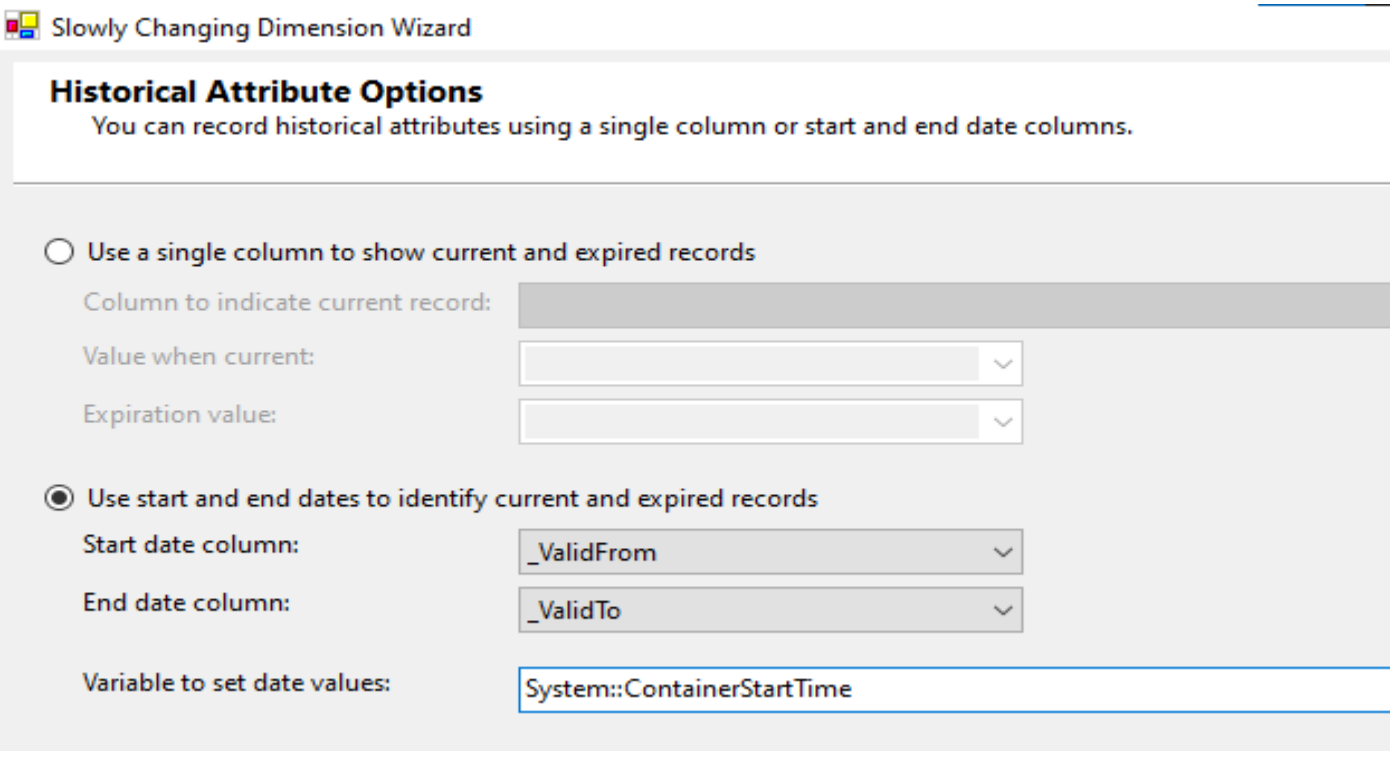
New...

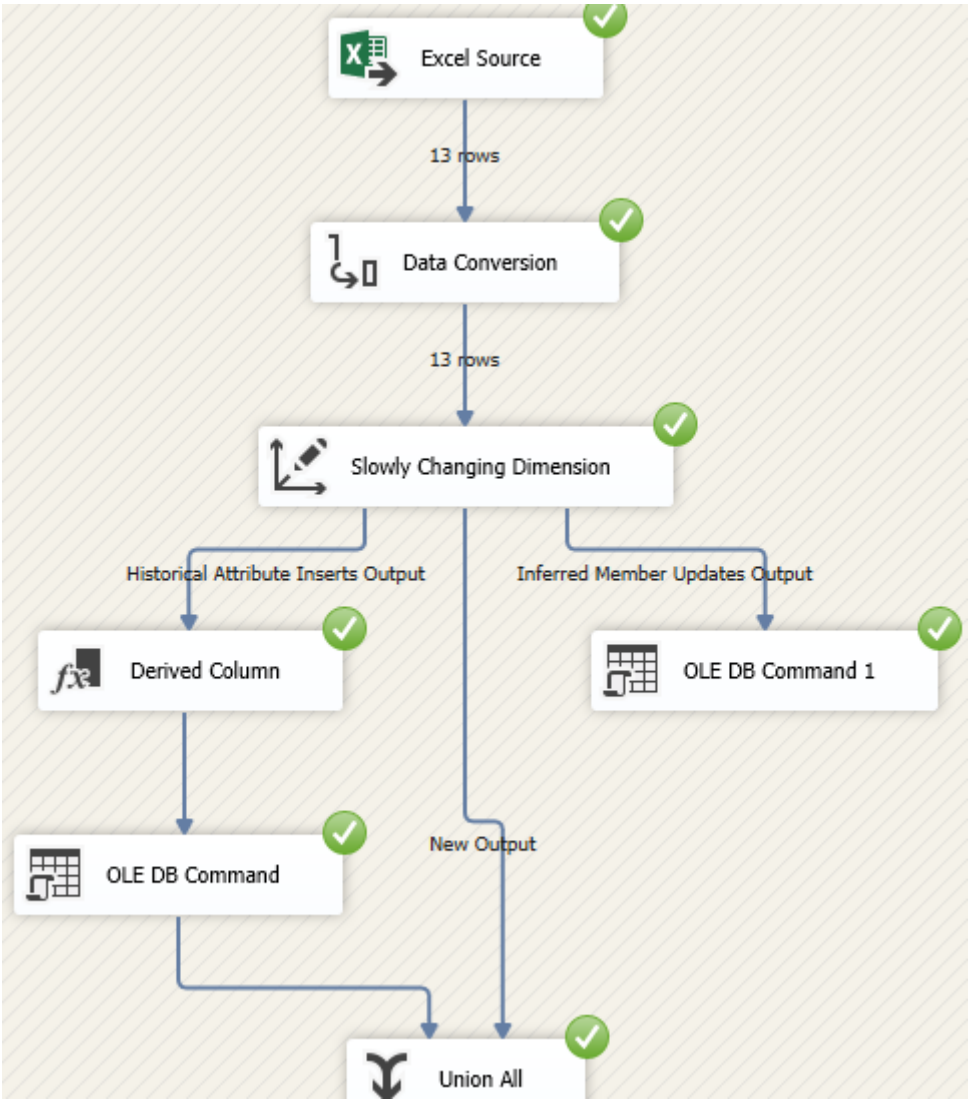
Table or view:

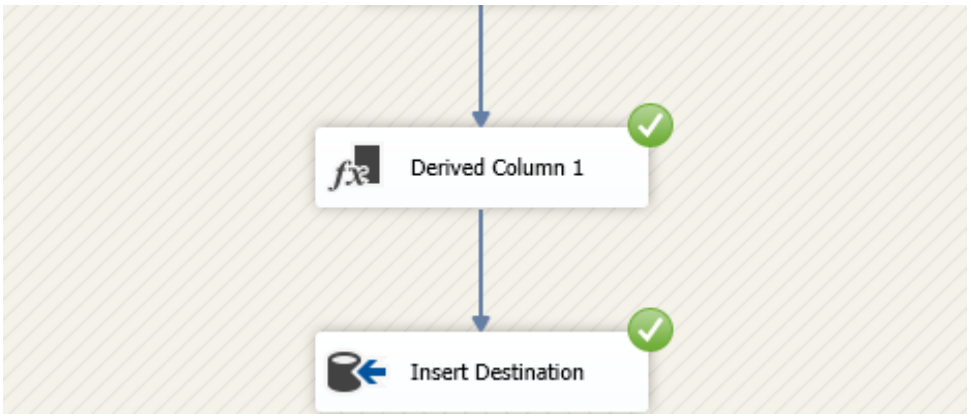
[dbo].[DimEpisodes]

Input Columns	Dimension Columns	Key Type
Con_Bed_Key	Bed_Key	Not a key column
Con_End Date	End_Date	Not a key column
Con_Episode...	Episode_Key_BK	Business key
Episode Type	Episode_Type	Not a key column
Con_Episode...	Episode_Type_Key	Not a key column
Con_Start Date	Start_Date	Not a key column


Slowly Changing Dimension Wizard
Slowly Changing Dimension Columns
Manage the changes to column data in your slowly changing dimensions by setting the change type for dimension columns.
Fixed Attribute
Select this type when the value in a column should not change. Changes are treated as errors.
Changing Attribute
Select this type when changed values should overwrite existing values. This is a Type 1 change.
Historical Attribute
Select this type when changes in column values are saved in new records. Previous values are saved in records marked as outdated. This is a Type 2 change.
Select a change type for slowly changing dimension columns:
Dimension Columns
Change Type
Bed_Clinic
Historical a...
Bed_Key
Historical a...
End_Date
Historical a...
Episode_Type
Historical a...
Episode_Type_Key
Historical a...
Start_Date
Historical a...


Slowly Changing Dimension Wizard
Historical Attribute Options
You can record historical attributes using a single column or start and end date columns.
Use a single column to show current and expired records
Column to indicate current record:
Value when current:
Expiration value:
Use start and end dates to identify current and expired records
Start date column:
_ValidFrom
End date column:
_ValidTo
Variable to set date values:
System::ContainerStartTime


Excel Source
13 rows
Data Conversion
13 rows
Slowly Changing Dimension
Historical Attribute Inserts Output
Inferred Member Updates Output
Derived Column
OLE DB Command 1
OLE DB Command
New Output
Union All


Derived Column 1
Insert Destination

Patients Table After Slowly Changing Dimension

```
      ,[English_Name]
      ,[PatType]
      ,[Patient_Type]
      ,[FileID]
      ,[mobile_phone]
      ,[district_key]
      ,[District_Name]
      ,[pat_nrst_relv]
      ,[pat_addr]
      ,[_ValidFrom]
      ,[_ValidTo]
  FROM [HealthcareDW].[dbo].[DimPatients]
```

133 %

Results | Messages

	patient_id_SK	patient_id_BK	Gov_Code	Governance	Gender_Key	gender	Nationality	Nationality_Code	Arabic_Name	English_Name	PatType	Pat
4	4	148259	1	ALEXANDRIA	2	Female	EGYPT	1	عايده فؤاد سليمان احمد	AYDA FOAD SELIMAN AHMED	1	عام
5	5	348617	1	ALEXANDRIA	1	Male	EGYPT	1	مختار عوض حسن اباظه	MOKHTAR AWAD HASSAN ABAZA	1	عام
6	6	361916	1	ALEXANDRIA	1	Male	EGYPT	1	يوسف مصطفى سعيد ندا	YOUSSEF MOSTAFA SAYED NADA	4	ذيد
7	7	363323	1	ALEXANDRIA	1	Male	EGYPT	1	محمد ايهاب شعبان احمد	MOHAMMED EHAB SHABAN AHMED	1	عام
8	8	442170	1	ALEXANDRIA	1	Male	EGYPT	1	ادهم كريم حسنى زيان	ADHAM KARIM HOSNY ZAYAN	1	عام
9	9	463333	1	ALEXANDRIA	1	Male	LB	123	محمد على حسين وفا	MOHAMMED ALI HUSSIN WAFA	1	عام
10	10	467593	1	ALEXANDRIA	2	Female	EGYPT	1	عايدة محمود حسونه ريان	AYDA MOHAMMED HASONAH RAYAN	1	عام
11	11	468546	1	ALEXANDRIA	2	Female	EGYPT	1	نواره عبدالله جمعه شحات	NAWARA ABDALLA GOMAA SHEHAT	1	عام
12	12	470006	1	ALEXANDRIA	1	Male	EGYPT	1	ابن انتصار السيد احمد	BABY ANTSAR EL SAYED AHMED	4	ذيد
13	13	471821	1	ALEXANDRIA	2	Female	EGYPT	1	فايزه احمد هويدى احمد	FAIZA Ahmed Hewdy Ahmed	1	عام

Services Table After Slowly Changing Dimension

```
SELECT TOP (1000) [Service_SK]
      ,[Category_Key_BK]
      ,[Category_Name]
      ,[Service_Key_BK]
      ,[Service_Name]
      ,[_ValidFrom]
      ,[_ValidTo]
  FROM [HealthcareDW].[dbo].[DimServices]
```

133 %

Results | Messages

	Service_SK	Category_Key_BK	Category_Name	Service_Key_BK	Service_Name	_ValidFrom	_ValidTo
288	288	12	Procedures	1000007748	CLEANSING ENEMA	2023-12-27 14:04:47.000	NULL
289	289	12	Procedures	1000007785	U/S GUIDED ASPIRATION OF BODY FLU...	2023-12-27 14:04:47.000	NULL
290	290	12	Procedures	1000008804	HOSPITAL RADIOLOGIST CALL FOR U/S...	2023-12-27 14:04:47.000	NULL
291	291	12	Procedures	1000009738	ER-CALL	2023-12-27 14:04:47.000	NULL
292	292	14	MEDICAL CARE	1000009982	MEDICAL CARE	2023-12-27 14:04:47.000	NULL
293	293	15	ESCORT ACCOM...	1000011092	ESCORT ACCOMMODATION	2023-12-27 14:04:47.000	NULL
294	294	28	NURSING CARE	1000011938	NURSING CARE	2023-12-27 14:04:47.000	NULL
295	295	90	DOWN PAYMENT	1000012460	DOWN PAYMENT	2023-12-27 14:04:47.000	NULL
296	296	99	DISCOUNT	1000013111	DISCOUNT	2023-12-27 14:04:47.000	NULL
297	297	7777	ADDITIONAL VAL...	1000013314	ADDITIONAL VALUES	2023-12-27 14:04:47.000	NULL
298	298	8888	DEBIT ACCOUNT	1000013353	DEBIT ACCOUNT	2023-12-27 14:04:47.000	NULL

Episode Table After Slowly Changing Dimension

```sql
SELECT TOP (1000) [Episode_Key_SK]
      ,[Episode_Key_BK]
      ,[Start_Date]
      ,[End_Date]
      ,[Episode_Type_Key]
      ,[Episode_Type]
      ,[Bed_Key]
      ,[Bed_Clinic]
      ,[_ValidFrom]
      ,[_ValidTo]
  FROM [HealthcareDW].[dbo].[DimEpisodes]
```

133 %

Results | Messages

	Episode_Key_SK	Episode_Key_BK	Start_Date	End_Date	Episode_Type_Key	Episode_Type	Bed_Key	Bed_Clinic	_ValidFrom
4	4	1063224	2011-11-25 00:00:00.000	2011-11-25 00:00:00.000	4	ER	10	EMERGENCY CLINIC	2023-12-27 14:04:48.000
5	5	1047628	2011-10-26 00:00:00.000	2011-10-30 00:00:00.000	1	Inpatient	4	BED 703 / 2	2023-12-27 14:04:48.000
6	6	1060235	2011-11-19 00:00:00.000	2011-11-22 00:00:00.000	1	Inpatient	9	BED 606/2	2023-12-27 14:04:48.000
7	7	1060543	2011-11-20 00:00:00.000	2011-11-21 00:00:00.000	4	ER	10	EMERGENCY CLINIC	2023-12-27 14:04:48.000
8	8	1043205	2011-10-18 00:00:00.000	2011-10-25 00:00:00.000	1	Inpatient	3	BED 201 -13(ICU)	2023-12-27 14:04:48.000
9	9	1058074	2011-11-15 00:00:00.000	2011-11-16 00:00:00.000	2	Outpatient	7	PAEDIATRIC (PROF) (New)	2023-12-27 14:04:48.000
10	10	1052323	2011-11-03 00:00:00.000	2011-11-04 00:00:00.000	1	Inpatient	6	WELL BABY BED (1)	2023-12-27 14:04:48.000
11	11	1051811	2011-11-02 00:00:00.000	2011-11-09 00:00:00.000	2	Outpatient	5	CARDIOLOGY (CONS) (New)	2023-12-27 14:04:48.000
12	12	1051036	2011-11-01 00:00:00.000	2011-11-04 00:00:00.000	3	Extrnal Servi...	10	BED 703/1	2023-12-27 14:04:48.000
13	13	1056691	2011-11-13 00:00:00.000	2011-11-16 00:00:00.000	3	Extrnal Servi...	11	WELL BABY BED	2023-12-27 14:04:48.000

5- Testing Update.

Will update the Service name for category key 1 and Category name ACCOMMODATION to HBS

And show What happens in the table after run slowly changing dimension once again

```
SELECT TOP (1000) [Service_SK]
      ,[Category_Key_BK]
      ,[Category_Name]
      ,[Service_Key_BK]
      ,[Service_Name]
      ,[_ValidFrom]
      ,[_ValidTo]
  FROM [HealthcareDW].[dbo].[DimServices]
  Where Category_Key_BK = 1 ;
```

133 %

Results | Messages

	Service_SK	Category_Key_BK	Category_Name	Service_Key_BK	Service_Name	_ValidFrom	_ValidTo
1	1	1	ACCOMMODATION	0	LDH	2023-12-27 14:04:47.000	2023-12-27 14:16:02.000
2	299	1	ACCOMMODATION	0	HBS	2023-12-27 14:16:02.000	NULL

Will update the Nationality for patient id 463333 and Nationality LB to Libya

And show What happens in the table after run slowly changing dimension once again

SQLQuery40.sql -...HREA33\ahos1 (55))*

```
      ,[Patient_Type]
      ,[FileID]
      ,[mobile_phone]
      ,[district_key]
      ,[District_Name]
      ,[pat_nrst_relv]
      ,[pat_addr]
      ,[_ValidFrom]
      ,[_ValidTo]
  FROM [HealthcareDW].[dbo].[DimPatients]
  WHERE patient_id_BK = 463333;
```

133 %

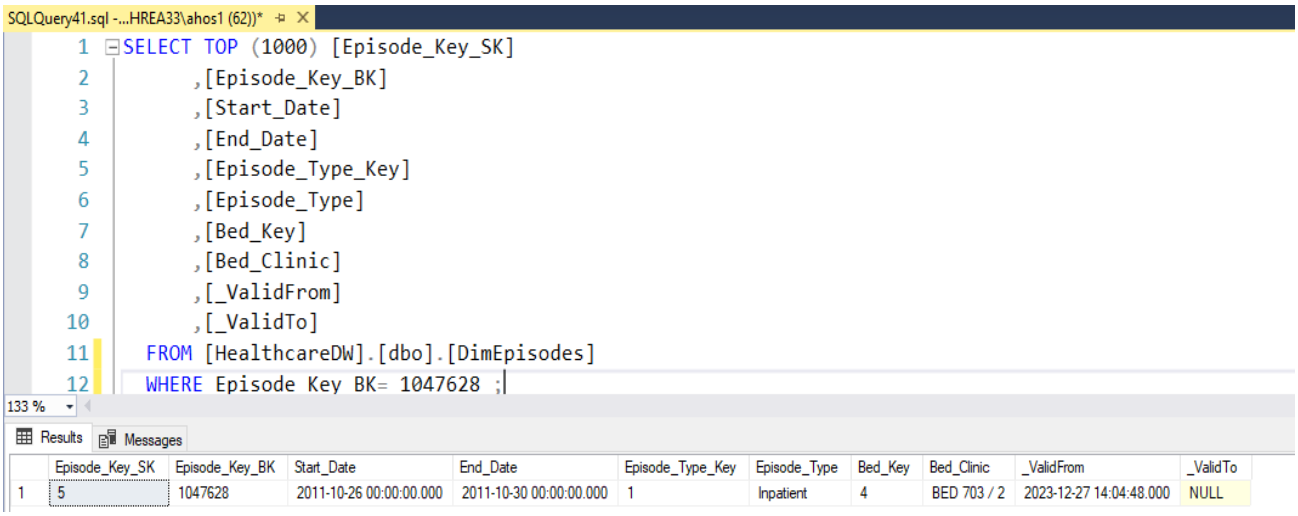
Results | Messages

	patient_id_SK	patient_id_BK	Gov_Code	Governance	Gender_Key	gender	Nationality	Nationality_Code	Arabic_Name	English_Name	PatType	Patient_Type	FileID
1	9	463333	1	ALEXANDRIA	1	Male	LB	123	محمد علي حسين وفا	MOHAMMED ALI HUSSIN WAFA	1	مريض عام	0
2	14	463333	1	ALEXANDRIA	1	Male	Libya	123	محمد علي حسين وفا	MOHAMMED ALI HUSSIN WAFA	1	مريض عام	0

Will update the Episode Type for Episode Key 1047628 and Episode Type inpatient to outpatient

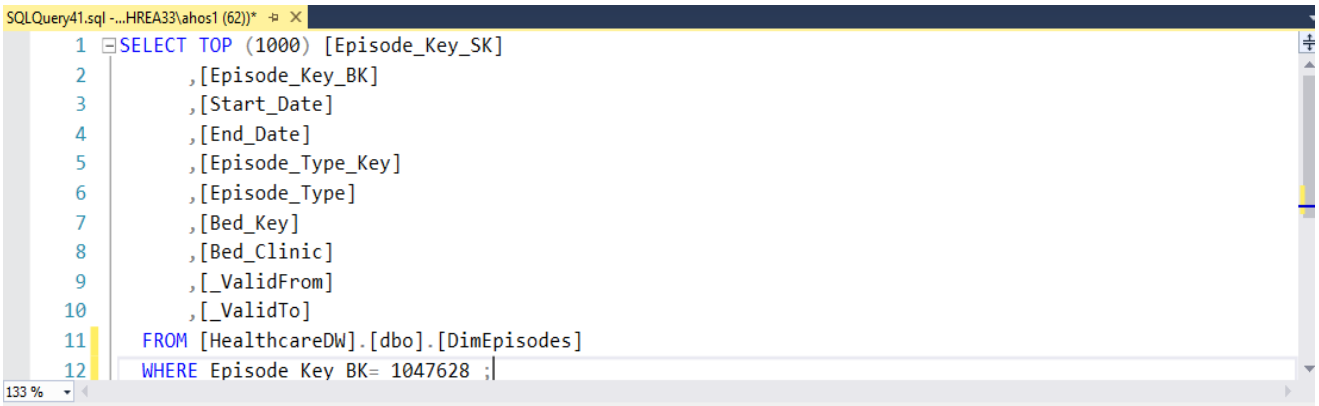
And show What happens in the table after run slowly changing dimension once again.

Before



	Episode_Key_SK	Episode_Key_BK	Start_Date	End_Date	Episode_Type_Key	Episode_Type	Bed_Key	Bed_Clinic	_ValidFrom	_ValidTo
1	5	1047628	2011-10-26 00:00:00.000	2011-10-30 00:00:00.000	1	Inpatient	4	BED 703 / 2	2023-12-27 14:04:48.000	NULL

After



	Episode_Key_SK	Episode_Key_BK	Start_Date	End_Date	Episode_Type_Key	Episode_Type	Bed_Key	Bed_Clinic	_ValidFrom	_ValidTo
1	5	1047628	2011-10-26 00:00:00.000	2011-10-30 00:00:00.000	1	Inpatient	4	BED 703 / 2	2023-12-27 14:04:48.000	2023-12-27 14:27:34.000
2	14	1047628	2011-10-26 00:00:00.000	2011-10-30 00:00:00.000	1	outpatient	4	BED 703 / 2	2023-12-27 14:27:34.000	NULL

For Reference :

You can find all codes in this GitHub link Repo :

https://github.com/ahmed-osama10/HealthCare-Data-Warehouse-