Project Information

Project Filename: **DatawarehouseProject_MariaRosarioSEBASTIAN_DSTI_A21.zip**

Github Repository: https://github.com/MSebastian2021/ETL_Group_Project_DSTI_A21 (private, link email to: emerick.duval@dsti.institute)

DB Scripts Folder: **ETL_Group_Project_DSTI_A21\db_scripts**

Operating System: **MS Windows 10 Home Edition, Version 10.0.19044 Build 19044**

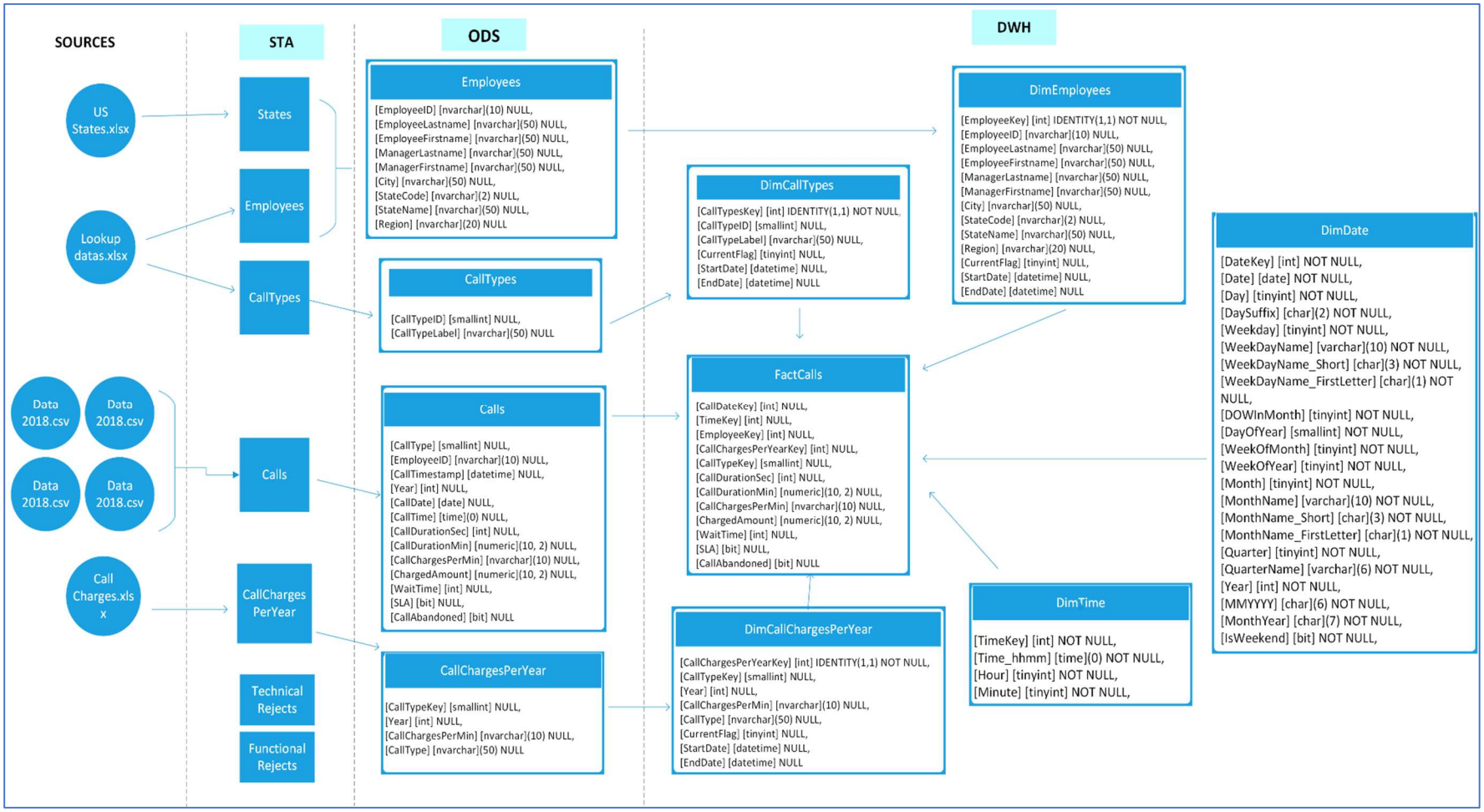
Visual Studio Specs:

Microsoft Visual Studio Community 2019
Version 16.11.15
VisualStudio.16.Release/16.11.15+32510.428
Microsoft .NET Framework
Version 4.8.04084

Database Specs:

SQL Server Management Studio	15.0.18390.0
SQL Server Management Objects (SMO)	16.100.46521.71
	15.0.19714.0
Microsoft Analysis Services Client Tools	10.0.19041.1
Microsoft Data Access Components	3.0 6.0
(MDAC)	4.0.30319.42000
Microsoft MSXML	10.0.19044
Microsoft .NET Framework	
Operating System	

Data Flow Diagram



Instructions to open and run the project

1. Unzip the file: **DatawarehouseProject_MariaRosarioSEBASTIAN_DSTI_A21.zip,** or clone the repository in github: **https://github.com/MSebastian2021/ETL_Group_Project_DSTI_A21** (private, link sent via email)
2. Run the DB Scripts in folder *db_scripts*.
3. Start with creating the *STA*, *ODS* and *DWH* Databases:
   - createdb_ssis_project_sta.sql
   - createdb_ssis_project_ods.sql
   - createdb_ssis_project_dwh.sql

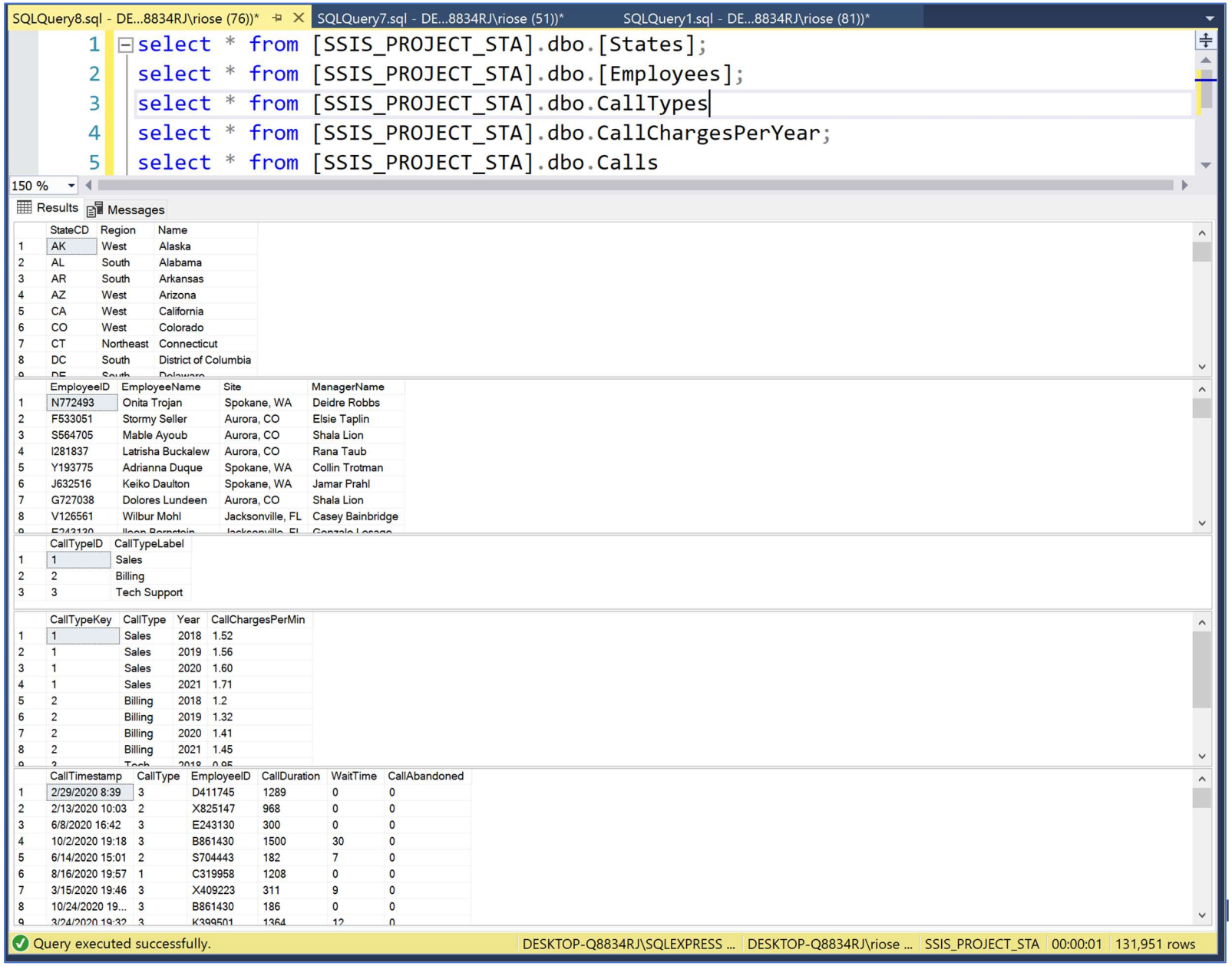
   and continue with the other scripts in *db_scripts* folder to create the tables.

4. Open the project, *ETL_Group_Project_DSTI_A21.sln* in Visual Studio.
5. Execute the STA packages to extract/retrieve data from the sources. If there is an issue extracting/opening .xlsx file, please see here for fix. You should be able to get the following number of rows for each table:

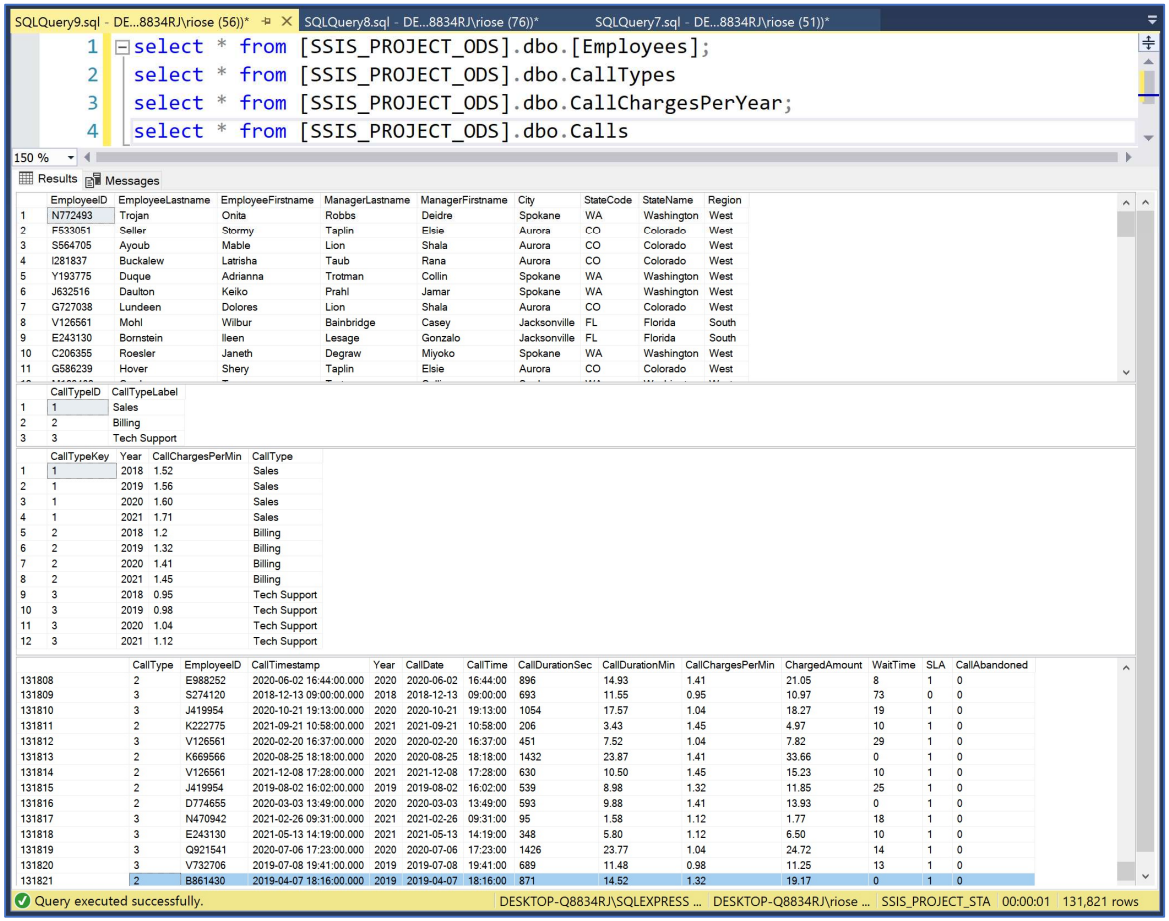
   - [SSIS_PROJECT_STA].dbo.[States] – 51 rows
   - [SSIS_PROJECT_STA].dbo.[CallTypes] – 3 rows
   - [SSIS_PROJECT_STA].dbo.[CallChargesPerYear] – 12 rows (NULL rows removed, and columns unpivoted)
   - [SSIS_PROJECT_STA].dbo.[Employees] – 64 rows
   - [SSIS_PROJECT_STA].dbo.[Calls] – 131,821 rows

   Running the following SELECT statement, should show the following results in below image:

```
select * from [SSIS_PROJECT_STA].dbo.[CallTypes];
select * from [SSIS_PROJECT_STA].dbo.[CallChargesPerYear];
select * from [SSIS_PROJECT_STA].dbo.[Employees];
select * from [SSIS_PROJECT_STA].dbo.[Calls]
```

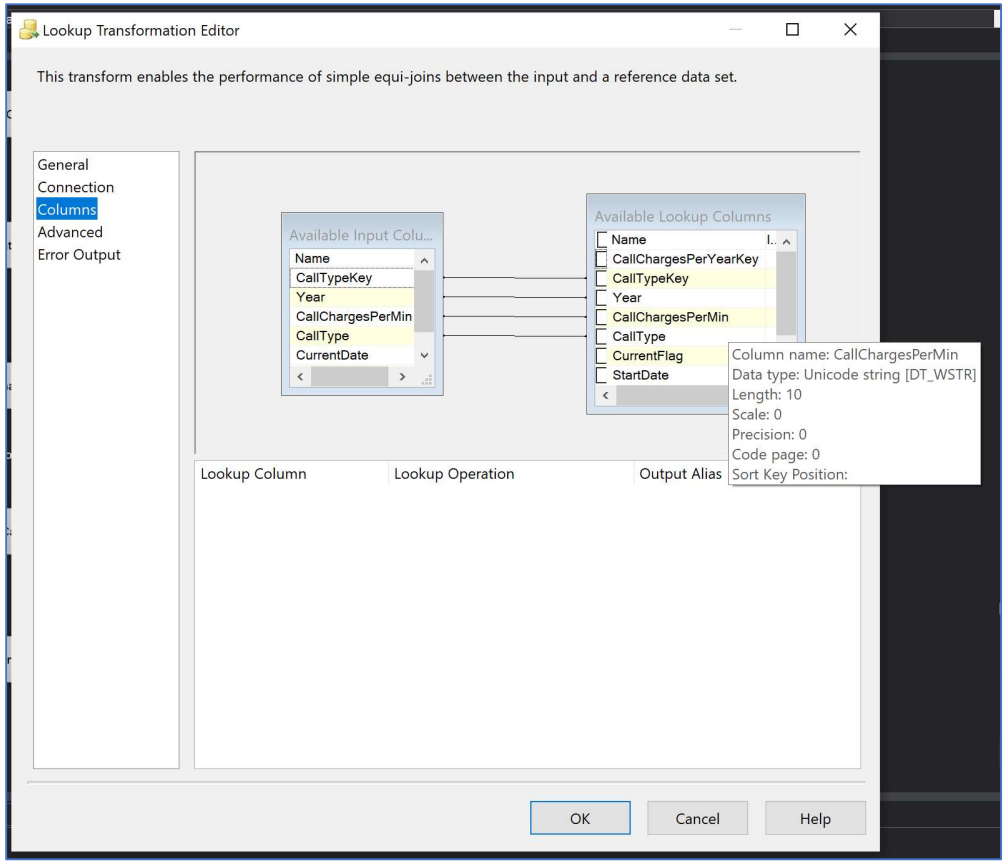

SQLQuery8.sql - DE...8834RJ\riose (76))*
SQLQuery7.sql - DE...8834RJ\riose (51))*
SQLQuery1.sql - DE...8834RJ\riose (81))*
1 select * from [SSIS_PROJECT_STA].dbo.[States];
2 select * from [SSIS_PROJECT_STA].dbo.[Employees];
3 select * from [SSIS_PROJECT_STA].dbo.CallTypes
4 select * from [SSIS_PROJECT_STA].dbo.CallChargesPerYear;
5 select * from [SSIS_PROJECT_STA].dbo.Calls
150 %
Results
Messages
StateCD Region Name
1 AK West Alaska
2 AL South Alabama
3 AR South Arkansas
4 AZ West Arizona
5 CA West California
6 CO West Colorado
7 CT Northeast Connecticut
8 DC South District of Columbia
EmployeeID EmployeeName Site ManagerName
1 N772493 Onita Trojan Spokane, WA Deidre Robbs
2 F533051 Stormy Seller Aurora, CO Elsie Taplin
3 S564705 Mable Ayoub Aurora, CO Shala Lion
4 I281837 Latrisha Buckalew Aurora, CO Rana Taub
5 Y193775 Adrianna Duque Spokane, WA Collin Trotman
6 J632516 Keiko Daulton Spokane, WA Jamar Prahl
7 G727038 Dolores Lundeen Aurora, CO Shala Lion
8 V126561 Wilbur Mohl Jacksonville, FL Casey Bainbridge
CallTypeID CallTypeLabel
1 1 Sales
2 2 Billing
3 3 Tech Support
CallTypeKey CallType Year CallChargesPerMin
1 1 Sales 2018 1.52
2 1 Sales 2019 1.56
3 1 Sales 2020 1.60
4 1 Sales 2021 1.71
5 2 Billing 2018 1.2
6 2 Billing 2019 1.32
7 2 Billing 2020 1.41
8 2 Billing 2021 1.45
CallTimestamp CallType EmployeeID CallDuration WaitTime CallAbandoned
1 2/29/2020 8:39 3 D411745 1289 0 0
2 2/13/2020 10:03 2 X825147 968 0 0
3 6/8/2020 16:42 3 E243130 300 0 0
4 10/2/2020 19:18 3 B861430 1500 30 0
5 6/14/2020 15:01 2 S704443 182 7 0
6 8/16/2020 19:57 1 C319958 1208 0 0
7 3/15/2020 19:46 3 X409223 311 9 0
8 10/24/2020 19... 3 B861430 186 0 0
Query executed successfully.
DESKTOP-Q8834RJ\SQLEXPRESS ...
DESKTOP-Q8834RJ\riose ...
SSIS_PROJECT_STA
00:00:01
131,951 rows

6. Execute the ODS packages to transform and transfer the data from STA to ODS. Running the same SELECT statements in *Step 5* from `[SSIS_PROJECT_ODS]` database this time, should have the same number of row results in all tables. *The **States** Table was not included in the ODS Database. Its data has been included in the Employees table instead. EmployeeName was split into EmployeeFirstname and EmployeeLastname, same with ManagerName.*



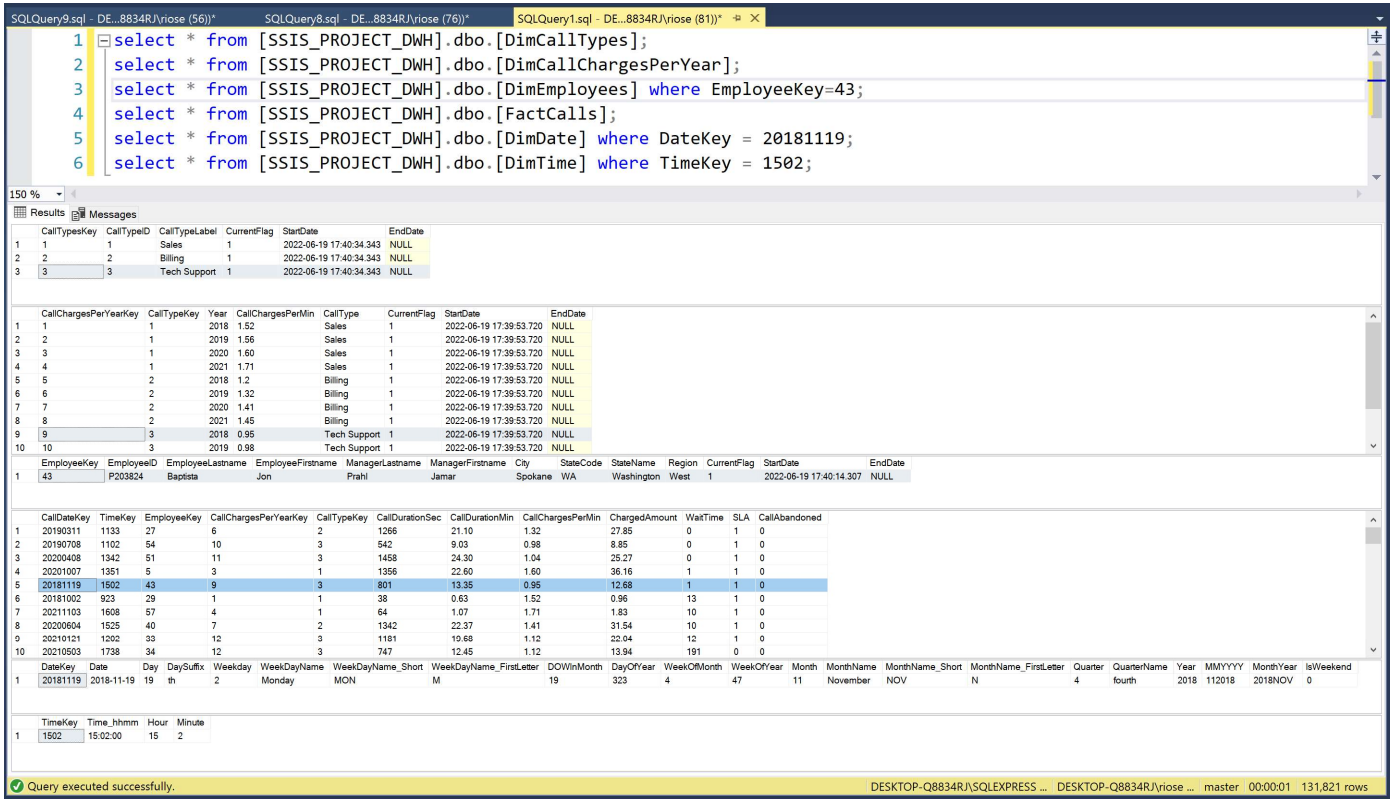
7. Execute the DWH Dimension packages to move data from ODS to DWH database.

   The *CallChargesPerMin* column across all tables was kept as *nvarchar(10)* and not converted to *float* datatype due to issues on lookup field comparison. It seems like float comparisons between fields are not allowed, as seen below in lookup transformation of DWH_DimCallChargesPerYear.



8. Execute the DWH_FactCalls.dstx to populate the data in the DWH Fact Table. The following SELECT statements, yield results in below image:

```
select * from [SSIS_PROJECT_DWH].dbo.[DimCallTypes];
select * from [SSIS_PROJECT_DWH].dbo.[DimCallChargesPerYear];
select * from [SSIS_PROJECT_DWH].dbo.[DimEmployees] where EmployeeKey=43;
select * from [SSIS_PROJECT_DWH].dbo.[FactCalls];
select * from [SSIS_PROJECT_DWH].dbo.[DimDate] where DateKey = 20181119;
select * from [SSIS_PROJECT_DWH].dbo.[DimTime] where TimeKey = 1502;
```



Resources

Fix to resolve issue on extracting Excel files with .xlsx on Visual Studio 2019

Time Dimension table inspiration