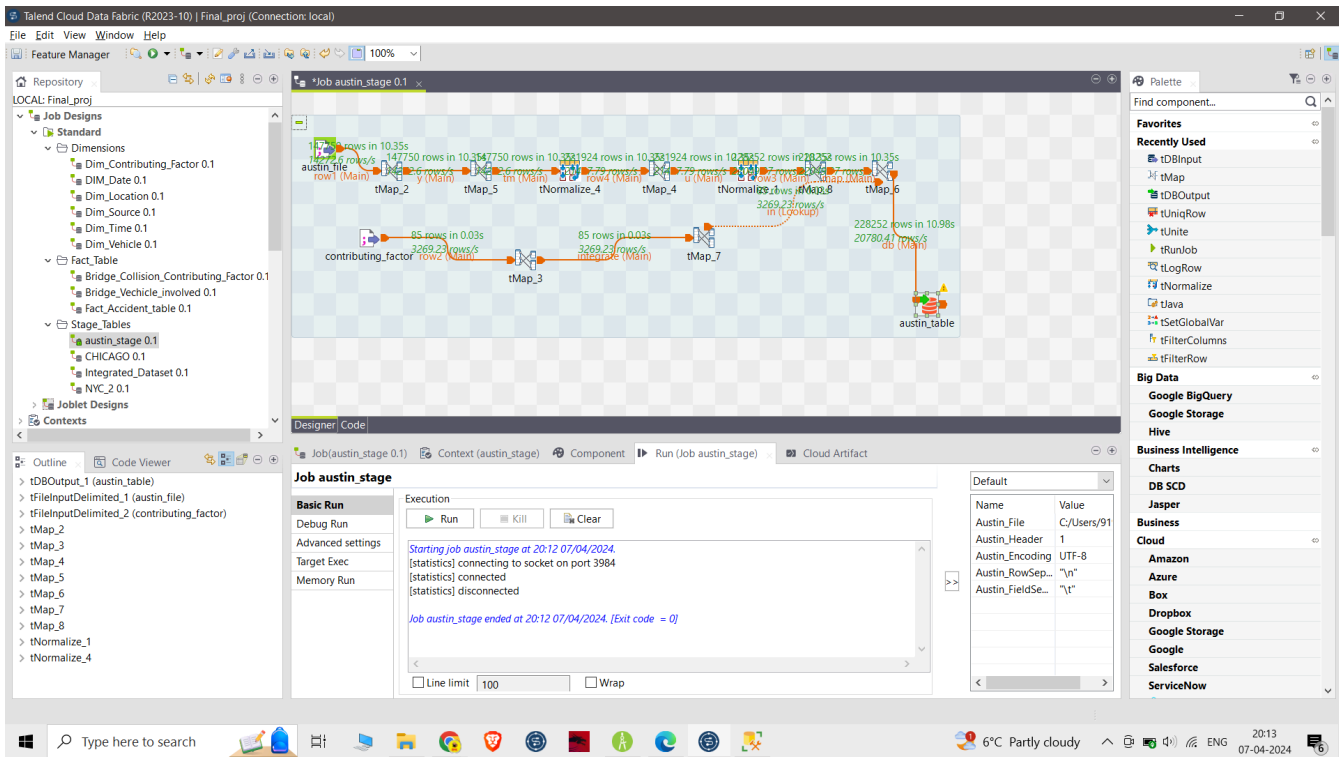
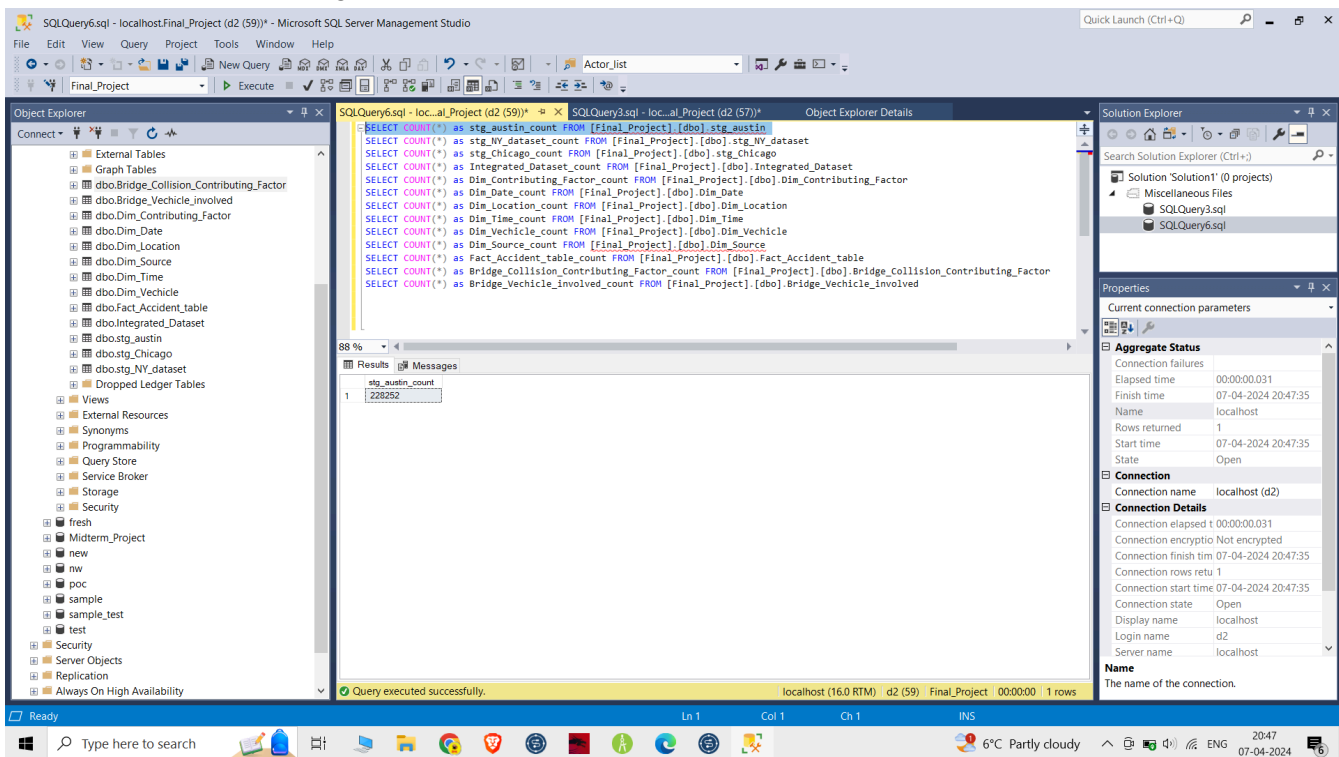


# Austin Stage Table

## Stage Table for Austin :



## No. of rows for Austin Stage Table :



# Chicago Stage Table

## Stage Table for Chicago :

The screenshot displays the Talend Cloud Data Fabric interface for a job named 'Job CHICAGO 0.1'. The main workspace shows a workflow diagram with components like 'tFileInputDelimited\_1', 'tMap\_1', 'tNormalize\_1', 'tMap\_2', 'tMap\_3', and 'stg\_Chicago'. The 'Designer' tab is active, showing the workflow code. The 'Execution' tab is also visible, displaying logs and a table of context parameters.

**Job CHICAGO 0.1 Execution Log:**

```
[WARN ] 2013:57 final_proj.chicago_0.1.CHICAGO- Null value will be used for context parameter Chicago_City  
[WARN ] 2013:57 final_proj.chicago_0.1.CHICAGO- Null value will be used for context parameter Chicago_City  
[statistics] connecting to socket on port 3369  
[statistics] connected  
[statistics] disconnected  
Job CHICAGO ended at 20:14 07/04/2024. [Exit code = 0]
```

**Context Parameters Table:**

Name	Value
Chicago_City_E...	
Chicago_City_F...	
Chicago_City_L...	
Chicago_City_R...	
Chicago_City_F...	
stg_Chicago_Cl...	
stg_Chicago_Cl...	
stg_Chicago_Cl...	
stg_Chicago_Cl...	

## No. of rows for Chicago Stage Table :

The screenshot shows the Microsoft SQL Server Enterprise Manager interface. A query is executed against the 'Final\_Project' database, returning the row count for the 'stg\_Chicago' table. The query is as follows:

```
SELECT COUNT(*) as stg_austin_count FROM [Final_Project].[dbo].[stg_austin]  
SELECT COUNT(*) as stg_NY_dataset_count FROM [Final_Project].[dbo].[stg_NY_dataset]  
SELECT COUNT(*) as stg_Chicago_count FROM [Final_Project].[dbo].[stg_Chicago]  
SELECT COUNT(*) as Integrated_Dataset_count FROM [Final_Project].[dbo].[Integrated_Dataset]  
SELECT COUNT(*) as Dim_Contributing_Factor_count FROM [Final_Project].[dbo].[Dim_Contributing_Factor]  
SELECT COUNT(*) as Dim_Date_count FROM [Final_Project].[dbo].[Dim_Date]  
SELECT COUNT(*) as Dim_Location_count FROM [Final_Project].[dbo].[Dim_Location]  
SELECT COUNT(*) as Dim_Time_count FROM [Final_Project].[dbo].[Dim_Time]  
SELECT COUNT(*) as Dim_Vehicle_count FROM [Final_Project].[dbo].[Dim_Vehicle]  
SELECT COUNT(*) as Dim_Source_count FROM [Final_Project].[dbo].[Dim_Source]  
SELECT COUNT(*) as Fact_Accident_table_count FROM [Final_Project].[dbo].[Fact_Accident_table]  
SELECT COUNT(*) as Bridge_Collision_Contributing_Factor_count FROM [Final_Project].[dbo].[Bridge_Collision_Contributing_Factor]  
SELECT COUNT(*) as Bridge_Vehicle_Involved_count FROM [Final_Project].[dbo].[Bridge_Vehicle_Involved]
```

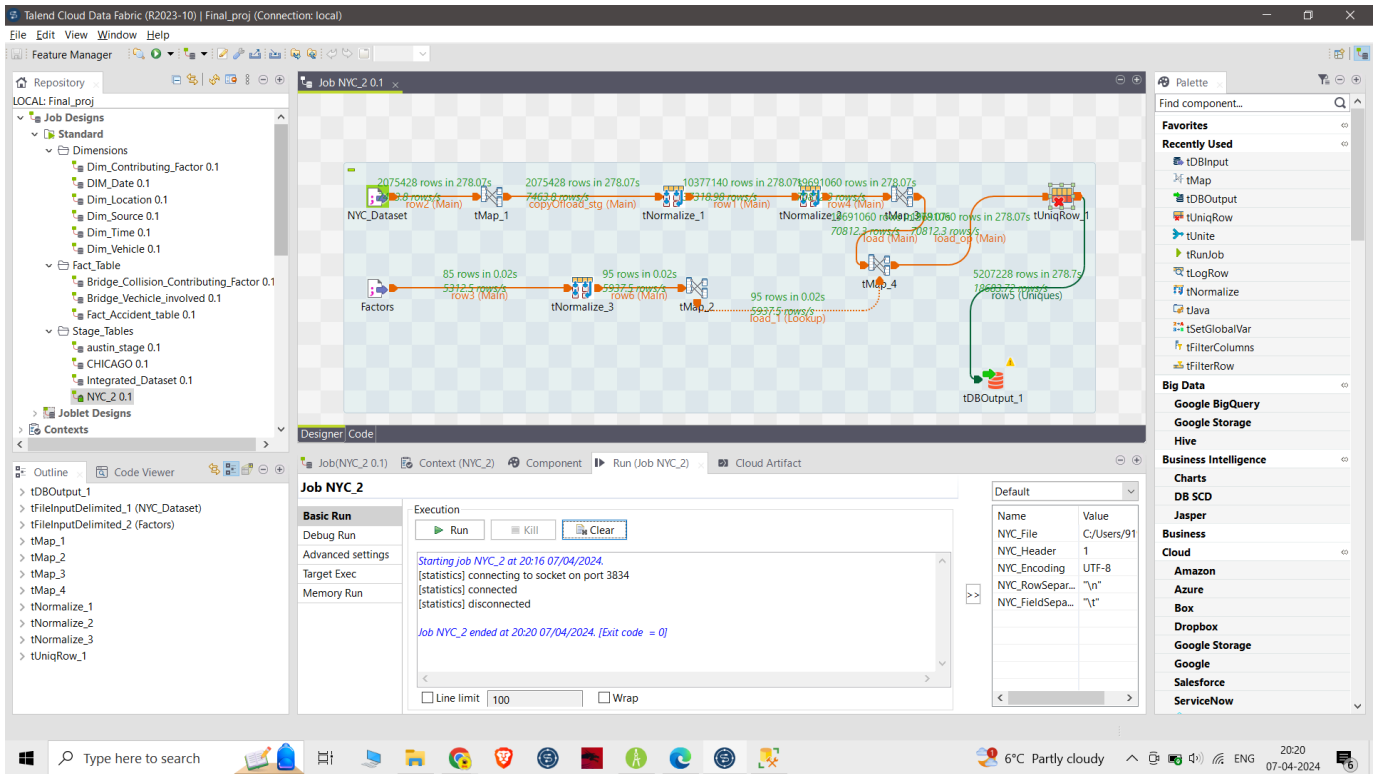
The query result shows 1 row with the value 1350572.

**Query Results:**

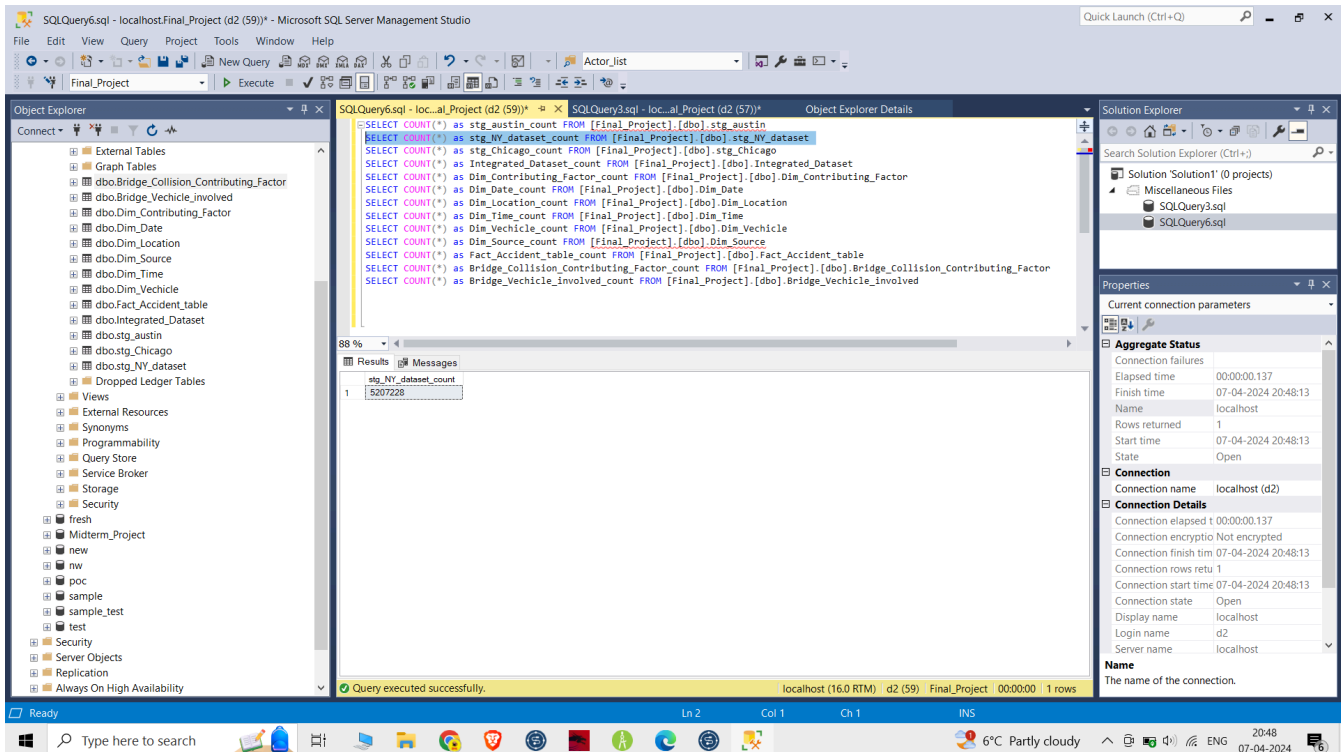
stg_Chicago_count
1350572

# New York Stage Table

Stage Table for New York :



No. of rows for New York Stage Table :



# SQL scripts for Dimensional model

The screenshot displays the Microsoft SQL Server Management Studio (SSMS) interface. The main window shows a SQL query script for creating a table named `Dim_Contributing_Factor`. The script includes comments about the code generation process and the table's structure. The table has the following columns: `Dim_Contributing_Factor(SK)` (int, NOT NULL), `Contributing_Factor_Code` (int, NOT NULL), `Contributing_Factor_Description` (text, NULL), `Pid` (varchar(500), NULL), `DI_CreateDate` (date, NULL), and `DI_WorkFlowFileName` (varchar(500), NULL). A primary key constraint is defined on the `Dim_Contributing_Factor(SK)` column.

The query is executed successfully, as indicated by the status bar and the Messages pane. The Messages pane shows the following output:

```
<<< CREATED TABLE Dim_Contributing_Factor >>>
<<< CREATED TABLE Dim_Date >>>
<<< CREATED TABLE Dim_Location >>>
<<< CREATED TABLE Dim_Source >>>
<<< CREATED TABLE Dim_Time >>>
<<< CREATED TABLE Fact_Accident_table >>>
<<< CREATED TABLE Bridge_Collision_Contributing_Factor >>>
<<< CREATED TABLE Bridge_Vehicle_Involved >>>
<<< CREATED TABLE Dim_Time >>>
```

The Properties pane on the right shows the connection details for the current connection (localhost (d2)).

Aggregate Status	
Connection failures	
Elapsed time	00:00:00.072
Finish time	07-04-2024 20:10:47
Name	localhost
Rows returned	0
Start time	07-04-2024 20:10:47
State	Open

Connection	
Connection name	localhost (d2)

Connection Details	
Connection elapsed t	00:00:00.072
Connection encryptio	Not encrypted
Connection finish tim	07-04-2024 20:10:47
Connection rows retu	0
Connection start time	07-04-2024 20:10:47
Connection state	Open
Display name	localhost
Login name	d2
Server name	localhost

The Name property is set to "The name of the connection."

# Dim\_Contributing\_Factor

The screenshot displays the Talend Cloud Data Fabric interface for a job named 'Dim\_Contributing\_Factor 0.1'. The interface is divided into several sections:

- Repository:** Shows a tree view of job designs under 'LOCAL: Final\_proj'. The 'Dimensions' folder contains 'Dim\_Contributing\_Factor 0.1', which includes sub-jobs for 'DIM\_Date 0.1', 'Dim\_Location 0.1', 'Dim\_Source 0.1', 'Dim\_Time 0.1', and 'Dim\_Vehicle 0.1'. The 'Fact\_Table' folder contains 'Bridge\_Collision\_Contributing\_Factor 0.1', 'Bridge\_Vehicle\_involved 0.1', and 'Fact\_Accident\_table 0.1'. The 'Stage\_Tables' folder contains 'austin\_stage 0.1', 'CHICAGO 0.1', 'Integrated\_Dataset 0.1', and 'NYC\_2 0.1'.
- Job Designer:** Shows a visual flow of the job. It starts with 'Final\_db' (28252 rows in 12.27s) and 'Final\_db' (207228 rows in 12.27s). These feed into 'tMerge' (18599.41 rows/s) and 'tMerge' (1350572 rows in 12.27s). The output of 'tMerge' goes to 'tUniqRow\_1' (6786.52 rows in 12.27s), which then feeds into 'tUniqRow\_1' (234 rows/s). The output of 'tUniqRow\_1' goes to 'tMap\_1' (85 rows in 12.74s), which then feeds into 'Final\_db' (667 rows/s).
- Execution Log:** Shows the job execution details. The job started at 2021/07/04/2024. The log indicates that the job is connecting to a socket on port 3934, and the statistics are disconnected. The job ended at 2021/07/04/2024 with an exit code of 0.
- Cloud Artifact:** Shows the default configuration for the job, including a table with columns 'Name' and 'Value'.

No. of rows for Dim\_Contributing\_Factor :

The screenshot displays the Microsoft SQL Server Management Studio interface. The 'SQLQuery6.sql' file is open, showing a query that counts the number of rows in the 'Dim\_Contributing\_Factor' table. The query is as follows:

```
SELECT COUNT(*) as stg_austin_count FROM [Final_Project].[dbo].stg_austin
SELECT COUNT(*) as stg_NY_dataset_count FROM [Final_Project].[dbo].stg_NY_dataset
SELECT COUNT(*) as stg_Chicago_count FROM [Final_Project].[dbo].stg_Chicago
SELECT COUNT(*) as Integrated_Dataset_count FROM [Final_Project].[dbo].Integrated_Dataset
SELECT COUNT(*) as Dim_Contributing_Factor_count FROM [Final_Project].[dbo].Dim_Contributing_Factor
SELECT COUNT(*) as Dim_Date_count FROM [Final_Project].[dbo].Dim_Date
SELECT COUNT(*) as Dim_Location_count FROM [Final_Project].[dbo].Dim_Location
SELECT COUNT(*) as Dim_Time_count FROM [Final_Project].[dbo].Dim_Time
SELECT COUNT(*) as Dim_Vehicle_count FROM [Final_Project].[dbo].Dim_Vehicle
SELECT COUNT(*) as Dim_Source_count FROM [Final_Project].[dbo].Dim_Source
SELECT COUNT(*) as Fact_Accident_table_count FROM [Final_Project].[dbo].Fact_Accident_table
SELECT COUNT(*) as Bridge_Collision_Contributing_Factor_count FROM [Final_Project].[dbo].Bridge_Collision_Contributing_Factor
SELECT COUNT(*) as Bridge_Vehicle_involved_count FROM [Final_Project].[dbo].Bridge_Vehicle_involved
```

The 'Results' pane shows the output of the query, which is a single row with the value 85.

The 'Object Explorer' pane on the left shows the database structure, including the 'Dim\_Contributing\_Factor' table. The 'Solution Explorer' pane on the right shows the project structure, including the 'SQLQuery6.sql' file.

# Dim\_Date

The screenshot displays the Talend Cloud Data Fabric (R2023-10) interface for a job named 'Job DIM\_Date 0.1'. The left sidebar shows a repository with various components like 'Dim\_Contributing\_Factor 0.1', 'DIM\_Date 0.1', 'Dim\_Location 0.1', 'Dim\_Source 0.1', 'Dim\_Time 0.1', 'Dim\_Vehicle 0.1', 'Fact\_Table', 'Bridge\_Collision\_Contributing\_Factor 0.1', 'Bridge\_Vehicle\_Involved 0.1', 'Fact\_Accident\_table 0.1', 'Stage Tables', 'austin\_stage 0.1', 'CHICAGO 0.1', 'Integrated\_Dataset 0.1', 'NYC\_2 0.1', 'Joblet Designs', and 'Contexts'. The main workspace shows a job design with components: 'tRowGenerator\_1' (10000 rows in 0.81s), 'tMap\_1' (12406.95 rows in 1.24s), 'tMap\_2' (10000 rows in 0.81s), and 'Load\_Dimension' (10000 rows in 1.18s). The 'Designer' tab is active, showing the job design. The 'Code' tab is also visible. The 'Execution' tab shows the job status: 'Starting Job DIM\_Date at 2022 07/04/2024', '[statistics] connecting to socket on port 3827', '[statistics] connected', '[statistics] disconnected', and 'Job DIM\_Date ended at 2022 07/04/2024. [Exit code = 0]'. The 'Basic Run' tab is selected, showing 'Run', 'Kill', and 'Clear' buttons. The 'Context (DIM\_Date)' tab is also visible. The 'Cloud Artifact' tab shows the job status. The bottom status bar indicates '6°C Partly cloudy' and '2022 07-04-2024'.

## No. of rows in Dim\_Date :

The screenshot displays the Microsoft SQL Server Management Studio (SSMS) interface. The 'Object Explorer' on the left shows the database structure, including tables like 'Dim\_Date'. The 'Query Editor' in the center shows a SQL query to count the number of rows in the 'Dim\_Date' table. The query is as follows:

```
SELECT COUNT(*) as stg_austin_count FROM [Final_Project].[dbo].[stg_austin]
SELECT COUNT(*) as stg_NY_dataset_count FROM [Final_Project].[dbo].[stg_NY_dataset]
SELECT COUNT(*) as stg_Chicago_count FROM [Final_Project].[dbo].[stg_Chicago]
SELECT COUNT(*) as Dim_Contributing_Factor_count FROM [Final_Project].[dbo].[Dim_Contributing_Factor]
SELECT COUNT(*) as Dim_Date_count FROM [Final_Project].[dbo].[Dim_Date]
SELECT COUNT(*) as Dim_Location_count FROM [Final_Project].[dbo].[Dim_Location]
SELECT COUNT(*) as Dim_Time_count FROM [Final_Project].[dbo].[Dim_Time]
SELECT COUNT(*) as Dim_Vehicle_count FROM [Final_Project].[dbo].[Dim_Vehicle]
SELECT COUNT(*) as Dim_Source_count FROM [Final_Project].[dbo].[Dim_Source]
SELECT COUNT(*) as Fact_Accident_table_count FROM [Final_Project].[dbo].[Fact_Accident_table]
SELECT COUNT(*) as Bridge_Collision_Contributing_Factor_count FROM [Final_Project].[dbo].[Bridge_Collision_Contributing_Factor]
SELECT COUNT(*) as Bridge_Vehicle_Involved_count FROM [Final_Project].[dbo].[Bridge_Vehicle_Involved]
```

The 'Results' tab shows the output of the query, which is a single row with the value '10000' for the 'Dim\_Date\_count' column. The 'Messages' tab shows the status 'Query executed successfully.' The bottom status bar indicates 'localhost (16.0 RTM) d2 (59) Final\_Project 00:00:00 1 rows'.



# Dim\_Location

The screenshot displays the Talend Cloud Data Fabric (R2023-10) interface for a job named 'Job Dim\_Location 0.1'. The job is configured with a 'Final\_db' input and output, and a 'Dim\_Location' context. The job flow includes components like 'tUniqRow\_1', 'tUniqRow\_1', 'tMap\_1', and 'Dim\_Location\_Output (Main)'. The 'Designer' code tab shows the job's execution details, including a 'Run' button and a 'Clear' button. The 'Execution' tab shows the job's status, indicating it was started on 2023/07/04/2024 and ended on 2024/07/04/2024. The 'Cloud Artifact' tab shows the job's artifacts, including 'Dim\_Location' and 'Dim\_Location\_Output (Main)'.

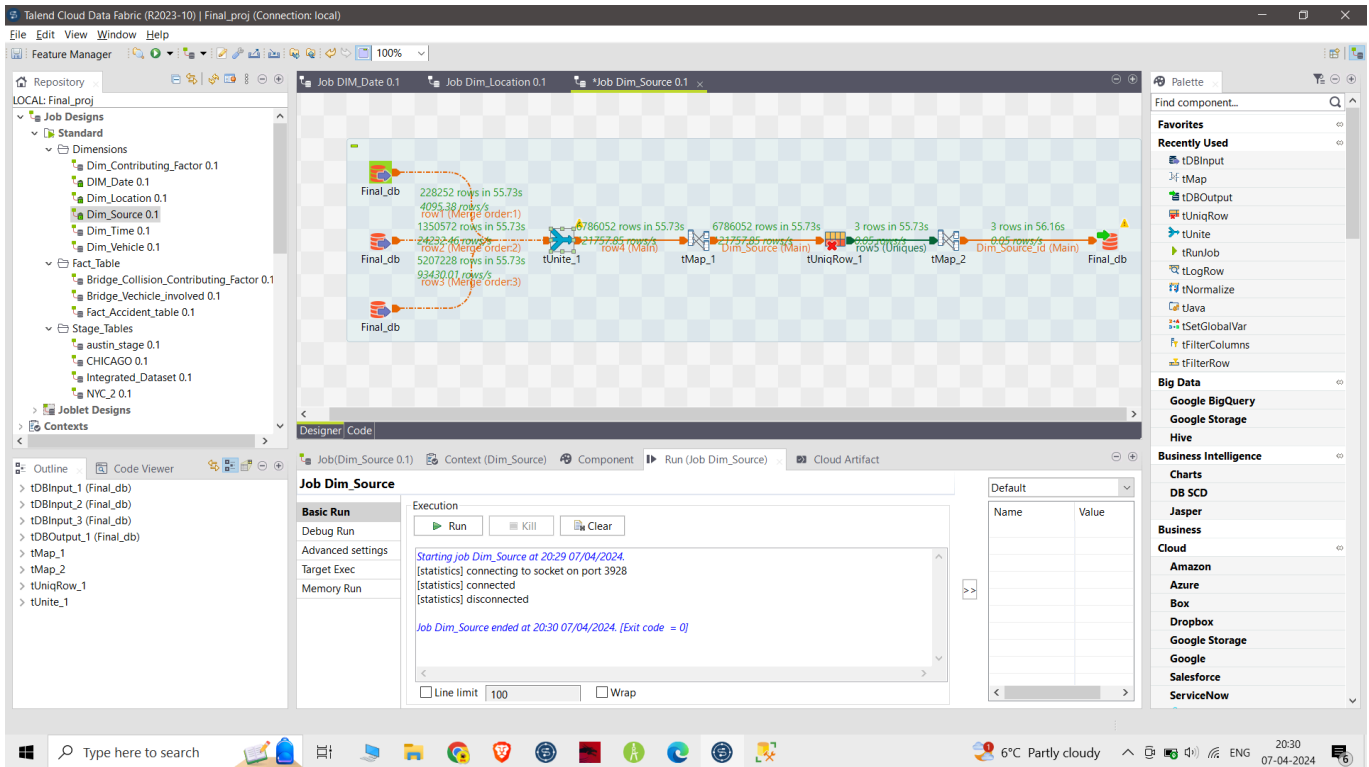
No. of rows in Dim\_Location :

The screenshot shows the Microsoft SQL Server Enterprise Manager interface. The 'Object Explorer' pane on the left displays the database structure, including tables, views, and security. The 'SQL Query Editor' pane in the center shows a query that counts the number of rows in the 'Dim\_Location' table. The 'Results' pane at the bottom shows the query results, indicating that there are 869738 rows in the 'Dim\_Location' table.

```
SELECT COUNT(*) as Dim_Location_count FROM [Final_Project].[dbo].[Dim_Location]
```

Dim_Location_count
869738

# Dim\_Source



No. of rows in Dim\_Source :

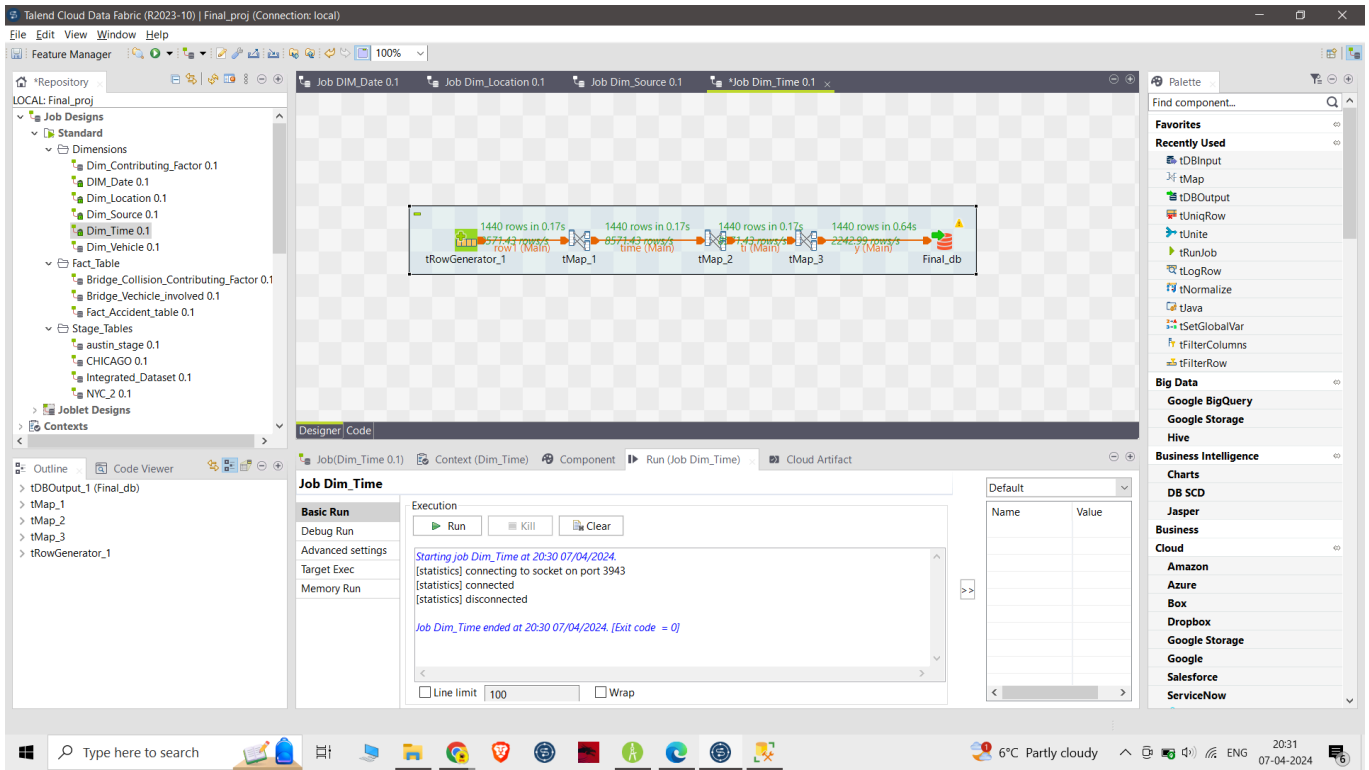
The screenshot displays the Microsoft SQL Server Management Studio interface. The 'Object Explorer' on the left shows the 'Final\_Project' database structure, including tables like 'Dim\_Source', 'Dim\_Location', 'Dim\_Time', 'Dim\_Vehicle', 'Fact\_Accident\_table', 'Bridge\_Vehicle\_Involved', 'Integrated\_Dataset', 'stg\_austin', 'stg\_Chicago', 'stg\_NY\_dataset', and 'Dropped Ledger Tables'. The 'Query Editor' in the center shows a SQL query that counts the number of rows in various tables. The query is as follows:

```
SELECT COUNT(*) as stg_austin_count FROM [Final_Project].[dbo].[stg_austin]
SELECT COUNT(*) as stg_NY_dataset_count FROM [Final_Project].[dbo].[stg_NY_dataset]
SELECT COUNT(*) as stg_Chicago_count FROM [Final_Project].[dbo].[stg_Chicago]
SELECT COUNT(*) as Integrated_Dataset_count FROM [Final_Project].[dbo].[Integrated_Dataset]
SELECT COUNT(*) as Dim_Contributing_Factor_count FROM [Final_Project].[dbo].[Dim_Contributing_Factor]
SELECT COUNT(*) as Dim_Date_count FROM [Final_Project].[dbo].[Dim_Date]
SELECT COUNT(*) as Dim_Location_count FROM [Final_Project].[dbo].[Dim_Location]
SELECT COUNT(*) as Dim_Time_count FROM [Final_Project].[dbo].[Dim_Time]
SELECT COUNT(*) as Dim_Vehicle_count FROM [Final_Project].[dbo].[Dim_Vehicle]
SELECT COUNT(*) as Dim_Source_count FROM [Final_Project].[dbo].[Dim_Source]
SELECT COUNT(*) as Fact_Accident_table_count FROM [Final_Project].[dbo].[Fact_Accident_table]
SELECT COUNT(*) as Bridge_Vehicle_Involved_count FROM [Final_Project].[dbo].[Bridge_Vehicle_Involved]
SELECT COUNT(*) as Integrated_Dataset_count FROM [Final_Project].[dbo].[Integrated_Dataset]
```

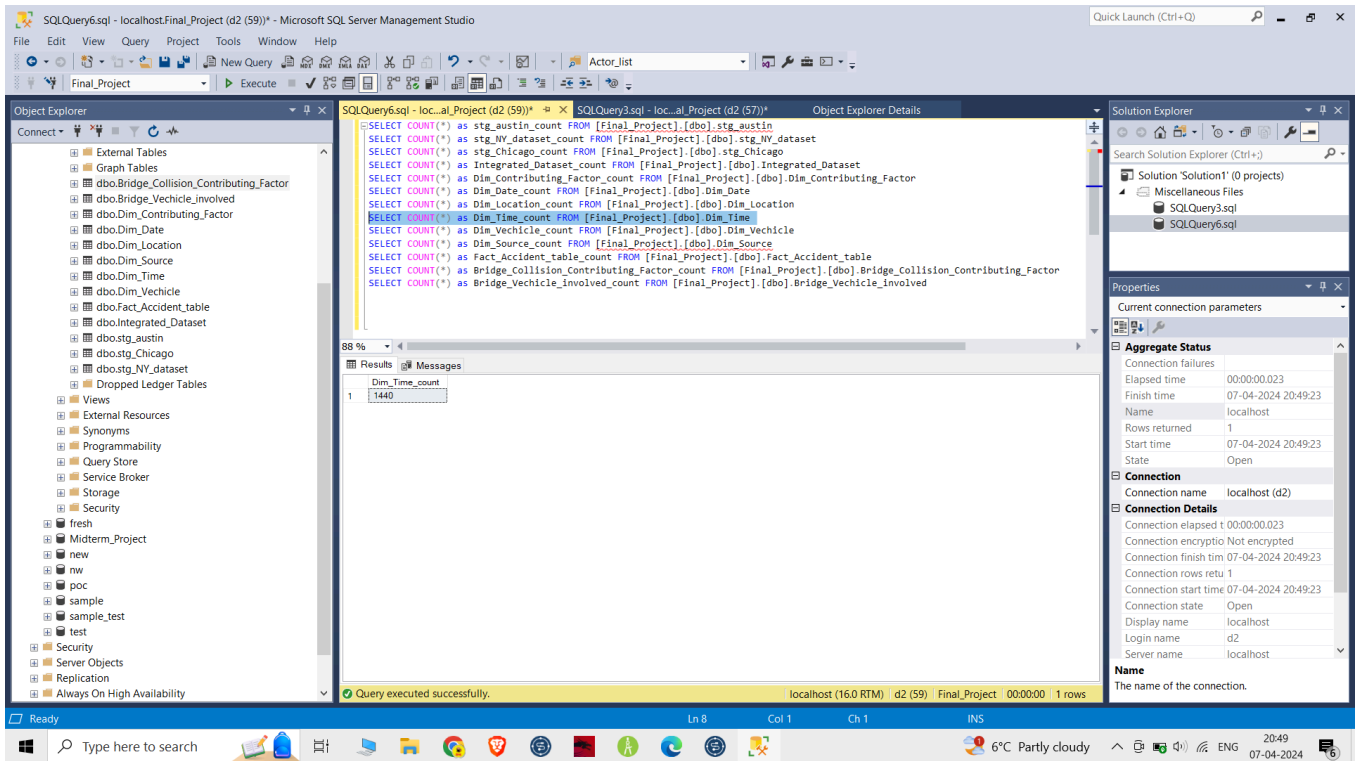
The 'Results' pane at the bottom shows the output of the query, which is a single row with the value 3 for 'Dim\_Source\_count'. The 'Properties' pane on the right shows the connection details for the 'localhost (d2)' connection.



## Dim\_Time



No. of rows in Dim\_Time :



# Dim\_Vehicle

The screenshot displays the Talend Cloud Data Fabric interface for a job named 'Job Dim\_Vehicle 0.1'. The workflow is visualized in the Designer view, showing a sequence of components: 'Final\_db' (tDBInput), 'tUnite\_1', 'tMap\_1', 'tUniqRow\_1', and 'tBSCD\_1'. The workflow is connected to a 'Final\_db' component. The execution details are shown in the 'Execution' tab, indicating that the job started at 20:31:07 on 07/04/2024 and ended at 20:31:07 on 07/04/2024. The job executed successfully, with 1773 rows returned. The 'Code' tab shows the SQL query used for the job.

**Job Dim\_Vehicle 0.1**

Execution

Starting job Dim\_Vehicle at 20:31:07/04/2024.  
[statistics] connecting to socket on port 3402  
[statistics] connected  
[statistics] disconnected  
Job Dim\_Vehicle ended at 20:31:07/04/2024. [Exit code = 0]

Line limit: 100

No. of rows in Dim\_Vehicle :

The screenshot shows the Microsoft SQL Server Management Studio interface. The 'Object Explorer' on the left displays the database structure, including tables like 'Dim\_Vehicle'. The 'SQL Query Editor' in the center contains a query that counts the number of rows in the 'Dim\_Vehicle' table. The 'Results' pane at the bottom shows the output of the query, which is 1773 rows. The 'Properties' pane on the right shows the connection details for the 'localhost (d2)' connection.

SQLQuery6.sql - localhost.Final\_Project (d2 (59)) - Microsoft SQL Server Management Studio

SQLQuery6.sql - localhost.Final\_Project (d2 (59))

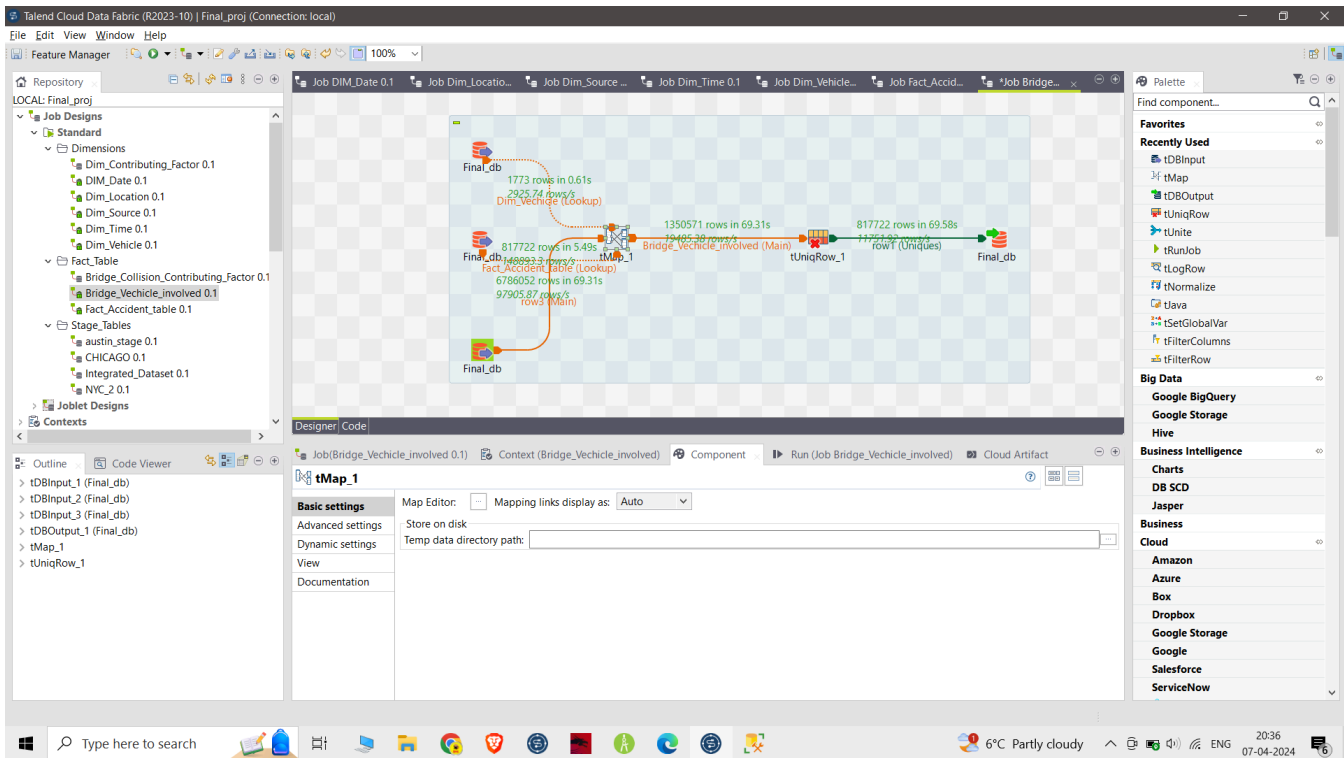
```
SELECT COUNT(*) as stg_austin_count FROM [Final_Project].[dbo].[stg_austin]
SELECT COUNT(*) as stg_NY_dataset_count FROM [Final_Project].[dbo].[stg_NY_dataset]
SELECT COUNT(*) as stg_Chicago_count FROM [Final_Project].[dbo].[stg_Chicago]
SELECT COUNT(*) as Integrated_Dataset_count FROM [Final_Project].[dbo].[Integrated_Dataset]
SELECT COUNT(*) as Dim_Contributing_Factor_count FROM [Final_Project].[dbo].[Dim_Contributing_Factor]
SELECT COUNT(*) as Dim_Date_count FROM [Final_Project].[dbo].[Dim_Date]
SELECT COUNT(*) as Dim_Location_count FROM [Final_Project].[dbo].[Dim_Location]
SELECT COUNT(*) as Dim_Time_count FROM [Final_Project].[dbo].[Dim_Time]
SELECT COUNT(*) as Dim_Vehicle_count FROM [Final_Project].[dbo].[Dim_Vehicle]
SELECT COUNT(*) as Dim_Source_count FROM [Final_Project].[dbo].[Dim_Source]
SELECT COUNT(*) as Fact_Accident_table_count FROM [Final_Project].[dbo].[Fact_Accident_table]
SELECT COUNT(*) as Bridge_Collision_Contributing_Factor_count FROM [Final_Project].[dbo].[Bridge_Collision_Contributing_Factor]
SELECT COUNT(*) as Bridge_Vehicle_Involved_count FROM [Final_Project].[dbo].[Bridge_Vehicle_Involved]
```

Results

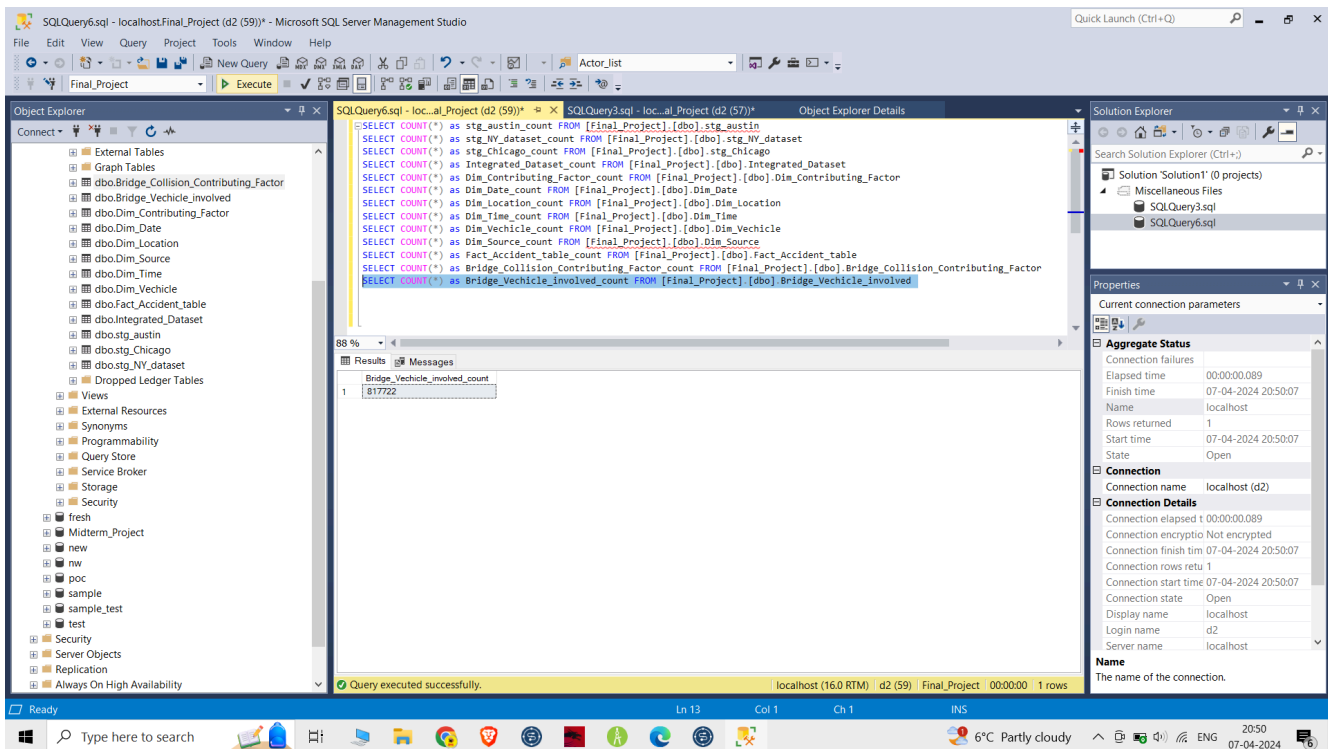
Dim_Vehicle_count
1773

Query executed successfully. localhost (16.0 RTM) d2 (59) Final\_Project 00:00:00 1 rows

# Bridge\_Vehicle\_involved



No. of rows in Bridge\_Vehicle\_involved :



# Bridge\_Collision\_Contributing\_Factor

The screenshot displays the Talend Cloud Data Fabric interface (R2023-10) for a project named 'Final\_proj'. The main workspace shows a data integration job titled 'Job Bridge\_Collision\_Contributing\_Factor'. The job flow includes several components: 'Final\_db' (817722 rows in 6.21s), 'Fact\_Accident\_table (Lookup)' (131720.68 rows/s), 'Dim\_Vehicle (Lookup)' (85 rows in 0.04s), 'Bridge\_Collision\_Contributing\_Factor (Join)' (1350571 rows in 89.65s), 'tUniqRow\_1' (1350459 rows in 89.71s), and 'Final\_db' (1350459 rows in 89.71s). The job is currently running, and the 'Designer' tab shows the job's execution details.

**Job Bridge\_Collision\_Contributing\_Factor**

**Basic Run**

Execution

Run Kill Clear

Advanced settings

Target Exec

Memory Run

Starting job Bridge\_Collision\_Contributing\_Factor at 20:37 07/04/2024.

[statistics] connecting to socket on port 3985

[statistics] connected

[statistics] disconnected

Job Bridge\_Collision\_Contributing\_Factor ended at 20:39 07/04/2024. [Exit code = 0]

No. of rows in Bridge\_Collision\_contributing\_Factor :

The screenshot displays the Microsoft SQL Server Management Studio interface. The 'Object Explorer' shows the 'Final\_Project' database structure, including tables like 'Bridge\_Collision\_Contributing\_Factor', 'Bridge\_Vehicle\_involved', 'Dim\_Date', 'Dim\_Location', 'Dim\_Time', 'Dim\_Vehicle', 'Fact\_Accident\_table', 'Integrated\_Dataset', 'stg\_austin', 'stg\_Chicago', 'stg\_NY\_dataset', and 'stg\_Vehicle\_involved'. The 'Query Editor' shows a SQL query that counts the number of rows in each of these tables. The 'Results' pane shows the output of the query, which is a single row with the value 1350459.

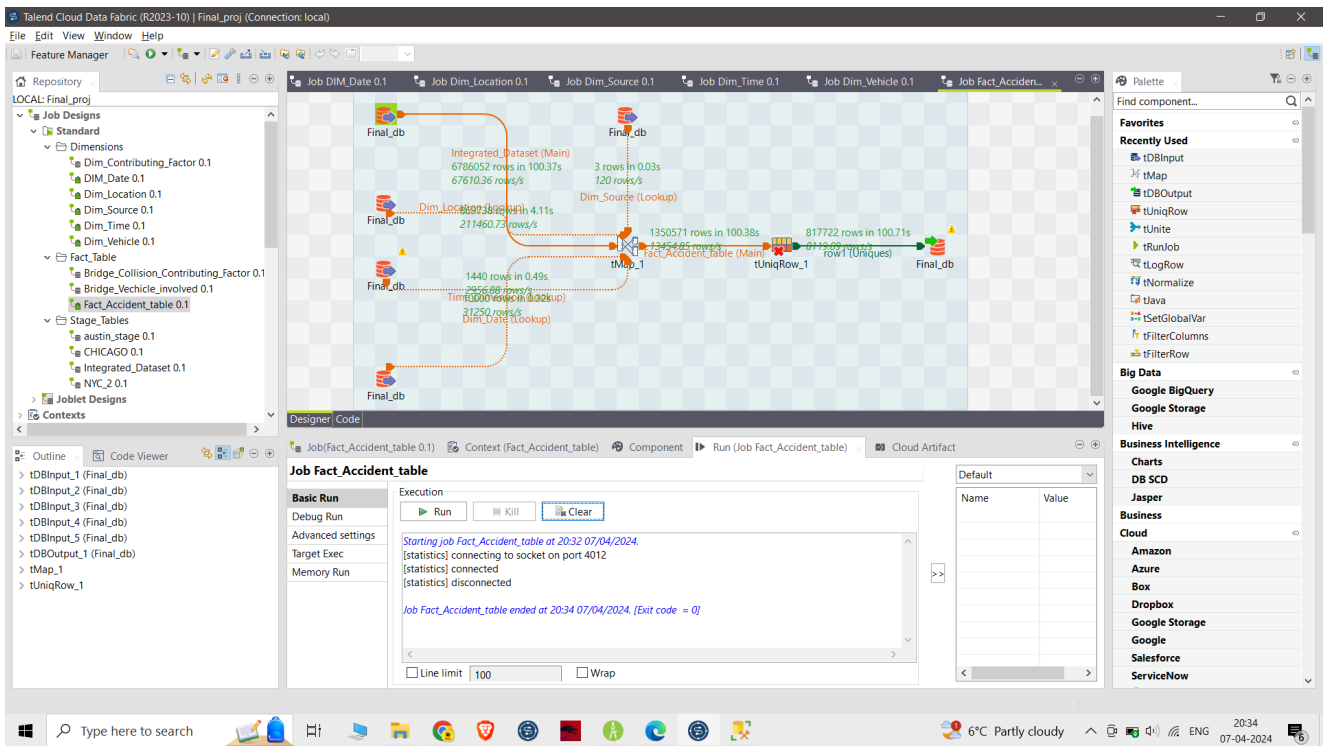
**SQLQuery6.sql - localhost\Final\_Project (d2 (59))\***

```
SELECT COUNT(*) as stg_austin_count FROM [Final_Project].[dbo].stg_austin
SELECT COUNT(*) as stg_Chicago_count FROM [Final_Project].[dbo].stg_Chicago
SELECT COUNT(*) as Integrated_Dataset_count FROM [Final_Project].[dbo].Integrated_Dataset
SELECT COUNT(*) as Dim_Contributing_Factor_count FROM [Final_Project].[dbo].Dim_Contributing_Factor
SELECT COUNT(*) as Dim_Date_count FROM [Final_Project].[dbo].Dim_Date
SELECT COUNT(*) as Dim_Location_count FROM [Final_Project].[dbo].Dim_Location
SELECT COUNT(*) as Dim_Time_count FROM [Final_Project].[dbo].Dim_Time
SELECT COUNT(*) as Dim_Vehicle_count FROM [Final_Project].[dbo].Dim_Vehicle
SELECT COUNT(*) as Dim_Source_count FROM [Final_Project].[dbo].Dim_Source
SELECT COUNT(*) as Fact_Accident_table_count FROM [Final_Project].[dbo].Fact_Accident_table
SELECT COUNT(*) as Bridge_Collision_Contributing_Factor_count FROM [Final_Project].[dbo].Bridge_Collision_Contributing_Factor
SELECT COUNT(*) as Bridge_Vehicle_involved_count FROM [Final_Project].[dbo].Bridge_Vehicle_involved
```

**Results**

Bridge_Collision_Contributing_Factor_count
1350459

# Fact\_Accident\_Table



No. of rows in Fact\_Accident\_Table\_Count :

