

# Documentation

## Scripts information :

### 1 Create \_Tables.sql :

- This script will be loaded first in case of initial data warehouse load. If the data warehouse table is already up and running then this script is optional. It defines the schema of final load table and final log table. For now, log table saves only error logs but script can be modified to store all the logs if necessary.

### 2 Email\_Configuration.sql :

- Send email task is one of the most essential task in any ETL process. It is an Email functionality configuration script that needs to be run prior to applying Email logging functionality either in SSIS or T-SQL.

### 3 Main.sql (Stored proc):

- Since the requirement was to use only SQL server and for the purpose of applying automation later on, All the data is bulk inserted into temporary tables keeping source files untouched. Tables will be terminated once the ETL load is finished. I have added Month and Day column for the purpose of creating composite key. But they can be hidden in the production environment.

- Tables for Financial\_D and Compensation files are first loaded into temp tables and Here necessary data validations are applied. In case of wrong file or corrupt file session will be terminated and error logs will be saved for monitoring purpose. In case of successful load they will be joined into final staging table. Here, I am using temp tables but we can use physical Staging tables based on Kimball's standard data warehousing approach once the dataset gets larger.

- Final Temp table is loaded into Final\_Table by applying data cleaning, transformation and aggregation. Since most of the records were numeric values, Transformation was easier. NULL values in the numeric columns are kept unchanged because data imputation might change the business logic. (For inst. 0 \$ in place of NULL). Any data type anomaly will trigger Email log and Table log.

*“Approach of FTP trigger can be applied by using Powershell script to fully automate the ETL. Once the new file arrives in the directory , Powershell will trigger the execution of Stored procedure. Powershell script can be automated using windows even scheduler or SSIS”.*

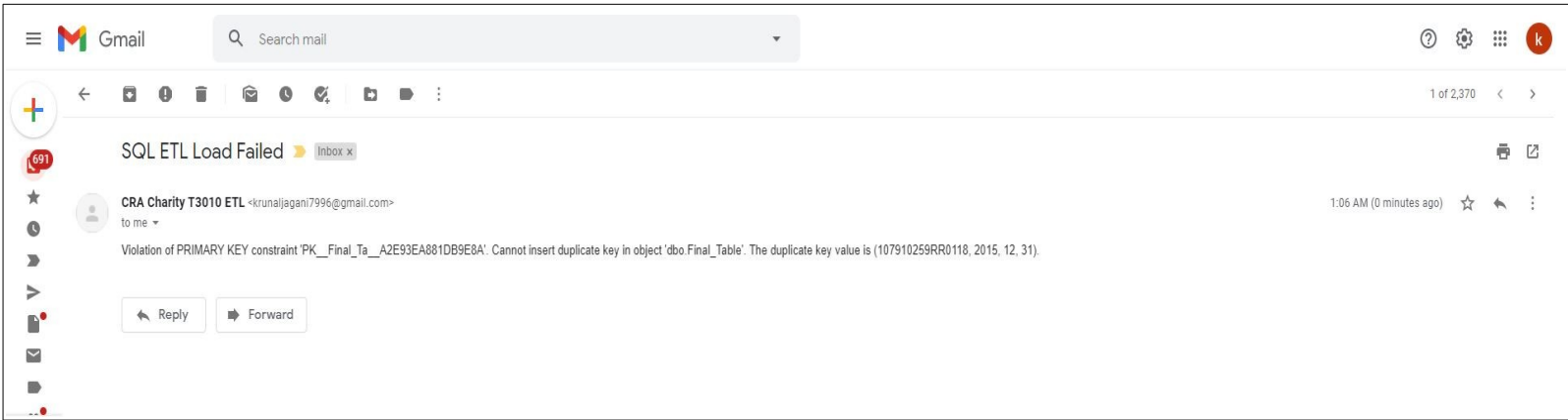
### 4 Testing\_DataCompleteness.sql (Stored proc):

- Automated stored procedure to check if Source and final tables have same number of records. This test is applied to check if all the values have been passing through the Data Processing pipeline. It is a dynamic procedure where tester can dynamically select year and source file that needs to be verified.

### 5: Testing\_DataCorrectness.sql (Stored Proc) :

- Automated stored procedure to check if Source and final tables have same number of records. Source and destination have different records with different names (TOTAL\_REVENUE vs “4700”) and datatypes (Money vs varchar). So, the data is compared by applying in-function transformation at run time. It is also a dynamic procedure where tester can select year, Attribute and source file at runtime to check if both source and destination have same aggregated amount in the given field.

Snippet 1 : (Email Log)



Snippet 2 : (Table log)

ErrorID	UserName	ErrorNumber	ErrorState	ErrorSeverity	ErrorLine	ErrorProcedure	ErrorMessage	ErrorDateTime
1	Microsoft Account\krunaljagani7996@gmail.com	235	0	16	1	NULL	Cannot convert a char value to money. The char val...	2020-12-13 13:23:19.877
2	Microsoft Account\krunaljagani7996@gmail.com	2627	1	14	1	NULL	Violation of PRIMARY KEY constraint 'PK_Final_T...	2020-12-13 13:46:41.207
3	Microsoft Account\krunaljagani7996@gmail.com	102	1	15	1	NULL	Incorrect syntax near 'Registration'.	2020-12-13 20:27:24.390
4	Microsoft Account\krunaljagani7996@gmail.com	208	1	16	1	NULL	Invalid object name 'final_table'.	2020-12-13 20:55:58.343
5	Microsoft Account\krunaljagani7996@gmail.com	208	1	16	1	NULL	Invalid object name 'final_table'.	2020-12-13 20:59:04.450
6	Microsoft Account\krunaljagani7996@gmail.com	8114	5	16	1	NULL	Error converting data type varchar to float.	2020-12-13 20:59:36.030
7	Microsoft Account\krunaljagani7996@gmail.com	8114	5	16	1	NULL	Error converting data type varchar to float.	2020-12-13 21:01:49.370
8	Microsoft Account\krunaljagani7996@gmail.com	2627	1	14	1	NULL	Violation of PRIMARY KEY constraint 'PK_Final_T...	2020-12-14 00:47:21.230

Query executed successfully.

LAPTOP-6HIVF9TH\SQLEXPRESS ... LAPTOP-6HIVF9TH\kruna ... Practice\_SCD 00:00:00 8 rows

### Snippet 3 : (Testing\_DataCompleteness)

```
121
122 -----
123
124 -- Execution of procedure.
125 -- Format : "EXEC sp_name year,Filepath".
126 EXEC Compare_YearWise_RowCount 2015, 'c:\users\kruna\onedrive\desktop\ch\data\2015\1.csv';
127
128 EXEC Compare_YearWise_RowCount 2016, 'c:\users\kruna\onedrive\desktop\ch\data\2016\1.csv';
129
130 EXEC Compare_YearWise_RowCount 2017, 'c:\users\kruna\onedrive\desktop\ch\data\2017\1.csv';
131
132 EXEC Compare_YearWise_RowCount 2018, 'c:\users\kruna\onedrive\desktop\ch\data\2018\1.csv';
```

100 %

Messages

```
(85276 rows affected)
2015
CHECK SUCCESS : 85276 == 85276

(84909 rows affected)
2016
CHECK SUCCESS : 84909 == 84909

(84418 rows affected)
2017
CHECK SUCCESS : 84418 == 84418

(36378 rows affected)
2018
CHECK SUCCESS : 36378 == 36378

Completion time: 2020-12-14T01:11:25.7536149-05:00
```

100 %

### Snippet 4 : (Testing\_DataCompleteness)

```
162
163 -----
164 -- Execution of procedure.
165 -- Format : "EXEC sp_name year,Attribute,Finaicial_D Filepath"
166 EXEC Compare_Data_correctness 2015,["5050"],'c:\users\kruna\onedrive\desktop\ch\data\2015\1.csv';
167 EXEC Compare_Data_correctness 2016,["5010"],'c:\users\kruna\onedrive\desktop\ch\data\2016\1.csv';
168 EXEC Compare_Data_correctness 2017,["5020"],'c:\users\kruna\onedrive\desktop\ch\data\2017\1.csv';
169 EXEC Compare_Data_correctness 2018,["5040"],'c:\users\kruna\onedrive\desktop\ch\data\2018\1.csv';
170
171
```

100 %

Messages

```
(85276 rows affected)

(85276 rows affected)
Warning: Null value is eliminated by an aggregate or other SET operation.
Warning: Null value is eliminated by an aggregate or other SET operation.

(1 row affected)
2015
["5050"]
Gift
CHECK SUCCESS : 7799538173.00 == 7799538173.00

(84909 rows affected)

(84909 rows affected)
Warning: Null value is eliminated by an aggregate or other SET operation.
Warning: Null value is eliminated by an aggregate or other SET operation.

(1 row affected)
2016
["5010"]
EMA
CHECK SUCCESS : 21020128267.00 == 21020128267.00

(84418 rows affected)

(84418 rows affected)
```

100 %

