

New York City (NYC) Motor Vehicles (MV)

Collisions - Final Project - Part 2

DIM TABLES:

1. Dim_Person_Contributing_Factor:

The screenshot displays the Talend Real-time Big Data Platform interface. On the left, the 'Repository' pane shows a project structure for 'LOCAL: NYC_MV_Collision' with various job designs and metadata. The main workspace shows a job design for 'Job NYC_Motor_Vehicle_Collisions_Dimension'. The job flow includes components like 'tUniqRow_1', 'tSortRow_1', and 'tMap_1' leading to 'Dim_Person_Contributing_Factor'. The 'Execution' tab shows the job running successfully, starting at 22:59 and ending at 23:00 on 20/04/2023.

Job NYC_Motor_Vehicle_Collisions_Dimension

Basic Run Execution

Run KILL Clear

Starting Job NYC_Motor_Vehicle_Collisions_Dimension at 22:59 20/04/2023
[statistics] connecting to socket on port 3540
[statistics] disconnected
Job NYC_Motor_Vehicle_Collisions_Dimension ended at 23:00 20/04/2023.

Line limit 100 Wrap

Time Taken: 1 min
Row Count: 556

2. Dim_Person_Type:

Talend Real-time Big Data Platform (8.0.1.20211109_1610) | ADW_Product_Price_History_SCD (Connection: Local)

Job NYC_Motor_Vehicle_Collisions_Dimension 1.1

Execution:

```
Starting job NYC_Motor_Vehicle_Collisions_Dimension at 00:21 20/04/2023.
[statistics] connecting to socket on port 3775
[statistics] connected
[statistics] disconnected
Job NYC_Motor_Vehicle_Collisions_Dimension ended at 00:59 20/04/2023. [Exit code = 0]
```

Time Taken: 19 mins

Row Count: 4863648

3. Dim_Vehicle_Contributing_Factor:

Talend Real-time Big Data Platform (8.0.1.20211109_1610) | NYC_MV_Collision (Connection: Local)

Job NYC_Motor_Vehicle_Collisions_Dimension

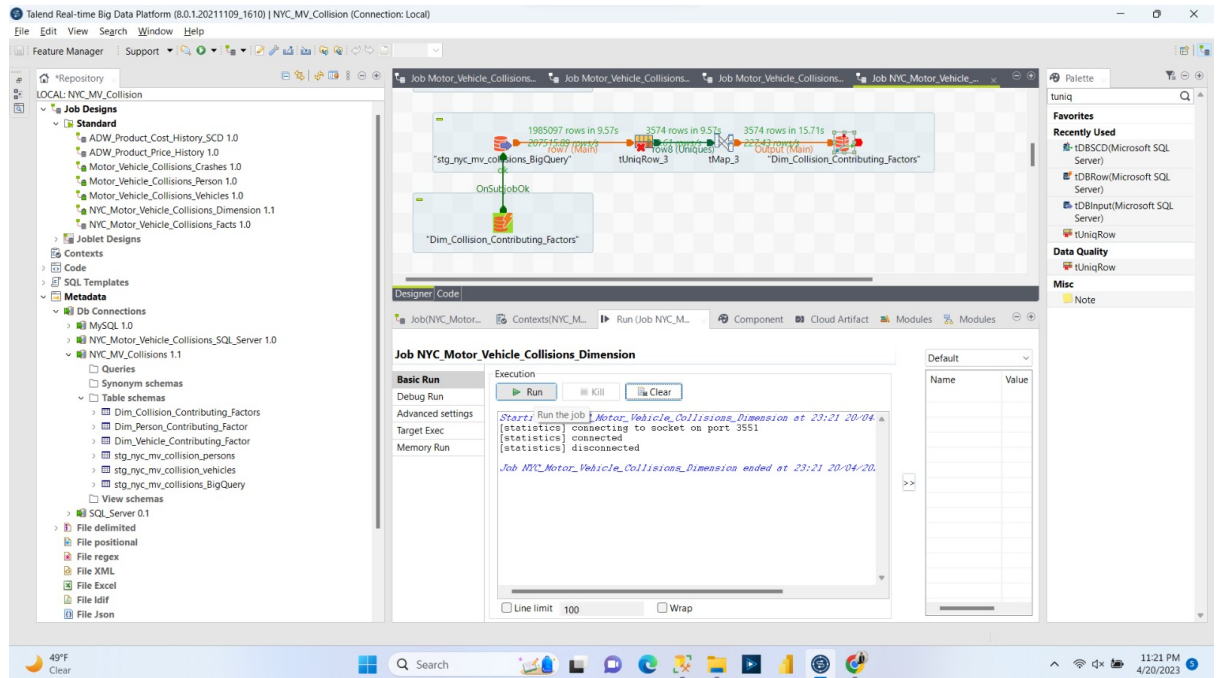
Execution:

```
Start: Run the job NYC_Motor_Vehicle_Collisions_Dimension at 23:15 20/04/2023.
[statistics] connecting to socket on port 3464
[statistics] connected
[statistics] disconnected
Job NYC_Motor_Vehicle_Collisions_Dimension ended at 23:15 20/04/2023.
```

Time Taken: 1 min

Row Count: 1871

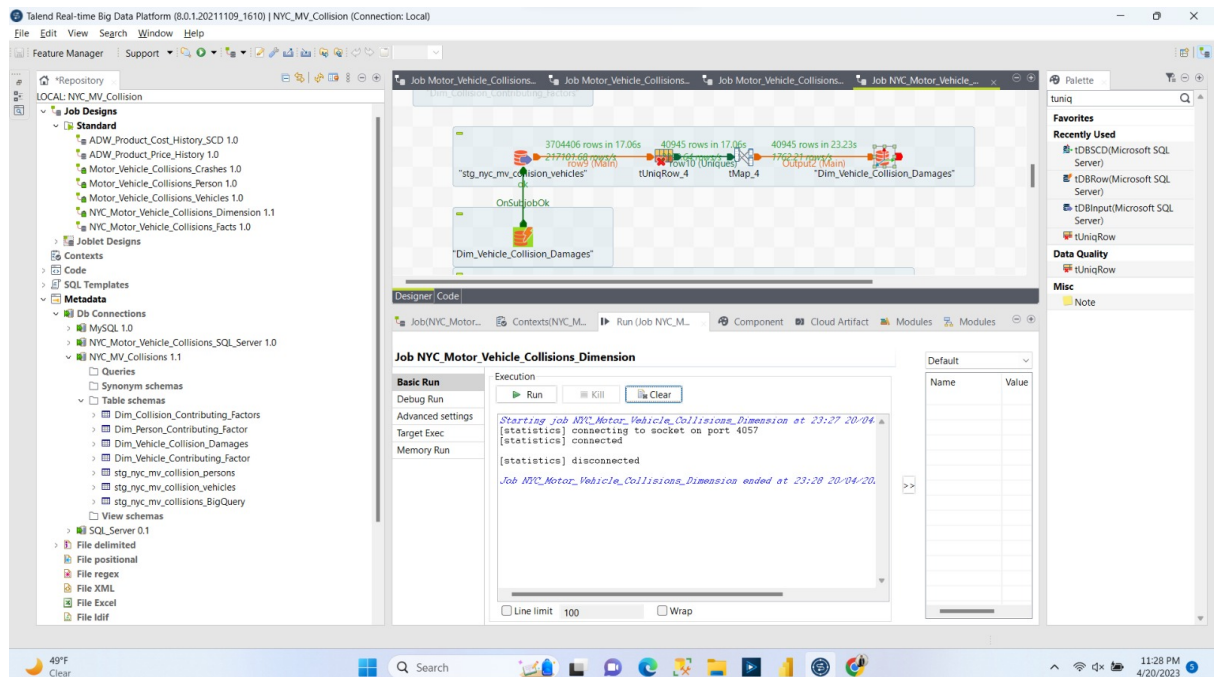
4. Dim_Collisions_Contributing_Factor:



Time Taken: 1 min

Row Count: 3574

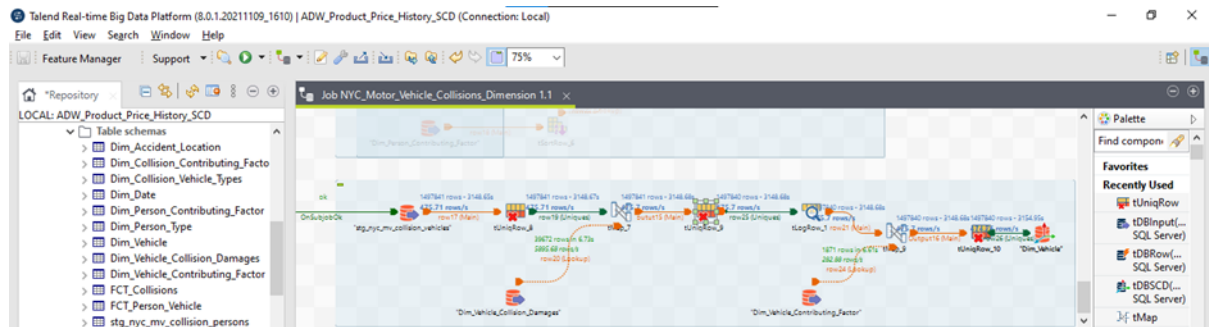
5. Dim_Vehicle_Collision_Damages



Time Taken: 2 min

Row Count: 40945

6. Dim_Vehicle:

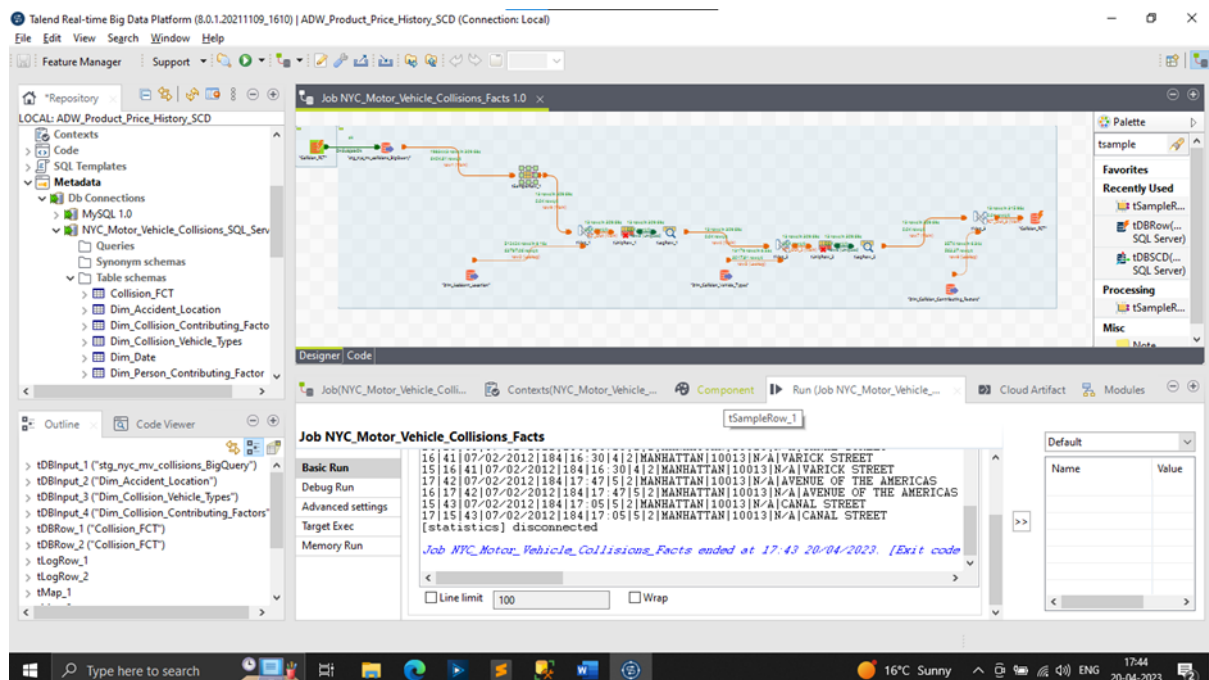


Time Taken: 14 mins

Row Count: 3704406 rows

FACT TABLES:

FCT_COLLISIONS:



Time Taken: 25 mins

Row Count: 1983443 rows

FCT_Person_Vehicle:

The screenshot displays the Talend Real-time Big Data Platform interface. The top menu bar includes File, Edit, View, Search, Window, and Help. Below the menu is a toolbar with various icons for file operations and job management. The main workspace is divided into several panels:

- Repository:** Shows a tree view of the local data warehouse schema, including tables like Collision_FCT, Dim_Accident_Location, Dim_Collision_Contributing_Factors, Dim_Collision_Vehicle_Types, Dim_Date, Dim_Person_Contributing_Factor, Dim_Person_Type, Dim_Vehicle, Dim_Vehicle_Collision_Damages, Dim_Vehicle_Contributing_Factor, FCT_Collisions, FCT_Person_Vehicle, Person_Vehicle_FCT, and stg_nyc_mv_collision_persons.
- Job NYC_Motor_Vehicle_Collisions_Facts 1.0:** A visual job design canvas showing a sequence of components: tStgNycMvCollision, tSortRow_25, tMap_3, tJoin_4, tLogRow_15, tJoin_6, tSortRow_3, and tSampleRow_3. Data flow is indicated by arrows and row counts.
- Code Viewer:** Displays the job configuration code, including tBlingput_1 through tBlingput_9.
- Execution Log:** Shows the job's execution status. It indicates that the job started at 19:15 on 20/04/2023. The log also shows an exception: `Exception: java.lang.OutOfMemoryError thrown from the UncaughtExceptionHandler`.

The bottom status bar shows the system clock at 20:00 on 20-04-2023 and the language set to ENG.

Time Taken: 30 mins

Row Count: 6413058 rows