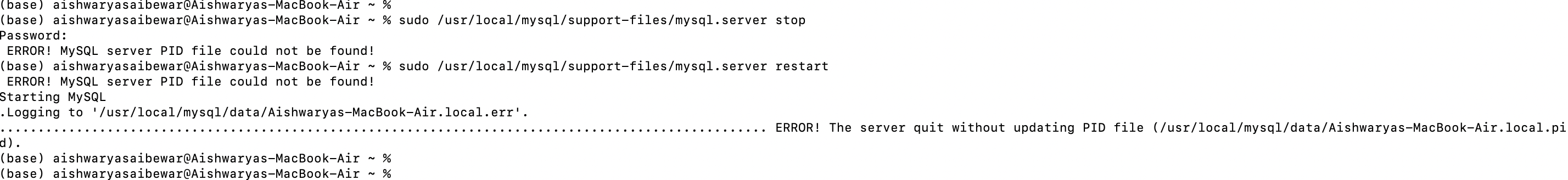
**Class Project1:**

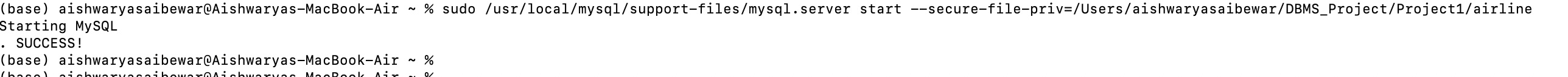
**Step 1: (Connecting to the database and accessing the folders)**

To load the data using command prompt navigate to command prompt and start the sql server.



1. sudo /usr/local/mysql/support-files/mysql.server stop
2. sudo /usr/local/mysql/support-files/mysql.server restart

Giving permission to access my folder where I had all the airline data:



1. sudo /usr/local/mysql/support-files/mysql.server start --secure-file-priv=/Users/aishwaryasaibewar/DBMS\_Project/Project1/airline

SETTING UP PERMISSION IN MYSQL SERVER:

1. /usr/local/mysql/bin/mysql --local-infile=1 -u root -p
2. SHOW VARIABLES LIKE "secure\_file\_priv";
3. SHOW GLOBAL VARIABLES LIKE 'local\_infile';
4. SET GLOBAL local\_infile = 'ON';

Text, letter

Description automatically generated

Graphical user interface, text

Description automatically generated

**Step 2: (Create table for storing data for years 1998 to 2016):**

Creating table for `airline\_reporting`using the below command. This tables will store data for the years 1998,1999,2015,2016. We added this data into a single table because there are similar columns in all these years.

*CREATE TABLE `airline\_reporting` (*

*`YEAR` int DEFAULT NULL,*

*`QUARTER` int DEFAULT NULL,*

*`MONTH` int DEFAULT NULL,*

*`DAY\_OF\_MONTH` int DEFAULT NULL,*

*`DAY\_OF\_WEEK` int DEFAULT NULL,*

*`FL\_DATE` DATETIME DEFAULT NULL,*

*`OP\_UNIQUE\_CARRIER` varchar(70),*

*`OP\_CARRIER\_AIRLINE\_ID` int DEFAULT NULL,*

*`OP\_CARRIER` varchar(70),*

*`TAIL\_NUM` varchar(70),*

*`OP\_CARRIER\_FL\_NUM` int DEFAULT NULL,*

*`ORIGIN\_AIRPORT\_ID` int DEFAULT NULL,*

*`ORIGIN\_AIRPORT\_SEQ\_ID` int DEFAULT NULL,*

*`ORIGIN\_CITY\_MARKET\_ID` int DEFAULT NULL,*

*`ORIGIN` varchar(70),*

*`ORIGIN\_CITY\_NAME` varchar(70),*

*`ORIGIN\_STATE\_ABR` varchar(70),*

*`ORIGIN\_STATE\_FIPS` int DEFAULT NULL,*

*`ORIGIN\_STATE\_NM` varchar(70),*

*`ORIGIN\_WAC` int DEFAULT NULL,*

*`DEST\_AIRPORT\_ID` int DEFAULT NULL,*

*`DEST\_AIRPORT\_SEQ\_ID` int DEFAULT NULL,*

*`DEST\_CITY\_MARKET\_ID` int DEFAULT NULL,*

*`DEST` varchar(70),*

*`DEST\_CITY\_NAME` varchar(70),*

*`DEST\_STATE\_ABR` varchar(70),*

*`DEST\_STATE\_FIPS` int DEFAULT NULL,*

*`DEST\_STATE\_NM` varchar(70),*

*`DEST\_WAC` int DEFAULT NULL,*

*`CRS\_DEP\_TIME` int DEFAULT NULL,*

*`DEP\_TIME` int DEFAULT NULL,*

*`DEP\_DELAY` int DEFAULT NULL,*

*`DEP\_DELAY\_NEW` int DEFAULT NULL,*

*`DEP\_DEL15` int DEFAULT NULL,*

*`DEP\_DELAY\_GROUP` int DEFAULT NULL,*

*`DEP\_TIME\_BLK` varchar(70),*

*`TAXI\_OUT` int DEFAULT NULL,*

*`WHEELS\_OFF` int DEFAULT NULL,*

*`WHEELS\_ON` int DEFAULT NULL,*

*`TAXI\_IN` int DEFAULT NULL,*

*`CRS\_ARR\_TIME` int DEFAULT NULL,*

*`ARR\_TIME` int DEFAULT NULL,*

*`ARR\_DELAY` int DEFAULT NULL,*

*`ARR\_DELAY\_NEW` int DEFAULT NULL,*

*`ARR\_DEL15` int DEFAULT NULL,*

*`ARR\_DELAY\_GROUP` int DEFAULT NULL,*

*`ARR\_TIME\_BLK` varchar(70),*

*`CANCELLED` int DEFAULT NULL,*

*`CANCELLATION\_CODE` varchar(70),*

*`DIVERTED` int DEFAULT NULL,*

*`CRS\_ELAPSED\_TIME` int DEFAULT NULL,*

*`ACTUAL\_ELAPSED\_TIME` int DEFAULT NULL,*

*`AIR\_TIME` int DEFAULT NULL,*

*`FLIGHTS` int DEFAULT NULL,*

*`DISTANCE` int DEFAULT NULL,*

*`DISTANCE\_GROUP` int DEFAULT NULL,*

*`CARRIER\_DELAY` varchar(70),*

*`WEATHER\_DELAY` varchar(70),*

*`NAS\_DELAY` varchar(70),*

*`SECURITY\_DELAY` varchar(70),*

*`LATE\_AIRCRAFT\_DELAY` varchar(70),*

*`FIRST\_DEP\_TIME` varchar(70),*

*`TOTAL\_ADD\_GTIME` varchar(70),*

*`LONGEST\_ADD\_GTIME` varchar(70),*

*`DIV\_AIRPORT\_LANDINGS` int DEFAULT NULL,*

*`DIV\_REACHED\_DEST` varchar(70),*

*`DIV\_ACTUAL\_ELAPSED\_TIME` varchar(70),*

*`DIV\_ARR\_DELAY` varchar(70),*

*`DIV\_DISTANCE` varchar(70),*

*`DIV1\_AIRPORT` varchar(70),*

*`DIV1\_AIRPORT\_ID` varchar(70),*

*`DIV1\_AIRPORT\_SEQ\_ID` varchar(70),*

*`DIV1\_WHEELS\_ON` varchar(70),*

*`DIV1\_TOTAL\_GTIME` varchar(70),*

*`DIV1\_LONGEST\_GTIME` varchar(70),*

*`DIV1\_WHEELS\_OFF` varchar(70),*

*`DIV1\_TAIL\_NUM` varchar(70),*

*`DIV2\_AIRPORT` varchar(70),*

*`DIV2\_AIRPORT\_ID` varchar(70),*

*`DIV2\_AIRPORT\_SEQ\_ID` varchar(70),*

*`DIV2\_WHEELS\_ON` varchar(70),*

*`DIV2\_TOTAL\_GTIME` varchar(70),*

*`DIV2\_LONGEST\_GTIME` varchar(70),*

*`DIV2\_WHEELS\_OFF` varchar(70),*

*`DIV2\_TAIL\_NUM` varchar(70),*

*`DIV3\_AIRPORT` varchar(70),*

*`DIV3\_AIRPORT\_ID` varchar(70),*

*`DIV3\_AIRPORT\_SEQ\_ID` varchar(70),*

*`DIV3\_WHEELS\_ON` varchar(70),*

*`DIV3\_TOTAL\_GTIME` varchar(70),*

*`DIV3\_LONGEST\_GTIME` varchar(70),*

*`DIV3\_WHEELS\_OFF` varchar(70),*

*`DIV3\_TAIL\_NUM` varchar(70),*

*`DIV4\_AIRPORT` varchar(70),*

*`DIV4\_AIRPORT\_ID` varchar(70),*

*`DIV4\_AIRPORT\_SEQ\_ID` varchar(70),*

*`DIV4\_WHEELS\_ON` varchar(70),*

*`DIV4\_TOTAL\_GTIME` varchar(70),*

*`DIV4\_LONGEST\_GTIME` varchar(70),*

*`DIV4\_WHEELS\_OFF` varchar(70),*

*`DIV4\_TAIL\_NUM` varchar(70),*

*`DIV5\_AIRPORT` varchar(70),*

*`DIV5\_AIRPORT\_ID` varchar(70),*

*`DIV5\_AIRPORT\_SEQ\_ID` varchar(70),*

*`DIV5\_WHEELS\_ON` varchar(70),*

*`DIV5\_TOTAL\_GTIME` varchar(70),*

*`DIV5\_LONGEST\_GTIME` varchar(70),*

*`DIV5\_WHEELS\_OFF` varchar(70),*

*`DIV5\_TAIL\_NUM` varchar(70)*

*);*

**Step 3: Loading data for the years 1998 to 2016**

Load the data for the years 1998,1999,2015,2016 into the *`airline\_reporting`* table via terminal using the following commands.

*mysql> LOAD DATA LOCAL INFILE '/Users/aishwaryasaibewar/DBMS\_Project/airline/Airline\_1998\_12.csv' INTO TABLE airline\_schema.airline\_reporting FIELDS TERMINATED BY ',' ENCLOSED BY '"' LINES TERMINATED BY '\n' IGNORE 1 ROWS;*

*Query OK, 455099 rows affected, 65535 warnings (6.19 sec)*

*Records: 455099 Deleted: 0 Skipped: 0 Warnings: 627751*

*mysql> LOAD DATA LOCAL INFILE '/Users/aishwaryasaibewar/DBMS\_Project/airline/Airline\_1999\_1.csv' INTO TABLE airline\_schema.airline\_reporting FIELDS TERMINATED BY ',' ENCLOSED BY '"' LINES TERMINATED BY '\n' IGNORE 1 ROWS;*

*Query OK, 453814 rows affected, 65535 warnings (6.05 sec)*

*Records: 453814 Deleted: 0 Skipped: 0 Warnings: 897087*

*mysql> LOAD DATA LOCAL INFILE '/Users/aishwaryasaibewar/DBMS\_Project/airline/Airline\_2015\_12.csv' INTO TABLE airline\_schema.airline\_reporting FIELDS TERMINATED BY ',' ENCLOSED BY '"' LINES TERMINATED BY '\n' IGNORE 1 ROWS;*

*Query OK, 479230 rows affected, 65535 warnings (6.31 sec)*

*Records: 479230 Deleted: 0 Skipped: 0 Warnings: 136754*

*mysql> LOAD DATA LOCAL INFILE '/Users/aishwaryasaibewar/DBMS\_Project/airline/Airline\_2016\_01.csv' INTO TABLE airline\_schema.airline\_reporting FIELDS TERMINATED BY ',' ENCLOSED BY '"' LINES TERMINATED BY '\n' IGNORE 1 ROWS;*

*Query OK, 445827 rows affected, 65535 warnings (6.23 sec)*

*Records: 445827 Deleted: 0 Skipped: 0 Warnings: 11782962*

Graphical user interface, text, application

Description automatically generated

**Step 4: (Create table for storing data for years 2020, 2021):**

Creating table for ‘airline\_marketing’ using the below command. This tables will store data for the years 2020 and 2021. We added this data into a single table because there are similar columns in these years.

*CREATE TABLE `airline\_marketing` (*

*`YEAR` int DEFAULT NULL,*

*`QUARTER` int DEFAULT NULL,*

*`MONTH` int DEFAULT NULL,*

*`DAY\_OF\_MONTH` int DEFAULT NULL,*

*`DAY\_OF\_WEEK` int DEFAULT NULL,*

*`FL\_DATE` DATETIME DEFAULT NULL,*

*`MKT\_UNIQUE\_CARRIER` varchar(70),*

*`BRANDED\_CODE\_SHARE` varchar(70),*

*`MKT\_CARRIER\_AIRLINE\_ID` int DEFAULT NULL,*

*`MKT\_CARRIER` varchar(70),*

*`MKT\_CARRIER\_FL\_NUM` int DEFAULT NULL,*

*`SCH\_OP\_UNIQUE\_CARRIER` varchar(70),*

*`SCH\_OP\_CARRIER\_AIRLINE\_ID` varchar(70),*

*`SCH\_OP\_CARRIER` varchar(70),*

*`SCH\_OP\_CARRIER\_FL\_NUM` varchar(70),*

*`OP\_UNIQUE\_CARRIER` varchar(70),*

*`OP\_CARRIER\_AIRLINE\_ID` int DEFAULT NULL,*

*`OP\_CARRIER` varchar(70),*

*`TAIL\_NUM` varchar(70),*

*`OP\_CARRIER\_FL\_NUM` int DEFAULT NULL,*

*`ORIGIN\_AIRPORT\_ID` int DEFAULT NULL,*

*`ORIGIN\_AIRPORT\_SEQ\_ID` int DEFAULT NULL,*

*`ORIGIN\_CITY\_MARKET\_ID` int DEFAULT NULL,*

*`ORIGIN` varchar(70),*

*`ORIGIN\_CITY\_NAME` varchar(70),*

*`ORIGIN\_STATE\_ABR` varchar(70),*

*`ORIGIN\_STATE\_FIPS` int DEFAULT NULL,*

*`ORIGIN\_STATE\_NM` varchar(70),*

*`ORIGIN\_WAC` int DEFAULT NULL,*

*`DEST\_AIRPORT\_ID` int DEFAULT NULL,*

*`DEST\_AIRPORT\_SEQ\_ID` int DEFAULT NULL,*

*`DEST\_CITY\_MARKET\_ID` int DEFAULT NULL,*

*`DEST` varchar(70),*

*`DEST\_CITY\_NAME` varchar(70),*

*`DEST\_STATE\_ABR` varchar(70),*

*`DEST\_STATE\_FIPS` int DEFAULT NULL,*

*`DEST\_STATE\_NM` varchar(70),*

*`DEST\_WAC` int DEFAULT NULL,*

*`CRS\_DEP\_TIME` int DEFAULT NULL,*

*`DEP\_TIME` int DEFAULT NULL,*

*`DEP\_DELAY` int DEFAULT NULL,*

*`DEP\_DELAY\_NEW` int DEFAULT NULL,*

*`DEP\_DEL15` int DEFAULT NULL,*

*`DEP\_DELAY\_GROUP` int DEFAULT NULL,*

*`DEP\_TIME\_BLK` varchar(70),*

*`TAXI\_OUT` int DEFAULT NULL,*

*`WHEELS\_OFF` int DEFAULT NULL,*

*`WHEELS\_ON` int DEFAULT NULL,*

*`TAXI\_IN` int DEFAULT NULL,*

*`CRS\_ARR\_TIME` int DEFAULT NULL,*

*`ARR\_TIME` int DEFAULT NULL,*

*`ARR\_DELAY` int DEFAULT NULL,*

*`ARR\_DELAY\_NEW` int DEFAULT NULL,*

*`ARR\_DEL15` int DEFAULT NULL,*

*`ARR\_DELAY\_GROUP` int DEFAULT NULL,*

*`ARR\_TIME\_BLK` varchar(70),*

*`CANCELLED` int DEFAULT NULL,*

*`CANCELLATION\_CODE` varchar(70),*

*`DIVERTED` int DEFAULT NULL,*

*`DUP` varchar(70),*

*`CRS\_ELAPSED\_TIME` int DEFAULT NULL,*

*`ACTUAL\_ELAPSED\_TIME` int DEFAULT NULL,*

*`AIR\_TIME` int DEFAULT NULL,*

*`FLIGHTS` int DEFAULT NULL,*

*`DISTANCE` int DEFAULT NULL,*

*`DISTANCE\_GROUP` int DEFAULT NULL,*

*`CARRIER\_DELAY` varchar(70),*

*`WEATHER\_DELAY` varchar(70),*

*`NAS\_DELAY` varchar(70),*

*`SECURITY\_DELAY` varchar(70),*

*`LATE\_AIRCRAFT\_DELAY` varchar(70),*

*`FIRST\_DEP\_TIME` varchar(70),*

*`TOTAL\_ADD\_GTIME` varchar(70),*

*`LONGEST\_ADD\_GTIME` varchar(70),*

*`DIV\_AIRPORT\_LANDINGS` int DEFAULT NULL,*

*`DIV\_REACHED\_DEST` varchar(70),*

*`DIV\_ACTUAL\_ELAPSED\_TIME` varchar(70),*

*`DIV\_ARR\_DELAY` varchar(70),*

*`DIV\_DISTANCE` varchar(70),*

*`DIV1\_AIRPORT` varchar(70),*

*`DIV1\_AIRPORT\_ID` varchar(70),*

*`DIV1\_AIRPORT\_SEQ\_ID` varchar(70),*

*`DIV1\_WHEELS\_ON` varchar(70),*

*`DIV1\_TOTAL\_GTIME` varchar(70),*

*`DIV1\_LONGEST\_GTIME` varchar(70),*

*`DIV1\_WHEELS\_OFF` varchar(70),*

*`DIV1\_TAIL\_NUM` varchar(70),*

*`DIV2\_AIRPORT` varchar(70),*

*`DIV2\_AIRPORT\_ID` varchar(70),*

*`DIV2\_AIRPORT\_SEQ\_ID` varchar(70),*

*`DIV2\_WHEELS\_ON` varchar(70),*

*`DIV2\_TOTAL\_GTIME` varchar(70),*

*`DIV2\_LONGEST\_GTIME` varchar(70),*

*`DIV2\_WHEELS\_OFF` varchar(70),*

*`DIV2\_TAIL\_NUM` varchar(70),*

*`DIV3\_AIRPORT` varchar(70),*

*`DIV3\_AIRPORT\_ID` varchar(70),*

*`DIV3\_AIRPORT\_SEQ\_ID` varchar(70),*

*`DIV3\_WHEELS\_ON` varchar(70),*

*`DIV3\_TOTAL\_GTIME` varchar(70),*

*`DIV3\_LONGEST\_GTIME` varchar(70),*

*`DIV3\_WHEELS\_OFF` varchar(70),*

*`DIV3\_TAIL\_NUM` varchar(70),*

*`DIV4\_AIRPORT` varchar(70),*

*`DIV4\_AIRPORT\_ID` varchar(70),*

*`DIV4\_AIRPORT\_SEQ\_ID` varchar(70),*

*`DIV4\_WHEELS\_ON` varchar(70),*

*`DIV4\_TOTAL\_GTIME` varchar(70),*

*`DIV4\_LONGEST\_GTIME` varchar(70),*

*`DIV4\_WHEELS\_OFF` varchar(70),*

*`DIV4\_TAIL\_NUM` varchar(70),*

*`DIV5\_AIRPORT` varchar(70),*

*`DIV5\_AIRPORT\_ID` varchar(70),*

*`DIV5\_AIRPORT\_SEQ\_ID` varchar(70),*

*`DIV5\_WHEELS\_ON` varchar(70),*

*`DIV5\_TOTAL\_GTIME` varchar(70),*

*`DIV5\_LONGEST\_GTIME` varchar(70),*

*`DIV5\_WHEELS\_OFF` varchar(70),*

*`DIV5\_TAIL\_NUM` varchar(70)*

*);*

**Step-5 : Loading data for the years 2020 and 2021**

Load the data for the years 2020 and 2021 into the *`airline\_marketing`* table via terminal using the following commands.

*mysql> LOAD DATA LOCAL INFILE '/Users/aishwaryasaibewar/DBMS\_Project/airline/Airline\_2020\_12.csv' INTO TABLE airline\_schema.airline\_marketing FIELDS TERMINATED BY ',' ENCLOSED BY '"' LINES TERMINATED BY '\n' IGNORE 1 ROWS;*

*Query OK, 397802 rows affected, 65535 warnings (6.23 sec)*

*Records: 397802 Deleted: 0 Skipped: 0 Warnings: 820258*

*mysql> LOAD DATA LOCAL INFILE '/Users/aishwaryasaibewar/DBMS\_Project/airline/Airline\_2021\_01.csv' INTO TABLE airline\_schema.airline\_marketing FIELDS TERMINATED BY ',' ENCLOSED BY '"' LINES TERMINATED BY '\n' IGNORE 1 ROWS;*

*Query OK, 379384 rows affected, 65535 warnings (5.51 sec)*

*Records: 379384 Deleted: 0 Skipped: 0 Warnings: 69717*

*A screenshot of a computer

Description automatically generated with medium confidence*

**Step-6: Join all the years data into a single table**

Assemble the ‘airline\_reporting’ data with “airline\_marketing” data into a single table AIRLINE\_DATA by uniting the fields which are not present in Marketing carrier with a NULL value.

Use the following command to unite two tables into a single table

*CREATE TABLE airline\_schema.AIRLINE\_DATA AS*

*SELECT `airline\_marketing`.`YEAR`,*

*`airline\_marketing`.`QUARTER`,*

*`airline\_marketing`.`MONTH`,*

*`airline\_marketing`.`DAY\_OF\_MONTH`,*

*`airline\_marketing`.`DAY\_OF\_WEEK`,*

*`airline\_marketing`.`FL\_DATE`,*

*`airline\_marketing`.`MKT\_UNIQUE\_CARRIER`,*

*`airline\_marketing`.`BRANDED\_CODE\_SHARE`,*

*`airline\_marketing`.`MKT\_CARRIER\_AIRLINE\_ID`,*

*`airline\_marketing`.`MKT\_CARRIER`,*

*`airline\_marketing`.`MKT\_CARRIER\_FL\_NUM`,*

*`airline\_marketing`.`SCH\_OP\_UNIQUE\_CARRIER`,*

*`airline\_marketing`.`SCH\_OP\_CARRIER\_AIRLINE\_ID`,*

*`airline\_marketing`.`SCH\_OP\_CARRIER`,*

*`airline\_marketing`.`SCH\_OP\_CARRIER\_FL\_NUM`,*

*`airline\_marketing`.`OP\_UNIQUE\_CARRIER`,*

*`airline\_marketing`.`OP\_CARRIER\_AIRLINE\_ID`,*

*`airline\_marketing`.`OP\_CARRIER`,*

*`airline\_marketing`.`TAIL\_NUM`,*

*`airline\_marketing`.`OP\_CARRIER\_FL\_NUM`,*

*`airline\_marketing`.`ORIGIN\_AIRPORT\_ID`,*

*`airline\_marketing`.`ORIGIN\_AIRPORT\_SEQ\_ID`,*

*`airline\_marketing`.`ORIGIN\_CITY\_MARKET\_ID`,*

*`airline\_marketing`.`ORIGIN`,*

*`airline\_marketing`.`ORIGIN\_CITY\_NAME`,*

*`airline\_marketing`.`ORIGIN\_STATE\_ABR`,*

*`airline\_marketing`.`ORIGIN\_STATE\_FIPS`,*

*`airline\_marketing`.`ORIGIN\_STATE\_NM`,*

*`airline\_marketing`.`ORIGIN\_WAC`,*

*`airline\_marketing`.`DEST\_AIRPORT\_ID`,*

*`airline\_marketing`.`DEST\_AIRPORT\_SEQ\_ID`,*

*`airline\_marketing`.`DEST\_CITY\_MARKET\_ID`,*

*`airline\_marketing`.`DEST`,*

*`airline\_marketing`.`DEST\_CITY\_NAME`,*

*`airline\_marketing`.`DEST\_STATE\_ABR`,*

*`airline\_marketing`.`DEST\_STATE\_FIPS`,*

*`airline\_marketing`.`DEST\_STATE\_NM`,*

*`airline\_marketing`.`DEST\_WAC`,*

*`airline\_marketing`.`CRS\_DEP\_TIME`,*

*`airline\_marketing`.`DEP\_TIME`,*

*`airline\_marketing`.`DEP\_DELAY`,*

*`airline\_marketing`.`DEP\_DELAY\_NEW`,*

*`airline\_marketing`.`DEP\_DEL15`,*

*`airline\_marketing`.`DEP\_DELAY\_GROUP`,*

*`airline\_marketing`.`DEP\_TIME\_BLK`,*

*`airline\_marketing`.`TAXI\_OUT`,*

*`airline\_marketing`.`WHEELS\_OFF`,*

*`airline\_marketing`.`WHEELS\_ON`,*

*`airline\_marketing`.`TAXI\_IN`,*

*`airline\_marketing`.`CRS\_ARR\_TIME`,*

*`airline\_marketing`.`ARR\_TIME`,*

*`airline\_marketing`.`ARR\_DELAY`,*

*`airline\_marketing`.`ARR\_DELAY\_NEW`,*

*`airline\_marketing`.`ARR\_DEL15`,*

*`airline\_marketing`.`ARR\_DELAY\_GROUP`,*

*`airline\_marketing`.`ARR\_TIME\_BLK`,*

*`airline\_marketing`.`CANCELLED`,*

*`airline\_marketing`.`CANCELLATION\_CODE`,*

*`airline\_marketing`.`DIVERTED`,*

*`airline\_marketing`.`DUP`,*

*`airline\_marketing`.`CRS\_ELAPSED\_TIME`,*

*`airline\_marketing`.`ACTUAL\_ELAPSED\_TIME`,*

*`airline\_marketing`.`AIR\_TIME`,*

*`airline\_marketing`.`FLIGHTS`,*

*`airline\_marketing`.`DISTANCE`,*

*`airline\_marketing`.`DISTANCE\_GROUP`,*

*`airline\_marketing`.`CARRIER\_DELAY`,*

*`airline\_marketing`.`WEATHER\_DELAY`,*

*`airline\_marketing`.`NAS\_DELAY`,*

*`airline\_marketing`.`SECURITY\_DELAY`,*

*`airline\_marketing`.`LATE\_AIRCRAFT\_DELAY`,*

*`airline\_marketing`.`FIRST\_DEP\_TIME`,*

*`airline\_marketing`.`TOTAL\_ADD\_GTIME`,*

*`airline\_marketing`.`LONGEST\_ADD\_GTIME`,*

*`airline\_marketing`.`DIV\_AIRPORT\_LANDINGS`,*

*`airline\_marketing`.`DIV\_REACHED\_DEST`,*

*`airline\_marketing`.`DIV\_ACTUAL\_ELAPSED\_TIME`,*

*`airline\_marketing`.`DIV\_ARR\_DELAY`,*

*`airline\_marketing`.`DIV\_DISTANCE`,*

*`airline\_marketing`.`DIV1\_AIRPORT`,*

*`airline\_marketing`.`DIV1\_AIRPORT\_ID`,*

*`airline\_marketing`.`DIV1\_AIRPORT\_SEQ\_ID`,*

*`airline\_marketing`.`DIV1\_WHEELS\_ON`,*

*`airline\_marketing`.`DIV1\_TOTAL\_GTIME`,*

*`airline\_marketing`.`DIV1\_LONGEST\_GTIME`,*

*`airline\_marketing`.`DIV1\_WHEELS\_OFF`,*

*`airline\_marketing`.`DIV1\_TAIL\_NUM`,*

*`airline\_marketing`.`DIV2\_AIRPORT`,*

*`airline\_marketing`.`DIV2\_AIRPORT\_ID`,*

*`airline\_marketing`.`DIV2\_AIRPORT\_SEQ\_ID`,*

*`airline\_marketing`.`DIV2\_WHEELS\_ON`,*

*`airline\_marketing`.`DIV2\_TOTAL\_GTIME`,*

*`airline\_marketing`.`DIV2\_LONGEST\_GTIME`,*

*`airline\_marketing`.`DIV2\_WHEELS\_OFF`,*

*`airline\_marketing`.`DIV2\_TAIL\_NUM`,*

*`airline\_marketing`.`DIV3\_AIRPORT`,*

*`airline\_marketing`.`DIV3\_AIRPORT\_ID`,*

*`airline\_marketing`.`DIV3\_AIRPORT\_SEQ\_ID`,*

*`airline\_marketing`.`DIV3\_WHEELS\_ON`,*

*`airline\_marketing`.`DIV3\_TOTAL\_GTIME`,*

*`airline\_marketing`.`DIV3\_LONGEST\_GTIME`,*

*`airline\_marketing`.`DIV3\_WHEELS\_OFF`,*

*`airline\_marketing`.`DIV3\_TAIL\_NUM`,*

*`airline\_marketing`.`DIV4\_AIRPORT`,*

*`airline\_marketing`.`DIV4\_AIRPORT\_ID`,*

*`airline\_marketing`.`DIV4\_AIRPORT\_SEQ\_ID`,*

*`airline\_marketing`.`DIV4\_WHEELS\_ON`,*

*`airline\_marketing`.`DIV4\_TOTAL\_GTIME`,*

*`airline\_marketing`.`DIV4\_LONGEST\_GTIME`,*

*`airline\_marketing`.`DIV4\_WHEELS\_OFF`,*

*`airline\_marketing`.`DIV4\_TAIL\_NUM`,*

*`airline\_marketing`.`DIV5\_AIRPORT`,*

*`airline\_marketing`.`DIV5\_AIRPORT\_ID`,*

*`airline\_marketing`.`DIV5\_AIRPORT\_SEQ\_ID`,*

*`airline\_marketing`.`DIV5\_WHEELS\_ON`,*

*`airline\_marketing`.`DIV5\_TOTAL\_GTIME`,*

*`airline\_marketing`.`DIV5\_LONGEST\_GTIME`,*

*`airline\_marketing`.`DIV5\_WHEELS\_OFF`,*

*`airline\_marketing`.`DIV5\_TAIL\_NUM`*

*FROM `airline\_schema`.`airline\_marketing`*

*UNION*

*SELECT `airline\_reporting`.`YEAR`,*

*`airline\_reporting`.`QUARTER`,*

*`airline\_reporting`.`MONTH`,*

*`airline\_reporting`.`DAY\_OF\_MONTH`,*

*`airline\_reporting`.`DAY\_OF\_WEEK`,*

*`airline\_reporting`.`FL\_DATE`,*

*NULL AS `MKT\_UNIQUE\_CARRIER`,*

*NULL AS `BRANDED\_CODE\_SHARE`,*

*NULL AS `MKT\_CARRIER\_AIRLINE\_ID`,*

*NULL AS `MKT\_CARRIER`,*

*NULL AS `MKT\_CARRIER\_FL\_NUM`,*

*NULL AS `SCH\_OP\_UNIQUE\_CARRIER`,*

*NULL AS `SCH\_OP\_CARRIER\_AIRLINE\_ID`,*

*NULL AS `SCH\_OP\_CARRIER`,*

*NULL AS `SCH\_OP\_CARRIER\_FL\_NUM`,*

*`airline\_reporting`.`OP\_UNIQUE\_CARRIER`,*

*`airline\_reporting`.`OP\_CARRIER\_AIRLINE\_ID`,*

*`airline\_reporting`.`OP\_CARRIER`,*

*`airline\_reporting`.`TAIL\_NUM`,*

*`airline\_reporting`.`OP\_CARRIER\_FL\_NUM`,*

*`airline\_reporting`.`ORIGIN\_AIRPORT\_ID`,*

*`airline\_reporting`.`ORIGIN\_AIRPORT\_SEQ\_ID`,*

*`airline\_reporting`.`ORIGIN\_CITY\_MARKET\_ID`,*

*`airline\_reporting`.`ORIGIN`,*

*`airline\_reporting`.`ORIGIN\_CITY\_NAME`,*

*`airline\_reporting`.`ORIGIN\_STATE\_ABR`,*

*`airline\_reporting`.`ORIGIN\_STATE\_FIPS`,*

*`airline\_reporting`.`ORIGIN\_STATE\_NM`,*

*`airline\_reporting`.`ORIGIN\_WAC`,*

*`airline\_reporting`.`DEST\_AIRPORT\_ID`,*

*`airline\_reporting`.`DEST\_AIRPORT\_SEQ\_ID`,*

*`airline\_reporting`.`DEST\_CITY\_MARKET\_ID`,*

*`airline\_reporting`.`DEST`,*

*`airline\_reporting`.`DEST\_CITY\_NAME`,*

*`airline\_reporting`.`DEST\_STATE\_ABR`,*

*`airline\_reporting`.`DEST\_STATE\_FIPS`,*

*`airline\_reporting`.`DEST\_STATE\_NM`,*

*`airline\_reporting`.`DEST\_WAC`,*

*`airline\_reporting`.`CRS\_DEP\_TIME`,*

*`airline\_reporting`.`DEP\_TIME`,*

*`airline\_reporting`.`DEP\_DELAY`,*

*`airline\_reporting`.`DEP\_DELAY\_NEW`,*

*`airline\_reporting`.`DEP\_DEL15`,*

*`airline\_reporting`.`DEP\_DELAY\_GROUP`,*

*`airline\_reporting`.`DEP\_TIME\_BLK`,*

*`airline\_reporting`.`TAXI\_OUT`,*

*`airline\_reporting`.`WHEELS\_OFF`,*

*`airline\_reporting`.`WHEELS\_ON`,*

*`airline\_reporting`.`TAXI\_IN`,*

*`airline\_reporting`.`CRS\_ARR\_TIME`,*

*`airline\_reporting`.`ARR\_TIME`,*

*`airline\_reporting`.`ARR\_DELAY`,*

*`airline\_reporting`.`ARR\_DELAY\_NEW`,*

*`airline\_reporting`.`ARR\_DEL15`,*

*`airline\_reporting`.`ARR\_DELAY\_GROUP`,*

*`airline\_reporting`.`ARR\_TIME\_BLK`,*

*`airline\_reporting`.`CANCELLED`,*

*`airline\_reporting`.`CANCELLATION\_CODE`,*

*`airline\_reporting`.`DIVERTED`,*

*NULL AS `DUP`,*

*`airline\_reporting`.`CRS\_ELAPSED\_TIME`,*

*`airline\_reporting`.`ACTUAL\_ELAPSED\_TIME`,*

*`airline\_reporting`.`AIR\_TIME`,*

*`airline\_reporting`.`FLIGHTS`,*

*`airline\_reporting`.`DISTANCE`,*

*`airline\_reporting`.`DISTANCE\_GROUP`,*

*`airline\_reporting`.`CARRIER\_DELAY`,*

*`airline\_reporting`.`WEATHER\_DELAY`,*

*`airline\_reporting`.`NAS\_DELAY`,*

*`airline\_reporting`.`SECURITY\_DELAY`,*

*`airline\_reporting`.`LATE\_AIRCRAFT\_DELAY`,*

*`airline\_reporting`.`FIRST\_DEP\_TIME`,*

*`airline\_reporting`.`TOTAL\_ADD\_GTIME`,*

*`airline\_reporting`.`LONGEST\_ADD\_GTIME`,*

*`airline\_reporting`.`DIV\_AIRPORT\_LANDINGS`,*

*`airline\_reporting`.`DIV\_REACHED\_DEST`,*

*`airline\_reporting`.`DIV\_ACTUAL\_ELAPSED\_TIME`,*

*`airline\_reporting`.`DIV\_ARR\_DELAY`,*

*`airline\_reporting`.`DIV\_DISTANCE`,*

*`airline\_reporting`.`DIV1\_AIRPORT`,*

*`airline\_reporting`.`DIV1\_AIRPORT\_ID`,*

*`airline\_reporting`.`DIV1\_AIRPORT\_SEQ\_ID`,*

*`airline\_reporting`.`DIV1\_WHEELS\_ON`,*

*`airline\_reporting`.`DIV1\_TOTAL\_GTIME`,*

*`airline\_reporting`.`DIV1\_LONGEST\_GTIME`,*

*`airline\_reporting`.`DIV1\_WHEELS\_OFF`,*

*`airline\_reporting`.`DIV1\_TAIL\_NUM`,*

*`airline\_reporting`.`DIV2\_AIRPORT`,*

*`airline\_reporting`.`DIV2\_AIRPORT\_ID`,*

*`airline\_reporting`.`DIV2\_AIRPORT\_SEQ\_ID`,*

*`airline\_reporting`.`DIV2\_WHEELS\_ON`,*

*`airline\_reporting`.`DIV2\_TOTAL\_GTIME`,*

*`airline\_reporting`.`DIV2\_LONGEST\_GTIME`,*

*`airline\_reporting`.`DIV2\_WHEELS\_OFF`,*

*`airline\_reporting`.`DIV2\_TAIL\_NUM`,*

*`airline\_reporting`.`DIV3\_AIRPORT`,*

*`airline\_reporting`.`DIV3\_AIRPORT\_ID`,*

*`airline\_reporting`.`DIV3\_AIRPORT\_SEQ\_ID`,*

*`airline\_reporting`.`DIV3\_WHEELS\_ON`,*

*`airline\_reporting`.`DIV3\_TOTAL\_GTIME`,*

*`airline\_reporting`.`DIV3\_LONGEST\_GTIME`,*

*`airline\_reporting`.`DIV3\_WHEELS\_OFF`,*

*`airline\_reporting`.`DIV3\_TAIL\_NUM`,*

*`airline\_reporting`.`DIV4\_AIRPORT`,*

*`airline\_reporting`.`DIV4\_AIRPORT\_ID`,*

*`airline\_reporting`.`DIV4\_AIRPORT\_SEQ\_ID`,*

*`airline\_reporting`.`DIV4\_WHEELS\_ON`,*

*`airline\_reporting`.`DIV4\_TOTAL\_GTIME`,*

*`airline\_reporting`.`DIV4\_LONGEST\_GTIME`,*

*`airline\_reporting`.`DIV4\_WHEELS\_OFF`,*

*`airline\_reporting`.`DIV4\_TAIL\_NUM`,*

*`airline\_reporting`.`DIV5\_AIRPORT`,*

*`airline\_reporting`.`DIV5\_AIRPORT\_ID`,*

*`airline\_reporting`.`DIV5\_AIRPORT\_SEQ\_ID`,*

*`airline\_reporting`.`DIV5\_WHEELS\_ON`,*

*`airline\_reporting`.`DIV5\_TOTAL\_GTIME`,*

*`airline\_reporting`.`DIV5\_LONGEST\_GTIME`,*

*`airline\_reporting`.`DIV5\_WHEELS\_OFF`,*

*`airline\_reporting`.`DIV5\_TAIL\_NUM`*

*FROM `airline\_schema`.`airline\_reporting`*

*;*

**Step 7: Import all the other tables except the years table using the table data import wizard.**

As these dataset sizes are much less compared to years datasets and we won’t be losing any data we can import them using table data import wizard.

Graphical user interface, text, application

Description automatically generated

Graphical user interface, text, application

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

**Step-8: Creating tables using MySQL queries**

* Creating the schema **airline\_data\_latest\_schema** (Note: The schema name needs to be in lower case as there will be a problem in generating dump file if its in upper case)

*create schema airline\_schema;*

*use airline\_schema;*

* Creating a backup of the table which has the data of all the years.

*Create table AIRLINE\_BACKUP as*

*select \**

*FROM airline \_schema.AIRLINE\_DATA;*

**Tables Creation:**

1. Creating **AIRLINE** table:

We have all the data in airline\_data table. Select the columns that are required in creating AIRLINE table and use the below command for it.

*Create table AIRLINE as*

*select FL\_DATE,OP\_CARRIER\_AIRLINE\_ID,TAIL\_NUM,OP\_CARRIER\_FL\_NUM,ORIGIN\_AIRPORT\_SEQ\_ID,DEST\_AIRPORT\_SEQ\_ID*

*FROM airline\_schema.AIRLINE\_DATA;*

1. Creating **AIRPORT** table:

*Create table AIRPORT as*

*select DISTINCT ORIGIN\_AIRPORT\_SEQ\_ID, ORIGIN\_CITY\_MARKET\_ID,ORIGIN,ORIGIN\_CITY\_NAME,ORIGIN\_STATE\_FIPS*

*FROM airline\_schema.AIRLINE\_DATA;*

*ALTER TABLE AIRPORT RENAME COLUMN ORIGIN\_AIRPORT\_SEQ\_ID TO AIRPORT\_SEQ\_ID;*

*ALTER TABLE AIRPORT RENAME COLUMN ORIGIN\_CITY\_MARKET\_ID TO CITY\_MARKET\_ID;*

*ALTER TABLE AIRPORT RENAME COLUMN ORIGIN TO AIRPORT\_CODE;*

*ALTER TABLE AIRPORT RENAME COLUMN ORIGIN\_CITY\_NAME TO CITY\_NAME;*

*ALTER TABLE AIRPORT RENAME COLUMN ORIGIN\_STATE\_FIPS TO STATE\_FIPS;*

1. Creating **STATE** table:

*Create table STATE as*

*select DISTINCT ORIGIN\_STATE\_FIPS, ORIGIN\_STATE\_ABR, ORIGIN\_STATE\_NM, ORIGIN\_WAC*

*FROM airline\_schema.AIRLINE\_DATA;*

*ALTER TABLE STATE RENAME COLUMN ORIGIN\_STATE\_FIPS TO STATE\_FIPS;*

*ALTER TABLE STATE RENAME COLUMN ORIGIN\_STATE\_ABR TO STATE\_ABR;*

*ALTER TABLE STATE RENAME COLUMN ORIGIN\_STATE\_NM TO STATE\_NM;*

*ALTER TABLE STATE RENAME COLUMN ORIGIN\_WAC TO WAC;*

1. Creating **CARRIER** table:

*Create table CARRIER as*

*select distinct OP\_CARRIER\_AIRLINE\_ID, OP\_UNIQUE\_CARRIER*

*FROM airline\_schema.AIRLINE\_DATA;*

1. Creating **DEPARTURE\_PERFORMANCE** table:

*CREATE TABLE DEPARTURE\_PERFORMANCE AS*

*SELECT FL\_DATE,*

*OP\_CARRIER\_FL\_NUM,*

*OP\_CARRIER\_AIRLINE\_ID,*

*ORIGIN\_AIRPORT\_SEQ\_ID,*

*DEST\_AIRPORT\_SEQ\_ID,*

*str\_to\_date(case when CRS\_DEP\_TIME = 2400 then '0000'*

*when length(cast(CRS\_DEP\_TIME as char)) = 4 then cast(CRS\_DEP\_TIME as char)*

*when length(cast(CRS\_DEP\_TIME as char)) = 3 then concat("0",cast(CRS\_DEP\_TIME as char))*

*when length(cast(CRS\_DEP\_TIME as char)) = 2 then concat("00",cast(CRS\_DEP\_TIME as char))*

*when length(cast(CRS\_DEP\_TIME as char)) = 1 then concat("000",cast(CRS\_DEP\_TIME as char))*

*end, '%H%i' )as CRS\_DEP\_TIME,*

*str\_to\_date(case when DEP\_TIME = 2400 then '0000'*

*when length(cast(DEP\_TIME as char)) = 4 then cast(DEP\_TIME as char)*

*when length(cast(DEP\_TIME as char)) = 3 then concat("0",cast(DEP\_TIME as char))*

*when length(cast(DEP\_TIME as char)) = 2 then concat("00",cast(DEP\_TIME as char))*

*when length(cast(DEP\_TIME as char)) = 1 then concat("000",cast(DEP\_TIME as char))*

*end, '%H%i' )as DEP\_TIME,*

*TAXI\_OUT,*

*WHEELS\_OFF*

*FROM airline\_schema.AIRLINE\_DATA;*

1. Creating **ARRIVAL\_PERFORMANCE** table:

*CREATE TABLE ARRIVAL\_PERFORMANCE AS*

*SELECT FL\_DATE,*

*OP\_CARRIER\_FL\_NUM,*

*OP\_CARRIER\_AIRLINE\_ID,*

*ORIGIN\_AIRPORT\_SEQ\_ID,*

*DEST\_AIRPORT\_SEQ\_ID,*

*str\_to\_date(case when CRS\_ARR\_TIME = 2400 then '0000'*

*when length(cast(CRS\_ARR\_TIME as char)) = 4 then cast(CRS\_ARR\_TIME as char)*

*when length(cast(CRS\_ARR\_TIME as char)) = 3 then concat("0",cast(CRS\_ARR\_TIME as char))*

*when length(cast(CRS\_ARR\_TIME as char)) = 2 then concat("00",cast(CRS\_ARR\_TIME as char))*

*when length(cast(CRS\_ARR\_TIME as char)) = 1 then concat("000",cast(CRS\_ARR\_TIME as char))*

*end, '%H%i') as CRS\_ARR\_TIME,*

*str\_to\_date(case when ARR\_TIME = 2400 then '0000'*

*when length(cast(ARR\_TIME as char)) = 4 then cast(ARR\_TIME as char)*

*when length(cast(ARR\_TIME as char)) = 3 then concat("0",cast(ARR\_TIME as char))*

*when length(cast(ARR\_TIME as char)) = 2 then concat("00",cast(ARR\_TIME as char))*

*when length(cast(ARR\_TIME as char)) = 1 then concat("000",cast(ARR\_TIME as char))*

*end, '%H%i' )as ARR\_TIME,*

*TAXI\_IN,*

*WHEELS\_ON*

*FROM airline\_schema.AIRLINE\_DATA;*

1. Creating **CANCEL\_DIVERT** table:

*CREATE TABLE CANCEL\_DIVERT AS*

*SELECT FL\_DATE,OP\_CARRIER\_FL\_NUM,OP\_CARRIER\_AIRLINE\_ID,ORIGIN\_AIRPORT\_SEQ\_ID,DEST\_AIRPORT\_SEQ\_ID, CANCELLED,CANCELLATION\_CODE,DIVERTED*

*FROM airline\_schema.AIRLINE\_DATA;*

1. Creating **FLIGHT\_SUMMARIES** table:

*CREATE TABLE FLIGHT\_SUMMARIES AS*

*SELECT FL\_DATE,OP\_CARRIER\_FL\_NUM,OP\_CARRIER\_AIRLINE\_ID,ORIGIN\_AIRPORT\_SEQ\_ID,DEST\_AIRPORT\_SEQ\_ID,CRS\_ELAPSED\_TIME, ACTUAL\_ELAPSED\_TIME, AIR\_TIME,FLIGHTS,DISTANCE,DISTANCE\_GROUP*

*FROM airline\_schema.AIRLINE\_DATA;*

*ALTER TABLE FLIGHT\_SUMMARIES*

*ADD PRIMARY KEY (FL\_DATE,OP\_CARRIER\_FL\_NUM,OP\_CARRIER\_AIRLINE\_ID,ORIGIN\_AIRPORT\_SEQ\_ID,DEST\_AIRPORT\_SEQ\_ID);*

1. Creating **DELAY** table:

*CREATE TABLE DELAY AS*

*SELECT FL\_DATE,OP\_CARRIER\_FL\_NUM,OP\_CARRIER\_AIRLINE\_ID,ORIGIN\_AIRPORT\_SEQ\_ID,DEST\_AIRPORT\_SEQ\_ID, CARRIER\_DELAY,WEATHER\_DELAY, NAS\_DELAY, SECURITY\_DELAY, LATE\_AIRCRAFT\_DELAY*

*FROM airline\_schema.AIRLINE\_DATA;*

1. Creating **GATE\_INFO** table:

*CREATE TABLE GATE\_INFO AS*

*SELECT FL\_DATE,OP\_CARRIER\_FL\_NUM,OP\_CARRIER\_AIRLINE\_ID,ORIGIN\_AIRPORT\_SEQ\_ID,DEST\_AIRPORT\_SEQ\_ID,FIRST\_DEP\_TIME,TOTAL\_ADD\_GTIME,LONGEST\_ADD\_GTIME*

*FROM airline\_schema.AIRLINE\_DATA;*

1. Creating **DIVERTED\_INFO** table:

*CREATE TABLE DIVERTED\_INFO AS*

*SELECT FL\_DATE,OP\_CARRIER\_FL\_NUM,OP\_CARRIER\_AIRLINE\_ID,ORIGIN\_AIRPORT\_SEQ\_ID,DEST\_AIRPORT\_SEQ\_ID,DIVERTED,DIV\_AIRPORT\_LANDINGS,DIV\_REACHED\_DEST,DIV\_ACTUAL\_ELAPSED\_TIME,DIV\_ARR\_DELAY,DIV\_DISTANCE,DIV1\_AIRPORT,DIV1\_AIRPORT\_ID,DIV1\_AIRPORT\_SEQ\_ID,DIV1\_WHEELS\_ON,DIV1\_TOTAL\_GTIME,DIV1\_LONGEST\_GTIME,DIV1\_WHEELS\_OFF,DIV1\_TAIL\_NUM,DIV2\_AIRPORT,DIV2\_AIRPORT\_ID,DIV2\_AIRPORT\_SEQ\_ID,DIV2\_WHEELS\_ON,DIV2\_TOTAL\_GTIME,DIV2\_LONGEST\_GTIME,DIV2\_WHEELS\_OFF,DIV2\_TAIL\_NUM,DIV3\_AIRPORT,DIV3\_AIRPORT\_ID,DIV3\_AIRPORT\_SEQ\_ID,DIV3\_WHEELS\_ON,DIV3\_TOTAL\_GTIME,DIV3\_LONGEST\_GTIME,DIV3\_WHEELS\_OFF,DIV3\_TAIL\_NUM,DIV4\_AIRPORT,DIV4\_AIRPORT\_ID,DIV4\_AIRPORT\_SEQ\_ID,DIV4\_WHEELS\_ON,DIV4\_TOTAL\_GTIME,DIV4\_LONGEST\_GTIME,DIV4\_WHEELS\_OFF,DIV4\_TAIL\_NUM,DIV5\_AIRPORT,DIV5\_AIRPORT\_ID,DIV5\_AIRPORT\_SEQ\_ID,DIV5\_WHEELS\_ON,DIV5\_TOTAL\_GTIME,DIV5\_LONGEST\_GTIME,DIV5\_WHEELS\_OFF,DIV5\_TAIL\_NUM*

*FROM airline\_schema.AIRLINE\_DATA;*

**Step-9: INCLUDING PRIMARY KEYS**

Note: If we have this combination of primary keys then we get a unique value

(FL\_DATE,OP\_CARRIER\_FL\_NUM,OP\_CARRIER\_AIRLINE\_ID,ORIGIN\_AIRPORT\_SEQ\_ID,DEST\_AIRPORT\_SEQ\_ID)

*ALTER TABLE AIRLINE*

*ADD PRIMARY KEY (FL\_DATE,OP\_CARRIER\_FL\_NUM,OP\_CARRIER\_AIRLINE\_ID,ORIGIN\_AIRPORT\_SEQ\_ID,DEST\_AIRPORT\_SEQ\_ID);*

*ALTER TABLE AIRPORT*

*ADD PRIMARY KEY (AIRPORT\_SEQ\_ID);*

*ALTER TABLE STATE*

*ADD PRIMARY KEY (STATE\_FIPS);*

*ALTER TABLE CARRIER*

*ADD PRIMARY KEY (OP\_CARRIER\_AIRLINE\_ID);*

*ALTER TABLE DEPARTURE\_PERFORMANCE*

*ADD PRIMARY KEY (FL\_DATE,OP\_CARRIER\_FL\_NUM,OP\_CARRIER\_AIRLINE\_ID,ORIGIN\_AIRPORT\_SEQ\_ID,DEST\_AIRPORT\_SEQ\_ID);*

*ALTER TABLE ARRIVAL\_PERFORMANCE*

*ADD PRIMARY KEY (FL\_DATE,OP\_CARRIER\_FL\_NUM,OP\_CARRIER\_AIRLINE\_ID,ORIGIN\_AIRPORT\_SEQ\_ID,DEST\_AIRPORT\_SEQ\_ID);*

*ALTER TABLE CANCEL\_DIVERT*

*ADD PRIMARY KEY (FL\_DATE,OP\_CARRIER\_FL\_NUM,OP\_CARRIER\_AIRLINE\_ID,ORIGIN\_AIRPORT\_SEQ\_ID,DEST\_AIRPORT\_SEQ\_ID);*

*ALTER TABLE FLIGHT\_SUMMARIES*

*ADD PRIMARY KEY (FL\_DATE,OP\_CARRIER\_FL\_NUM,OP\_CARRIER\_AIRLINE\_ID,ORIGIN\_AIRPORT\_SEQ\_ID,DEST\_AIRPORT\_SEQ\_ID);*

*ALTER TABLE DELAY*

*ADD PRIMARY KEY (FL\_DATE,OP\_CARRIER\_FL\_NUM,OP\_CARRIER\_AIRLINE\_ID,ORIGIN\_AIRPORT\_SEQ\_ID,DEST\_AIRPORT\_SEQ\_ID);*

*ALTER TABLE GATE\_INFO*

*ADD PRIMARY KEY (FL\_DATE,OP\_CARRIER\_FL\_NUM,OP\_CARRIER\_AIRLINE\_ID,ORIGIN\_AIRPORT\_SEQ\_ID,DEST\_AIRPORT\_SEQ\_ID);*

*ALTER TABLE DIVERTED\_INFO*

*ADD PRIMARY KEY (FL\_DATE,OP\_CARRIER\_FL\_NUM,OP\_CARRIER\_AIRLINE\_ID,ORIGIN\_AIRPORT\_SEQ\_ID,DEST\_AIRPORT\_SEQ\_ID);*

**Step-10: INCLUDING FOREIGN KEYS**

-- Adding Foreign keys in AIRLINE table

*ALTER TABLE AIRLINE*

*ADD FOREIGN KEY (ORIGIN\_AIRPORT\_SEQ\_ID) REFERENCES AIRPORT(AIRPORT\_SEQ\_ID);*

*ALTER TABLE AIRLINE*

*ADD FOREIGN KEY (DEST\_AIRPORT\_SEQ\_ID) REFERENCES AIRPORT(AIRPORT\_SEQ\_ID);*

*ALTER TABLE AIRLINE*

*ADD FOREIGN KEY (OP\_CARRIER\_AIRLINE\_ID) REFERENCES CARRIER(OP\_CARRIER\_AIRLINE\_ID);*

-- Adding Foreign keys in AIRPORT table

*ALTER TABLE AIRPORT*

*ADD FOREIGN KEY (STATE\_FIPS) REFERENCES STATE(STATE\_FIPS);*

*--* Adding Foreign keys in AIRLINE table

*ALTER TABLE AIRLINE*

*ADD FOREIGN KEY (FL\_DATE,OP\_CARRIER\_FL\_NUM,OP\_CARRIER\_AIRLINE\_ID,ORIGIN\_AIRPORT\_SEQ\_ID,DEST\_AIRPORT\_SEQ\_ID) REFERENCES ARRIVAL\_PERFORMANCE(FL\_DATE,OP\_CARRIER\_FL\_NUM,OP\_CARRIER\_AIRLINE\_ID,ORIGIN\_AIRPORT\_SEQ\_ID,DEST\_AIRPORT\_SEQ\_ID);*

*ALTER TABLE AIRLINE*

*ADD FOREIGN KEY (FL\_DATE,OP\_CARRIER\_FL\_NUM,OP\_CARRIER\_AIRLINE\_ID,ORIGIN\_AIRPORT\_SEQ\_ID,DEST\_AIRPORT\_SEQ\_ID) REFERENCES DEPARTURE\_PERFORMANCE(FL\_DATE,OP\_CARRIER\_FL\_NUM,OP\_CARRIER\_AIRLINE\_ID,ORIGIN\_AIRPORT\_SEQ\_ID,DEST\_AIRPORT\_SEQ\_ID);*

*ALTER TABLE AIRLINE*

*ADD FOREIGN KEY (FL\_DATE,OP\_CARRIER\_FL\_NUM,OP\_CARRIER\_AIRLINE\_ID,ORIGIN\_AIRPORT\_SEQ\_ID,DEST\_AIRPORT\_SEQ\_ID) REFERENCES CANCEL\_DIVERT(FL\_DATE,OP\_CARRIER\_FL\_NUM,OP\_CARRIER\_AIRLINE\_ID,ORIGIN\_AIRPORT\_SEQ\_ID,DEST\_AIRPORT\_SEQ\_ID);*

*ALTER TABLE AIRLINE*

*ADD FOREIGN KEY (FL\_DATE,OP\_CARRIER\_FL\_NUM,OP\_CARRIER\_AIRLINE\_ID,ORIGIN\_AIRPORT\_SEQ\_ID,DEST\_AIRPORT\_SEQ\_ID) REFERENCES FLIGHT\_SUMMARIES(FL\_DATE,OP\_CARRIER\_FL\_NUM,OP\_CARRIER\_AIRLINE\_ID,ORIGIN\_AIRPORT\_SEQ\_ID,DEST\_AIRPORT\_SEQ\_ID);*

*ALTER TABLE AIRLINE*

*ADD FOREIGN KEY (FL\_DATE,OP\_CARRIER\_FL\_NUM,OP\_CARRIER\_AIRLINE\_ID,ORIGIN\_AIRPORT\_SEQ\_ID,DEST\_AIRPORT\_SEQ\_ID) REFERENCES DELAY(FL\_DATE,OP\_CARRIER\_FL\_NUM,OP\_CARRIER\_AIRLINE\_ID,ORIGIN\_AIRPORT\_SEQ\_ID,DEST\_AIRPORT\_SEQ\_ID);*

*ALTER TABLE AIRLINE*

*ADD FOREIGN KEY (FL\_DATE,OP\_CARRIER\_FL\_NUM,OP\_CARRIER\_AIRLINE\_ID,ORIGIN\_AIRPORT\_SEQ\_ID,DEST\_AIRPORT\_SEQ\_ID) REFERENCES GATE\_INFO(FL\_DATE,OP\_CARRIER\_FL\_NUM,OP\_CARRIER\_AIRLINE\_ID,ORIGIN\_AIRPORT\_SEQ\_ID,DEST\_AIRPORT\_SEQ\_ID);*

*ALTER TABLE AIRLINE*

*ADD FOREIGN KEY (FL\_DATE,OP\_CARRIER\_FL\_NUM,OP\_CARRIER\_AIRLINE\_ID,ORIGIN\_AIRPORT\_SEQ\_ID,DEST\_AIRPORT\_SEQ\_ID) REFERENCES DIVERTED\_INFO(FL\_DATE,OP\_CARRIER\_FL\_NUM,OP\_CARRIER\_AIRLINE\_ID,ORIGIN\_AIRPORT\_SEQ\_ID,DEST\_AIRPORT\_SEQ\_ID);*

**Step-11: Queries for the deleted columns**

These are the columns that were deleted as they can be calculated rather than stored. So they are not included in the database.

Query to fetch dep\_delay,arr\_delay, DepDel15,ArrDel15:

*select*

*FL\_DATE,*

*OP\_CARRIER\_FL\_NUM,*

*OP\_CARRIER\_AIRLINE\_ID,*

*ORIGIN\_AIRPORT\_ID,*

*DEST\_AIRPORT\_ID,*

*dep\_time,*

*crs\_dep\_time,*

*case when crs\_dep\_time = '00:00:00' then 0 else (dep\_time - crs\_dep\_time)/100 end as dep\_delay,*

*case when crs\_arr\_time = '00:00:00' then 0 else (arr\_time - crs\_arr\_time)/100 end as arr\_delay,*

*case when dep\_delay > 15 then 1 else 0 end dep\_del15,*

*case when arr\_delay > 15 then 1 else 0 end arr\_del15*

*from flight\_performance;*

For find year,month,day:

*select \* ,*

*year(FL\_DATE) as YEAR ,*

*month(FL\_DATE)) as MONTH,*

*Day(FL\_DATE) as QUATER*

*from Airline ;*

Columns which are identical:

IATA\_CODE\_Reporting\_Airline is similar to DOT\_ID\_Reporting\_Airline

DestCityName is similar to Dest

DepDelayMinutes is similar to DepDelay

ArrDelayMinutes is similar to ArrDelay

Fetch the ORIGIN\_AIRPORT\_ID FROM ORIGIN\_AIRPORT\_SEQ\_ID

*SELECT SUBSTRING(ORIGIN\_AIRPORT\_SEQ\_ID, 1, 5) as ORIGIN\_AIRPORT\_ID from AIRLINE ;*

Fetch the DEST\_AIRPORT\_ID FROM DEST\_AIRPORT\_SEQ\_ID

*SELECT SUBSTRING(DEST\_AIRPORT\_SEQ\_ID, 1, 5) as DEST\_AIRPORT\_ID from AIRLINE ;*

Query to fetch departure delay groups

*SELECT DEP\_TIME, CRS\_DEP\_TIME,*

*CASE WHEN (DEP\_TIME-CRS\_DEP\_TIME) < -15 THEN -2*

*WHEN -15< (DEP\_TIME-CRS\_DEP\_TIME) < -1 THEN -1*

*WHEN 0 < (DEP\_TIME-CRS\_DEP\_TIME) < -14 THEN 0*

*WHEN 15< (DEP\_TIME-CRS\_DEP\_TIME) < 29 THEN 1*

*WHEN 30< (DEP\_TIME-CRS\_DEP\_TIME) < 44 THEN 2*

*WHEN 45< (DEP\_TIME-CRS\_DEP\_TIME) < 59 THEN 3*

*WHEN 60< (DEP\_TIME-CRS\_DEP\_TIME) < 74 THEN 4*

*WHEN 75< (DEP\_TIME-CRS\_DEP\_TIME) < 89 THEN 5*

*WHEN 90< (DEP\_TIME-CRS\_DEP\_TIME) < 104 THEN 6*

*WHEN 105< (DEP\_TIME-CRS\_DEP\_TIME) < 119 THEN 7*

*WHEN 120< (DEP\_TIME-CRS\_DEP\_TIME) < 134 THEN 8*

*WHEN 135< (DEP\_TIME-CRS\_DEP\_TIME) < 149 THEN 9*

*WHEN 150< (DEP\_TIME-CRS\_DEP\_TIME) < 164 THEN 10*

*WHEN 165< (DEP\_TIME-CRS\_DEP\_TIME) < 179 THEN 11*

*ELSE 12*

*END AS DEP\_DELAY\_GRP*

*FROM DEPARTURE\_PERFORMANCE;*

Note: I have included ER Diagram, SQL Queries, Dependency diagram in separate files.