

#### Project Information

Project Filename: **DatawarehouseProject\_MariaRosarioSEBASTIAN\_DSTI\_A21.zip**

**Github Repository:** [https://github.com/MSebastian2021/ETL\\_Group\\_Project\\_DSTI\\_A21](https://github.com/MSebastian2021/ETL_Group_Project_DSTI_A21) (private, link email to: [emerick.duval@dsti.institute](mailto:emerick.duval@dsti.institute))

DB Scripts Folder: **ETL\_Group\_Project\_DSTI\_A21\db\_scripts**

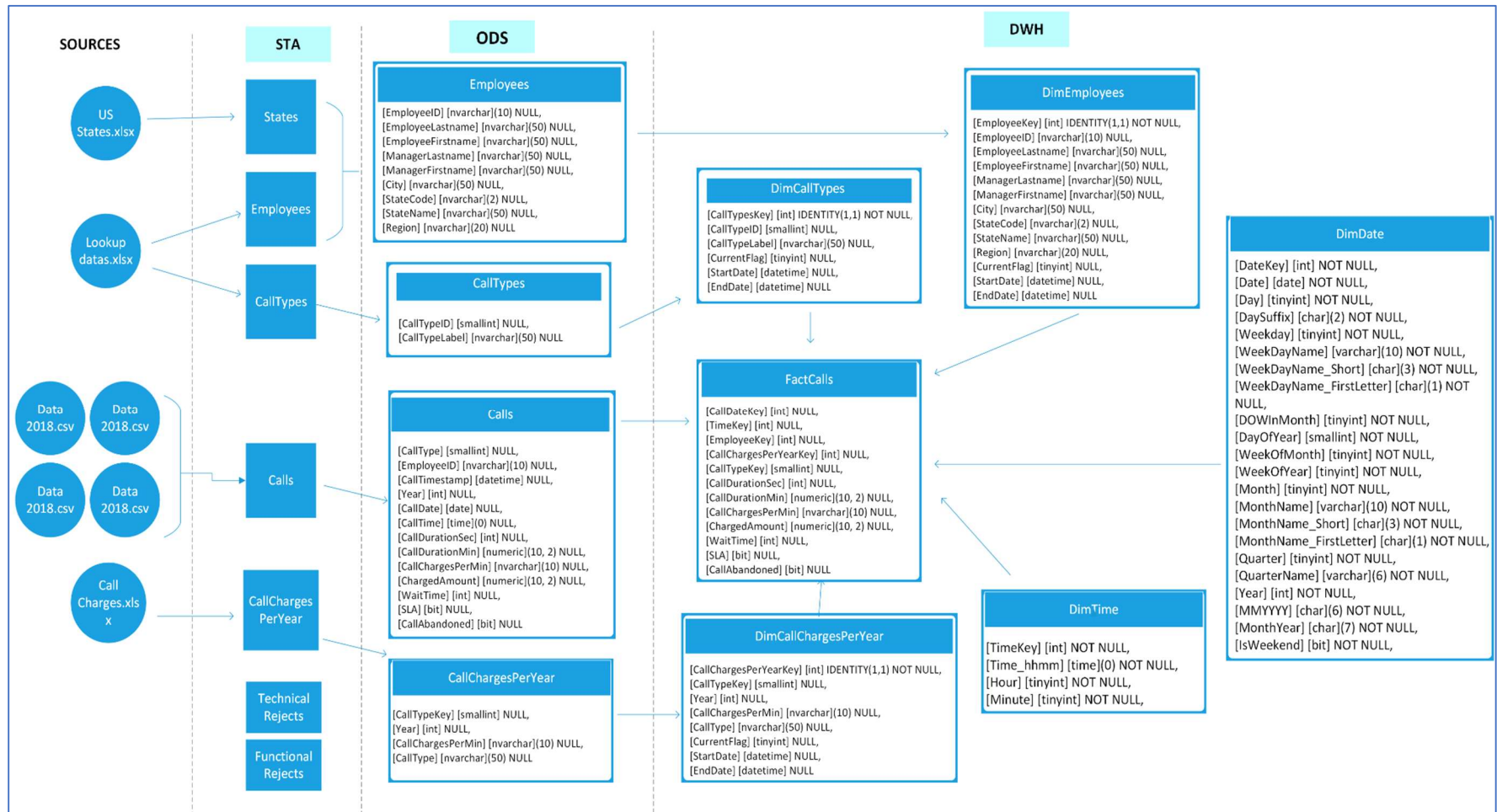
Operating System: **MS Windows 10 Home Edition, Version 10.0.19044 Build 19044**

Visual Studio Specs:

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

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: **DatwarehouseProject\_MariaRosarioSEBASTIAN\_DSTI\_A21.zip**, or clone the repository in github: [https://github.com/MSebastian2021/ETL\\_Group\\_Project\\_DSTI\\_A21](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.sqland 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))\* X 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
9	DE	South	Delaware

	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	Latisha Buckalew	Aurora, CO	Rana Taub
5	Y193775	Adrianna Duque	Spokane, WA	Collin Trotman
6	J632516	Keiko Daulton	Spokane, WA	Jamar Prael
7	G727038	Dolores Lundeen	Aurora, CO	Shala Lion
8	V126561	Wilbur Mohl	Jacksonville, FL	Casey Bainbridge
9	E242120	Ilona Bernstein	Jacksonville, FL	Conzole Legee

	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
9	2	Billing	2019	0.95

	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
9	3/24/2020 19:32	3	K399501	1364	12	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.*

The screenshot displays the SQL Server Enterprise Manager interface with three SQL queries executed in the [SSIS\_PROJECT\_ODS] database. The results are shown in three separate panes.

**Query 1: Employees**

```
1 select * from [SSIS_PROJECT_ODS].dbo.[Employees];
```

EmployeeID	EmployeeLastname	EmployeeFirstname	ManagerLastname	ManagerFirstname	City	StateCode	StateName	Region
1	N772493	Trojan	Robbs	Deidre	Spokane	WA	Washington	West
2	F633051	Saier	Stewmy	Elise	Aurora	CO	Colorado	West
3	S564705	Ayoub	Mable	Lion	Aurora	CO	Colorado	West
4	I281837	Buckalew	Latrisha	Taub	Aurora	CO	Colorado	West
5	Y193775	Duque	Adrianna	Trotman	Spokane	WA	Washington	West
6	J632516	Daulton	Keiko	Prahl	Spokane	WA	Washington	West
7	G727038	Lundeen	Dolores	Lion	Aurora	CO	Colorado	West
8	V126561	Mohl	Wilbur	Bainbridge	Jacksonville	FL	Florida	South
9	E243130	Bornstein	Ileen	Lesage	Jacksonville	FL	Florida	South
10	C206355	Roesler	Janeth	Degraw	Spokane	WA	Washington	West
11	G586239	Hover	Shery	Taplin	Aurora	CO	Colorado	West

**Query 2: CallTypes**

```
2 select * from [SSIS_PROJECT_ODS].dbo.CallTypes
```

CallTypeID	CallTypeLabel
1	Sales
2	Billing
3	Tech Support

**Query 3: CallChargesPerYear**

```
3 select * from [SSIS_PROJECT_ODS].dbo.CallChargesPerYear;
```

CallTypeKey	Year	CallChargesPerMin	CallType
1	2018	1.52	Sales
2	2019	1.56	Sales
3	2020	1.60	Sales
4	2021	1.71	Sales
5	2018	1.2	Billing
6	2019	1.32	Billing
7	2020	1.41	Billing
8	2021	1.45	Billing
9	2018	0.95	Tech Support
10	2019	0.98	Tech Support
11	2020	1.04	Tech Support
12	2021	1.12	Tech Support

**Query 4: Calls**

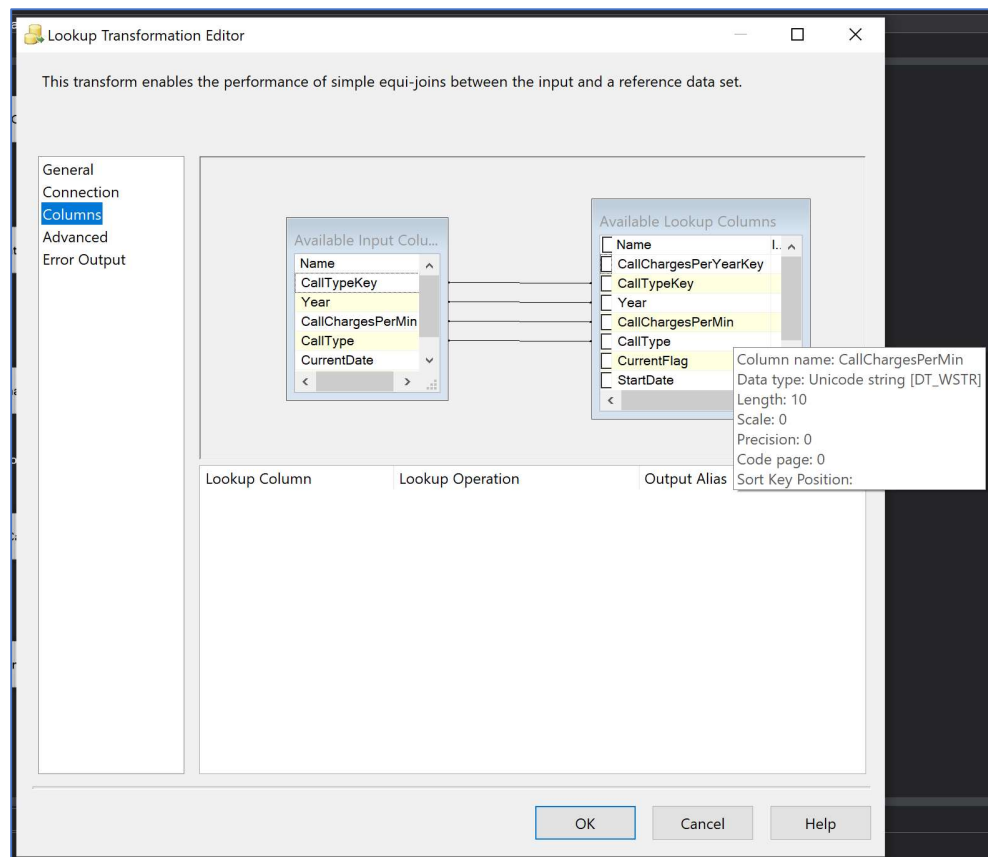
```
4 select * from [SSIS_PROJECT_ODS].dbo.Calls
```

CallType	EmployeeID	CallTimestamp	Year	CallDate	CallTime	CallDurationSec	CallDurationMin	CallChargesPerMin	ChargedAmount	WaitTime	SLA	CallAbandoned	
131808	2	E996252	2020-06-02 16:44:00.000	2020	2020-06-02	16:44:00	896	14.93	1.41	21.05	8	1	0
131809	3	S274120	2018-12-13 09:00:00.000	2018	2018-12-13	09:00:00	693	11.55	0.95	10.97	73	0	0
131810	3	J419954	2020-10-21 09:13:00.000	2020	2020-10-21	19:13:00	1054	17.57	1.04	18.27	19	1	0
131811	2	K222775	2021-09-21 10:58:00.000	2021	2021-09-21	10:58:00	206	3.43	1.45	4.97	10	1	0
131812	3	V126561	2020-02-20 16:37:00.000	2020	2020-02-20	16:37:00	451	7.52	1.04	7.82	29	1	0
131813	2	K669566	2020-08-25 18:18:00.000	2020	2020-08-25	18:18:00	1432	23.87	1.41	33.66	0	1	0
131814	2	V126561	2021-12-08 17:28:00.000	2021	2021-12-08	17:28:00	630	10.50	1.45	15.23	10	1	0
131815	2	J419954	2019-08-02 16:02:00.000	2019	2019-08-02	16:02:00	539	8.98	1.32	11.85	25	1	0
131816	2	D774655	2020-03-03 13:49:00.000	2020	2020-03-03	13:49:00	593	9.88	1.41	13.93	0	1	0
131817	3	N470942	2021-02-26 09:31:00.000	2021	2021-02-26	09:31:00	95	1.58	1.12	1.77	18	1	0
131818	3	E243130	2021-05-13 14:19:00.000	2021	2021-05-13	14:19:00	348	5.80	1.12	6.50	10	1	0
131819	3	Q921541	2020-07-06 17:23:00.000	2020	2020-07-06	17:23:00	1426	23.77	1.04	24.72	14	1	0
131820	3	V732706	2019-07-08 19:41:00.000	2019	2019-07-08	19:41:00	689	11.48	0.98	11.25	13	1	0
131821	2	B861430	2019-04-07 18:16:00.000	2019	2019-04-07	18:16:00	871	14.52	1.32	19.17	0	1	0

Query executed successfully. DESKTOP-Q8834RJ\SQLEXPRESS ... DESKTOP-Q8834RJ\riose ... SSIS\_PROJECT\_STA 00:00:01 131,821 rows

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.dtsx 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;
```

SQLQuery9.sql - DE...8834R\riose (56)\*    SQLQuery8.sql - DE...8834R\riose (76)\*    SQLQuery1.sql - DE...8834R\riose (81)\*

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

Results

CallTypesKey	CallTypeID	CallTypeLabel	CurrentFlag	StartDate	EndDate
1	1	Sales	1	2022-06-19 17:40:34.343	NULL
2	2	Billing	1	2022-06-19 17:40:34.343	NULL
3	3	Tech Support	1	2022-06-19 17:40:34.343	NULL

CallChargesPerYearKey	CallTypeKey	Year	CallChargesPerMin	CallType	CurrentFlag	StartDate	EndDate
1	1	2018	1.52	Sales	1	2022-06-19 17:39:53.720	NULL
2	1	2019	1.56	Sales	1	2022-06-19 17:39:53.720	NULL
3	1	2020	1.50	Sales	1	2022-06-19 17:39:53.720	NULL
4	1	2021	1.71	Sales	1	2022-06-19 17:39:53.720	NULL
5	2	2018	1.2	Billing	1	2022-06-19 17:39:53.720	NULL
6	2	2019	1.32	Billing	1	2022-06-19 17:39:53.720	NULL
7	2	2020	1.41	Billing	1	2022-06-19 17:39:53.720	NULL
8	2	2021	1.45	Billing	1	2022-06-19 17:39:53.720	NULL
9	3	2018	0.95	Tech Support	1	2022-06-19 17:39:53.720	NULL
10	3	2019	0.98	Tech Support	1	2022-06-19 17:39:53.720	NULL

EmployeeKey	EmployeeID	EmployeeLastName	EmployeeFirstName	ManagerLastName	ManagerFirstName	City	StateCode	StateName	Region	CurrentFlag	StartDate	EndDate
43	P203824	Baptista	Jon	Prahl	Jamar	Spokane	WA	Washington	West	1	2022-06-19 17:40:14.307	NULL

CallDateKey	TimeKey	EmployeeKey	CallChargesPerYearKey	CallTypeKey	CallDurationSec	CallDurationMin	CallChargesPerMin	ChargedAmount	WaitTime	SLA	CallAbandoned
20190311	1133	27	6	2	1266	21.10	1.32	27.85	0	1	0
20190708	1102	54	10	3	542	9.03	0.98	8.85	0	1	0
20200408	1342	51	11	3	1458	24.30	1.04	25.27	0	1	0
20201007	1351	5	3	1	1356	22.60	1.60	36.16	1	1	0
20181119	1502	43	9	3	801	13.35	0.95	12.68	1	1	0
20181002	923	29	1	1	38	0.63	1.52	0.96	13	1	0
20211103	1608	57	4	1	64	1.07	1.71	1.83	10	1	0
20200604	1525	40	7	2	1342	22.37	1.41	31.54	10	1	0
20201011	1202	33	12	3	1181	19.68	1.12	22.04	12	1	0
20210503	1738	34	12	3	747	12.45	1.12	13.94	191	0	0

DateKey	Date	Day	DaySuffix	Weekday	WeekDayName	WeekDayName_Short	WeekDayName_FirstLetter	DOWInMonth	DayOfYear	WeekOfMonth	WeekOfYear	Month	MonthName	MonthName_Short	MonthName_FirstLetter	Quarter	QuarterName	Year	MMYYYY	MonthYear	IsWeekend
20181119	2018-11-19	19	th	2	Monday	MON	M	19	323	4	47	11	November	NOV	N	4	fourth	2018	112018	2018NOV	0

TimeKey	Time_hhmm	Hour	Minute
1502	15:02:00	15	2

Query executed successfully.    DESKTOP-Q8834R\SQLXPRESS ...    DESKTOP-Q8834R\riose ...    master    00:00:01    131,821 rows

## Resources

[Fix](#) to resolve issue on extracting Excel files with .xlsx on Visual Studio 2019

[Time Dimension table inspiration](#)