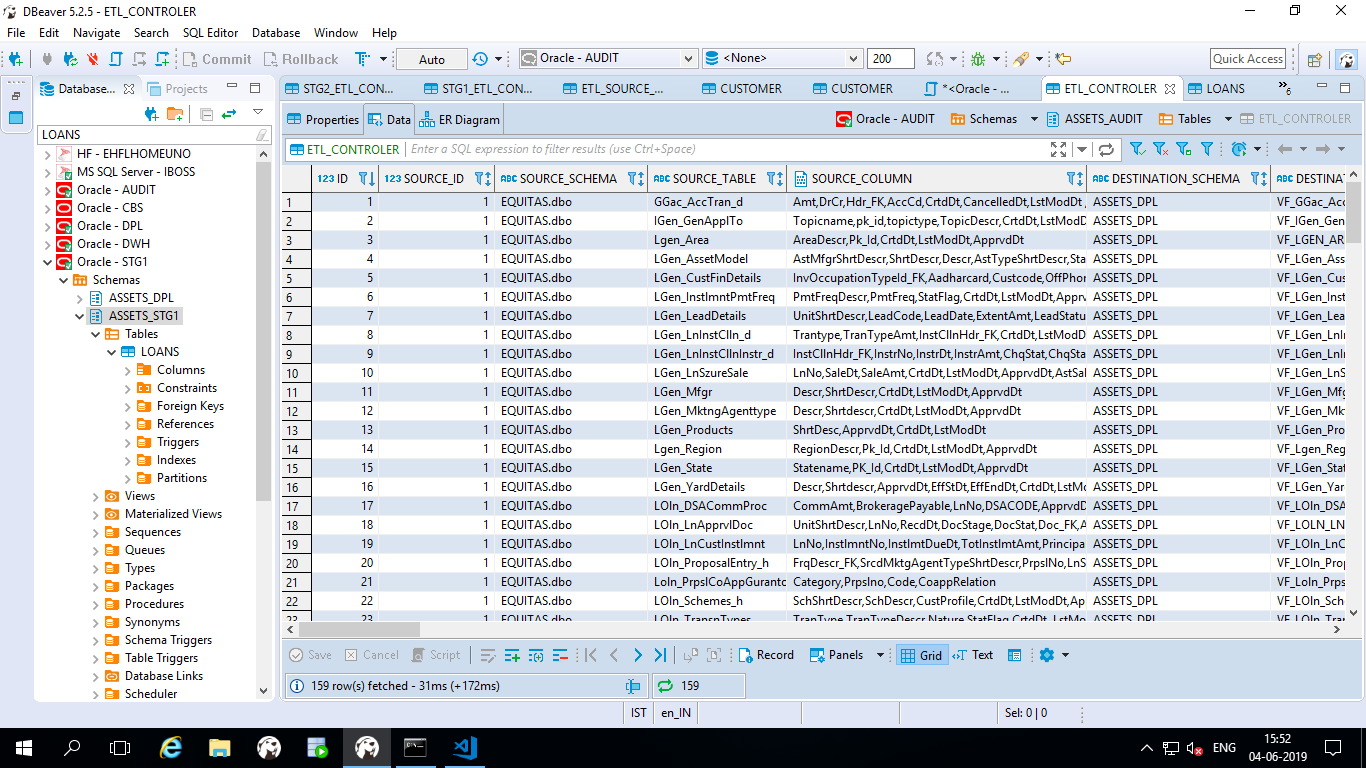
PYTHON ETL PROCESS

1. Pre-Requirements for ETL Process
2. Source to Blue
3. Blue to Grey
4. Grey to Green
5. Green to Red

**Pre-Requirements for ETL Process:**

Step1: CreateETL\_CONTROLER Table for Source to Blue Zone in AUDIT schema.

Step1.1: Create ETL\_CONTROLER table with all source tables, columns, destination tables, columns and full load conditions.



**ETL\_CONTROLER**

**Full-Load:**

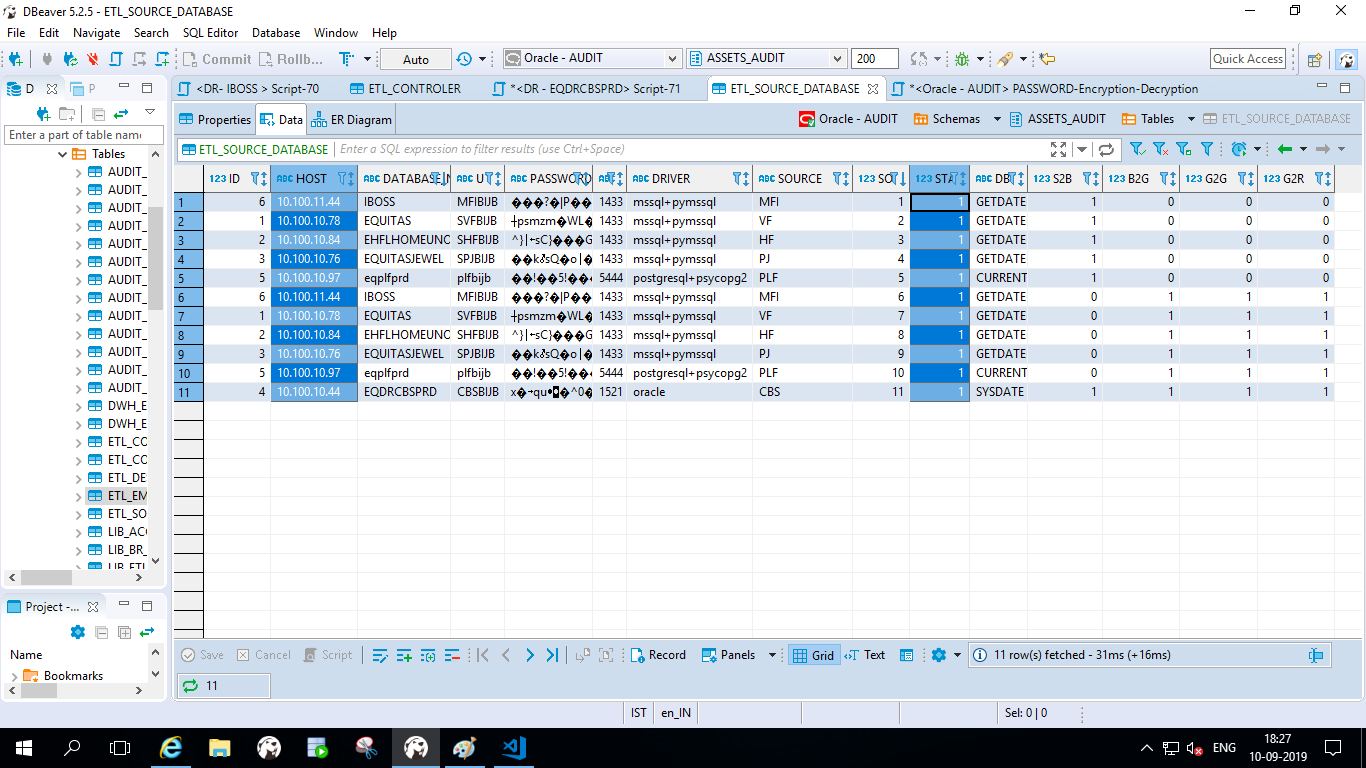
Step1.2: Update full load condition and status is one for all tables in Controller table while loading Full data push.

**Incremental Load:**

Step1.3: Please be careful while updating the Incremental conditions and ETL\_TIME to run the Incremental data.

Step1.4: Update status as one for daily incremental tables and remaining should be zero.

**Step2: Creating the ETL\_SOURCE\_DATABASE table.**



**ETL\_SOURCE\_DATABASE**

Step2.1: In **ETL\_SOURCE\_DATABASE** table having all Source database credentials.

Step2.2: Here you have to update status as **“One” or “Zero”**

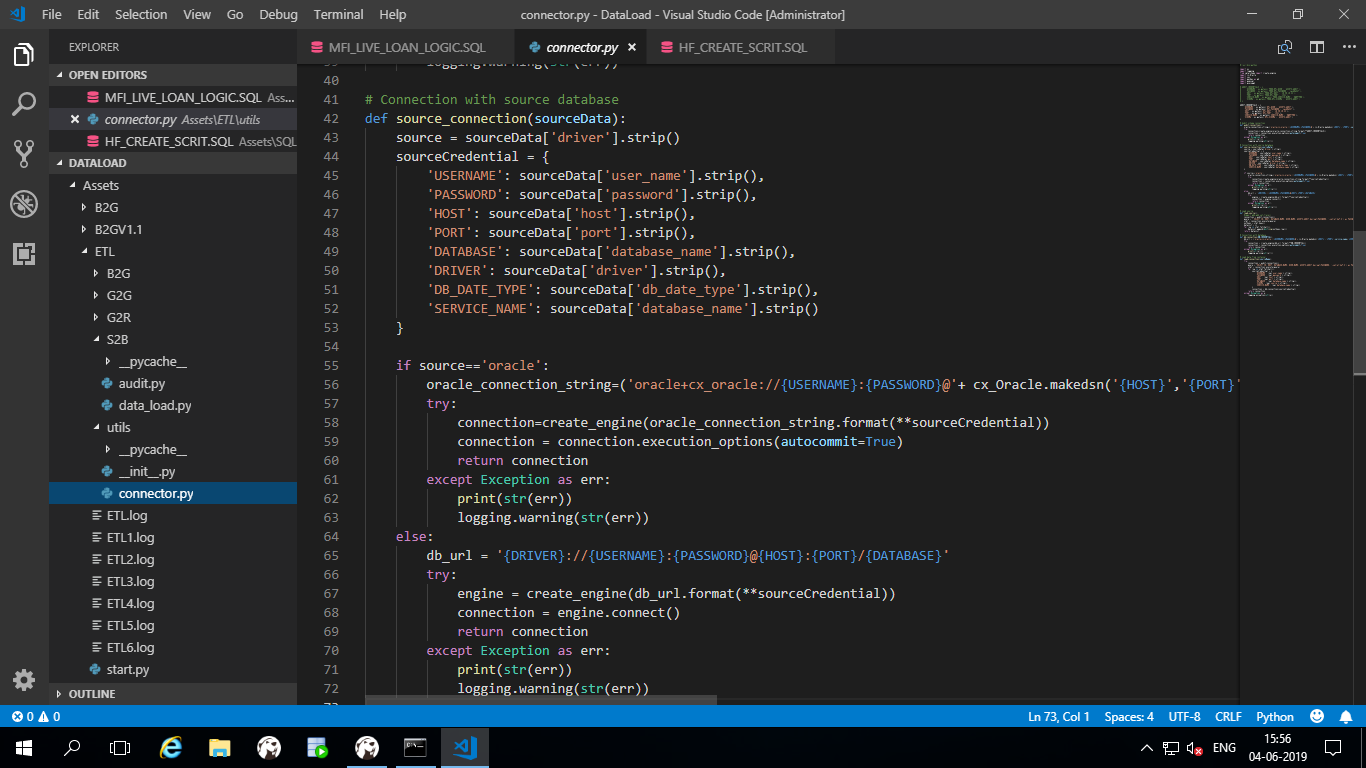
Step2.3: We have a flexibility here to run **Source wise or Zone wise** ETL process.

Step2.4: Update status is 1 which are the sources you are going to run in zone wise.

Step2.5:First it will establish the audit connection and fetch the all the Source credential records which we updated status as one.

Step3: The below one is the python connector file path location .This will give you the audit connection, Source connection, Destination connection and data load Source Details.

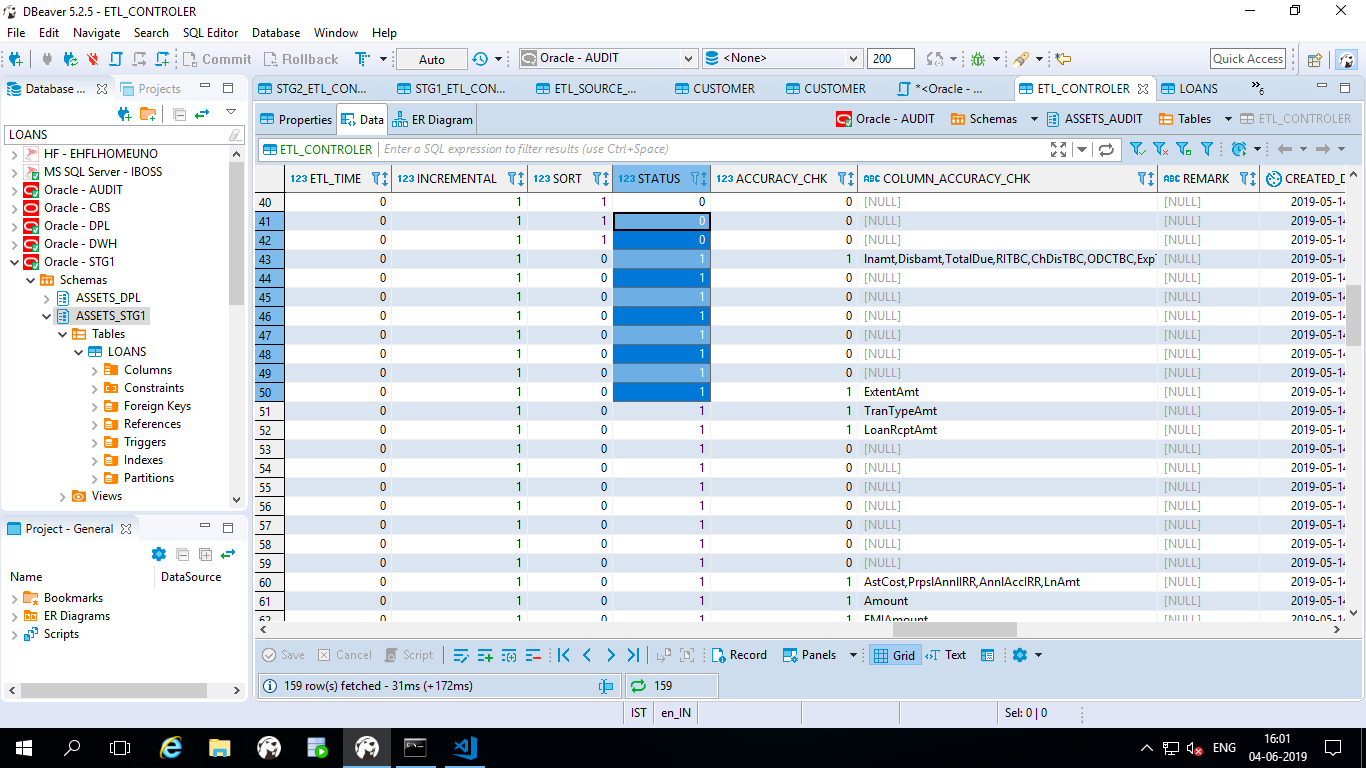
**D:\DataLoad\Assets\ETL\utils\Connector.py**



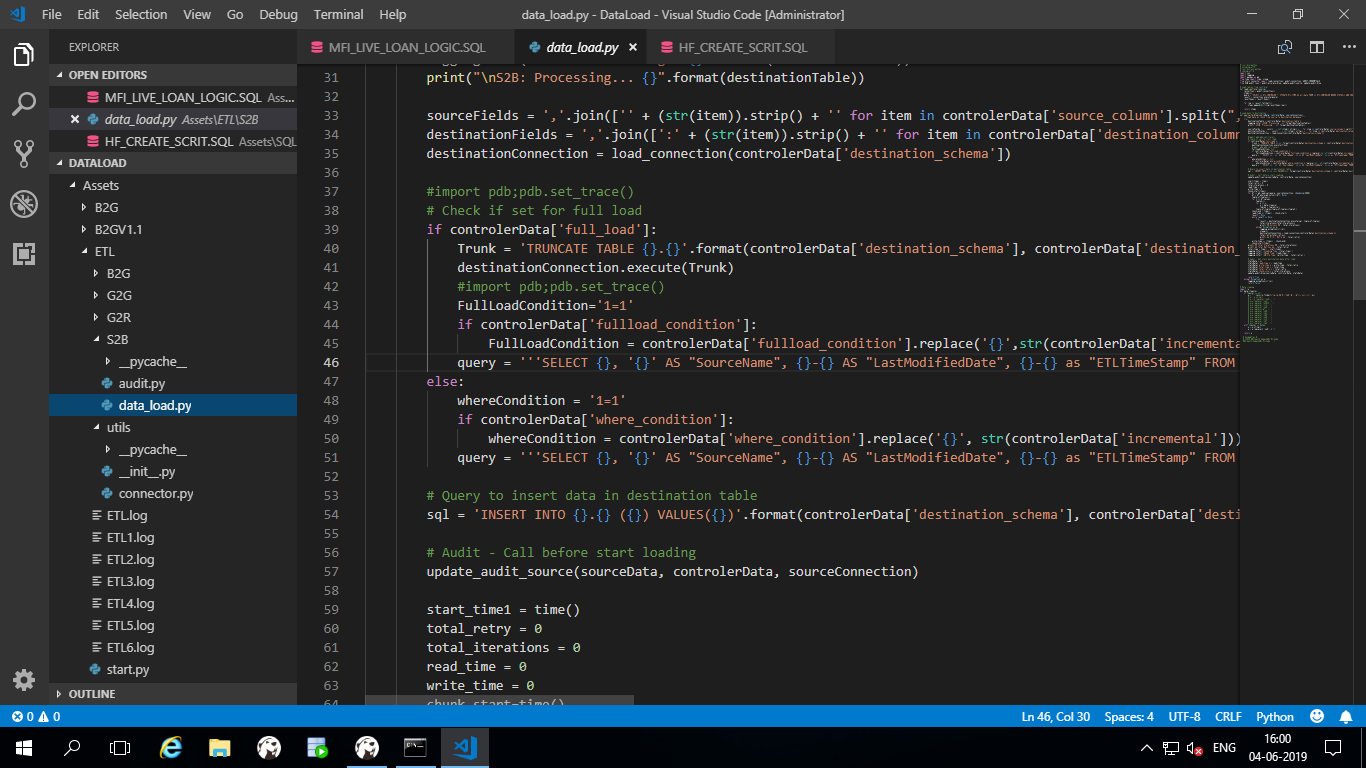
Step4:

Here you have to check whether you are running for full load or incremental load.If it is full load you can see full load column condition should be “One “.

Step5: Based on the source credentials it will fetch all the tables’ information from the ETL\_CONTROLER Table.

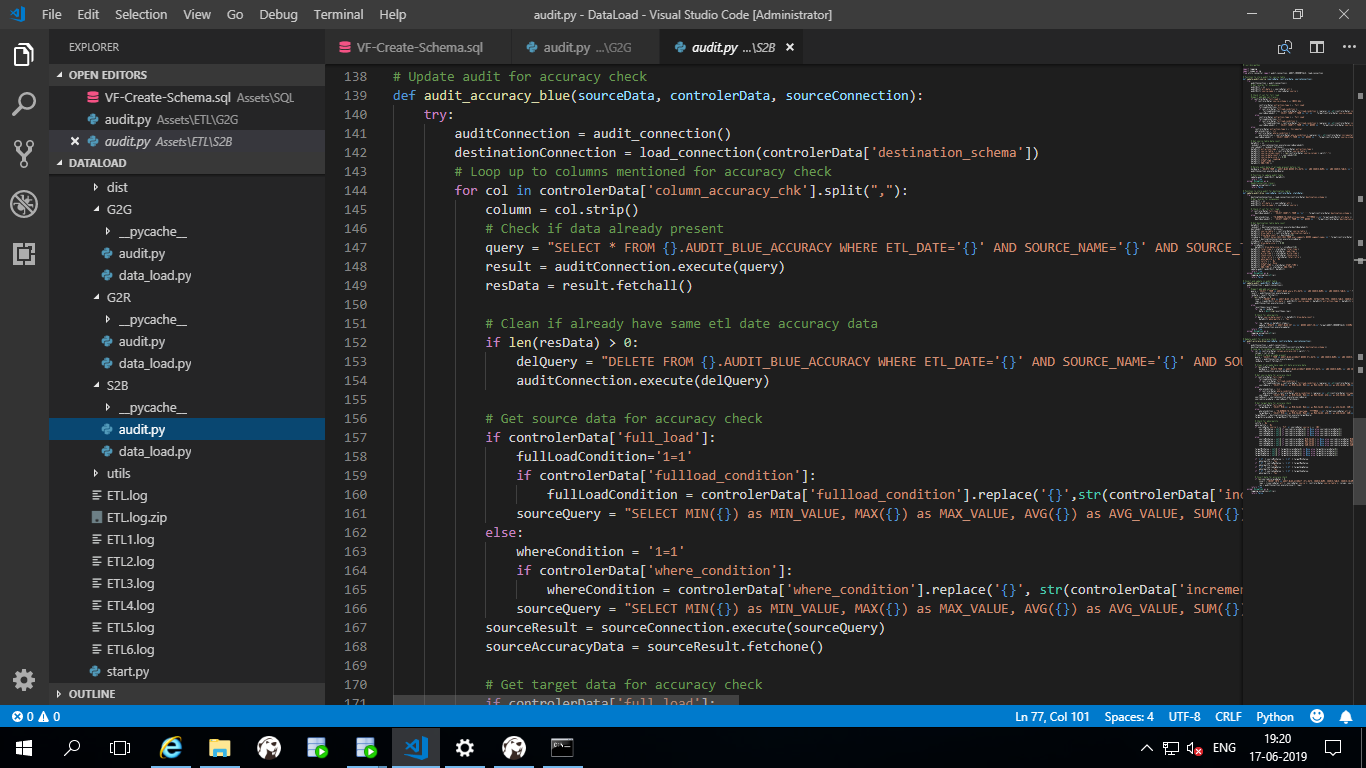


Step6: Here Data\_load.py file will push the data based on the chunk size: **\DataLoad\Assets\ETL\S2B\data\_load.py**

Step7: Then it will start loading data from Source to Blue

Step8: After loading the data from source to blue check the Audit Blue table in Audit schema whether all tables are loaded or not.

**\DataLoad\Assets\ETL\S2B\audit.py**

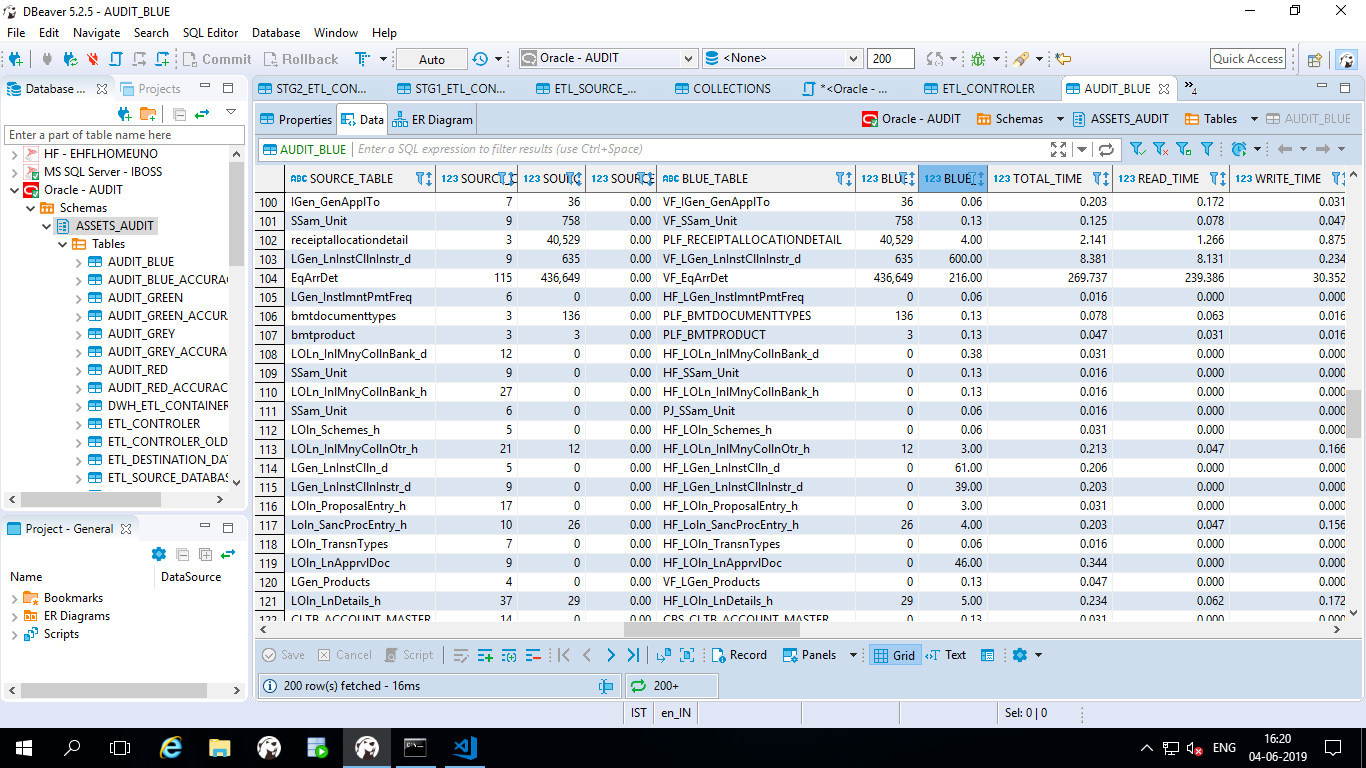


Step8.1 Here you can check all the statistics of each table means Start time ,End\_time,Total\_time,Read\_time,Write\_time,Abnormality and Status

Step8.2: In Audit blue table you have to check weather abnormality column is YES or No. If you found any yes then you have to check logic condition of the particular table in controller table.

Step9: The audit.py will tell you what type of logics we are running to take the statistics for table wise and zone wise.

Step9.1: We can see all the statistics of Source to blue zone table count level information in Audit Blue table.



**AUDIT\_BLUE**

**Blue to Grey**

Step10: Next zone is Blue to Grey check whether The STG1\_ETL\_CONTAINER table is updated or not. If it is updated go for data push between Blue to Grey.

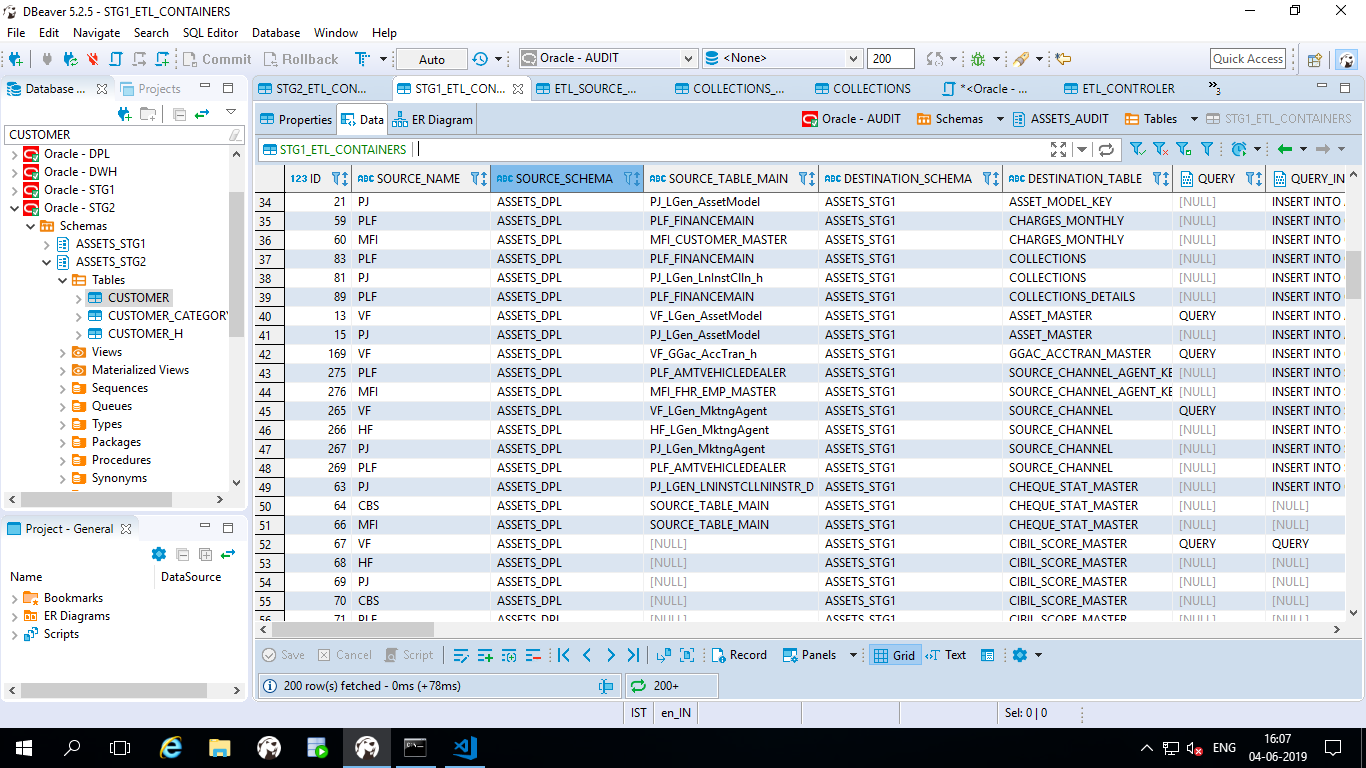
Step10.1: Ifnot update the STG1\_ETL\_CONTAINER .Here we have to match with ETL\_DATE as ETLTIMESTAMP of incremental load.

Step10.2: In stg1\_etl\_container we can set priority of the containers by using the sort column in the stg1-container table.

Step10.3: Here we are using same query for both incremental and full load.

Step10.4: Statistic condition and joins are using for updating the audit table in container wise. If you want add any new requirements in future you can add at any point of time.

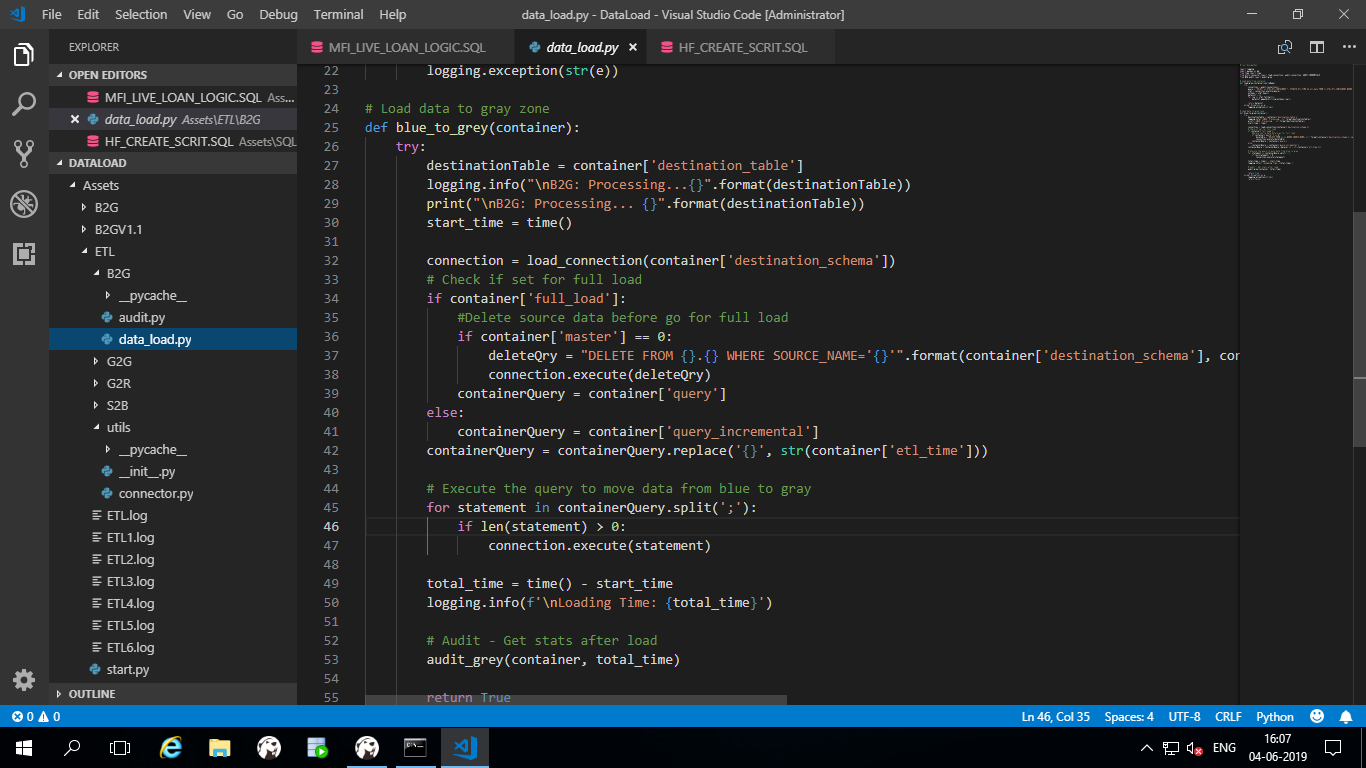
The below screen short will give you the reference of the process.



Step11: Python script for loading the Blue to Grey

**D:\DataLoad\Assets\ETL\B2G\data\_load.py**

Step11.1: The below code will tell you the process of pushing the data between Blue to Grey



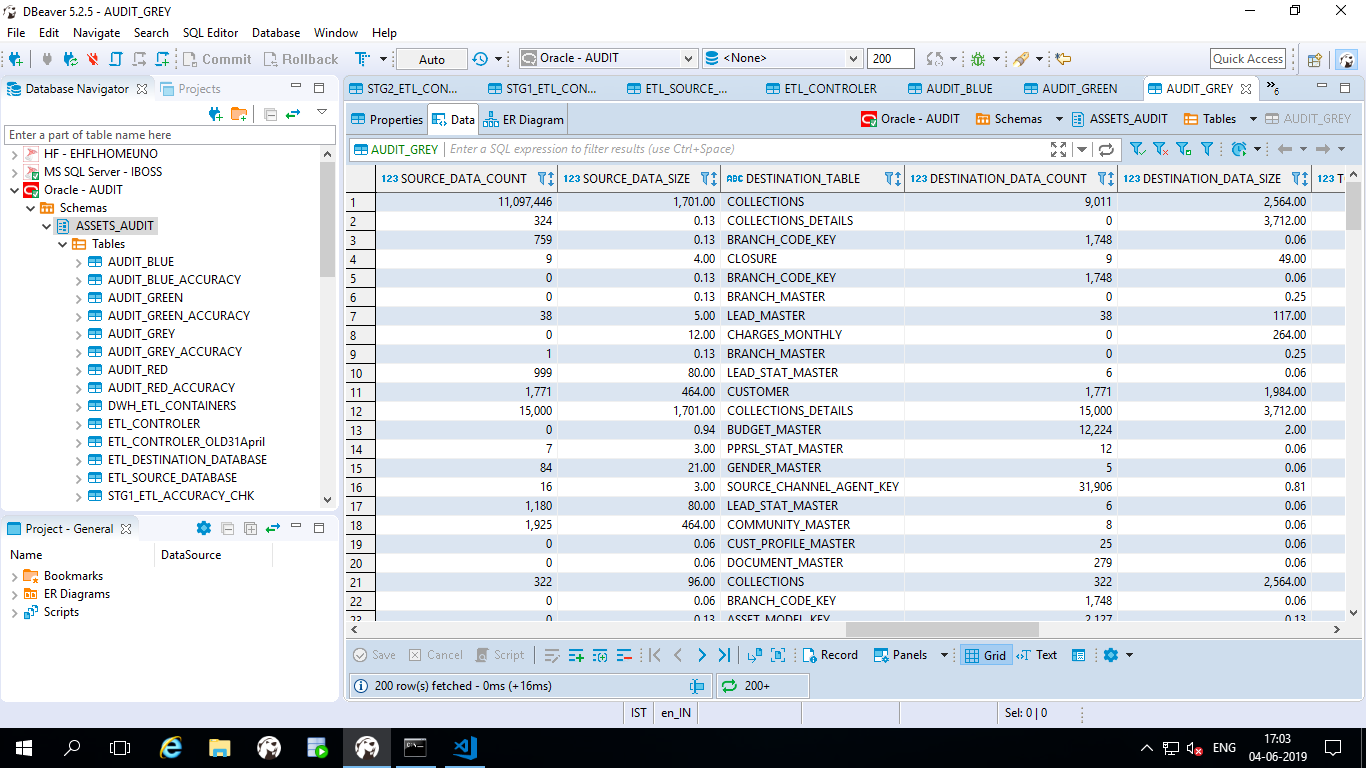
**Blue to Grey**

Step12: After loading all containers blue to Grey you can see all the transformed data available in the containers in STG1 Zone i.e Grey Zone and also you can check Audit Grey table in Audit schema.

**D:\DataLoad\Assets\ETL\B2G\audit.py**

Step13: The audit.py will tell you what type of logics we are running to take the statistics for table wise and zone wise.

Step14: We can see all the statistics of Blue to Grey zone table level information in Audit Grey table.



**Audit\_Grey**

**Grey to Green:**

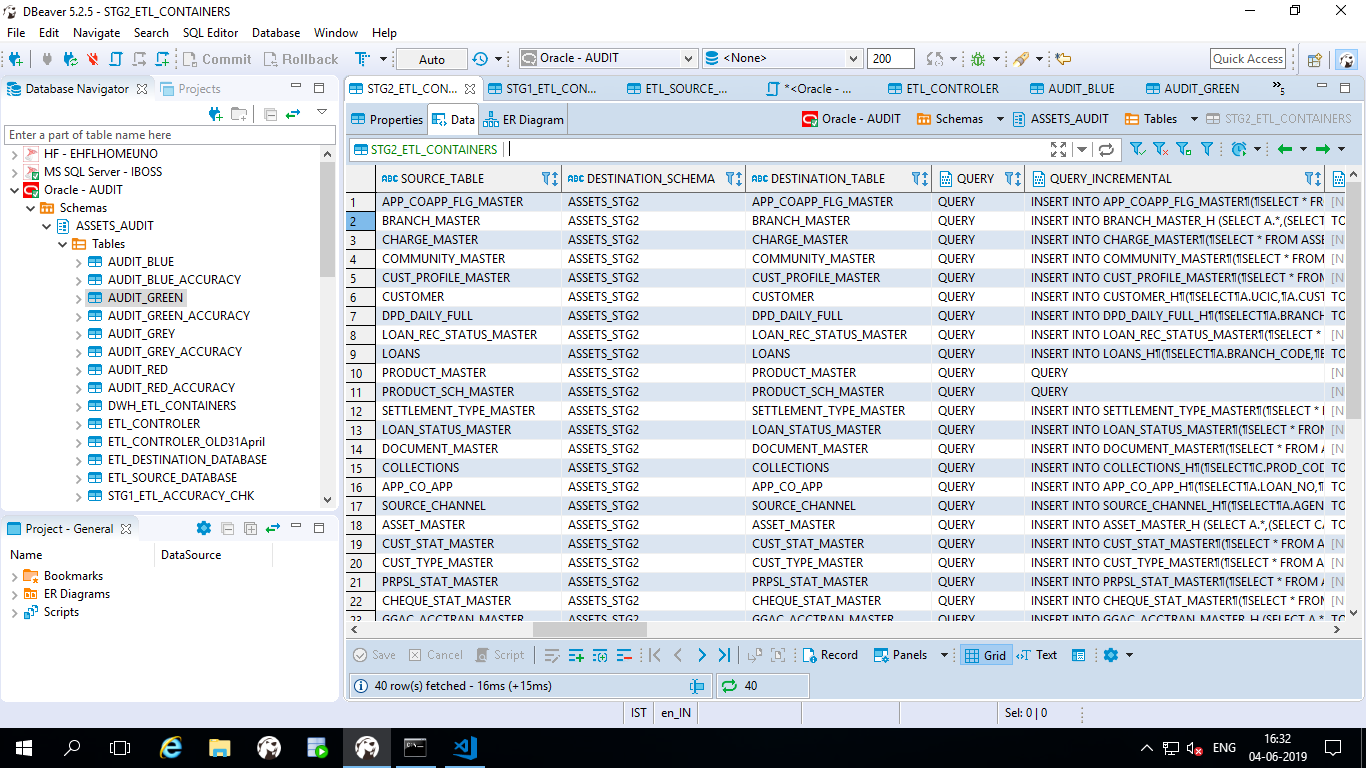
1. **Main Container**
2. **Version Container**

The process of grey to red zone, we are procedures to load data from Grey to Green. The procedure will insert new records to main container directly. If any old records came then it will generate version and update the records and move to Version containers.

**Main Container:**

Step15: Creating STG2\_ETL\_CONTAINER table in Audit Schema.

Step16: In G2G process will start based on the status column for the particular etl\_time.

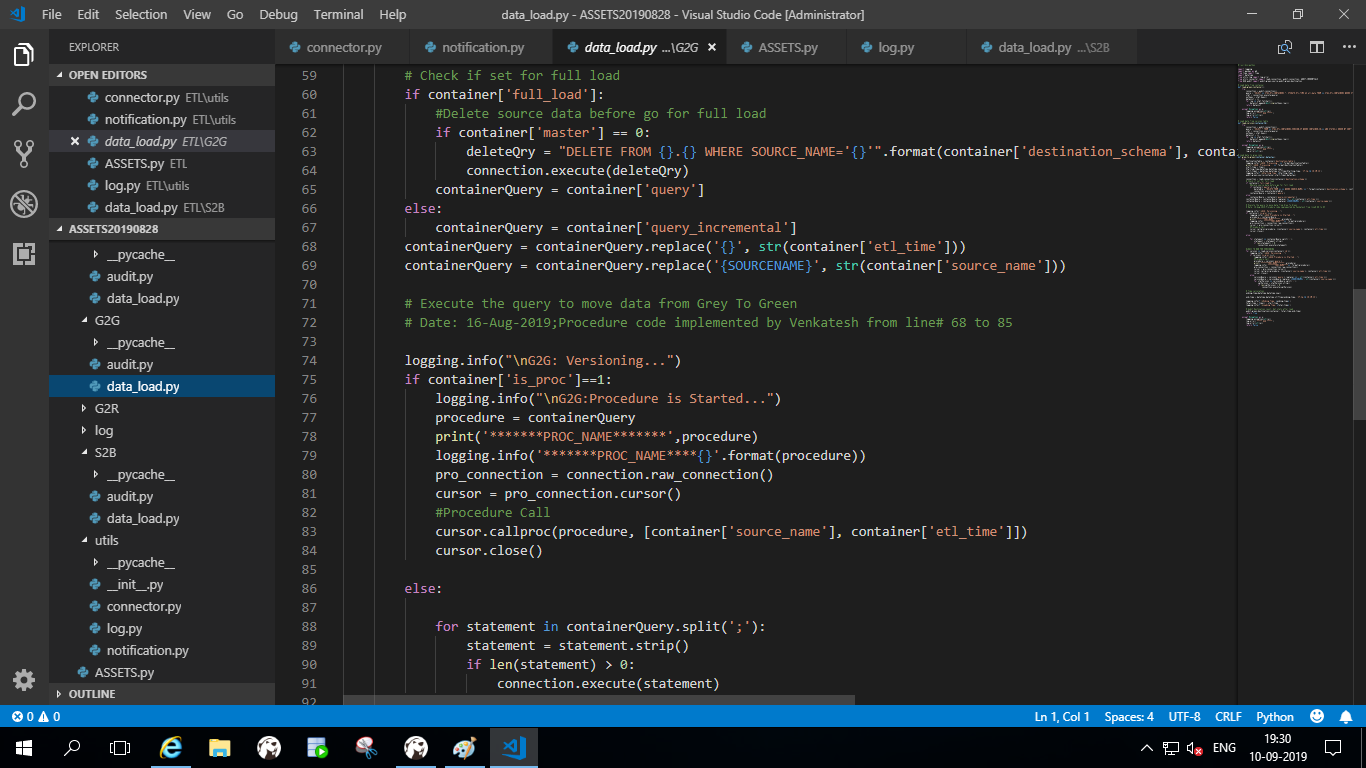


**STG2\_ETL\_CONTAINER**

Step17: In the STG2 ETL\_CONTAINER you have to update all Queries in container wise for full load and incremental as well as update status as “ONE “which are the containers you are going to load and If it is procedure we have to update Is\_Proc column status should be one .

Step19:The below python code will push the data from Grey2Green.

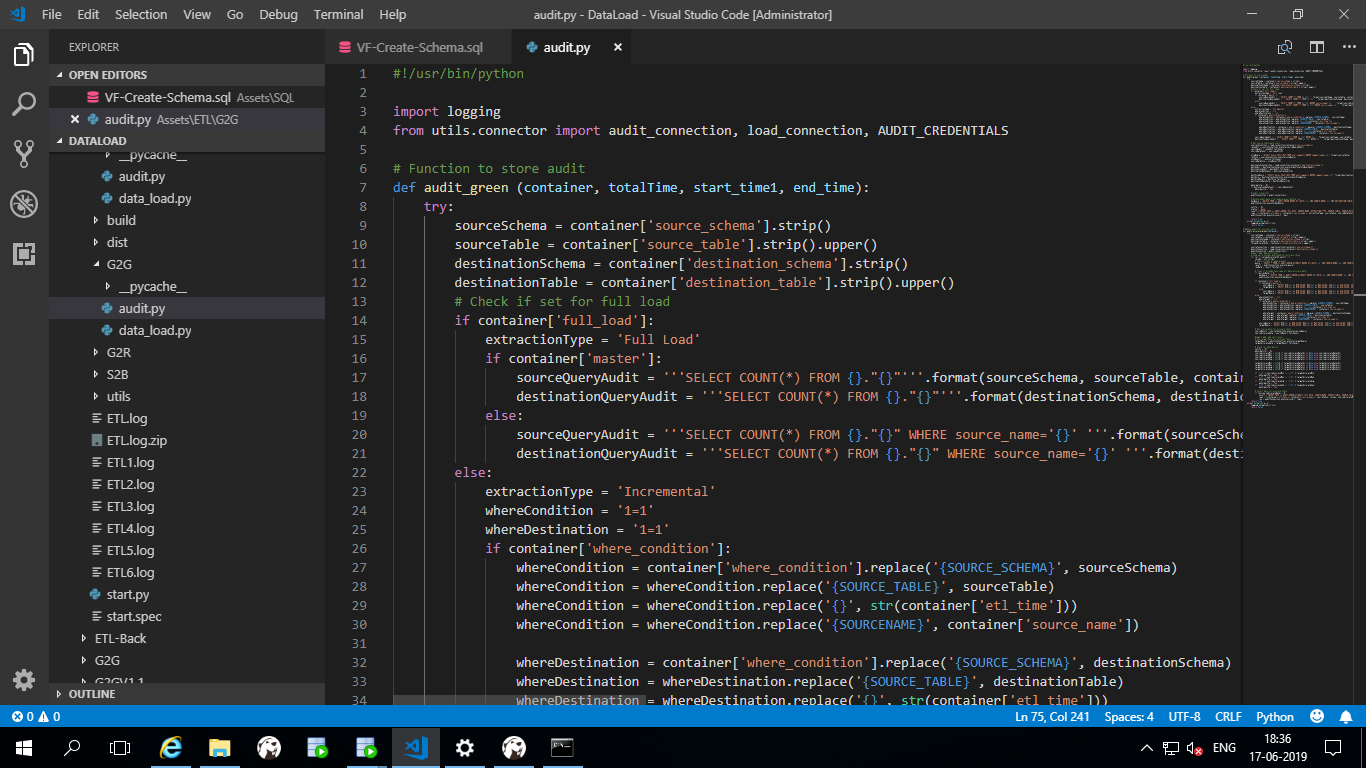
**D:\DataLoad\Assets\ETL\G2G\data\_load.py**



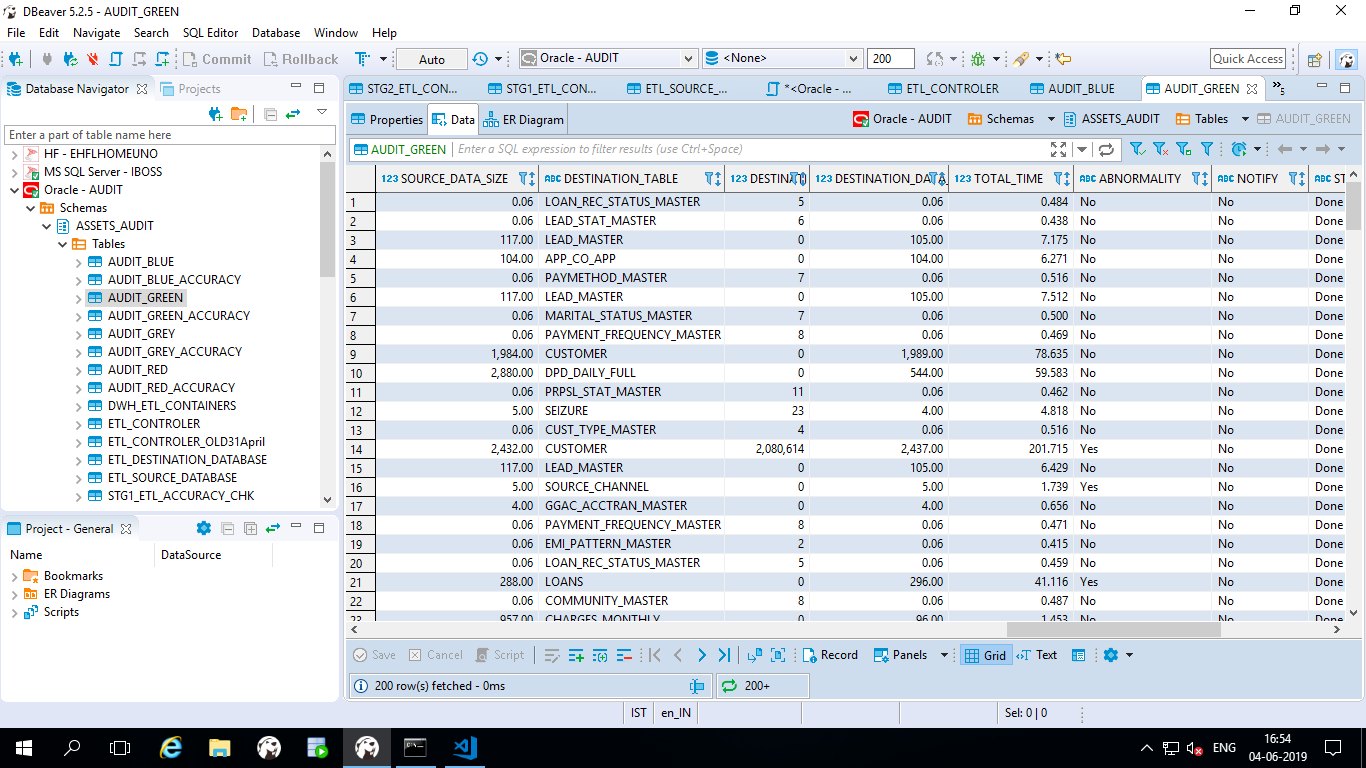
Step20:

The below python code will update the Grey to Green statistics of the data.

**D:\DataLoad\Assets\ETL\G2G\audit.py**



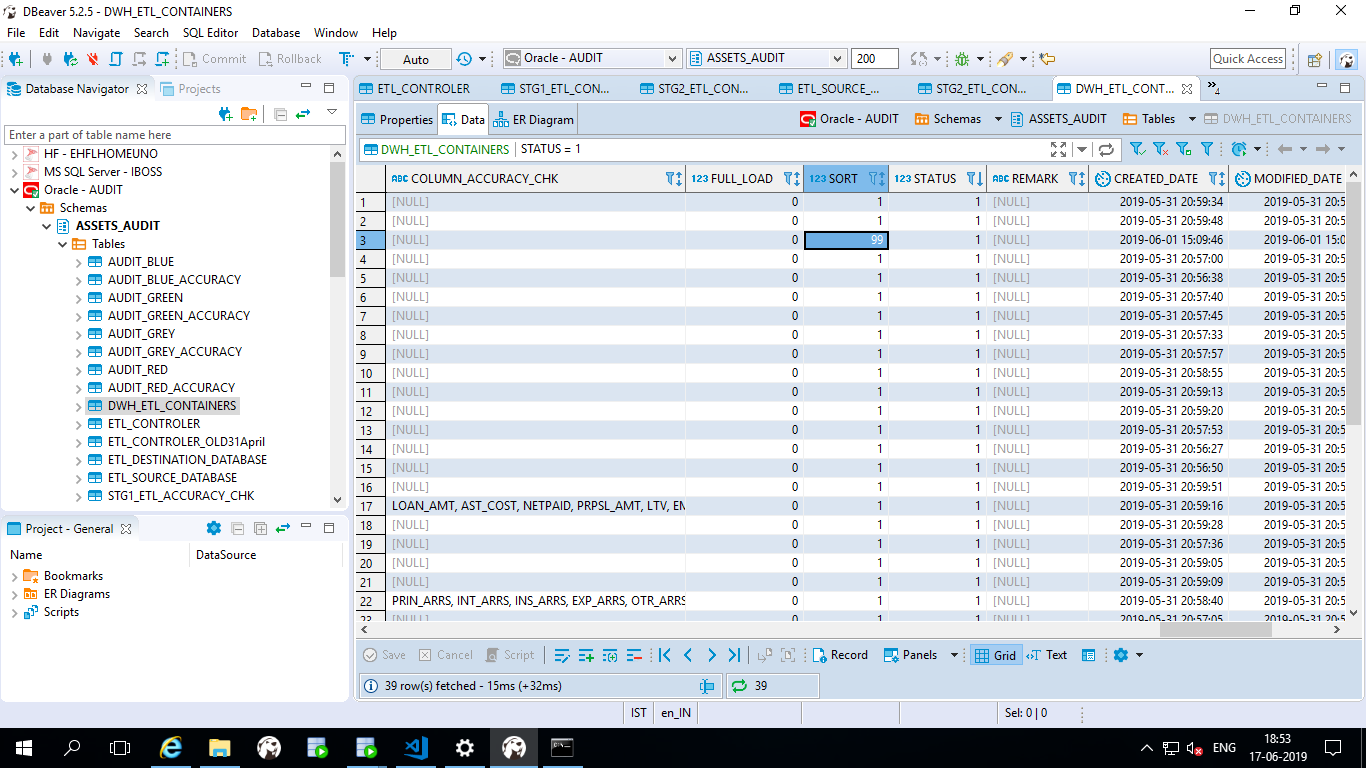
Step21: Once container data pushed to green container table you can see all the container level statistics in AUDIT\_GREEN Table.

 AUDIT\_GREEN

**Green to Red:**

Step22: Creating the DWH\_ETL\_CONTAINERS table in audit schema.

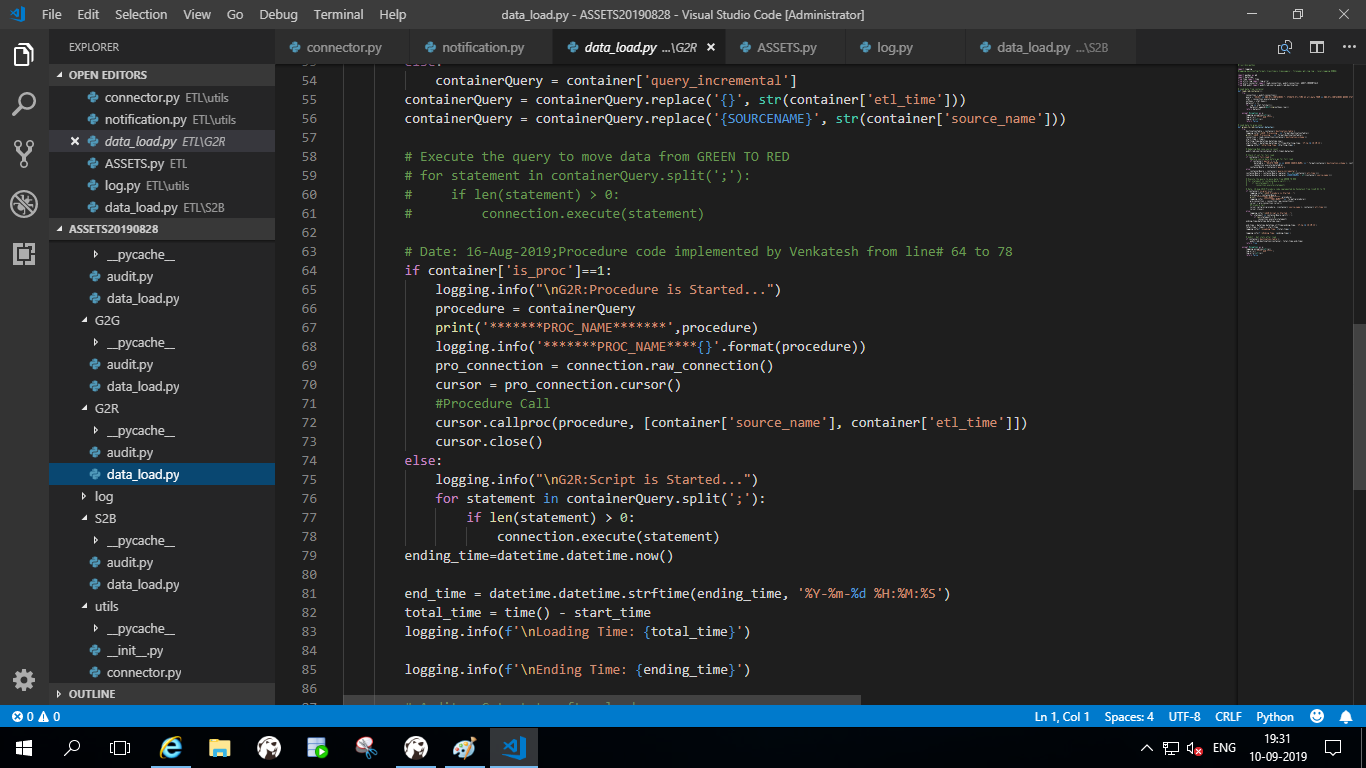
Step22.1:If you’re going for full load, you have to update full load column as “one”and status should be “one”



Step23:

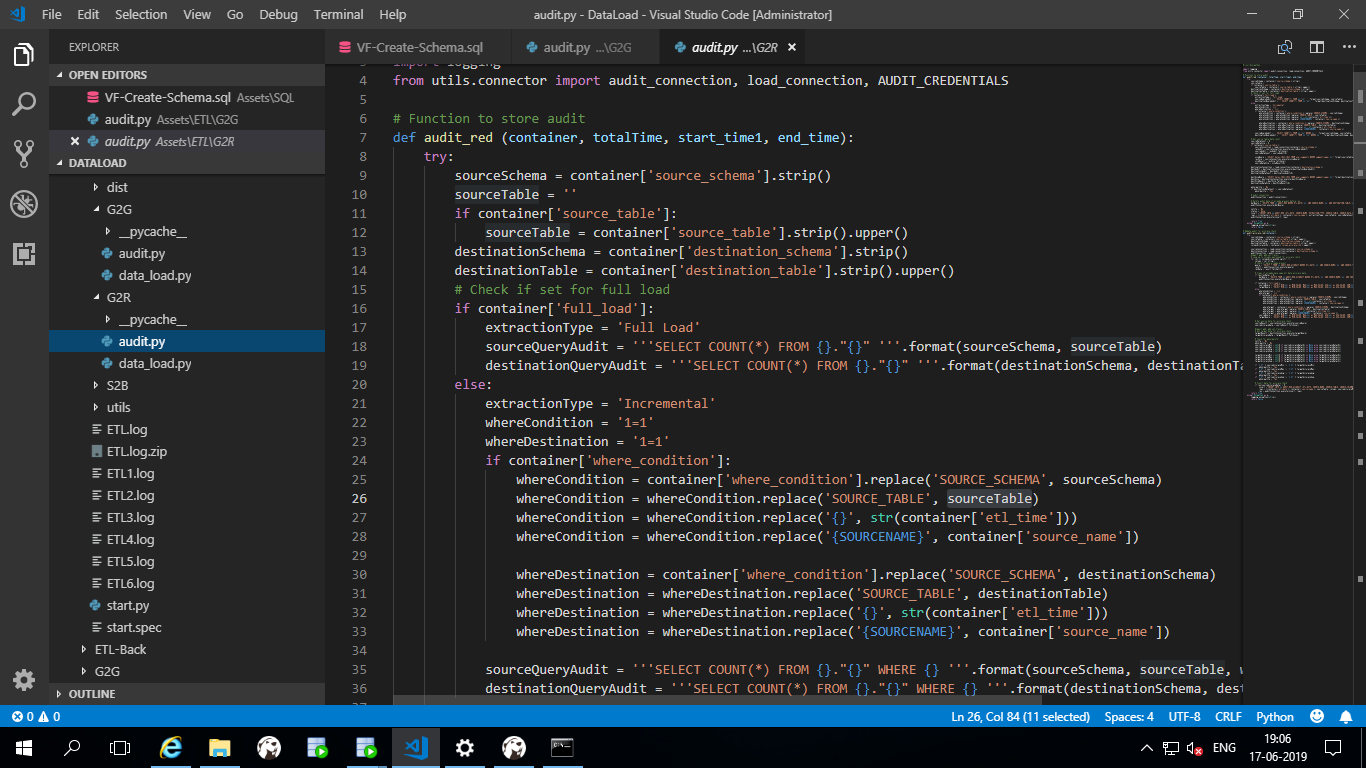
Using the below python code just we are copying the data from Green zone to Red zone.

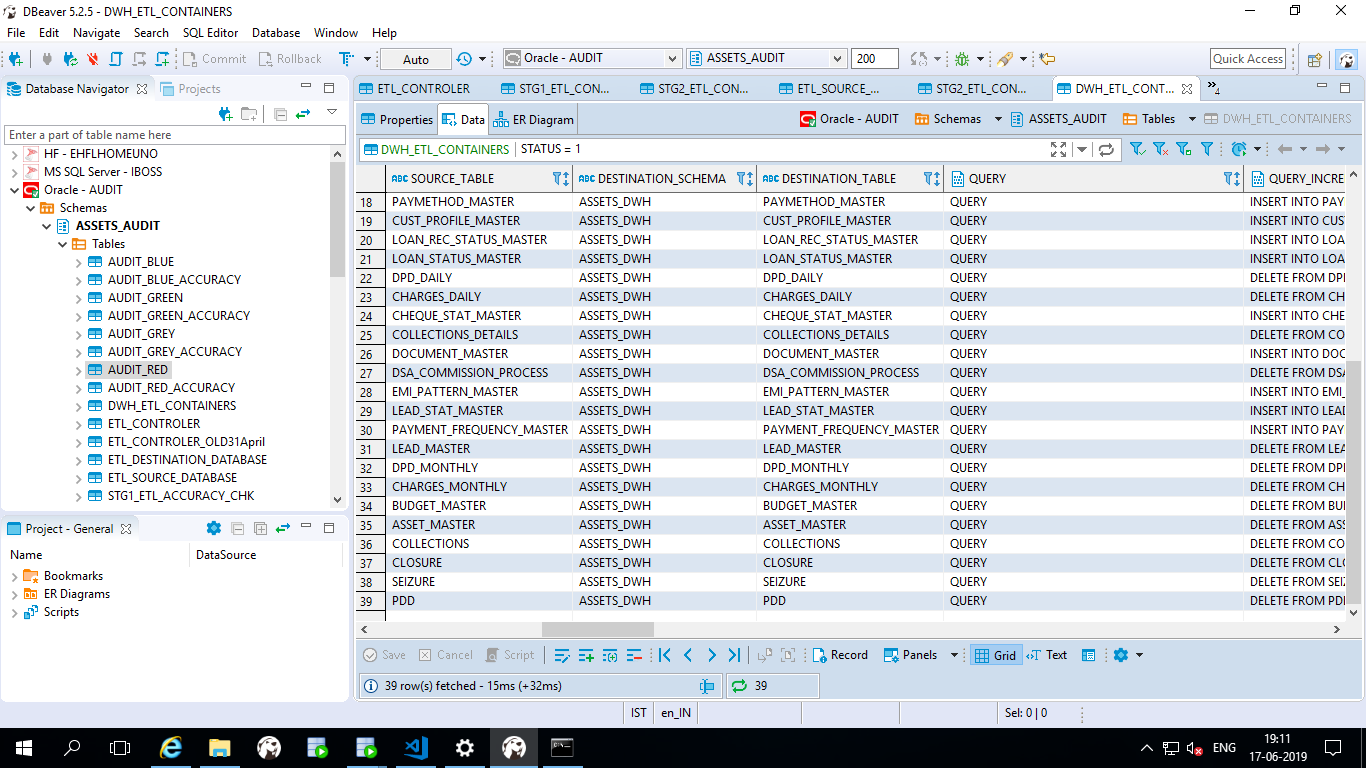
Every time the previous data will be flushed out and insert updated data in Red zone.

D:\DataLoad\Assets\ETL\G2R\data\_load.py

Step24: The below script will update statistics of the Grey to Green zone

D:\DataLoad\Assets\ETL\G2R\audit.py



Step25: After successfully completion of data pushing you can see the red zone audit results.

Step26: After completion of all the zones you can check log file also weather your process is successfully completed are not in the location **D:\DataLoad\Assets\ETL\ETL.log**