

## Y-DATA PROFILING

Python library was used to profile all 3 datasets and understand the inconsistencies in the 3 datasets.

```

❶ import pandas as pd
from pandas_profiling import ProfileReport

# Function to read a TSV file and generate a profiling report
def read_and_generate_profile_report(file_name, city_name):
    try:
        # Reading the TSV file
        df = pd.read_csv(file_name, sep='\t', on_bad_lines='skip')

        # Generating the profiling report
        profile = ProfileReport(df, title=f"Profiling Report of {city_name}")

        # Saving the report to an HTML file
        profile.to_file(f"{city_name}_Profiling_Report.html")
        print(f"Profiling Report for {city_name} saved successfully.")

    except Exception as e:
        print(f"An error occurred while generating the report for {city_name}: {e}")

# List of TSV files and city names
files_and_cities = [
    ('Austin.tsv', 'Austin'),
    ('Chicago.tsv', 'Chicago'),
    ('NY.tsv', 'New York')
]

# Read each TSV file and generate its profiling report
for file_name, city_name in files_and_cities:
    read_and_generate_profile_report(file_name, city_name)

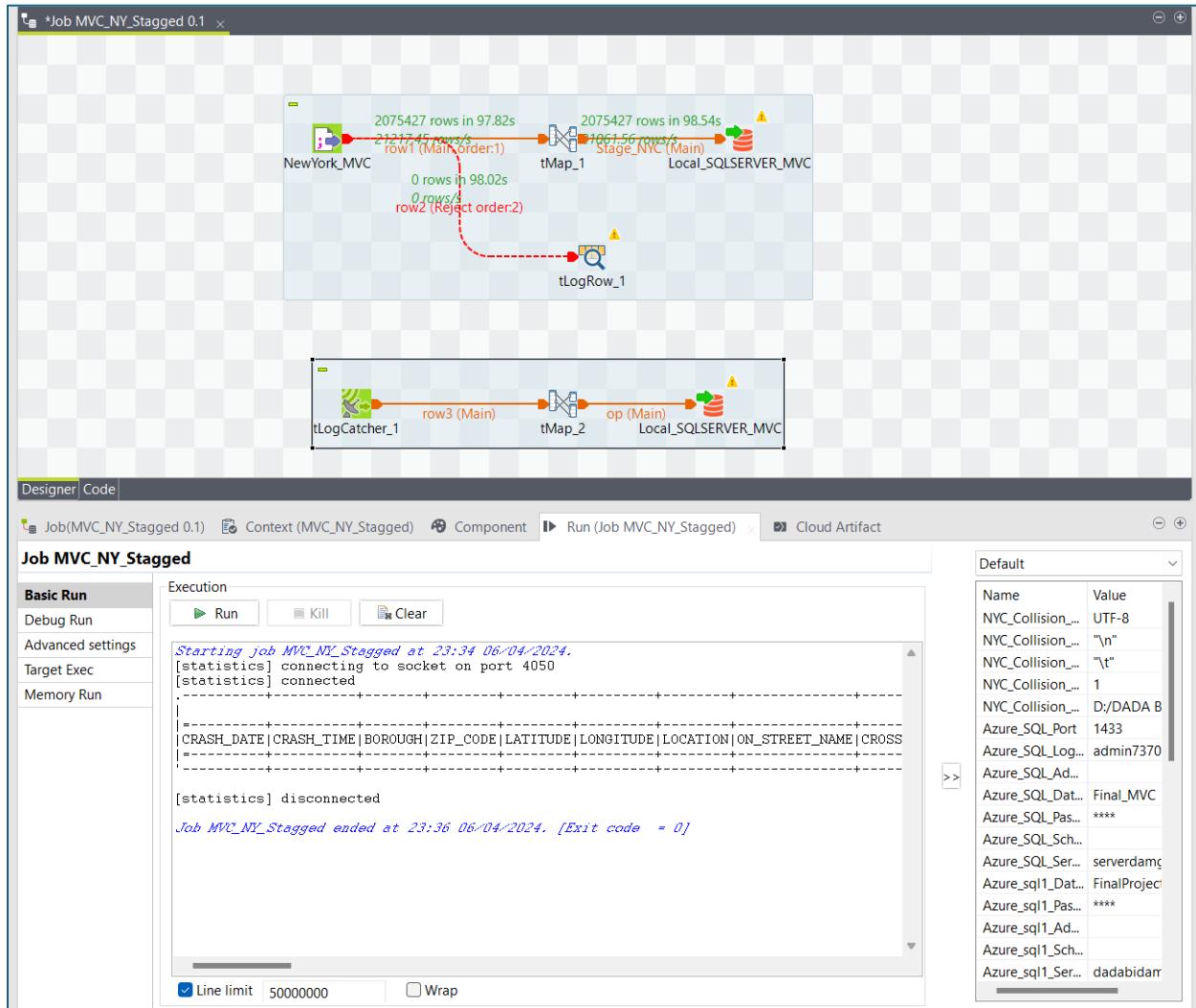
❷ /var/folders/v4/zbs69z_s0j77jdv8gdrmjw0000gn/T/ipykernel_15104/3843584403.py:2: DeprecationWarning: 'import pandas_profiling' is going to be deprecated by April 1st. Please use `import ydata_profiling` from pandas_profiling import ProfileReport
/var/folders/v4/zbs69z_s0j77jdv8gdrmjw0000gn/T/ipykernel_15104/3843584403.py:8: DtypeWarning: Columns (7,53) have mixed types. Specify dtype option on import or set low_memory=False.
df = pd.read_csv(file_name, sep='\t', on_bad_lines='skip')
/Users/pradyashinde/anaconda3/lib/python3.11/site-packages/ydata_profiling/profile_report.py:354: UserWarning: Try running command: 'pip install --upgrade Pillow' to avoid ValueError
warnings.warn(
    Summarize dataset:  0%|          | 0/5 [00:00<?, ?it/s]
    /Users/pradyashinde/anaconda3/lib/python3.11/site-packages/scipy/stats/_stats_py.py:5445: ConstantInputWarning: An input array is constant; the correlation coefficient is not defined.
    warnings.warn(stats.ConstantInputWarning(warn_msg))
    /Users/pradyashinde/anaconda3/lib/python3.11/site-packages/scipy/stats/_stats_py.py:5445: ConstantInputWarning: An input array is constant; the correlation coefficient is not defined.
    warnings.warn(stats.ConstantInputWarning(warn_msg))
    /Users/pradyashinde/anaconda3/lib/python3.11/site-packages/scipy/stats/_stats_py.py:5445: ConstantInputWarning: An input array is constant; the correlation coefficient is not defined.
    warnings.warn(stats.ConstantInputWarning(warn_msg))
    Generate report structure:  0%|          | 0/1 [00:00<?, ?it/s]
    Render HTML:  0%|          | 0/1 [00:00<?, ?it/s]
    Export report to file:  0%|          | 0/1 [00:00<?, ?it/s]
    Profiling Report for Austin saved successfully.
    /Users/pradyashinde/anaconda3/lib/python3.11/site-packages/ydata_profiling/profile_report.py:354: UserWarning: Try running command: 'pip install --upgrade Pillow' to avoid ValueError
    warnings.warn(
    Summarize dataset:  0%|          | 0/5 [00:00<?, ?it/s]
    Generate report structure:  0%|          | 0/1 [00:00<?, ?it/s]
    Render HTML:  0%|          | 0/1 [00:00<?, ?it/s]
    Export report to file:  0%|          | 0/1 [00:00<?, ?it/s]
    Profiling Report for Chicago saved successfully.
    /var/folders/v4/zbs69z_s0j77jdv8gdrmjw0000gn/T/ipykernel_15104/3843584403.py:8: DtypeWarning: Columns (3) have mixed types. Specify dtype option on import or set low_memory=False.
    df = pd.read_csv(file_name, sep='\t', on_bad_lines='skip')
    /Users/pradyashinde/anaconda3/lib/python3.11/site-packages/ydata_profiling/profile_report.py:354: UserWarning: Try running command: 'pip install --upgrade Pillow' to avoid ValueError
    warnings.warn(
    Summarize dataset:  0%|          | 0/5 [00:00<?, ?it/s]
    Generate report structure:  0%|          | 0/1 [00:00<?, ?it/s]
    Render HTML:  0%|          | 0/1 [00:00<?, ?it/s]
    Export report to file:  0%|          | 0/1 [00:00<?, ?it/s]
    Profiling Report for New York saved successfully.

```

## TALEND WORKFLOWS

The **New York dataset** was initially staged in its original form using a Talend workflow. Subsequently, this staged table served as the input for another Talend workflow, where all required transformations were performed based on the profiling report that was generated as well as the additional requirements. Finally, the transformed and normalized table was staged in SQL Server. Screenshots for the same as below:

### NEW YORK STAGED- AS IS WORKFLOW



-- NEW YORK CLEANED

```
SELECT * FROM stg_MVC_NY
SELECT COUNT(*) FROM stg_MVC_NY
```

-- NEW YORK NORMALIZED

	CRASH_DATE	CRASH_TIME	BOROUGH	ZIP_CODE	LATITUDE	LONGITUDE	LOCATION	ON_STREET_NAME	CROSS_STREET_NAME	OFF_STREET_NAME	NUMBER_OF_PERSONS_INJU
1	2016-04-04 00:00:00.000	14:46			40.797	-73.93778	(40.797,-73.93778)			EAST 116 STREET	0
2	2016-04-04 00:00:00.000	15:35	QUEENS	11356	40.78167	-73.83152	(40.78167,-73.83152)			136-03 20 AVENUE	0
3	2016-03-29 00:00:00.000	10:37			NULL	NULL				12 AVENUE	0
4	2016-03-22 00:00:00.000	8:54	MANHATTAN	10025	NULL	NULL		WEST END AVENUE	WEST 107 STREET		
5	2016-04-10 00:00:00.000	21:15	MANHATTAN	10018	40.7545	-73.9918	(40.7545,-73.9918)			557 8 AVENUE	1
6	2016-03-17 00:00:00.000	8:00	QUEENS	11106	40.76546	-73.92957	(40.76546,-73.92957)			31-41 23 STREET	0
7	2016-03-21 00:00:00.000	12:00			NULL	NULL		HORACE HARDING EXPRESSWAY	SPRINGFIELD BOULEVARD		
8	2016-03-29 00:00:00.000	20:25	BROOKLYN	11228	NULL	NULL		CROPSEY AVENUE	15 AVENUE		
9	2016-03-24 00:00:00.000	13:55	BROOKLYN	11228	40.60928	-74.00935	(40.60928,-74.00935)			69 BAY 11 STREET	0
10	2016-03-22 00:00:00.000	11:00			NULL	NULL		NEW LOTS AVENUE	ASHFORD STREET		
11	2016-03-31 00:00:00.000	16:30			NULL	NULL				VICTORY BOULEVARD	3
12	2016-04-03 00:00:00.000	19:30	QUEENS	11368	NULL	NULL		34 AVENUE	113 STREET		
13	2016-04-08 00:00:00.000	6:30	QUEENS	11413	NULL	NULL				147-50 231 street	0
14	2016-03-28 00:00:00.000	12:30			NULL	NULL		staten island expy			
15	2016-03-19 00:00:00.000	22:00	QUEENS	11432	NULL	NULL		HILLSIDE AVENUE	MIDLAND PARKWAY		
16	2016-04-07 00:00:00.000	11:36			40.7553	-73.97536	(40.755302,-73.97536)			EAST 47 STREET	0
17	2016-03-24 00:00:00.000	0:00			40.50883	-74.00677	(40.508827,-74.00677)				
18	2016-03-23 00:00:00.000	10:30	BRONX	10458	40.85434	-73.89133	(40.85434,-73.89133)	Belt Pkwy		540 EAST 183 STREET	0
19	2016-03-22 00:00:00.000	6:22			40.86174	-73.9118	(40.86174,-73.9118)			WEST FORDHAM ROAD	0
20	2016-03-17 00:00:00.000	8:30			40.67447	-73.84942	(40.674465,-73.84942)			133 AVENUE	0
21	2016-04-11 00:00:00.000	20:45	BRONX	10470	NULL	NULL		ONEIDA AVENUE	EAST 236 STREET		
22	2016-04-03 00:00:00.000	2:10	BROOKLYN	11211	40.70558	-73.95047	(40.705578,-73.95047)			502 BROADWAY	1
23	2016-04-10 00:00:00.000	17:09	BRONX	10451	NULL	NULL		MAJOR DEEGAN EXPRESSWAY	EAST 138 STREET		
24	2016-04-12 00:00:00.000	14:00	MANHATTAN	10003	NULL	NULL		2 AVENUE	EAST 9 STREET		
25	2016-03-30 00:00:00.000	16:15	BROOKLYN	11226	NULL	NULL		CHURCH AVENUE	LLOYD STREET		
26	2016-04-04 00:00:00.000	15:35	BROOKLYN	11221	NULL	NULL		STUYVESANT AVENUE	DE KALB AVENUE		
27	2016-04-07 00:00:00.000	7:30	STATEN ISLAND	10306	NULL	NULL		TYSENS LANE	CLAWSON STREET		
28	2016-03-29 00:00:00.000	17:30			40.8047	-73.91243	(40.8047,-73.91243)	BRUCKNER BOULEVARD	EAST 138 STREET		
29	2016-03-26 00:00:00.000	11:00	BRONX	10458	40.85363	-73.88881	(40.853634,-73.88881)			614 CRESCENT AVENUE	0
30	2016-03-21 00:00:00.000	8:30	BROOKLYN	11234	NULL	NULL		PRESTON COURT	RALPH AVENUE		
31	2016-04-12 00:00:00.000	15:21	MANHATTAN	10032	NULL	NULL		EDGEcombe AVENUE	WEST 166 STREET		
32	2016-04-02 00:00:00.000	16:13	BROOKLYN	11216	NULL	NULL		ROGERS AVENUE	SAINT JOHNS PLACE		
33	2016-03-30 00:00:00.000	12:25	BROOKLYN	11222	40.73312	-73.9598	(40.733116,-73.9598)			178 WEST STREET	0
34	2016-04-01 00:00:00.000	14:32			NULL	NULL		VANDERVORT AVENUE	MEEKER AVENUE		
35	2016-03-22 00:00:00.000	13:30			NULL	NULL		Harlem River Dr			
36	2016-03-20 00:00:00.000	15:25	BROOKLYN	11230	40.61158	-73.9681	(40.611576,-73.9681)			1510 OCEAN PARKWAY	0
37	2016-04-09 00:00:00.000	13:55	MANHATTAN	10029	NULL	NULL		EAST 116 STREET	PARK AVENUE		
38	2016-03-18 00:00:00.000	18:50	BROOKLYN	11225	NULL	NULL		BEDFORD AVENUE	SULLIVAN PLACE		

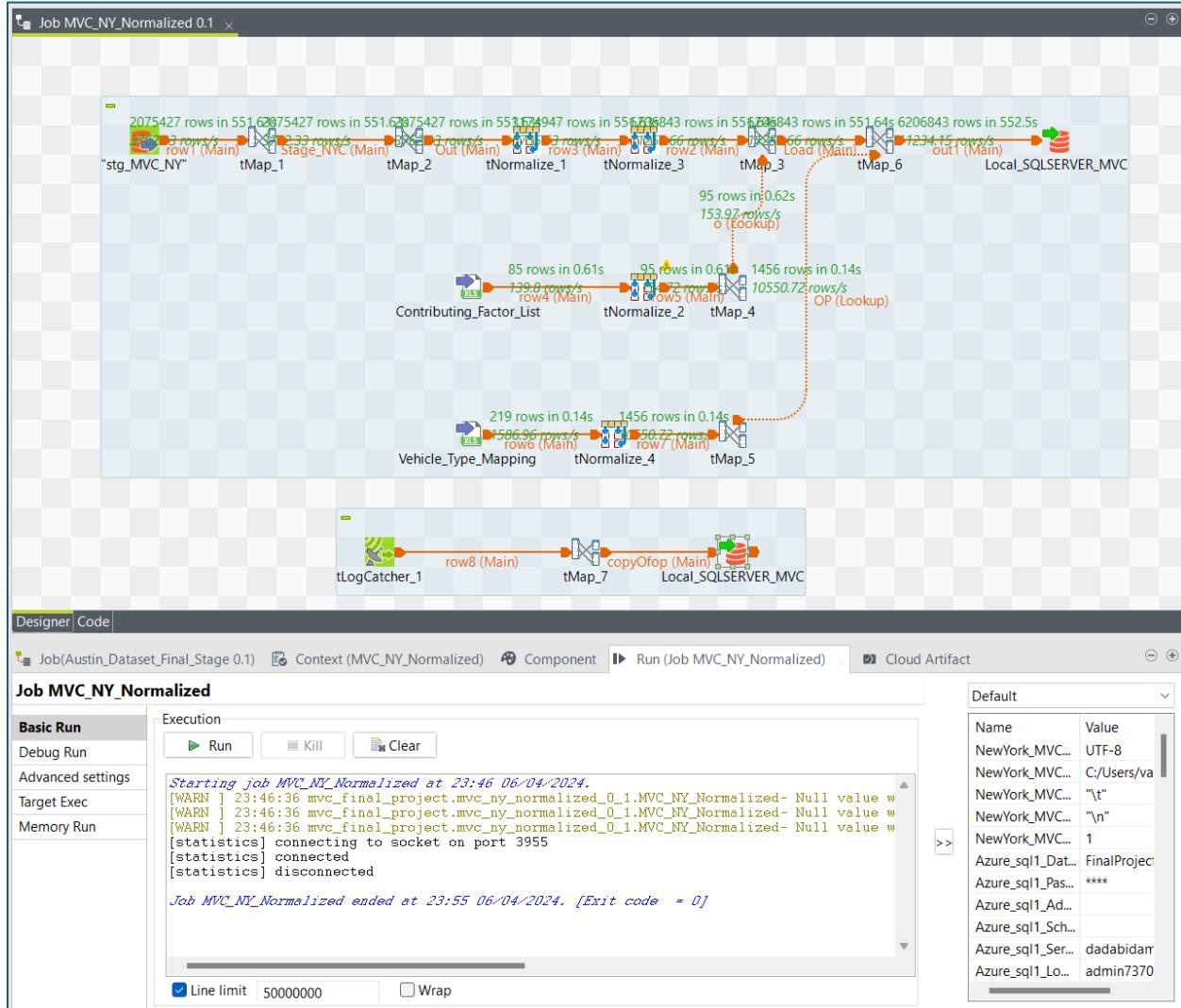
Executing query... VAISHVEER (16.0 RTM) VAISHVEER\vaish (57) MVC\_FINAL\_PROJECT 00:00:03 | 0 rows

-- NEW YORK CLEANED

```
SELECT * FROM stg_MVC_NY
SELECT COUNT(*) FROM stg_MVC_NY
```

-- NEW YORK NORMALIZED

	(No column name)
1	2075427

**NEW YORK NORMALIZED WORKFLOW**

## GROUP5

SQL QUERIES.sql - V...h (57) Executing...\* → X SQLQuery1.sql - VA...SHVEER\vaish (54)\*

```

SELECT COUNT(*) FROM stg_MVC_NY

--NEW YORK NORMALIZED
SELECT * FROM stg_MVC_NY_normalized
SELECT COUNT(*) FROM stg_MVC_NY_normalized
SELECT * FROM stg_MVC_NY_normalized WHERE CONTRIBUTING_FACTOR_VEHICLE_OPR='UNINSPECTED'

```

110 %

Results Messages

	COLLISION_ID	CRASH_DATE	Source	CRASH_TIME	CRASH_HOUR	BOROUGH	ZIP_CODE	LATITUDE	LONGITUDE	LOCATION	ON_STREET_NAME	CROSS_STREET_NAME	OFF_STREET_N
1	3409835	2016-04-04 00:00:00.000	NEW YORK	14:46:00	14	NA	-1	40.797	-73.93778	(40.797,-73.93778)	NA	NA	EAST 116 STRE
2	3409835	2016-04-04 00:00:00.000	NEW YORK	14:46:00	14	NA	-1	40.797	-73.93778	(40.797,-73.93778)	NA	NA	EAST 116 STRE
3	3409835	2016-04-04 00:00:00.000	NEW YORK	14:46:00	14	NA	-1	40.797	-73.93778	(40.797,-73.93778)	NA	NA	EAST 116 STRE
4	3409835	2016-04-04 00:00:00.000	NEW YORK	14:46:00	14	NA	-1	40.797	-73.93778	(40.797,-73.93778)	NA	NA	EAST 116 STRE
5	3426390	2016-04-04 00:00:00.000	NEW YORK	15:35:00	15	QUEENS	11356	40.78167	-73.83152	(40.78167,-73.83152)	NA	NA	136-03 20 AVE
6	3426390	2016-04-04 00:00:00.000	NEW YORK	15:35:00	15	QUEENS	11356	40.78167	-73.83152	(40.78167,-73.83152)	NA	NA	136-03 20 AVE
7	3408171	2016-03-29 00:00:00.000	NEW YORK	10:37:00	10	NA	-1	-1	-1	-1	NA	NA	12 AVENUE
8	3408171	2016-03-29 00:00:00.000	NEW YORK	10:37:00	10	NA	-1	-1	-1	-1	NA	NA	12 AVENUE
9	3409569	2016-03-22 00:00:00.000	NEW YORK	8:54:00	8	MANHATTAN	10025	-1	-1	-1	WEST END AVENUE	WEST 107 STREET	NA
10	3409569	2016-03-22 00:00:00.000	NEW YORK	8:54:00	8	MANHATTAN	10025	-1	-1	-1	WEST END AVENUE	WEST 107 STREET	NA
11	3409569	2016-03-22 00:00:00.000	NEW YORK	8:54:00	8	MANHATTAN	10025	-1	-1	-1	WEST END AVENUE	WEST 107 STREET	NA
12	3409569	2016-03-22 00:00:00.000	NEW YORK	8:54:00	8	MANHATTAN	10025	-1	-1	-1	WEST END AVENUE	WEST 107 STREET	NA
13	3407404	2016-04-10 00:00:00.000	NEW YORK	21:15:00	21	MANHATTAN	10018	40.7545	-73.9918	(40.7545,-73.9918)	NA	NA	557 8 AVENU
14	3407404	2016-04-10 00:00:00.000	NEW YORK	21:15:00	21	MANHATTAN	10018	40.7545	-73.9918	(40.7545,-73.9918)	NA	NA	557 8 AVENU
15	3407404	2016-04-10 00:00:00.000	NEW YORK	21:15:00	21	MANHATTAN	10018	40.7545	-73.9918	(40.7545,-73.9918)	NA	NA	557 8 AVENU
16	3407404	2016-04-10 00:00:00.000	NEW YORK	21:15:00	21	MANHATTAN	10018	40.7545	-73.9918	(40.7545,-73.9918)	NA	NA	557 8 AVENU
17	3428680	2016-03-17 00:00:00.000	NEW YORK	8:00:00	8	QUEENS	11106	40.76546	-73.92957	(40.76546,-73.92957)	NA	NA	31-41 23 STRI
18	3428680	2016-03-17 00:00:00.000	NEW YORK	8:00:00	8	QUEENS	11106	40.76546	-73.92957	(40.76546,-73.92957)	NA	NA	31-41 23 STRI
19	3427320	2016-03-21 00:00:00.000	NEW YORK	12:00:00	12	NA	-1	-1	-1	-1	HORACE HARDING EXPRESSWAY	SPRINGFIELD BOULEVARD	NA
20	3427320	2016-03-21 00:00:00.000	NEW YORK	12:00:00	12	NA	-1	-1	-1	-1	HORACE HARDING EXPRESSWAY	SPRINGFIELD BOULEVARD	NA
21	3415223	2016-03-29 00:00:00.000	NEW YORK	20:25:00	20	BROOKLYN	11228	-1	-1	-1	CROPSEY AVENUE	15 AVENUE	NA
22	3415172	2016-03-24 00:00:00.000	NEW YORK	13:55:00	13	BROOKLYN	11228	40.60928	-74.00935	(40.609276,-74.009346)	NA	NA	69 BAY 11 S'
23	3415172	2016-03-24 00:00:00.000	NEW YORK	13:55:00	13	BROOKLYN	11228	40.60928	-74.00935	(40.609276,-74.009346)	NA	NA	69 BAY 11 S'
24	3415172	2016-03-24 00:00:00.000	NEW YORK	13:55:00	13	BROOKLYN	11228	40.60928	-74.00935	(40.609276,-74.009346)	NA	NA	69 BAY 11 S'
25	3415172	2016-03-24 00:00:00.000	NEW YORK	13:55:00	13	BROOKLYN	11228	40.60928	-74.00935	(40.609276,-74.009346)	NA	NA	69 BAY 11 S'
26	3418611	2016-03-22 00:00:00.000	NEW YORK	11:00:00	11	NA	-1	-1	-1	-1	NEW LOTS AVENUE	ASHFORD STREET	NA
27	3418611	2016-03-22 00:00:00.000	NEW YORK	11:00:00	11	NA	-1	-1	-1	-1	NEW LOTS AVENUE	ASHFORD STREET	NA
28	3430092	2016-03-31 00:00:00.000	NEW YORK	16:30:00	16	NA	-1	-1	-1	-1	NA	NA	VICTORY BOULE
29	3430092	2016-03-31 00:00:00.000	NEW YORK	16:30:00	16	NA	-1	-1	-1	-1	NA	NA	VICTORY BOULE
30	3430092	2016-03-31 00:00:00.000	NEW YORK	16:30:00	16	NA	-1	-1	-1	-1	NA	NA	VICTORY BOULE
31	3430092	2016-03-31 00:00:00.000	NEW YORK	16:30:00	16	NA	-1	-1	-1	-1	NA	NA	VICTORY BOULE
32	3429401	2016-04-02 00:00:00.000	NEW YORK	19:30:00	19	QUEENS	11368	-1	-1	-1	34 AVENUE	113 STREET	NA
33	3429401	2016-04-03 00:00:00.000	NEW YORK	19:30:00	19	QUEENS	11368	-1	-1	-1	34 AVENUE	113 STREET	NA
34	3424672	2016-04-08 00:00:00.000	NEW YORK	6:30:00	6	QUEENS	11413	-1	-1	-1	NA	NA	147-50 231 str
35	3424672	2016-04-08 00:00:00.000	NEW YORK	6:30:00	6	QUEENS	11413	-1	-1	-1	NA	NA	147-50 231 str
36	3424672	2016-04-08 00:00:00.000	NEW YORK	6:30:00	6	QUEENS	11413	-1	-1	-1	NA	NA	147-50 231 str
37	3424672	2016-04-08 00:00:00.000	NEW YORK	6:30:00	6	QUEENS	11413	-1	-1	-1	NA	NA	147-50 231 str
38	3430040	2016-03-28 00:00:00.000	NEW YORK	12:30:00	12	NA	-1	-1	-1	-1	staten island expy	NA	NA

Executing query... VAISHVEER (16.0 RTM) VAISHVEER\vaish (57) MVC\_FINAL\_PROJECT 00:00:02 0 rows

SQL QUERIES.sql - ...ISHVEER\vaish (57)\* → X SQLQuery1.sql - VA...SHVEER\vaish (54)\*

```

SELECT COUNT(*) FROM stg_MVC_NY

--NEW YORK NORMALIZED
SELECT * FROM stg_MVC_NY_normalized
SELECT COUNT(*) FROM stg_MVC_NY_normalized
SELECT * FROM stg_MVC_NY_normalized WHERE CONTRIBUTING_FACTOR_VEHICLE_OPR='UNINSPECTED'

```

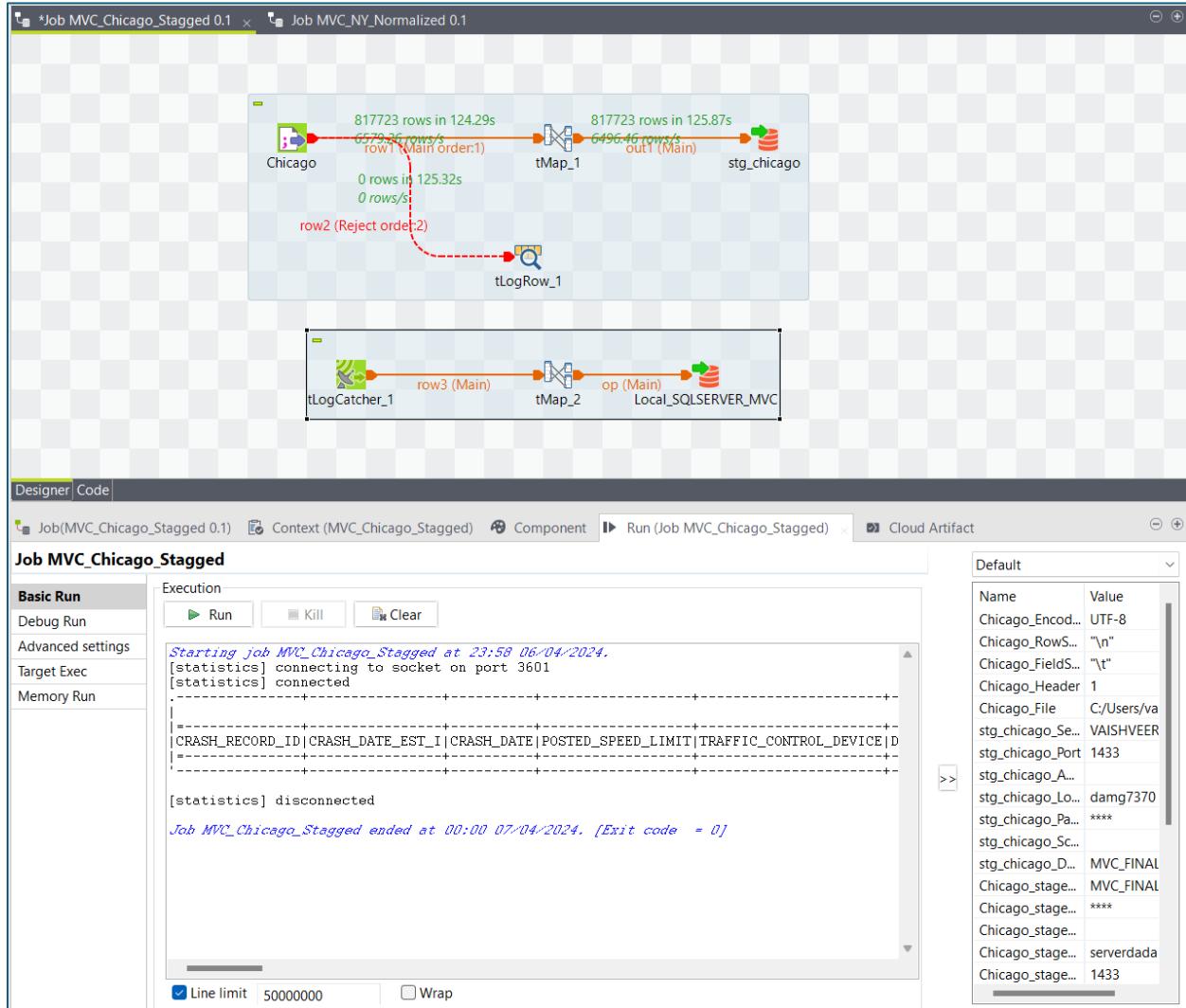
110 %

Results Messages

(No column name)
1 6206843

The **Chicago dataset** was initially staged in its original form using a Talend workflow. Subsequently, this staged table served as the input for another Talend workflow, where all required transformations were performed based on the profiling report that was generated. This staged table then served as the input for a final Talend workflow where the additional requirements were catered to. Finally, the transformed and normalized table was staged in SQL Server. Screenshots for the same as below:

### **CHICAGO STAGED- AS IS WORKFLOW**



--CHICAGO CLEANED

```
SELECT * FROM stg_MVC_Chicago
SELECT COUNT(*) FROM stg_MVC_Chicago
```

-- CHTCAGO NORMAI T7FD

Results Messages

CRASH_RECORD_ID	CRASH_DATE_EST_J	CRASH_DATE	POSTED_SPEED_LIMIT	TRAFFIC_CONTROL_DEVICE	DEVICE_CONDITION	WEATHER_CONDITION	LIGHTING_CONDITION
d98e01df3fb2a0bcc201c06bf89253ebc22e0adb2e20c2012...		2019-11-19 07:41:00.000	30	NO CONTROLS	NO CONTROLS	CLEAR	DAYLIGHT
d98ef0cbed1a68149f3e0524535d494b4b5246e5b50f5149393...		2022-09-28 18:04:00.000	30	NO CONTROLS	NO CONTROLS	CLEAR	DAYLIGHT
d98e23e61536e9305577873d53b6e2846735696b3adc7e8a...		2017-10-14 19:20:00.000	30	NO CONTROLS	NO CONTROLS	RAIN	DARKNESS
d98e3cadce881778404093e614e41e5497e347957e85...		2023-10-28 12:20:00.000	30	NO CONTROLS	NO CONTROLS	CLEAR	UNKNOWN
d98e424e096662a9334e2a2e7885ac669919c1bf8dd7b993...		2019-08-03 17:50:00.000	30	TRAFFIC SIGNAL	FUNCTIONING PROPERLY	CLEAR	DAYLIGHT
d98e5028196570b0321977eaa33a6dd8d502a09393e1e2a...	Y	2018-05-13 22:00:00.000	30	NO CONTROLS	NO CONTROLS	UNKNOWN	DARKNESS, LIGHTED ROAD
d98e59e0196575ab218aae0b7e07474b88e7e150d...		2022-06-22 22:10:00.000	30	TRAFFIC SIGNAL	FUNCTIONING PROPERLY	CLEAR	DARKNESS, LIGHTED ROAD
d98e681b7a41fb4c1044ba1c193a06b691504e8b3d0aaf9...		2019-06-26 21:40:00.000	30	NO CONTROLS	NO CONTROLS	CLEAR	DARKNESS, LIGHTED ROAD
d98e8659d96f3cd3636a099333e714274a8143ca110cc01...		2021-06-16 13:20:00.000	30	NO CONTROLS	NO CONTROLS	CLEAR	DAYLIGHT
d98e8d21166c944d4de01428a6490698a08fb0b...		2019-11-20 13:20:00.000	30	NO CONTROLS	NO CONTROLS	CLEAR	DAYLIGHT
d98e9cc75dbb5061d3ac4e8b72d553705a79deb4e044be...		2016-12-05 16:00:00.000	30	NO CONTROLS	NO CONTROLS	CLEAR	DAYLIGHT
d98e1368b4154d9b4c219af0b6c8d9d31000933d427c5...		2019-05-21 09:20:00.000	30	STOP SIGN/FLASHER	FUNCTIONING PROPERLY	CLEAR	DAYLIGHT
d98e9262b495bd241b5c1ff611e2355223990094e085d...		2019-01-23 16:00:00.000	30	NO CONTROLS	NO CONTROLS	CLEAR	DUSK
d98d2186621c1876654205384b0b18a4e1e202548cb1660...		2019-05-11 11:08:00.000	30	STOP SIGN/FLASHER	NO CONTROLS	CLOUDY/OVERCAST	DAYLIGHT
d98d372662204aa3149546e1c3d3d6ee01969686a08fb0b...		2022-03-23 21:30:00.000	30	NO CONTROLS	NO CONTROLS	CLEAR	DARKNESS, LIGHTED ROAD
d98d3d55a9244d748897840bda123b01197693916599e6d...		2017-07-15 09:00:00.000	30	NO CONTROLS	NO CONTROLS	CLEAR	DAYLIGHT
d98d422acc032896a5bb6a31e396a057929512c2ee0821e...	Y	2018-02-15 00:01:00.000	20	NO CONTROLS	NO CONTROLS	RAIN	DARKNESS, LIGHTED ROAD
d98e555b0c8194293d2b376b6d1f1a7407d12e969c06...		2023-01-11 08:47:00.000	30	NO CONTROLS	NO CONTROLS	CLEAR	DAYLIGHT
d98f1368b4154d9b4c219af0b6c8d9d31000933d427c5...		2019-03-26 23:25:00.000	30	STOP SIGN/FLASHER	NO CONTROLS	CLEAR	DARKNESS, LIGHTED ROAD
d98f7003a732a2d465d46ab708256e08249046175daff0...		2018-07-26 17:30:00.000	30	NO CONTROLS	NO CONTROLS	CLEAR	DAYLIGHT
d98f71c4fd75c03ebca90d8a09239b52ae9529a420261b...		2021-06-19 18:00:00.000	30	NO CONTROLS	NO CONTROLS	CLEAR	DAYLIGHT
d98f7d566a6669e6c27e81b2e77ae50f61d3e0e3fe9fa8f...		2020-03-29 15:45:00.000	35	STOP SIGN/FLASHER	NO CONTROLS	RAIN	DAYLIGHT
d98f9399d9fb6c8533650a72d95b612c7fb41075419d9ef...		2016-11-09 12:15:00.000	30	STOP SIGN/FLASHER	FUNCTIONING PROPERLY	CLEAR	DAYLIGHT
d98f9444a44be0bfff626225e228738b09727a78e4280de3b177b...		2019-04-27 18:10:00.000	30	STOP SIGN/FLASHER	NO CONTROLS	SNOW	DAYLIGHT
d98f95d511ed7a5f792aee9ce4d2e2e979e022ed9976813...		2021-05-01 13:30:00.000	10	NO CONTROLS	NO CONTROLS	CLEAR	DAYLIGHT
d98dd33cb394a98b83327484d0e9a3e2b073d21003...		2017-02-20 18:30:00.000	15	NO CONTROLS	NO CONTROLS	CLEAR	DUSK
d98dddec164521224bc455625e3d3be0f99071598ae4313ea...		2020-03-26 10:44:00.000	30	NO CONTROLS	NO CONTROLS	CLEAR	DAYLIGHT
d98e17441a731c851a80b40a3392d560d775424e041969...		2023-07-10 12:57:00.000	30	TRAFFIC SIGNAL	FUNCTIONING PROPERLY	CLEAR	DAYLIGHT
d98e33f585a9711641a9e3b073d2f363e4330e062ea3...		2019-06-26 08:16:00.000	30	NO CONTROLS	NO CONTROLS	CLEAR	DAYLIGHT
d98e3deddb80039606669f19279bf6279250445a066d61...		2019-11-08 22:00:00.000	25	NO CONTROLS	NO CONTROLS	CLEAR	DARKNESS
d98e5a58823db0301ae22445693dabe71d700fbfa7546cd...		2020-06-13 15:00:00	35	STOP SIGN/FLASHER	FUNCTIONING PROPERLY	CLEAR	DAYLIGHT
d98e64d2fa7c17c1850b12509396eas59fb0eca6551dae1289c...		2022-09-29 09:40:00.000	30	TRAFFIC SIGNAL	FUNCTIONING PROPERLY	CLEAR	DAYLIGHT
d98edc106774634e971b18aaad06034e2b073d21003...		2021-04-21 01:30:00.000	30	TRAFFIC SIGNAL	FUNCTIONING PROPERLY	CLEAR	DARKNESS, LIGHTED ROAD
d98ef3456486778915940d037cd41d1879495627cc28612...		2019-12-08 02:00:00.000	25	NO CONTROLS	NO CONTROLS	CLEAR	DARKNESS, LIGHTED ROAD
d98efc125dbd08c5b624b3d08fb5a5482e99ccc728144eb...		2018-09-10 17:14:00.000	30	STOP SIGN/FLASHER	FUNCTIONING PROPERLY	CLEAR	DAYLIGHT
d98f0471ef60015e50d40a3aabb65b182e6f10897e22a38c...		2018-01-20 08:55:00.000	30	STOP SIGN/FLASHER	FUNCTIONING PROPERLY	CLEAR	DAYLIGHT
d98f29731a810b3a305928dade1b26d7c48120c8677b...		2018-01-26 11:08:00.000	30	NO CONTROLS	NO CONTROLS	CLEAR	DAYLIGHT
d98f39fbea19d341a4c64833f698dab643f22472830ff8...		2017-06-14 23:33:00.000	5	NO CONTROLS	NO CONTROLS	CLEAR	DARKNESS

Executing query... VAISHVEER (16.0 RTM) VAISHVEER\vaish (57) MVC FINAL PROJECT 00:00:02 | 0 rows

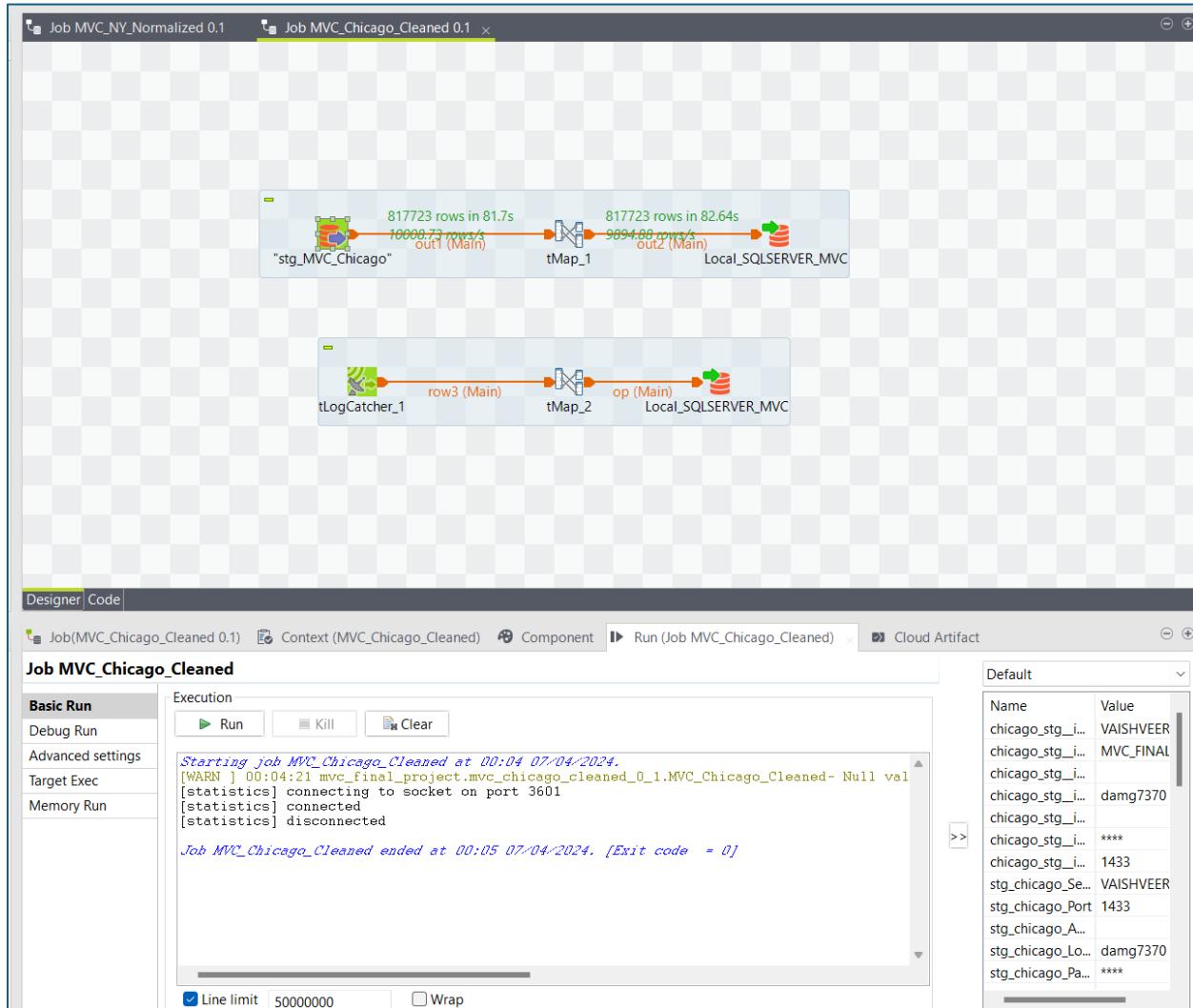
--CHICAGO CLEANED

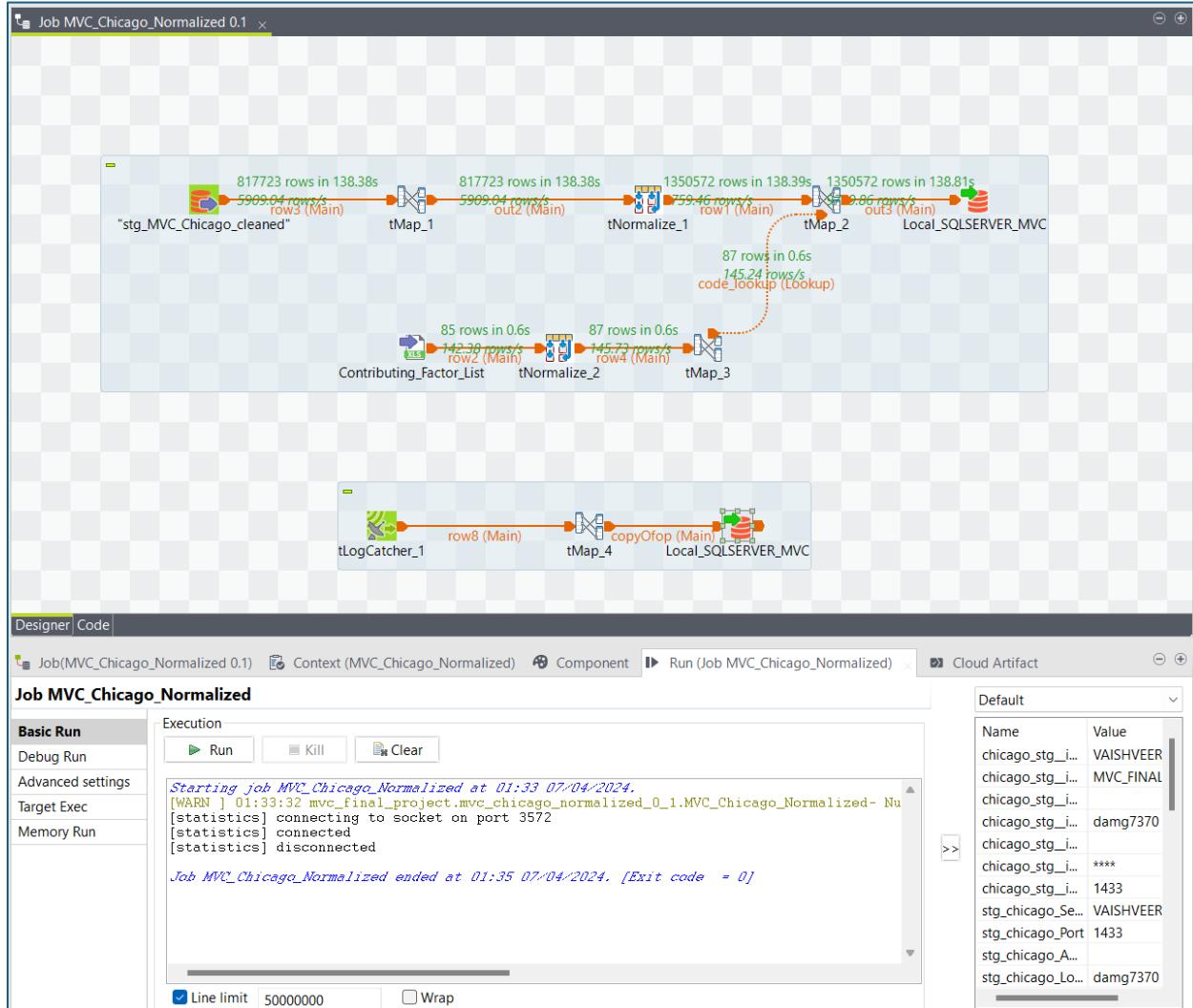
```
SELECT * FROM stg_MVC_Chicago
SELECT COUNT(*) FROM stg_MVC_Chicago
```

-- CHTCAGO NORMAI T7FD

Results Messages

(No column name)
1 817723

**CHICAGO CLEANED WORKFLOW**

**CHICAGO NORMALIZED WORKFLOW**

SQL QUERIES.sql - V...h (57) Executing... SQLQuery1.sql - VA...SHVEER(vaish (54)\*

```
--CHICAGO NORMALIZED
SELECT * FROM stg_MVC_Chicago_normalized
SELECT COUNT(*) FROM stg_MVC_Chicago_normalized
SELECT * FROM stg_MVC_Chicago_normalized WHERE CONTRTRITORY_CAUSE_CODE=72
```

110 %

Results Messages

CRASH_ID	SOURCE	CRASH_RECORD_ID	CRASH_DATE_EST_I	CRASH_DATE	CRASH_DATE_ONLY	CRASH_TIME	POSTED_SPEED_LIMIT	TRAFFIC_CONTROL_DEVICE	DEVICE_CONDITION
100	CHICAGO	711d632dd484b30e63dfcd068edf788a561d734588d5e07...	NA	2017-06-15 12:00:00.000	2017-06-15 00:00:00.000	12:00:00.000	35	TRAFFIC SIGNAL	FUNCTIONING PF
101	CHICAGO	711d64a4b4705bdd083c1b8556866d3a6104ee0ed2a50f...	NA	2017-10-14 14:10:00.000	2017-10-14 00:00:00.000	14:10:00.000	30	TRAFFIC SIGNAL	FUNCTIONING PF
102	CHICAGO	71248bae100b1360c1a2c419a317d5a87887878f14163acd8...	NA	2024-01-18 09:00:00.000	2024-01-18 00:00:00.000	09:00:00.000	30	NO CONTROLS	NO CONTROLS
103	CHICAGO	71248bae100b1360c1a2c419a317d5a87887878f14163acd8...	NA	2024-01-18 09:00:00.000	2024-01-18 00:00:00.000	09:00:00.000	30	NO CONTROLS	NO CONTROLS
104	CHICAGO	71248bae100b1360c1a2c419a317d5a87887878f14163acd8...	NA	2024-01-18 09:00:00.000	2024-01-18 00:00:00.000	09:00:00.000	30	NO CONTROLS	NO CONTROLS
105	CHICAGO	71248bae100b1360c1a2c419a317d5a87887878f14163acd8...	NA	2024-01-18 09:00:00.000	2024-01-18 00:00:00.000	09:00:00.000	30	NO CONTROLS	NO CONTROLS
106	CHICAGO	71248bae100b1360c1a2c419a317d5a87887878f14163acd8...	NA	2024-01-18 09:00:00.000	2024-01-18 00:00:00.000	09:00:00.000	30	NO CONTROLS	NO CONTROLS
107	CHICAGO	71248bae100b1360c1a2c419a317d5a87887878f14163acd8...	NA	2024-01-18 09:00:00.000	2024-01-18 00:00:00.000	09:00:00.000	30	NO CONTROLS	NO CONTROLS
108	CHICAGO	71248bae100b1360c1a2c419a317d5a87887878f14163acd8...	NA	2024-01-18 09:00:00.000	2024-01-18 00:00:00.000	09:00:00.000	30	NO CONTROLS	NO CONTROLS
109	CHICAGO	71248bae100b1360c1a2c419a317d5a87887878f14163acd8...	NA	2024-01-18 09:00:00.000	2024-01-18 00:00:00.000	09:00:00.000	30	NO CONTROLS	NO CONTROLS
110	CHICAGO	71248bae100b1360c1a2c419a317d5a87887878f14163acd8...	NA	2024-01-18 09:00:00.000	2024-01-18 00:00:00.000	09:00:00.000	30	NO CONTROLS	NO CONTROLS
111	CHICAGO	71248bae100b1360c1a2c419a317d5a87887878f14163acd8...	NA	2024-01-18 09:00:00.000	2024-01-18 00:00:00.000	09:00:00.000	30	NO CONTROLS	NO CONTROLS
112	CHICAGO	71248bae100b1360c1a2c419a317d5a87887878f14163acd8...	NA	2024-01-18 09:00:00.000	2024-01-18 00:00:00.000	09:00:00.000	30	NO CONTROLS	NO CONTROLS
113	CHICAGO	71248bae100b1360c1a2c419a317d5a87887878f14163acd8...	NA	2024-01-18 09:00:00.000	2024-01-18 00:00:00.000	09:00:00.000	30	NO CONTROLS	NO CONTROLS
114	CHICAGO	71248bae100b1360c1a2c419a317d5a87887878f14163acd8...	NA	2024-01-18 09:00:00.000	2024-01-18 00:00:00.000	09:00:00.000	30	NO CONTROLS	NO CONTROLS
115	CHICAGO	71248bae100b1360c1a2c419a317d5a87887878f14163acd8...	NA	2024-01-18 09:00:00.000	2024-01-18 00:00:00.000	09:00:00.000	30	NO CONTROLS	NO CONTROLS
116	CHICAGO	71248bae100b1360c1a2c419a317d5a87887878f14163acd8...	NA	2024-01-18 09:00:00.000	2024-01-18 00:00:00.000	09:00:00.000	30	NO CONTROLS	NO CONTROLS
117	CHICAGO	71248bae100b1360c1a2c419a317d5a87887878f14163acd8...	NA	2024-01-18 09:00:00.000	2024-01-18 00:00:00.000	09:00:00.000	30	NO CONTROLS	NO CONTROLS
118	CHICAGO	71248bae100b1360c1a2c419a317d5a87887878f14163acd8...	NA	2024-01-18 09:00:00.000	2024-01-18 00:00:00.000	09:00:00.000	30	NO CONTROLS	NO CONTROLS
119	CHICAGO	71248bae100b1360c1a2c419a317d5a87887878f14163acd8...	NA	2024-01-18 09:00:00.000	2024-01-18 00:00:00.000	09:00:00.000	30	NO CONTROLS	NO CONTROLS
120	CHICAGO	71248bae100b1360c1a2c419a317d5a87887878f14163acd8...	NA	2024-01-18 09:00:00.000	2024-01-18 00:00:00.000	09:00:00.000	30	NO CONTROLS	NO CONTROLS
121	CHICAGO	71248bae100b1360c1a2c419a317d5a87887878f14163acd8...	NA	2024-01-18 09:00:00.000	2024-01-18 00:00:00.000	09:00:00.000	30	NO CONTROLS	NO CONTROLS

Executing query... VAISHVEER (16.0 RTM) | VAISHVEER(vaish (57) | MVC\_FINAL\_PROJECT | 00:00:04 | 0 rows

--CHICAGO NORMALIZED

```
SELECT * FROM stg_MVC_Chicago_normalized
SELECT COUNT(*) FROM stg_MVC_Chicago_normalized
SELECT * FROM stg_MVC_Chicago_normalized WHERE CONTRTRITORY_CAUSE_CODE=72
```

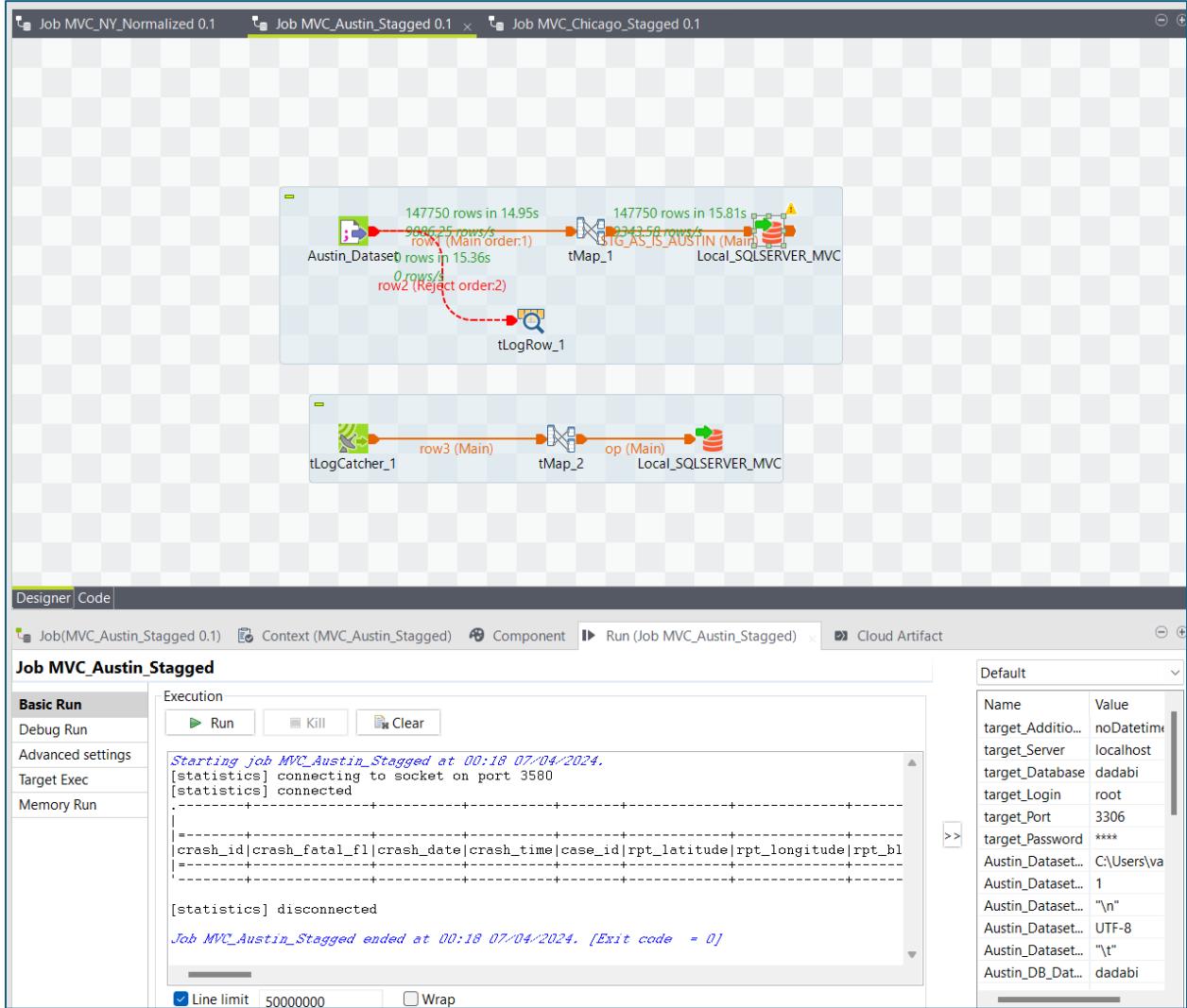
110 %

Results Messages

(No column name)
1350572

The **Austin dataset** was initially staged in its original form using a Talend workflow. Subsequently, this staged table served as the input for another Talend workflow, where all required transformations and the additional requirements were performed based on the profiling report that was generated. Finally, the transformed and normalized table was staged in SQL Server. Screenshots for the same as below:

### **AUSTIN STAGED- AS IS WORKFLOW**



## GROUP5

```

SELECT * FROM stg_MVC_Austin
SELECT COUNT(*) FROM stg_MVC_Austin
SELECT * FROM stg_MVC_Austin where Contributing_Factor_Code=72

```

110 %

Results Messages

	crash_id	crash_fatal_flg	crash_date	crash_time	case_id	rpt_latitude	rpt_longitude	rpt_block_num	rpt_street_pfx	rpt_street_name	rpt_street_sfx	crash_speed_limit	road_constr_zone_flg	latitude	longitude
1	13762420	N	2014-03-30 10:58:00.000	10:58:00	140890874	NULL	NULL	3400		PALM WAY TO MOPAC NB RAMP	RD	10	N	30.404	-97.724
2	13777334	N	2014-03-27 13:07:00.000	13:07:00	140860852	NULL	NULL	8704		BALCONES CLUB DR	DR	-1	N	30.43798	-97.780
3	13777441	N	2014-03-28 15:42:00.000	15:42:00	140871196	NULL	NULL	8704		E US 290 HWY SRV EB	BLVD	60	N	30.32762	-97.662
4	13797332	N	2014-04-09 14:09:00.000	14:09:00	140991015	NULL	NULL	8000		BEN WHITE	BLVD	-1	N	30.22516	-97.767
5	13795604	N	2014-04-07 18:00:00.000	18:00:00	140971248	NULL	NULL	200	W	IH 35 SVRD	RD	50	N	30.16984	-97.784
6	13765070	N	2014-03-31 03:26:00.000	03:26:00	140900191	NULL	NULL	8700	S	S FM 973 RD	RD	-1	N	30.1932	-97.647
7	13790426	N	2014-04-04 14:34:00.000	14:34:00	140941160	NULL	NULL	4000		ANDERSON	LN	65	N	30.33279	-97.686
8	13795213	N	2014-04-18 02:06:00.000	02:06:00	141080164	NULL	NULL	1500	E	MOPAC	EXPY	65	N	30.29613	-97.718
9	13786430	N	2014-04-04 00:11:00.000	00:11:00	140940020	NULL	NULL	3400	N	IH 35 NB	RD	55	N	30.26578	-97.782
10	13792606	N	2014-04-18 10:05:00.000	10:05:00	141080445	NULL	NULL	700	S	N IH 35 SRV DB	BLVD	55	N	30.27857	-97.730
11	13803033	N	2014-04-13 13:06:00.000	13:06:00	141090908	NULL	NULL	2100		MIDDLE FISKVILLE	RD	40	N	30.37493	-97.678
12	13765043	N	2014-03-28 16:04:00.000	16:04:00	140871241	NULL	NULL			E BEN WHITE BLVD EB	BLVD	50	N	30.22219	-97.683
13	13800672	N	2014-04-11 12:13:00.000	12:13:00	141010879	NULL	NULL	8300		1ST	ST	35	N	30.236	-97.763
14	13792523	N	2014-04-08 10:00:00.000	10:00:00	140980602	NULL	NULL	2900	S	E ST JOHNS AVE	AVE	35	N	30.33423	-97.707
15	13781584	N	2014-03-31 08:59:00.000	08:59:00	140900474	NULL	NULL	400	E	SLAUGHTER	LN	40	N	30.17245	-97.799
16	13777523	N	2014-04-09 04:14:00.000	04:14:00	140940563	NULL	NULL	600	W	IH 35 SB	HWY	65	N	30.23571	-97.740
17	13797628	N	2014-04-10 12:30:00.000	12:30:00	141000856	NULL	NULL	2200	S	WILLIAM CANNON	DR	40	N	30.23373	-97.684
18	13767418	N	2014-04-01 10:10:00.000	10:10:00	140910083	NULL	NULL	6300	W	IH 35	FWY	70	N	30.3782	-97.675
19	13788264	N	2014-04-15 08:30:00.000	08:30:00	141050418	NULL	NULL	11400	N	RESEARCH ST AT N MOPAC R...	RD	30	N	30.36387	-97.702
20	13781499	N	2014-03-29 09:25:00.000	09:25:00	140880700	NULL	NULL			LAKELINE	BLVD	45	N	30.47948	-97.803
21	13802545	N	2014-04-08 20:18:00.000	20:18:00	140981718	NULL	NULL			MANCHACA RD	RD	35	N	30.20284	-97.806
22	13803139	N	2014-04-14 21:35:00.000	21:35:00	141041715	NULL	NULL			RESEARCB ST AT N MOPAC R...	RD	-1	N	30.38445	-97.742
23	13781338	N	2014-04-10 13:30:00.000	13:30:00	141000953	NULL	NULL			BRAKER LN	LN	40	N	30.32822	-97.688
24	13792522	N	2014-04-09 16:59:00.000	16:59:00	140991324	NULL	NULL	888	W	MANCHACA	RD	55	N	30.16418	-97.830
25	13777582	N	2014-04-06 18:00:00.000	18:00:00	140961051	NULL	NULL			IH 35	RD	65	N	30.39929	-97.673
26	13776685	N	2014-03-26 22:01:00.000	22:01:00	140851833	NULL	NULL	12100	N	FRANCIA	TRL	-1	N	30.34085	-97.701
27	13803040	N	2014-04-13 13:00:00.000	13:00:00	141030973	NULL	NULL	7700		CEDAR AVE	AVE	-1	N	30.27605	-97.713
28	13790527	N	2014-04-04 17:37:00.000	17:37:00	140941351	NULL	NULL	6300		BURNET RD	RD	45	N	30.21916	-97.690
29	13805688	N	2014-04-25 17:37:00.000	17:37:00	141151531	NULL	NULL	11700	S	BERKMAN DR	DR	35	N	30.31671	-97.690
30	13786908	N	2014-04-07 01:04:00.000	01:04:00	140970067	NULL	NULL			IH 35 SVRD	RD	-1	N	30.10925	-97.808
31	13805308	N	2014-04-18 21:12:00.000	21:12:00	140900361	NULL	NULL	9008		E ANDERSON EB TO N IH 35 S...	RD	-1	N	30.17068	-97.794
32	1001	Y	2014-03-22 22:26:00.000	22:26:00	240621579	NULL	NULL			NOT REPORTED	RD	1	N	30.36666	-97.693
33	13799205	N	2014-04-16 17:37:00.000	17:37:00	141061423	NULL	NULL	1301		IH 35 SRV DB	BLVD	45	N	30.19545	-97.767
34	13790942	N	2014-04-06 13:30:00.000	13:30:00	140960863	NULL	NULL	8000		IH 35	RD	70	N	30.34686	-97.711
35	13765712	N	2014-03-29 02:46:00.000	02:46:00	140880296	NULL	NULL	6200	N	LAMAR	BLVD	45	N	30.26811	-97.733
36	13788308	N	2014-04-15 08:23:00.000	08:23:00	141050412	30.34692	-97.7116	8000				-1	N	30.40583	-97.678
37	13781529	N	2014-03-29 17:09:00.000	17:09:00	140881373	NULL	NULL	800				-1	N		
38	13784546	N	2014-04-09 17:20:00.000	17:20:00	140991369	NULL	NULL	12400	N			-1	N		

Query canceled.

VAISHVEER (16.0 RTM) VAISHVEER(vaish) MVC\_FINAL\_PROJECT 00:00:02 133,064 rows

```

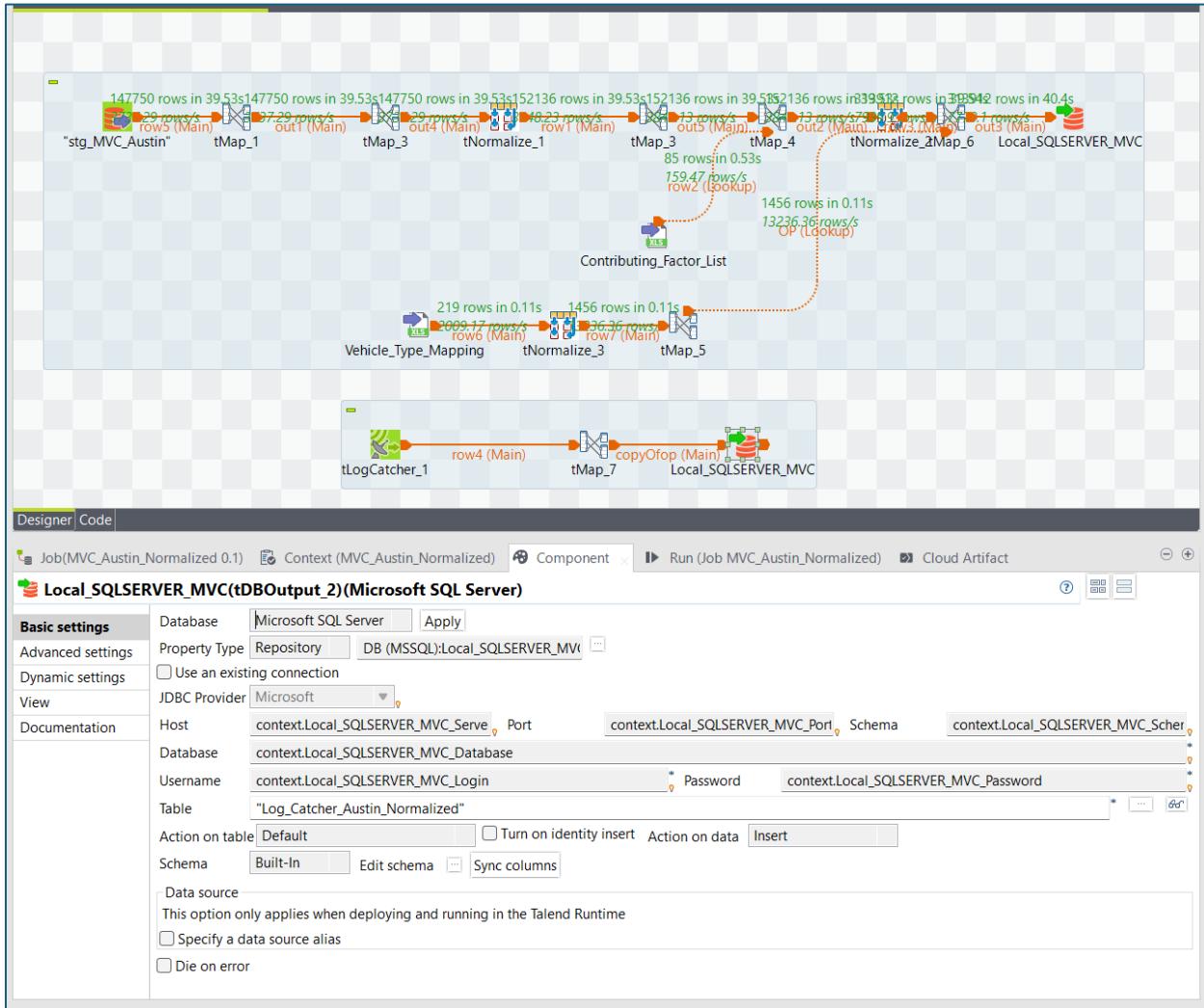
--AUSTIN CLEANED
SELECT * FROM stg_MVC_Austin
SELECT COUNT(*) FROM stg_MVC_Austin
SELECT * FROM stg_MVC_Austin where Contributing_Factor_Code=72

```

110 %

Results Messages

(No column name)
1 147750

AUSTIN NORMALIZED WORKFLOW

## GROUP5

SQL QUERIES.sql - V...h (57) Executing... SQLQuery1.sql - VA...SHVEER\vaish (54)\*

```
--AUSTIN NORMALIZED
SELECT * FROM stg_MVC_Austin_normalized
SELECT COUNT(*) FROM stg_MVC_Austin_normalized
SELECT * FROM stg_MVC_Austin_normalized WHERE Contributing_Factor_Code =-1
SELECT * FROM stg_MVC_Austin_normalized WHERE Contributing_Factor_Description=NA
```

110 %

	crash_id	Source	crash_fatal_flg	crash_date	crash_date_only	crash_time	crash_hour	case_id	rpt_latitude	rpt_longitude	rpt_block_num	rpt_street_pfx	rpt_street_name	rpt_street_sfx	crash_speed
1	13762420	AUSTIN	N	2014-03-30 10:58:00.000	2014-03-30 00:00:00.000	10:58:00	10	1408990874	-1	-1	NA	3707 MANCHACA	RD	10	
2	1377334	AUSTIN	N	2014-03-27 13:07:00.000	2014-03-27 00:00:00.000	13:07:00	13	140860852	-1	-1	3400	NA	PALM WAY TO MOPAC NB RAMP	NA	50
3	13777441	AUSTIN	N	2014-03-28 15:42:00.000	2014-03-28 00:00:00.000	15:42:00	15	140871196	-1	-1	8704	NA	BALCONES CLUB DR	DR	-1
4	13797332	AUSTIN	N	2014-04-09 14:09:00.000	2014-04-09 00:00:00.000	14:09:00	14	140991015	-1	-1	8000	NA	E US 290 HWY SVRD EB	NA	60
5	13795804	AUSTIN	N	2014-04-07 18:00:00.000	2014-04-07 00:00:00.000	18:00:00	18	140971248	-1	-1	200	W	BEN WHITE	BLVD	-1
6	13765070	AUSTIN	N	2014-03-31 03:26:00.000	2014-03-31 00:00:00.000	03:26:00	3	140900191	-1	-1	8700	S	IH 35 SVRD	NA	50
7	13790426	AUSTIN	N	2014-04-14 14:34:00.000	2014-04-04 00:00:00.000	14:34:00	14	140941160	-1	-1	4000	NA	S FM 973 RD	RD	-1
8	13795213	AUSTIN	N	2014-04-18 02:06:00.000	2014-04-18 00:00:00.000	02:06:00	2	141080164	-1	-1	1500	E	ANDERSON	LN	65
9	13786430	AUSTIN	N	2014-04-04 00:11:00.000	2014-04-04 00:00:00.000	00:11:00	0	140940020	-1	-1	3400	N	IH 35 NB	NA	55
10	13792606	AUSTIN	N	2014-04-18 10:05:00.000	2014-04-18 00:00:00.000	10:05:00	10	141080445	-1	-1	700	S	MOPAC	EXPY	65
11	13792606	AUSTIN	N	2014-04-18 10:05:00.000	2014-04-18 00:00:00.000	10:05:00	10	141080445	-1	-1	700	S	MOPAC	EXPY	65
12	13803033	AUSTIN	N	2014-04-13 13:06:00.000	2014-04-13 00:00:00.000	13:06:00	13	141030908	-1	-1	2100	NA	N IH 35 SVRD SB	NA	55
13	13803033	AUSTIN	N	2014-04-13 13:06:00.000	2014-04-13 00:00:00.000	13:06:00	13	141030908	-1	-1	2100	NA	N IH 35 SVRD SB	NA	55
14	13765043	AUSTIN	N	2014-03-28 16:04:00.000	2014-03-28 00:00:00.000	16:04:00	16	140871241	-1	-1	-1	NA	MIDDLE FISKVILLE	RD	40
15	13765043	AUSTIN	N	2014-03-28 16:04:00.000	2014-03-28 00:00:00.000	16:04:00	16	140871241	-1	-1	-1	NA	MIDDLE FISKVILLE	RD	40
16	13800672	AUSTIN	N	2014-04-11 12:13:00.000	2014-04-11 00:00:00.000	12:13:00	12	141010879	-1	-1	8300	NA	E BEN WHITE BLVD EB	BLVD	50
17	13800672	AUSTIN	N	2014-04-11 12:13:00.000	2014-04-11 00:00:00.000	12:13:00	12	141010879	-1	-1	8300	NA	E BEN WHITE BLVD EB	BLVD	50
18	13800672	AUSTIN	N	2014-04-11 12:13:00.000	2014-04-11 00:00:00.000	12:13:00	12	141010879	-1	-1	8300	NA	E BEN WHITE BLVD EB	BLVD	50
19	13792523	AUSTIN	N	2014-04-08 10:00:00.000	2014-04-08 00:00:00.000	10:00:00	10	140980602	-1	-1	2900	S	1ST	ST	35
20	13792523	AUSTIN	N	2014-04-08 10:00:00.000	2014-04-08 00:00:00.000	10:00:00	10	140980602	-1	-1	2900	S	1ST	ST	35
21	13781584	AUSTIN	N	2014-03-31 08:59:00.000	2014-03-31 00:00:00.000	08:59:00	8	140900474	-1	-1	400	E	E ST JOHNS AVE	AVE	35
22	13781584	AUSTIN	N	2014-03-31 08:59:00.000	2014-03-31 00:00:00.000	08:59:00	8	140900474	-1	-1	400	E	E ST JOHNS AVE	AVE	35
23	13773523	AUSTIN	N	2014-04-04 09:14:00.000	2014-04-04 00:00:00.000	09:14:00	9	140940563	-1	-1	600	W	SLAUGHTER	LN	40
24	13773523	AUSTIN	N	2014-04-04 09:14:00.000	2014-04-04 00:00:00.000	09:14:00	9	140940563	-1	-1	600	W	SLAUGHTER	LN	40
25	13797628	AUSTIN	N	2014-04-10 12:30:00.000	2014-04-10 00:00:00.000	12:30:00	12	141000856	-1	-1	2200	S	S IH 35 SB	HWY	65
26	13797628	AUSTIN	N	2014-04-10 12:30:00.000	2014-04-10 00:00:00.000	12:30:00	12	141000856	-1	-1	2200	S	S IH 35 SB	HWY	65
27	13767418	AUSTIN	N	2014-04-01 01:10:00.000	2014-04-01 00:00:00.000	01:10:00	1	140910083	-1	-1	6300	W	WILLIAM CANNON	DR	40
28	13788264	AUSTIN	N	2014-04-15 08:30:00.000	2014-04-15 00:00:00.000	08:30:00	8	141050418	-1	-1	11400	N	IH 35	FWY	70
29	13788264	AUSTIN	N	2014-04-15 08:30:00.000	2014-04-15 00:00:00.000	08:30:00	8	141050418	-1	-1	11400	N	IH 35	FWY	70
30	13788264	AUSTIN	N	2014-04-15 08:30:00.000	2014-04-15 00:00:00.000	08:30:00	8	141050418	-1	-1	11400	N	IH 35	FWY	70
31	13781499	AUSTIN	N	2014-03-29 09:25:00.000	2014-03-29 00:00:00.000	09:25:00	9	140880700	-1	-1	-1	NA	QUAIL PARK DR	DR	30
32	13781499	AUSTIN	N	2014-03-29 09:25:00.000	2014-03-29 00:00:00.000	09:25:00	9	140880700	-1	-1	-1	NA	QUAIL PARK DR	DR	30
33	13781499	AUSTIN	N	2014-03-29 09:25:00.000	2014-03-29 00:00:00.000	09:25:00	9	140880700	-1	-1	-1	NA	QUAIL PARK DR	DR	30
34	13802545	AUSTIN	N	2014-04-08 20:18:00.000	2014-04-08 00:00:00.000	20:18:00	20	140981718	-1	-1	-1	NA	LAKELINE	BLVD	45
35	13803139	AUSTIN	N	2014-04-14 21:35:00.000	2014-04-14 00:00:00.000	21:35:00	21	141041715	-1	-1	-1	NA	MANCHACA RD	RD	35
36	13803139	AUSTIN	N	2014-04-14 21:35:00.000	2014-04-14 00:00:00.000	21:35:00	21	141041715	-1	-1	-1	NA	MANCHACA RD	RD	35
37	13803139	AUSTIN	N	2014-04-14 21:35:00.000	2014-04-14 00:00:00.000	21:35:00	21	141041715	-1	-1	-1	NA	MANCHACA RD	RD	35
38	13781338	AUSTIN	N	2014-04-10 13:30:00.000	2014-04-10 00:00:00.000	13:30:00	13	141000953	-1	-1	-1	NA	RESEARCHSB AT N MOPAC R.	NA	-1

Executing query... VAISHVEER (16.0 RTM) VAISHVEER\vaish (57) MVC\_FINAL\_PROJECT 00:00:02 0 rows

SQL QUERIES.sql - ISHVEER\vaish (57)\* SQLQuery1.sql - VA...SHVEER\vaish (54)\*

```
--AUSTIN NORMALIZED
SELECT * FROM stg_MVC_Austin_normalized
SELECT COUNT(*) FROM stg_MVC_Austin_normalized
SELECT * FROM stg_MVC_Austin_normalized WHERE Contributing_Factor_Code =-1
SELECT * FROM stg_MVC_Austin_normalized WHERE Contributing_Factor_Description=NA
```

110 %

(No column name)	
1	313912

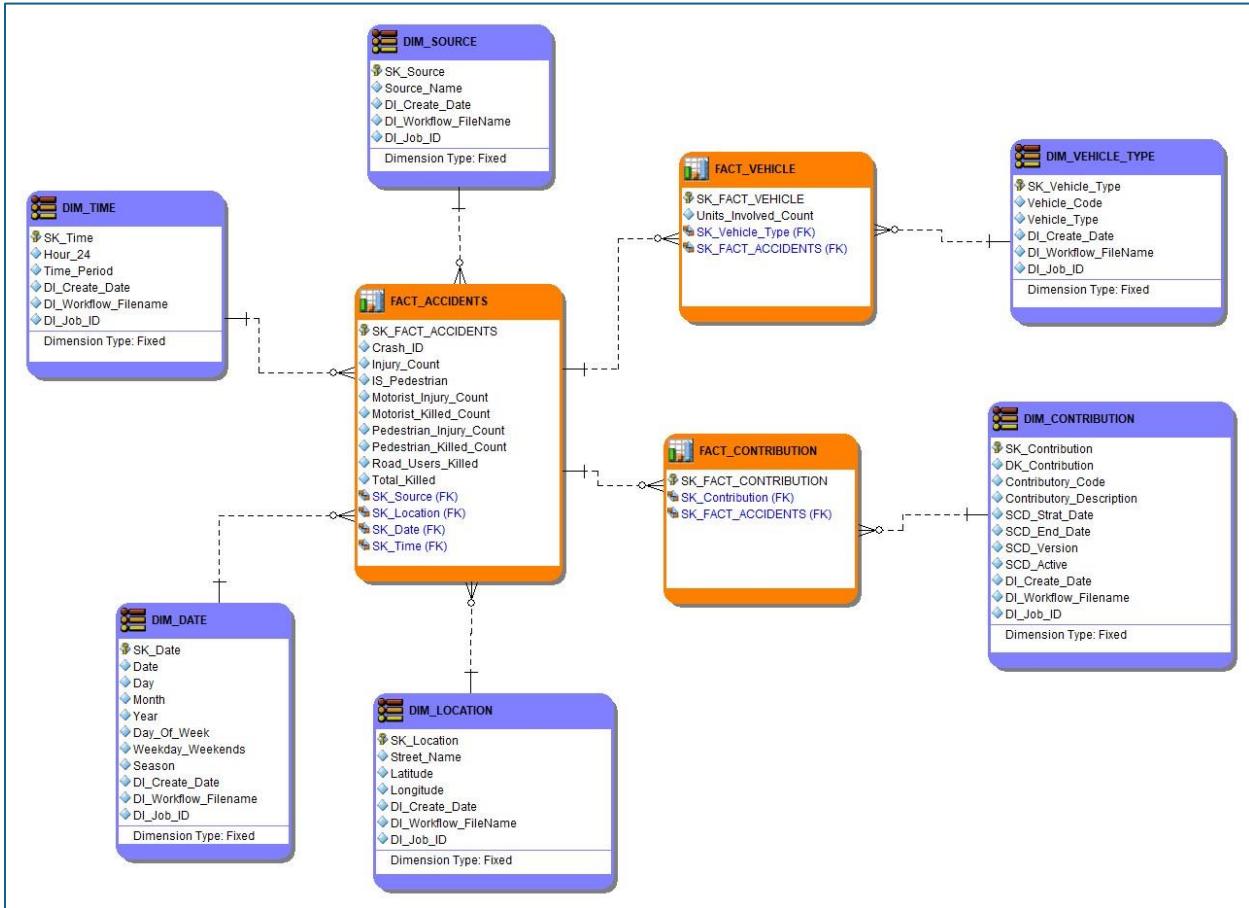
## CONTEXT VARIABLE AND ENVIRONMENTS

The screenshot shows the Talend Data Integration environment with the following components:

- Left Sidebar:** Repository tree showing "LOCAL:MVC\_FINAL\_PROJECT" with "Job Designs" (Standard, MVC), "Fact", "Joblet Designs", and "Contexts" (MVC\_Context, MVC\_Austin\_context, MVC\_Chicago\_context).
- Middle Panel:** A job design titled "Job Load\_Dim\_Location 0.1" and "Job Load\_Dim\_Contribution 0.1". It contains a flow with components like tDBInput, tMap, and tLogCatcher.
- Right Panel:**
  - Manage context environments:** A dialog box titled "Manage context environments" with the "Default (Default)" checkbox checked. Other options include Prod, Dev, and Test.
  - Manage variables:** A table listing variables with their names, types, comments, and enable prompt status. Some variables are marked as "noDate" or "\*\*\*\*\*".
  - Bottom Right:** A sidebar with various cloud and data integration service icons (Jasper, Business, Cloud, Amazon, Azure, Box, Dropbox, Google Storage, Google, Salesforce, ServiceNow, iCloudStart, iCloudStop, Custom Code, Data Quality, Matching, etc.).

Name	Type	Comment	Enable prompt
1 Local_Mysql (from repository context)	String		<input type="checkbox"/>
2 Local_Mysql_Database	String		<input type="checkbox"/>
3 Local_Mysql_AdditionalParams	String		<input type="checkbox"/>
4 Local_Mysql_Server	String		<input type="checkbox"/>
5 Local_Mysql_Password	Password		<input type="checkbox"/>
6 Local_Mysql_Login	String		<input type="checkbox"/>
7 Local_Mysql_Port	String		<input type="checkbox"/>
8 NY_staged (from repository context)	String		<input type="checkbox"/>
9 dallas_staged_Database	String		<input type="checkbox"/>
10 dallas_staged_Password	Password		<input type="checkbox"/>
11 dallas_staged_AdditionalParams	String		<input type="checkbox"/>
12 dallas_staged_Login	String		<input type="checkbox"/>
13 dallas_staged_Schema	String		<input type="checkbox"/>
14 dallas_staged_Port	String		<input type="checkbox"/>
15 dallas_staged_Server	String		<input type="checkbox"/>
16 austin_staged (from repository context)	String		<input type="checkbox"/>
17 austin_staged_Database	String		<input type="checkbox"/>
18 austin_staged_Password	Password		<input type="checkbox"/>
19 austin_staged_Schema	String		<input type="checkbox"/>
20 austin_staged_AdditionalParams	String		<input type="checkbox"/>
21 austin_staged_Server	String		<input type="checkbox"/>
22 austin_staged_Login	String		<input type="checkbox"/>
23 austin_staged_Port	String		<input type="checkbox"/>

## DIMENSIONAL MODEL LOADING



The dimensional model consists of a central fact table named **FACT\_ACCIDENTS**, which contains keys to dimensional tables, as well as accident-related metrics like injury and fatality counts. The fact table is linked to four dimension tables: **DIM\_TIME**, **DIM\_DATE**, **DIM\_SOURCE**, and **DIM\_LOCATION**, which provide detailed attributes for time, date, data source, and location of accidents, respectively.

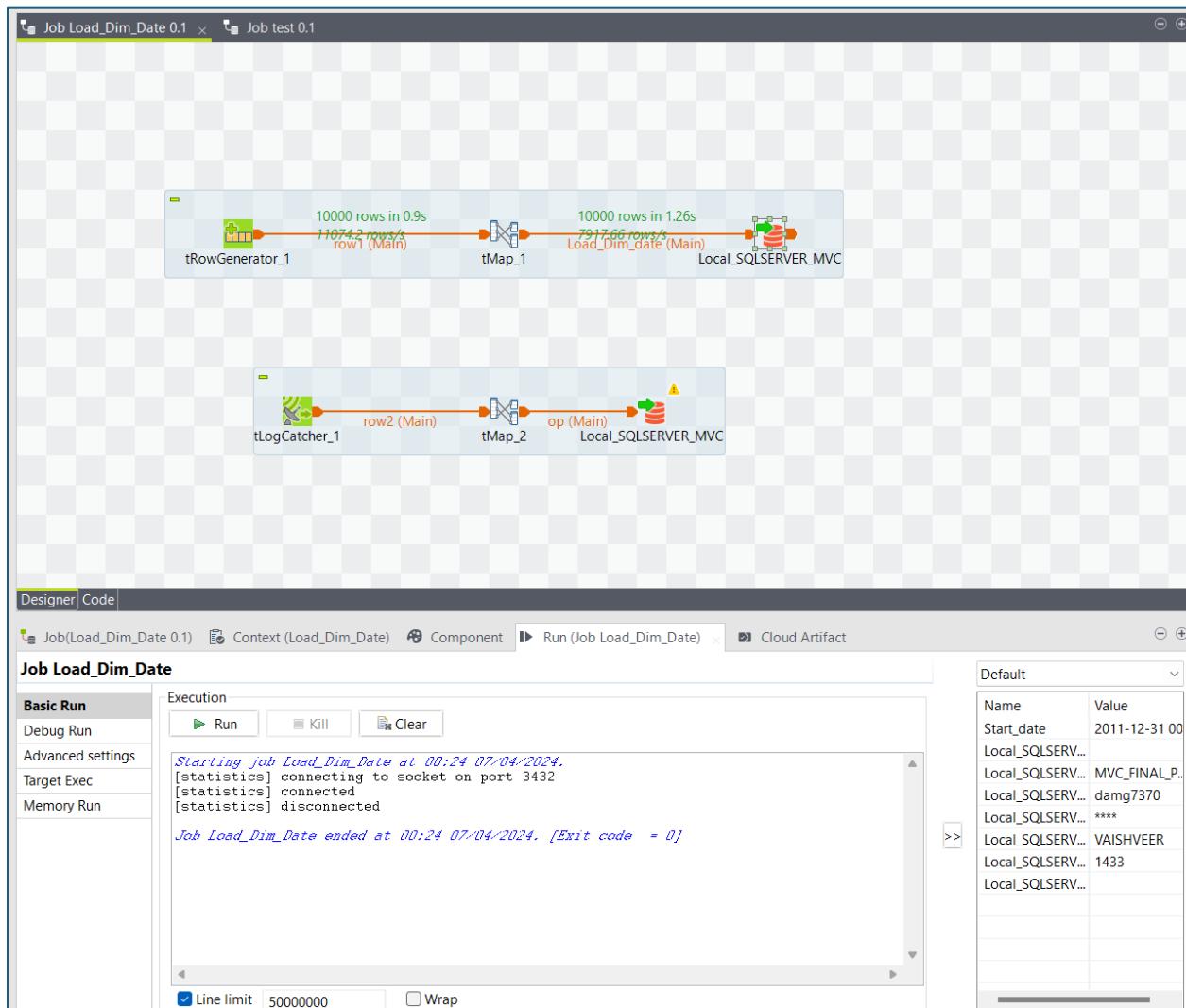
Additionally, there are two other fact tables, **FACT\_VEHICLE** and **FACT\_CONTRIBUTION**, that relate to the main **FACT\_ACCIDENTS** table. The **FACT\_VEHICLE** table includes details about the vehicles involved, like the count of units and types of vehicles, linked to its own dimension table, **DIM\_VEHICLE\_TYPE**. The **FACT\_CONTRIBUTION** table tracks contributing factors to accidents and links to the **DIM\_CONTRIBUTION** dimension table, which includes descriptions and codes for these factors.

The next step involved loading the dimensional model that was created to cater to the business requirements. Dimension created as below:

### DIM\_DATE WORKFLOW

Here, we have used a tRowGenerator to generate 10000 rows of date. This is because based on the profiling report the dataset contains date from 2012. So we have generated 10000 rows to cater to the needs.

From this generated date the day,month,year are further extracted.



```
SELECT COUNT(*) FROM Dim_Date
SELECT * FROM Dim_Date
```

110 % ▶

Results Messages

(No column name)	1	10000									
SK_Date	Date	Day	Month	Year	Season	Day_Of_The_Week	Weekdays_Weekends	DL_CreateDate	DL_JobID	DL_WorkflowName	
1	20120101	2012-01-01 00:00:00.000	1	1	2012	Winter	Sunday	Weekend	2024-04-11 16:27:29.397	LCIFHN	Load_Dim_Date
2	20120102	2012-01-02 00:00:00.000	2	1	2012	Winter	Monday	Weekday	2024-04-11 16:27:29.400	LCIFHN	Load_Dim_Date
3	20120103	2012-01-03 00:00:00.000	3	1	2012	Winter	Tuesday	Weekday	2024-04-11 16:27:29.400	LCIFHN	Load_Dim_Date
4	20120104	2012-01-04 00:00:00.000	4	1	2012	Winter	Wednesday	Weekday	2024-04-11 16:27:29.400	LCIFHN	Load_Dim_Date
5	20120105	2012-01-05 00:00:00.000	5	1	2012	Winter	Thursday	Weekday	2024-04-11 16:27:29.400	LCIFHN	Load_Dim_Date
6	20120106	2012-01-06 00:00:00.000	6	1	2012	Winter	Friday	Weekday	2024-04-11 16:27:29.400	LCIFHN	Load_Dim_Date
7	20120107	2012-01-07 00:00:00.000	7	1	2012	Winter	Saturday	Weekend	2024-04-11 16:27:29.400	LCIFHN	Load_Dim_Date
8	20120108	2012-01-08 00:00:00.000	8	1	2012	Winter	Sunday	Weekend	2024-04-11 16:27:29.403	LCIFHN	Load_Dim_Date
9	20120109	2012-01-09 00:00:00.000	9	1	2012	Winter	Monday	Weekday	2024-04-11 16:27:29.403	LCIFHN	Load_Dim_Date
10	20120110	2012-01-10 00:00:00.000	10	1	2012	Winter	Tuesday	Weekday	2024-04-11 16:27:29.403	LCIFHN	Load_Dim_Date
11	20120111	2012-01-11 00:00:00.000	11	1	2012	Winter	Wednesday	Weekday	2024-04-11 16:27:29.403	LCIFHN	Load_Dim_Date
12	20120112	2012-01-12 00:00:00.000	12	1	2012	Winter	Thursday	Weekday	2024-04-11 16:27:29.403	LCIFHN	Load_Dim_Date
13	20120113	2012-01-13 00:00:00.000	13	1	2012	Winter	Friday	Weekday	2024-04-11 16:27:29.403	LCIFHN	Load_Dim_Date
14	20120114	2012-01-14 00:00:00.000	14	1	2012	Winter	Saturday	Weekend	2024-04-11 16:27:29.403	LCIFHN	Load_Dim_Date
15	20120115	2012-01-15 00:00:00.000	15	1	2012	Winter	Sunday	Weekend	2024-04-11 16:27:29.403	LCIFHN	Load_Dim_Date
16	20120116	2012-01-16 00:00:00.000	16	1	2012	Winter	Monday	Weekday	2024-04-11 16:27:29.403	LCIFHN	Load_Dim_Date
17	20120117	2012-01-17 00:00:00.000	17	1	2012	Winter	Tuesday	Weekday	2024-04-11 16:27:29.403	LCIFHN	Load_Dim_Date
18	20120118	2012-01-18 00:00:00.000	18	1	2012	Winter	Wednesday	Weekday	2024-04-11 16:27:29.407	LCIFHN	Load_Dim_Date
19	20120119	2012-01-19 00:00:00.000	19	1	2012	Winter	Thursday	Weekday	2024-04-11 16:27:29.407	LCIFHN	Load_Dim_Date
20	20120120	2012-01-20 00:00:00.000	20	1	2012	Winter	Friday	Weekday	2024-04-11 16:27:29.407	LCIFHN	Load_Dim_Date
21	20120121	2012-01-21 00:00:00.000	21	1	2012	Winter	Saturday	Weekend	2024-04-11 16:27:29.407	LCIFHN	Load_Dim_Date
22	20120122	2012-01-22 00:00:00.000	22	1	2012	Winter	Sunday	Weekend	2024-04-11 16:27:29.407	LCIFHN	Load_Dim_Date
23	20120123	2012-01-23 00:00:00.000	23	1	2012	Winter	Monday	Weekday	2024-04-11 16:27:29.407	LCIFHN	Load_Dim_Date
24	20120124	2012-01-24 00:00:00.000	24	1	2012	Winter	Tuesday	Weekday	2024-04-11 16:27:29.407	LCIFHN	Load_Dim_Date
25	20120125	2012-01-25 00:00:00.000	25	1	2012	Winter	Wednesday	Weekday	2024-04-11 16:27:29.407	LCIFHN	Load_Dim_Date
26	20120126	2012-01-26 00:00:00.000	26	1	2012	Winter	Thursday	Weekday	2024-04-11 16:27:29.407	LCIFHN	Load_Dim_Date
27	20120127	2012-01-27 00:00:00.000	27	1	2012	Winter	Friday	Weekday	2024-04-11 16:27:29.407	LCIFHN	Load_Dim_Date
28	20120128	2012-01-28 00:00:00.000	28	1	2012	Winter	Saturday	Weekend	2024-04-11 16:27:29.407	LCIFHN	Load_Dim_Date
29	20120129	2012-01-29 00:00:00.000	29	1	2012	Winter	Sunday	Weekend	2024-04-11 16:27:29.407	LCIFHN	Load_Dim_Date
30	20120130	2012-01-30 00:00:00.000	30	1	2012	Winter	Monday	Weekday	2024-04-11 16:27:29.410	LCIFHN	Load_Dim_Date
31	20120131	2012-01-31 00:00:00.000	31	1	2012	Winter	Tuesday	Weekday	2024-04-11 16:27:29.410	LCIFHN	Load_Dim_Date
32	20120201	2012-02-01 00:00:00.000	1	2	2012	Winter	Wednesday	Weekday	2024-04-11 16:27:29.410	LCIFHN	Load_Dim_Date
33	20120202	2012-02-02 00:00:00.000	2	2	2012	Winter	Thursday	Weekday	2024-04-11 16:27:29.410	LCIFHN	Load_Dim_Date
34	20120203	2012-02-03 00:00:00.000	3	2	2012	Winter	Friday	Weekday	2024-04-11 16:27:29.410	LCIFHN	Load_Dim_Date
35	20120204	2012-02-04 00:00:00.000	4	2	2012	Winter	Saturday	Weekend	2024-04-11 16:27:29.410	LCIFHN	Load_Dim_Date
36	20120205	2012-02-05 00:00:00.000	5	2	2012	Winter	Sunday	Weekend	2024-04-11 16:27:29.410	LCIFHN	Load_Dim_Date
37	20120206	2012-02-06 00:00:00.000	6	2	2012	Winter	Monday	Weekday	2024-04-11 16:27:29.410	LCIFHN	Load_Dim_Date

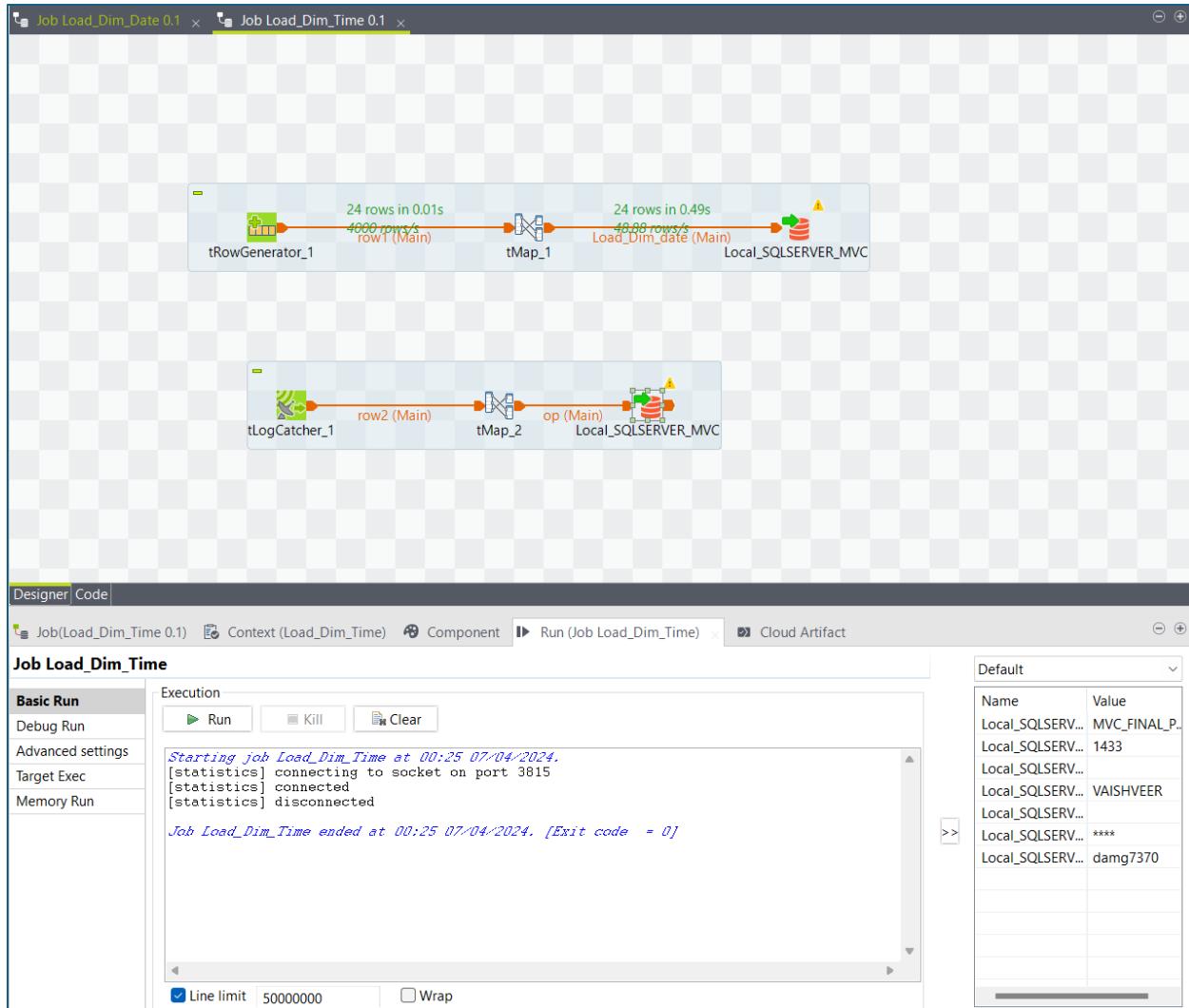
Query executed successfully.

VAISHVEER (16.0 RTM) | VAISHVEER\vaish (59) | MVC\_FINAL\_PROJECT | 00:00:00 | 10,001 rows

## DIM TIME WORKFLOW

Here, we have used a tRowGenerator to generate 24 rows of time. This is because we have considered only a granularity of 24hrs in our Time dimension based on the business requirements.

From this generated Time, the hour and time\_period was further extracted.



SQL QUERIES.sql -...ISHVEER\vaish (57)\* ⇑ × SQLQuery1.sql - VA...SHVEER\vaish (54)\*

```
SELECT COUNT(*) FROM DIM_DATE

--DIM TIME
SELECT * FROM DIM_TIME
SELECT COUNT(*) FROM DIM_TIME
```

110 %

Results Messages

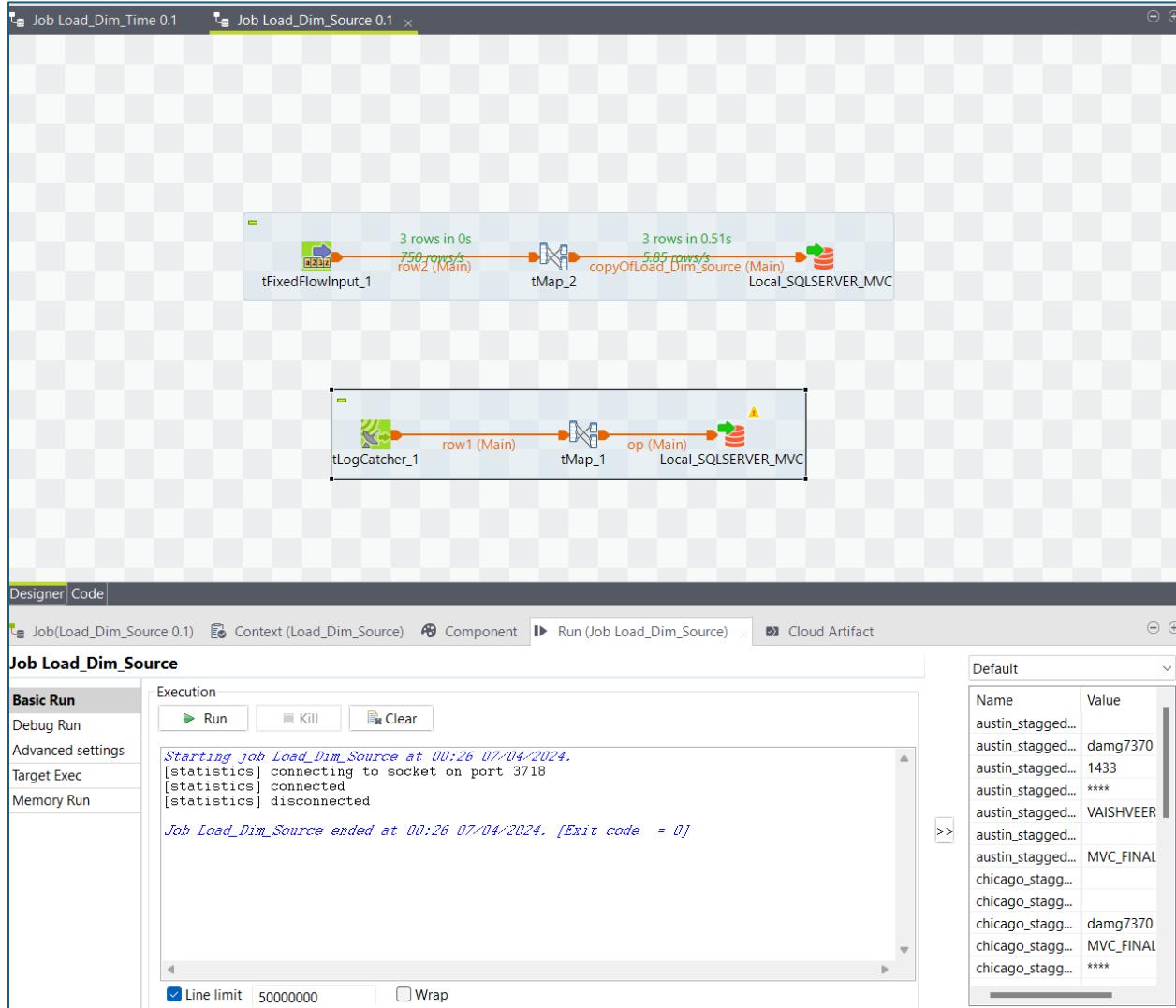
	Time_SK	Hour_24	Time_Period	DL_CreateDate	DL_JobID	DL_WorkflowName
1	1	0	Late Night	2024-04-07 09:48:28.677	kSYUWb	Load_Dim_Time
2	2	1	Late Night	2024-04-07 09:48:28.677	kSYUWb	Load_Dim_Time
3	3	2	Late Night	2024-04-07 09:48:28.677	kSYUWb	Load_Dim_Time
4	4	3	Late Night	2024-04-07 09:48:28.677	kSYUWb	Load_Dim_Time
5	5	4	Late Night	2024-04-07 09:48:28.677	kSYUWb	Load_Dim_Time
6	6	5	Early Morning	2024-04-07 09:48:28.677	kSYUWb	Load_Dim_Time
7	7	6	Early Morning	2024-04-07 09:48:28.677	kSYUWb	Load_Dim_Time
8	8	7	Early Morning	2024-04-07 09:48:28.677	kSYUWb	Load_Dim_Time
9	9	8	Morning	2024-04-07 09:48:28.677	kSYUWb	Load_Dim_Time
10	10	9	Morning	2024-04-07 09:48:28.677	kSYUWb	Load_Dim_Time
11	11	10	Morning	2024-04-07 09:48:28.677	kSYUWb	Load_Dim_Time
12	12	11	Morning	2024-04-07 09:48:28.677	kSYUWb	Load_Dim_Time
13	13	12	Early Afternoon	2024-04-07 09:48:28.677	kSYUWb	Load_Dim_Time
14	14	13	Early Afternoon	2024-04-07 09:48:28.677	kSYUWb	Load_Dim_Time
15	15	14	Afternoon	2024-04-07 09:48:28.677	kSYUWb	Load_Dim_Time
16	16	15	Afternoon	2024-04-07 09:48:28.677	kSYUWb	Load_Dim_Time
17	17	16	Afternoon	2024-04-07 09:48:28.677	kSYUWb	Load_Dim_Time
18	18	17	Evening	2024-04-07 09:48:28.677	kSYUWb	Load_Dim_Time
19	19	18	Evening	2024-04-07 09:48:28.677	kSYUWb	Load_Dim_Time
20	20	19	Evening	2024-04-07 09:48:28.677	kSYUWb	Load_Dim_Time
21	21	20	Evening	2024-04-07 09:48:28.677	kSYUWb	Load_Dim_Time
22	22	21	Night	2024-04-07 09:48:28.677	kSYUWb	Load_Dim_Time
23	23	22	Night	2024-04-07 09:48:28.677	kSYUWb	Load_Dim_Time
24	24	23	Night	2024-04-07 09:48:28.677	kSYUWb	Load_Dim_Time

(No column name)

1 24

## DIM\_SOURCE WORKFLOW

Here, we have used a tFixedFlowInput to generate 3 rows to represent the 3 different sources- Chicago, Austin and New York.



SQL QUERIES.sql -...ISHVEER\vaish (57) X SQLQuery1.sql - VA...SHVEER\vaish (54)\*

```
--DIM SOURCE

SELECT * FROM DIM_Source
SELECT COUNT(*) FROM DIM_Source
```

110 %

Results Messages

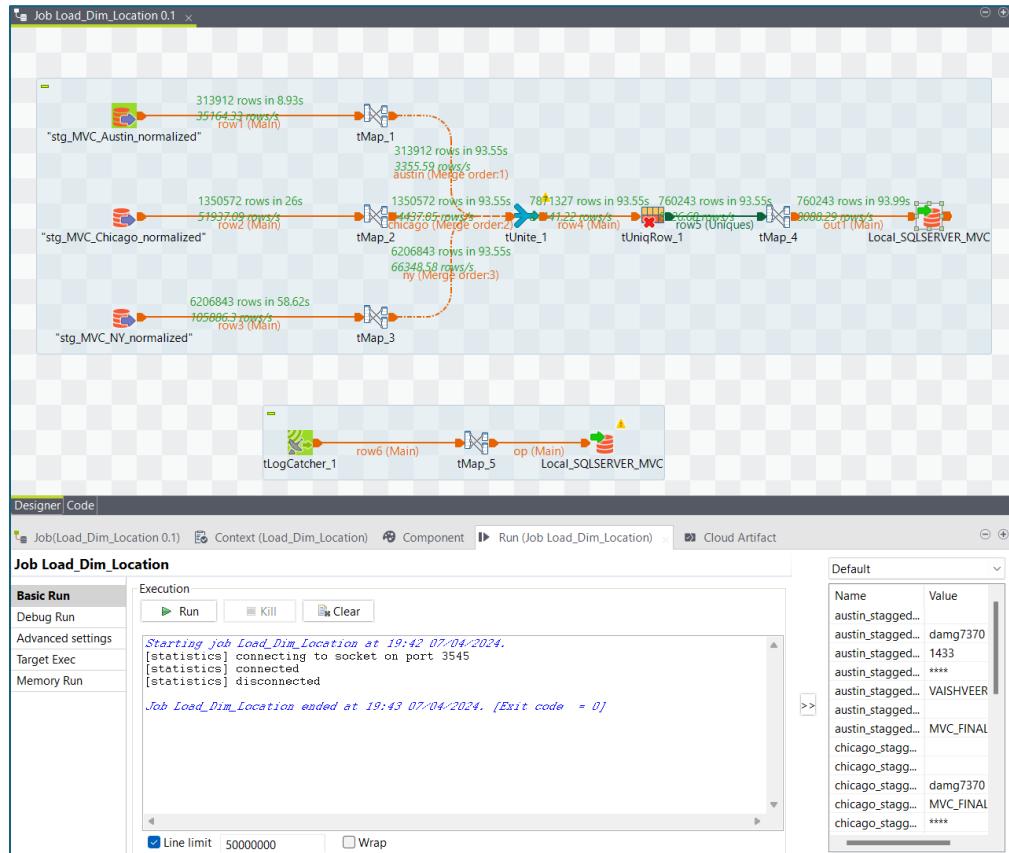
	SK_Source	Source	DI_CreateDate	DI_JobID	DI_WorkflowName
1	1	CHICAGO	2024-04-07 00:26:23.967	6y73tA	Load_Dim_Source
2	2	AUSTIN	2024-04-07 00:26:23.970	6y73tA	Load_Dim_Source
3	3	NEW YORK	2024-04-07 00:26:23.970	6y73tA	Load_Dim_Source

(No column name)

1	3
---	---

## DIM LOCATION WORKFLOW

The 3 staged final normalized tables for Chicago,Austin and New York cities are taken as input , united and then their unique combination of Street Name, Latitude and Longitude are extracted and staged in DIM\_Location



```
--DIM LOCATION
SELECT * FROM DIM_Location
SELECT COUNT(*) FROM DIM_Location

--DIM VEHICLE TYPE
```

Results Messages

	SK_Location	Street_Name	Latitude	Longitude	DL_CreateDate	DL_JobID	DL_WorkflowFileName
1	3707 MANCHACA RD	-1	-1	2024-04-07 15:00:42.877	s5roa9	Load_Dim_Location	
2	PALM WAY TO MOPAC NB RAMP	30.404	-97.72445	2024-04-07 15:00:42.877	s5roa9	Load_Dim_Location	
3	BALCONES CLUB DR	30.43798	-97.78855	2024-04-07 15:00:42.877	s5roa9	Load_Dim_Location	
4	US0290	30.32761	-97.66269	2024-04-07 15:00:42.877	s5roa9	Load_Dim_Location	
5	US0290	30.22516	-97.76762	2024-04-07 15:00:42.877	s5roa9	Load_Dim_Location	
6	IH0035	30.16984	-97.78426	2024-04-07 15:00:42.877	s5roa9	Load_Dim_Location	
7	FM0973	30.1932	-97.6474	2024-04-07 15:00:42.877	s5roa9	Load_Dim_Location	
8	E ANDERSON LN	30.33279	-97.68612	2024-04-07 15:00:42.877	s5roa9	Load_Dim_Location	
9	IH0035	30.29613	-97.71183	2024-04-07 15:00:42.877	s5roa9	Load_Dim_Location	
10	SL0001	30.26578	-97.78259	2024-04-07 15:00:42.877	s5roa9	Load_Dim_Location	
11	IH0035	30.27857	-97.73046	2024-04-07 15:00:42.877	s5roa9	Load_Dim_Location	
12	MIDDLE FISKVILLE RD	30.37493	-97.67831	2024-04-07 15:00:42.877	s5roa9	Load_Dim_Location	
13	US0183	30.22219	-97.68314	2024-04-07 15:00:42.877	s5roa9	Load_Dim_Location	
14	S 1ST ST	30.231	-97.76349	2024-04-07 15:00:42.893	s5roa9	Load_Dim_Location	
15	E ST JOHNS AVE	30.33423	-97.70716	2024-04-07 15:00:42.893	s5roa9	Load_Dim_Location	
16	W SLAUGHTER LN	30.17245	-97.77996	2024-04-07 15:00:42.893	s5roa9	Load_Dim_Location	
17	IH0035	30.23571	-97.74075	2024-04-07 15:00:42.893	s5roa9	Load_Dim_Location	
18	W WILLIAM CANNON DR	30.23374	-97.86477	2024-04-07 15:00:42.893	s5roa9	Load_Dim_Location	
19	IH0035	30.3782	-97.67532	2024-04-07 15:00:42.893	s5roa9	Load_Dim_Location	
20	QUAIL PARK DR	30.36387	-97.70285	2024-04-07 15:00:42.893	s5roa9	Load_Dim_Location	
21	LAKELINE BLVD	30.47948	-97.803	2024-04-07 15:00:42.893	s5roa9	Load_Dim_Location	
22	MENCHACA RD	30.20284	-97.80688	2024-04-07 15:00:42.893	s5roa9	Load_Dim_Location	
23	US0183	30.38445	-97.74255	2024-04-07 15:00:42.893	s5roa9	Load_Dim_Location	
24	W BRAKER LN	30.38282	-97.68813	2024-04-07 15:00:42.893	s5roa9	Load_Dim_Location	
25	FM2304	30.16418	-97.83003	2024-04-07 15:00:42.893	s5roa9	Load_Dim_Location	

✖ Query canceled. VAISHVEER (16.0 RTM) VAISHVEER(vaish) (57) MVC\_FINAL\_PROJECT 00:00:04 643,080 rows

```
--DIM LOCATION
SELECT * FROM DIM_Location
SELECT COUNT(*) FROM DIM_Location

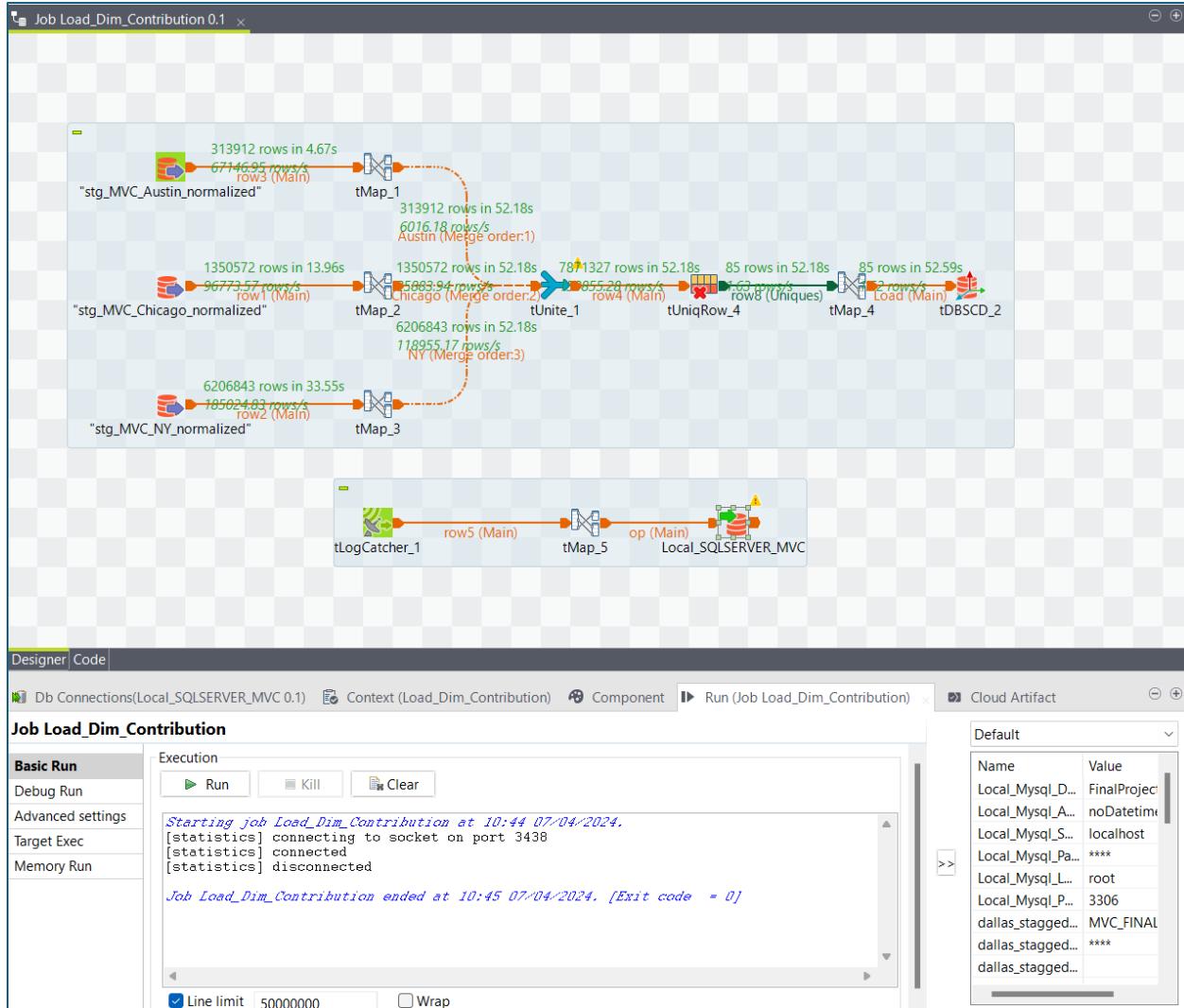
--DIM VEHICLE TYPE
```

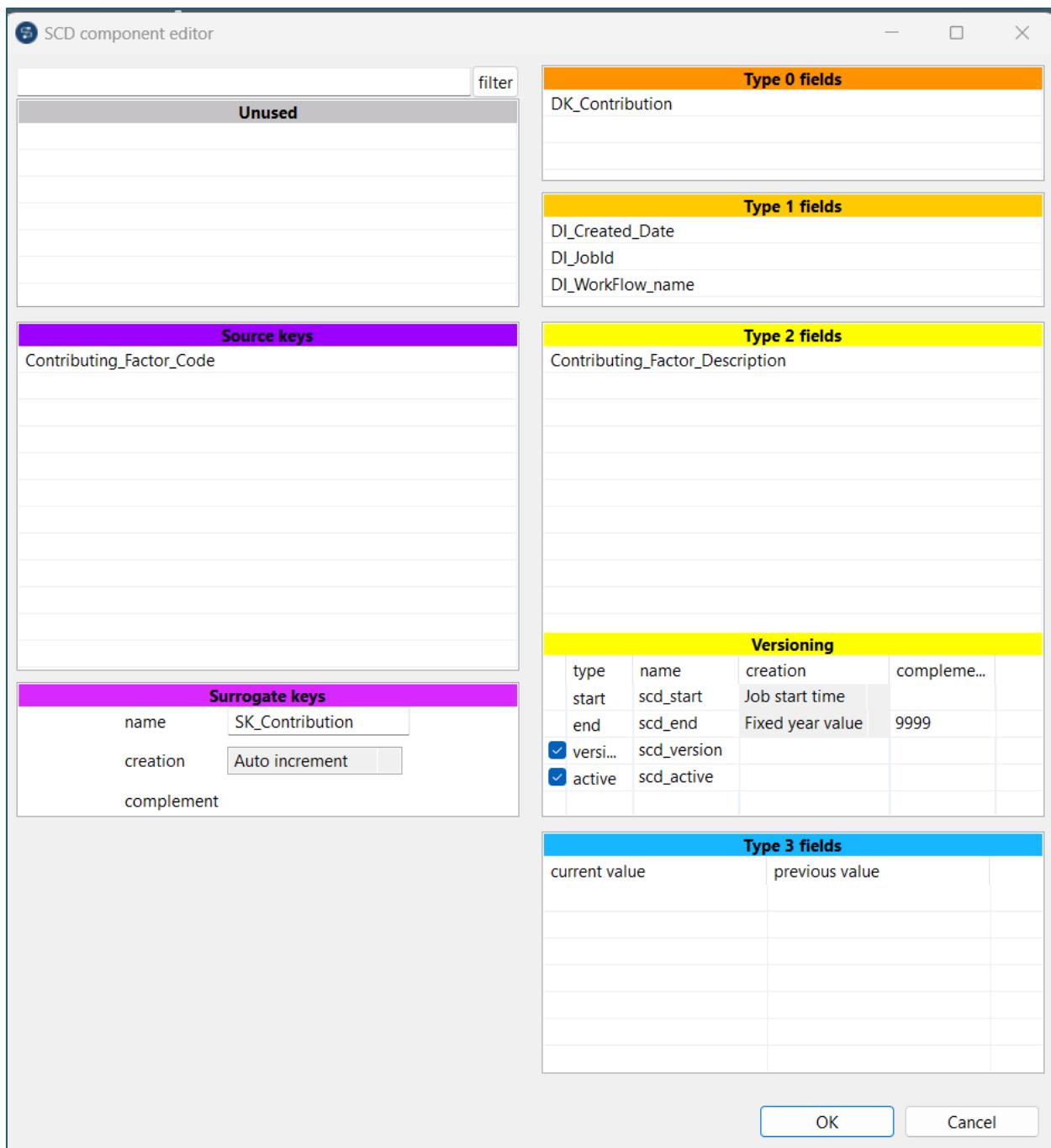
Results Messages

	(No column name)
1	760243

## DIM CONTRIBUTION WORKFLOW

The 3 staged final normalized tables for Chicago,Austin and New York cities are taken as input , united and then their unique Contribution Codes along with their Description are extracted and staged in DIM\_Contribution





```

CREATE TABLE Dim_Contribution (
    SK_Contribution int IDENTITY(1,1) NOT NULL, -- Auto-increment in MSSQL
    DK_Contribution int NOT NULL,
    Contributing_Factor_code int NULL,
    Contributing_Factor_Description varchar(400) NULL,
    scd_start datetime NOT NULL,
    scd_end datetime NULL,
    scd_version int NOT NULL,
    scd_active bit NOT NULL, -- bit is commonly used for boolean values in MSSQL
    DI_JobId varchar(20) NULL,
    DI_WorkFlow_Name varchar(80) NULL,
    DI_Created_Date date NOT NULL -- GETDATE() for current date in MSSQL
)

--DIM CONTRIBUTION
SELECT * FROM Dim_Contribution
SELECT COUNT(*) FROM Dim_Contribution
  
```

The screenshot shows a SQL query window in SSMS. The code creates a table named 'Dim\_Contribution' with various columns including auto-incrementing primary key 'SK\_Contribution', 'DK\_Contribution', 'Contributing\_Factor\_code', 'Contributing\_Factor\_Description', timestamp columns 'scd\_start' and 'scd\_end', integer 'scd\_version', bit 'scd\_active', and date 'DI\_Created\_Date'. It then performs a self-select to show all rows and a count of rows.

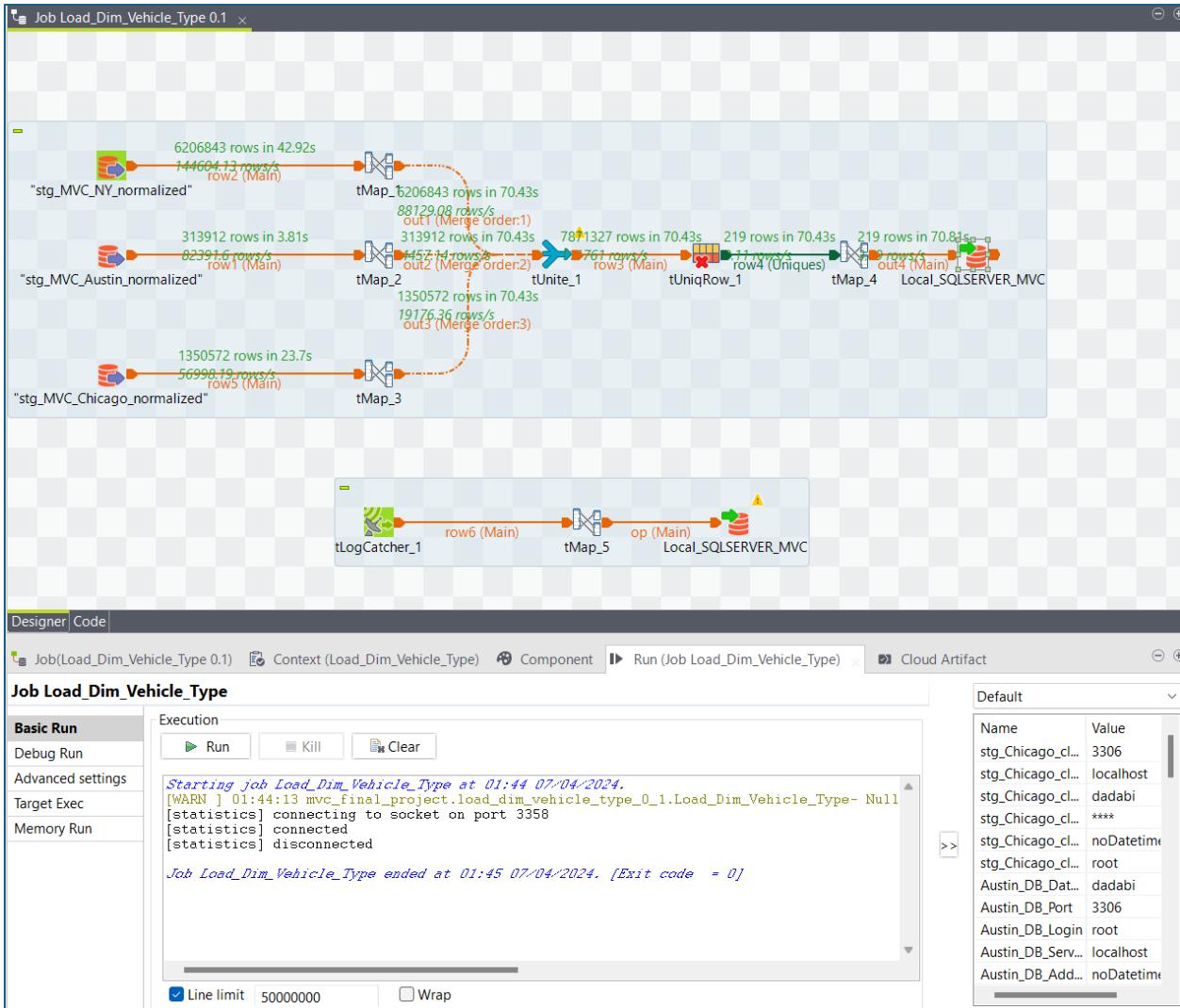
	SK_Contribution	DK_Contribution	Contributing_Factor_code	Contributing_Factor_Description	scd_start	scd_end	scd_version	scd_active	DI_JobId	DI_WorkFlow_Name	DI_Created_Date
1	1	-1	-1	NA	2024-04-07 15:04:05.287	9999-01-01 12:00:00.000	1	1	OBaSFz	Load_Dim_Contribution	2024-04-07
2	2	74	74	CELL/MOBILE DEVICE USE - TALKING	2024-04-07 15:04:05.287	9999-01-01 12:00:00.000	1	1	OBaSFz	Load_Dim_Contribution	2024-04-07
3	3	44	44	FOLLOWED TOO CLOSELY	2024-04-07 15:04:05.287	9999-01-01 12:00:00.000	1	1	OBaSFz	Load_Dim_Contribution	2024-04-07
4	4	20	20	DRIVER INATTENTION	2024-04-07 15:04:05.287	9999-01-01 12:00:00.000	1	1	OBaSFz	Load_Dim_Contribution	2024-04-07
5	5	41	41	FAULTY EVASIVE ACTION	2024-04-07 15:04:05.287	9999-01-01 12:00:00.000	1	1	OBaSFz	Load_Dim_Contribution	2024-04-07
6	6	60	60	UNSAFE SPEED	2024-04-07 15:04:05.287	9999-01-01 12:00:00.000	1	1	OBaSFz	Load_Dim_Contribution	2024-04-07
7	7	67	67	INTOXICATED - ALCOHOL	2024-04-07 15:04:05.287	9999-01-01 12:00:00.000	1	1	OBaSFz	Load_Dim_Contribution	2024-04-07
8	8	22	22	FAILED TO CONTROL SPEED	2024-04-07 15:04:05.287	9999-01-01 12:00:00.000	1	1	OBaSFz	Load_Dim_Contribution	2024-04-07
9	9	68	68	INTOXICATED - DRUG	2024-04-07 15:04:05.287	9999-01-01 12:00:00.000	1	1	OBaSFz	Load_Dim_Contribution	2024-04-07
10	10	45	45	HAD BEEN DRINKING	2024-04-07 15:04:05.287	9999-01-01 12:00:00.000	1	1	OBaSFz	Load_Dim_Contribution	2024-04-07
11	11	46	46	HANDICAPPED DRIVER (EXPLAIN IN NARRATIVE)	2024-04-07 15:04:05.287	9999-01-01 12:00:00.000	1	1	OBaSFz	Load_Dim_Contribution	2024-04-07
12	12	64	64	TURNED IMPROPERLY - WIDE RIGHT	2024-04-07 15:04:05.287	9999-01-01 12:00:00.000	1	1	OBaSFz	Load_Dim_Contribution	2024-04-07
13	13	23	23	FAILED TO DRIVE IN SINGLE LANE	2024-04-07 15:04:05.287	9999-01-01 12:00:00.000	1	1	OBaSFz	Load_Dim_Contribution	2024-04-07
14	14	48	48	IMPAIRED VISIBILITY (EXPLAIN IN NARRATIVE)	2024-04-07 15:04:05.287	9999-01-01 12:00:00.000	1	1	OBaSFz	Load_Dim_Contribution	2024-04-07
15	15	19	19	DISTRACTION IN VEHICLE	2024-04-07 15:04:05.287	9999-01-01 12:00:00.000	1	1	OBaSFz	Load_Dim_Contribution	2024-04-07
				FAILED TO ADHERE TO SIGNALS	2024-04-07 15:04:05.287	9999-01-01 12:00:00.000	4	1	OoOoO	Load_Dim_Contribution	2024-04-07

(No column name)  
1 85

## DIM VEHICLE TYPE WORKFLOW

The 3 staged final normalized tables for Chicago,Austin and New York cities are taken as input , united and then their unique Vehicle Codes along with their Description are extracted and staged in DIM\_Vehicle\_Type

The mapping document for Vehicle\_Type was created based on our understanding of the combined Vehicle\_Type data.



```
--DIM VEHICLE TYPE
SELECT * FROM DIM_Vehicle_Type
SELECT COUNT(*) FROM DIM_Vehicle_Type

--DIM SOURCE
SELECT * FROM DIM_Source
SELECT COUNT(*) FROM DIM_Source

--FACT ACCIDENTS
SELECT * FROM FACT_ACCIDENTS
SELECT COUNT(*) FROM FACT_ACCIDENTS
```

110 % ▶

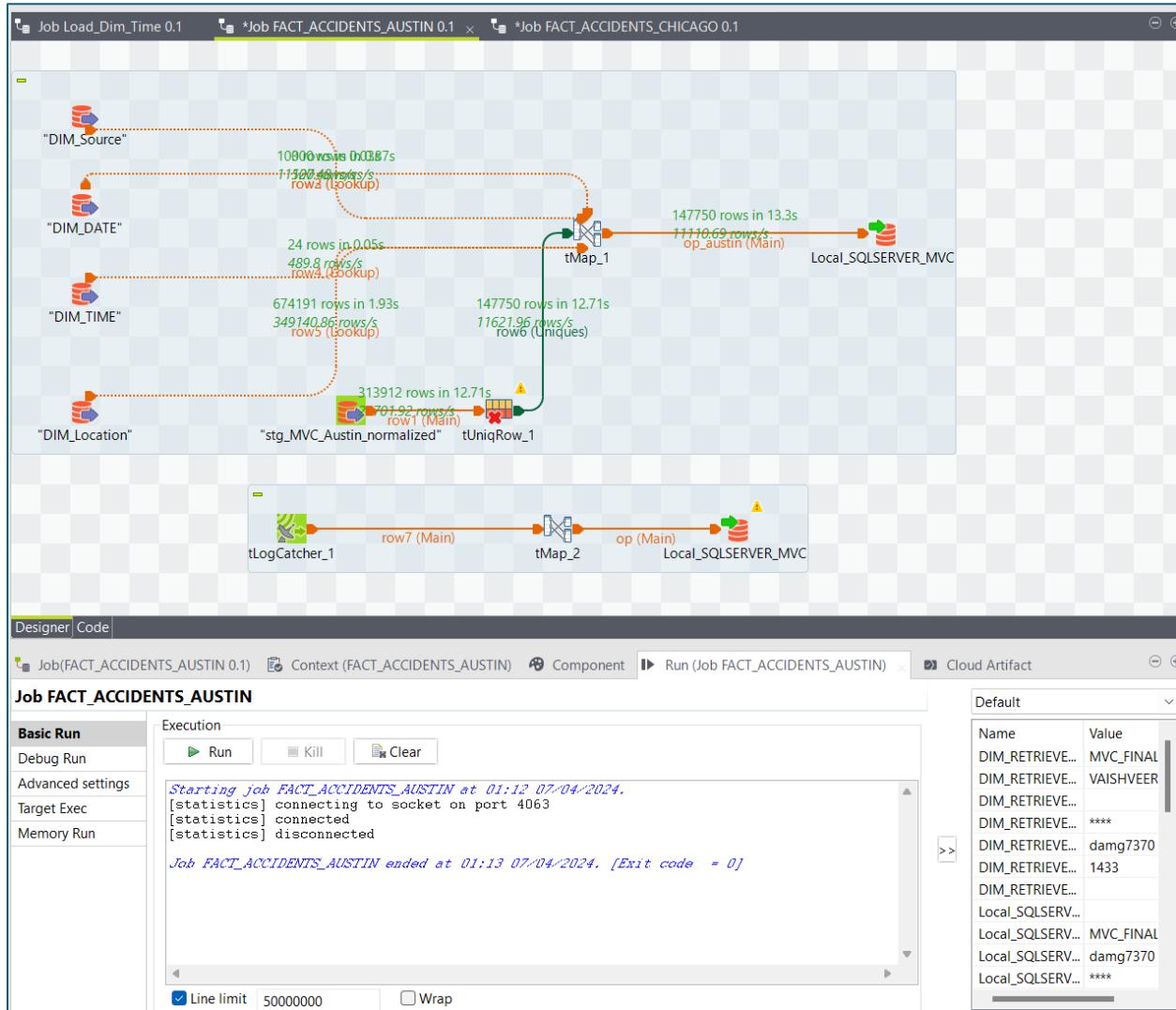
Results Messages

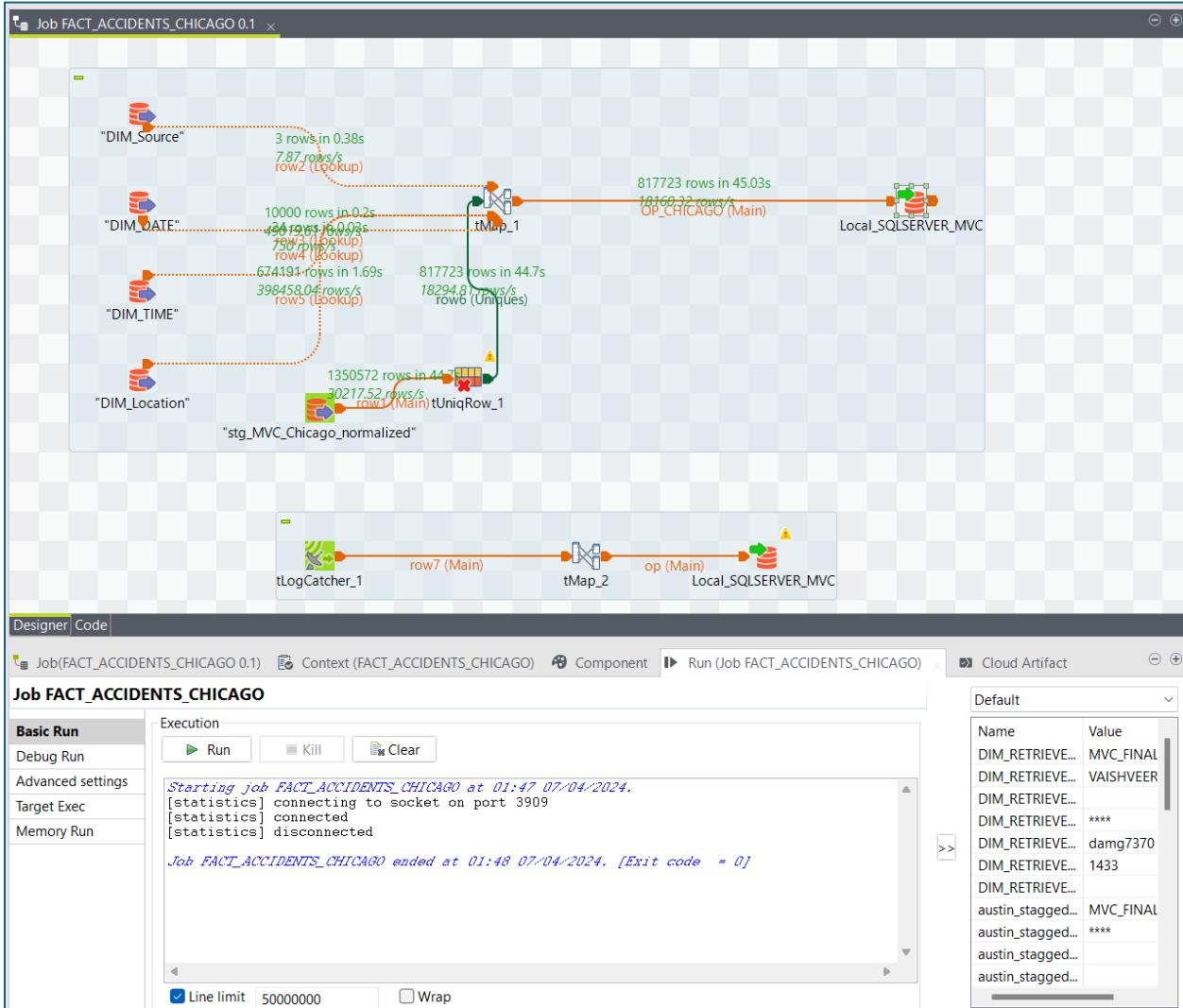
	SK_Vehicle_Type	VEHICLE_CODE	Vehicle_Type	DI_CreateDate	DI_JobId	DI_WorkflowName
1	1	4	4 DOOR	2024-04-07 01:44:14.013	gAVYVn	Load_Dim_Vehicle_Type
2	2	184	STATION WAGON	2024-04-07 01:44:14.017	gAVYVn	Load_Dim_Vehicle_Type
3	3	168	SEDAN	2024-04-07 01:44:14.017	gAVYVn	Load_Dim_Vehicle_Type
4	4	1	2 DOOR	2024-04-07 01:44:14.017	gAVYVn	Load_Dim_Vehicle_Type
5	5	188	TAXI	2024-04-07 01:44:14.017	gAVYVn	Load_Dim_Vehicle_Type
6	6	23	BIKE	2024-04-07 01:44:14.020	gAVYVn	Load_Dim_Vehicle_Type
7	7	147	PICK UP TRUCK	2024-04-07 01:44:14.023	gAVYVn	Load_Dim_Vehicle_Type
8	8	219	NA	2024-04-07 01:44:14.030	gAVYVn	Load_Dim_Vehicle_Type
9	9	205	VAN	2024-04-07 01:44:14.033	gAVYVn	Load_Dim_Vehicle_Type
10	10	33	BUS	2024-04-07 01:44:14.040	gAVYVn	Load_Dim_Vehicle_Type
11	11	142	OTHER	2024-04-07 01:44:14.053	gAVYVn	Load_Dim_Vehicle_Type
12	12	29	BOX TRUCK	2024-04-07 01:44:14.057	gAVYVn	Load_Dim_Vehicle_Type
13	13	192	TOW TRUCK	2024-04-07 01:44:14.060	gAVYVn	Load_Dim_Vehicle_Type
14	14	187	TANK TRUCK	2024-04-07 01:44:14.063	gAVYVn	Load_Dim_Vehicle_Type
15	15	57	DUMP TRUCK	2024-04-07 01:44:14.070	gAVYVn	Load_Dim_Vehicle_Type
16	16	100	MOTOR VEHICLE	2024-04-07 01:44:14.073	gAVYVn	Load_Dim_Vehicle_Type
	(No column name)					
1	219					

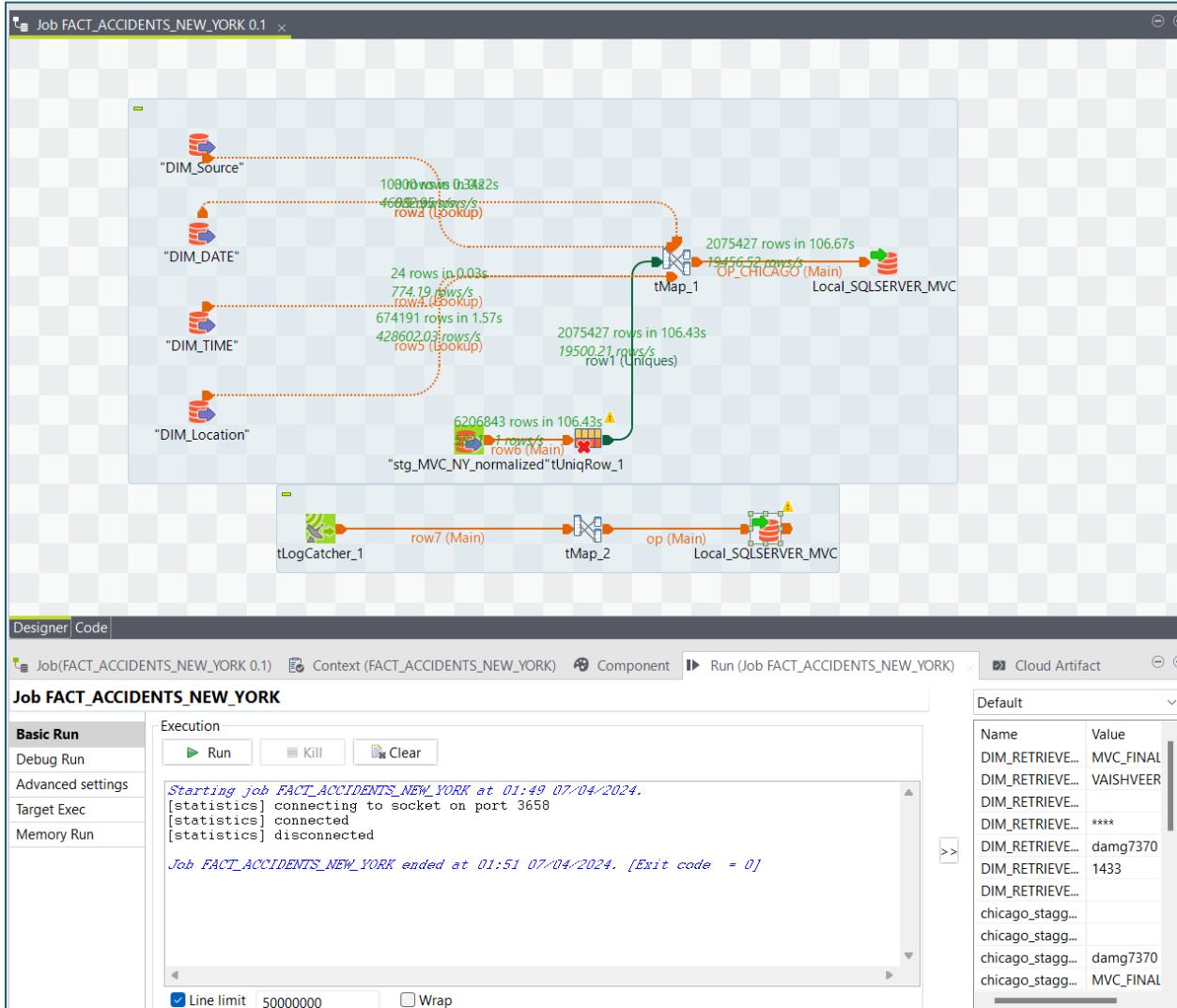
## Once the dimensions were created the next step was to create the FACTS

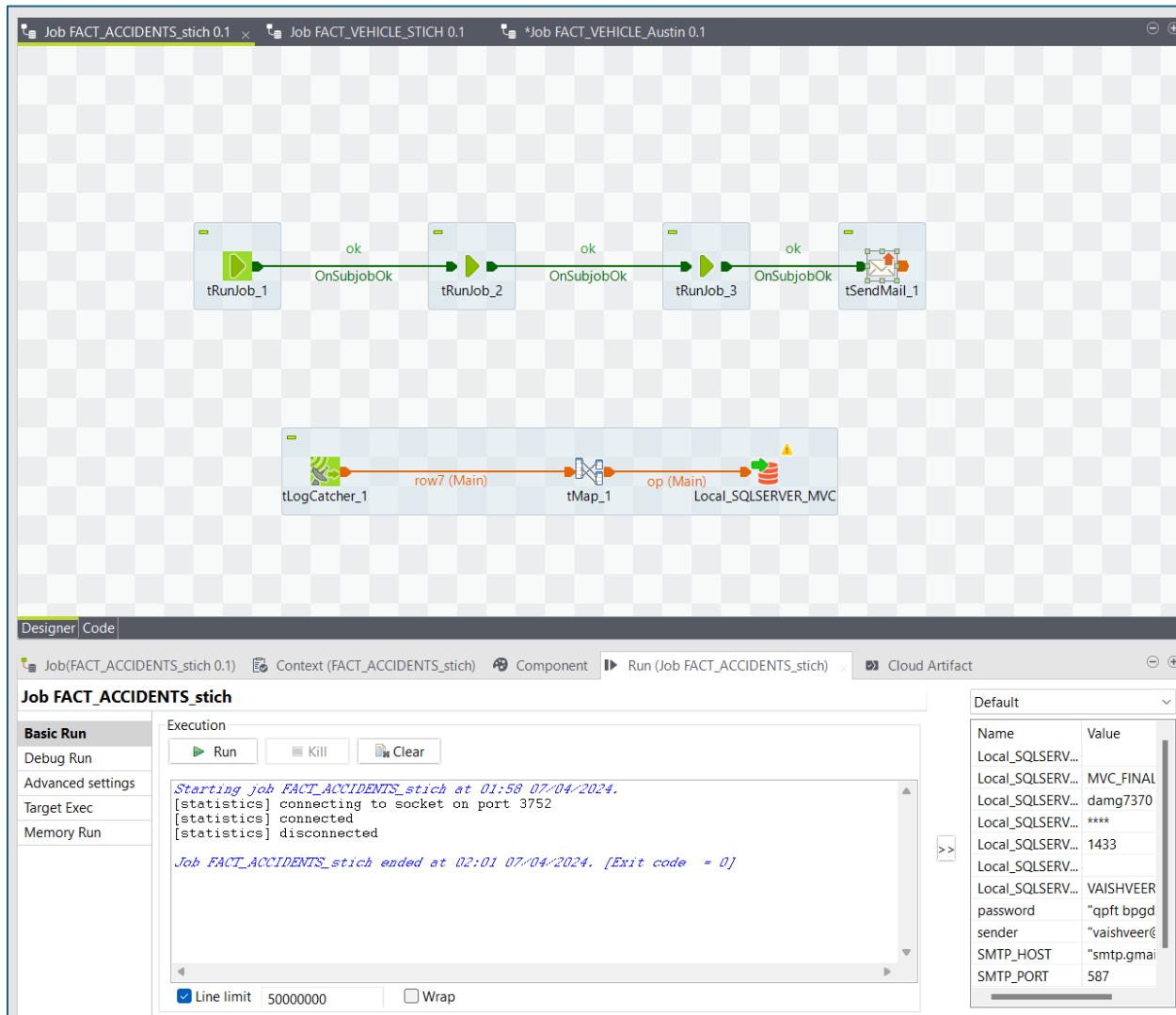
**For FACT\_ACCIDENTS** 3 individual workflows were created for each city and then all 3 jobs were connected in an orchestration job where the FACT ACCIDENTS was loaded successfully, and an email notification was sent out upon completion.

### FACT\_ACCIDENTS- AUSTIN INDIVIDUAL



**FACT\_ACCIDENTS- CHICAGO INDIVIDUAL**

FACT\_ACCIDENTS- NEW YORK INDIVIDUAL

**FACT\_ACCIDENTS- FINAL STICH JOB**

The screenshot shows the configuration dialog for the tSendMail\_1 component. The "Basic settings" tab is selected.

**Basic settings**

- To: "vaishveer@gmail.com,anusreemohanani@gmail.com"
- From: context.sender
- Show sender's name? (checkbox)
- Cc:
- Bcc:
- Subject: "Talend Open Studio Notification"
- Message:
 

"Hello,

FACT\_ACCIDENTS HAS BEEN LOADED SUCCESSFULLY!"
- Die if the attachment file doesn't exist. (checkbox)
- Attachments: File
- Content Transfer Encoding:

## Talend Open Studio Notification ➔ Inbox ✎

vaishveer@gmail.com

to me, anusreemohan11 ✎

2:01AM (1 minute ago)

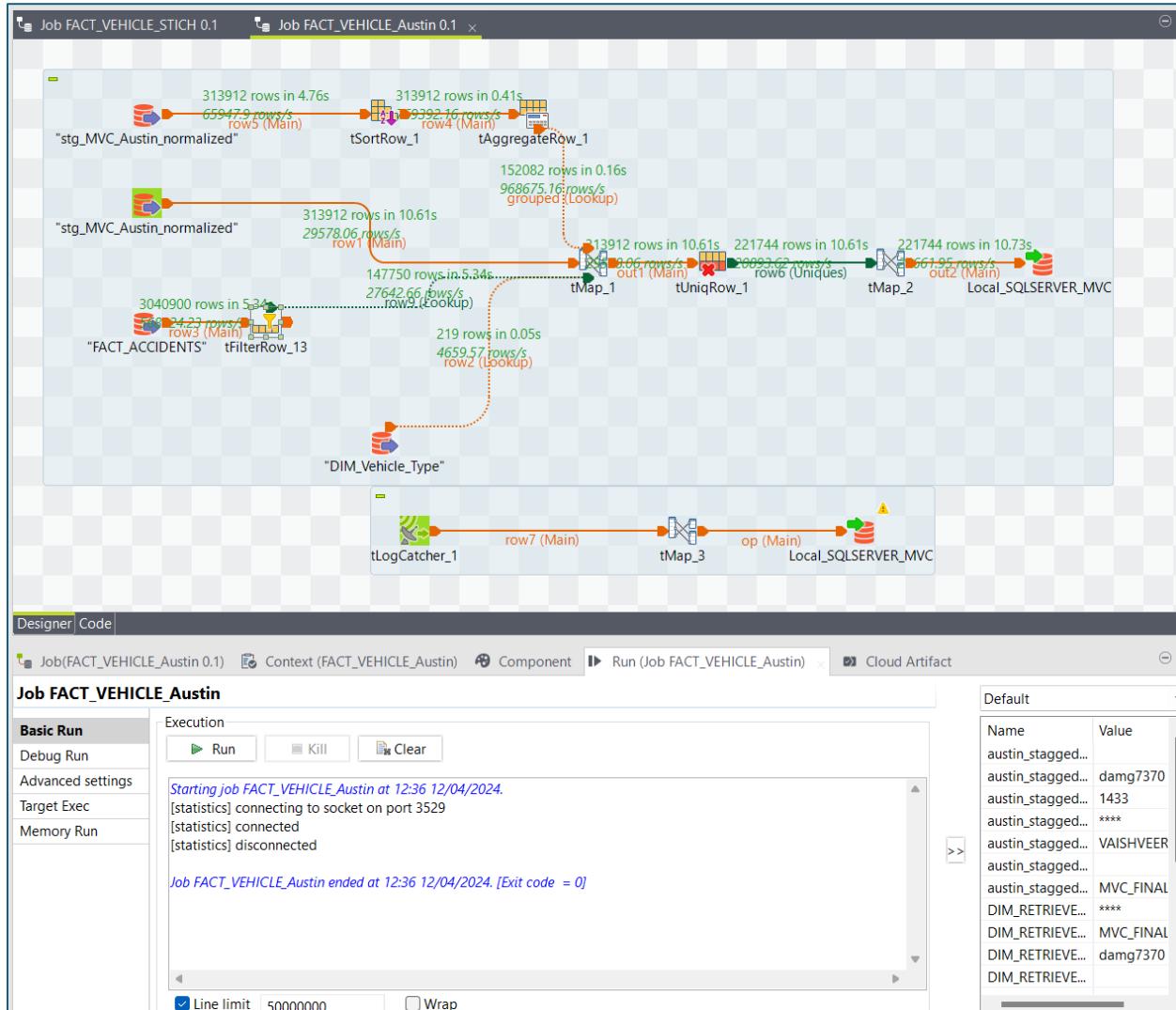


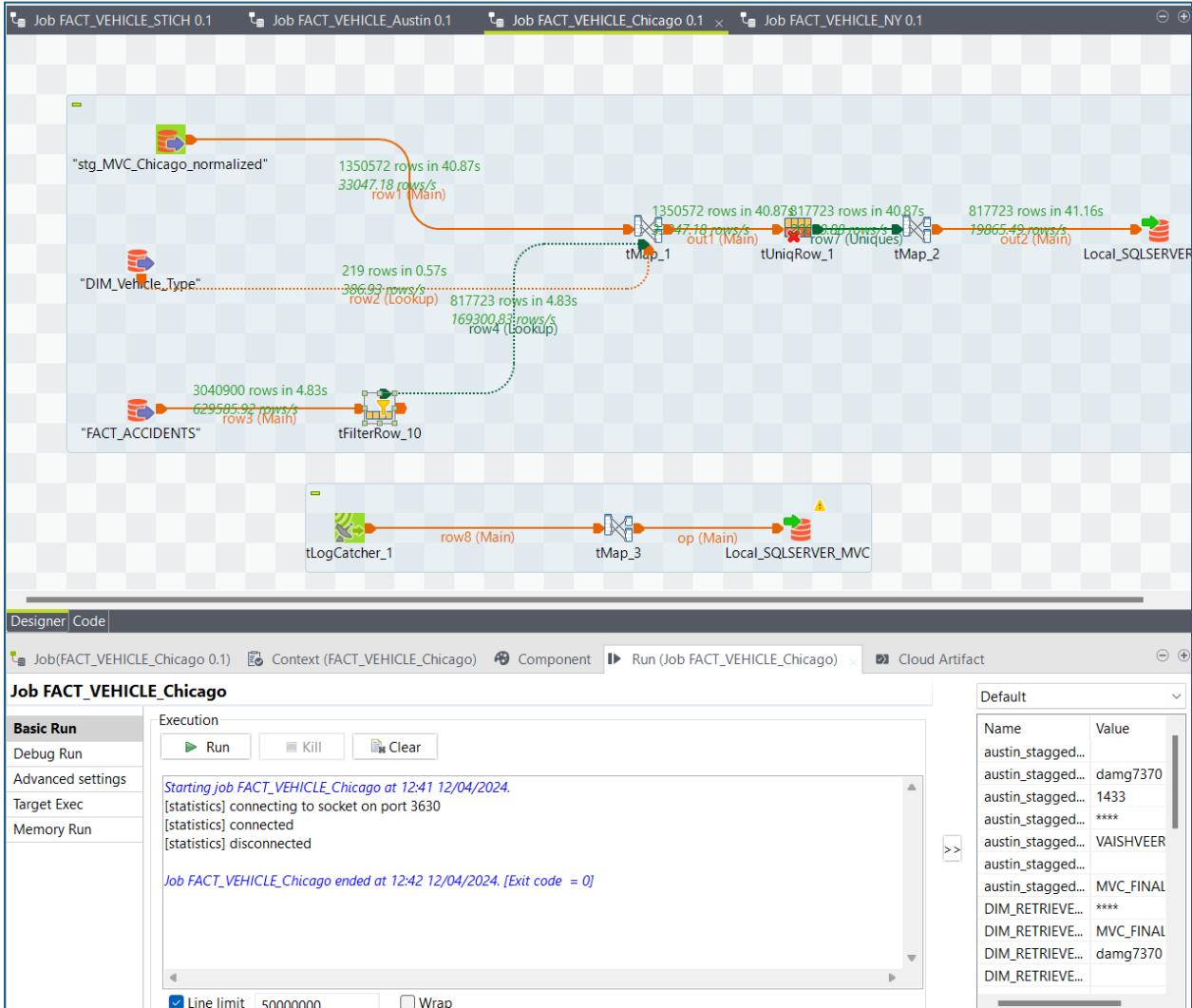
Hello,

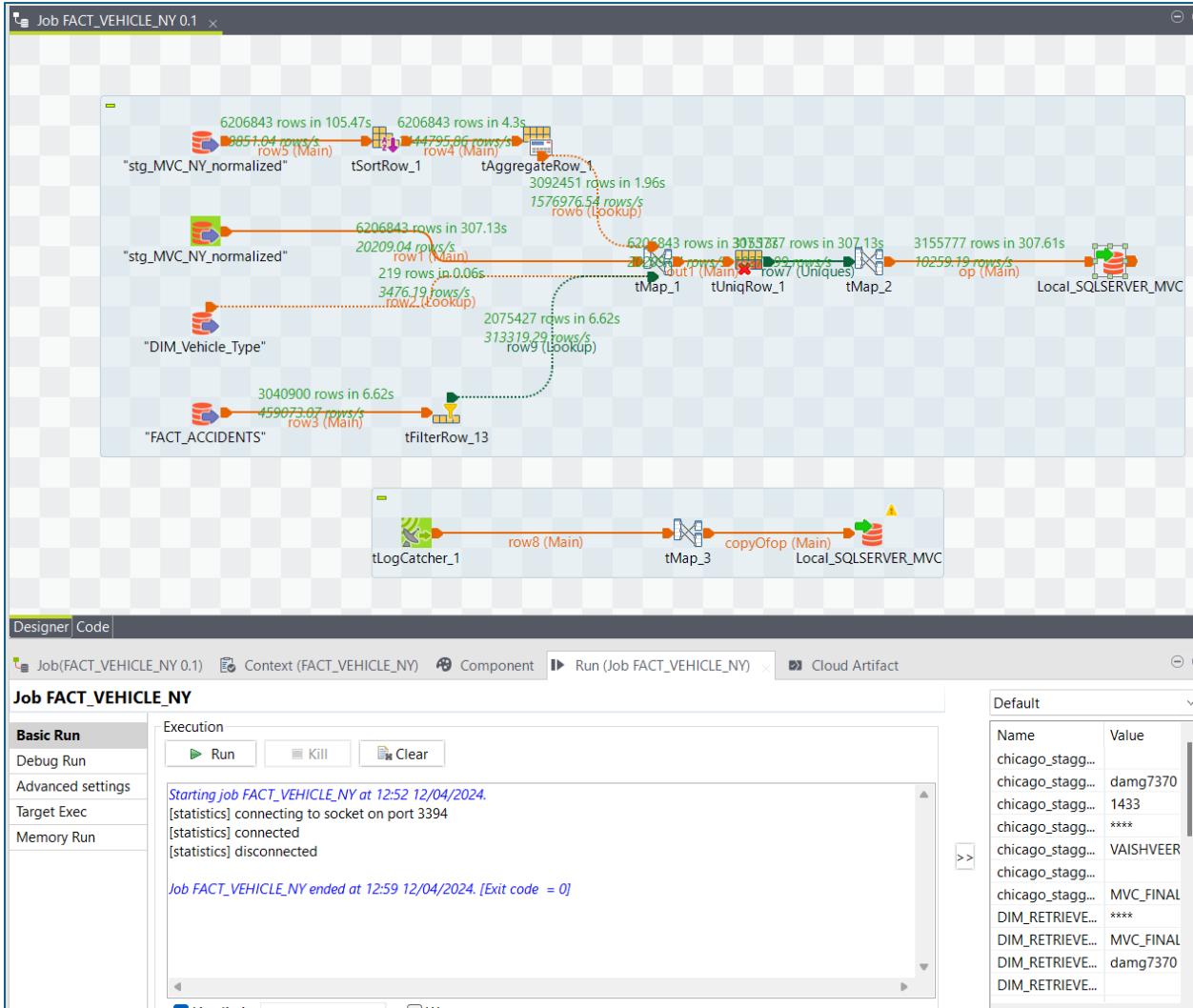
FACT\_ACCIDENTS HAS BEEN LOADED SUCCESSFULLY!

**For FACT\_VEHICLE** 3 individual workflows were created for each city and then all 3 jobs were connected in an orchestration job where the FACT VEHICLE was loaded successfully, and an email notification was sent out upon completion.

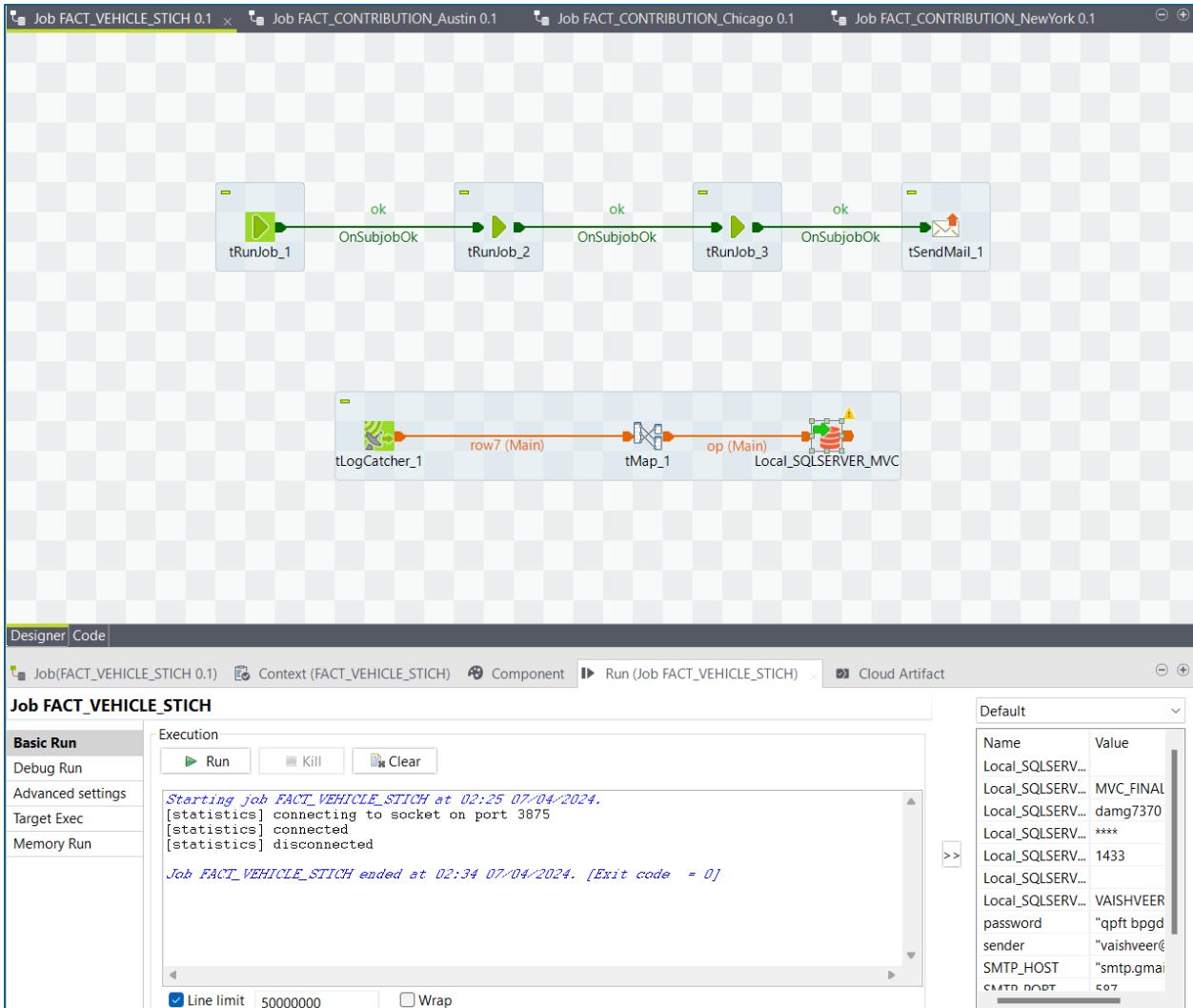
### FACT\_VEHICLE- AUSTIN INDIVIDUAL



FACT VEHICLE- CHICAGO INDIVIDUALFACT VEHICLE- NEW YORK INDIVIDUAL



## FACT VEHICLE- FINAL STICH JOB



This screenshot shows the configuration dialog for the "tSendMail\_1" component. The left sidebar lists "Basic settings", "Advanced settings", "Dynamic settings", "View", and "Documentation". The "Basic settings" section contains fields for "To" (value: "vaishveer@gmail.com,anusreemohan11@gmail.com,banerjee.c@northeastern.edu"), "From" (value: "context.sender"), "Subject" (value: "Talend Open Studio Notification"), and "Message" (value: "Hello,  
FACT\_VEHICLE HAS BEEN LOADED SUCCESSFULLY!"). There is also a checkbox for "Die if the attachment file doesn't exist." and tabs for "Attachments" (File) and "Content Transfer Encoding".

vaishveer@gmail.com  
to banerjee.c, me, anusreemohanani11

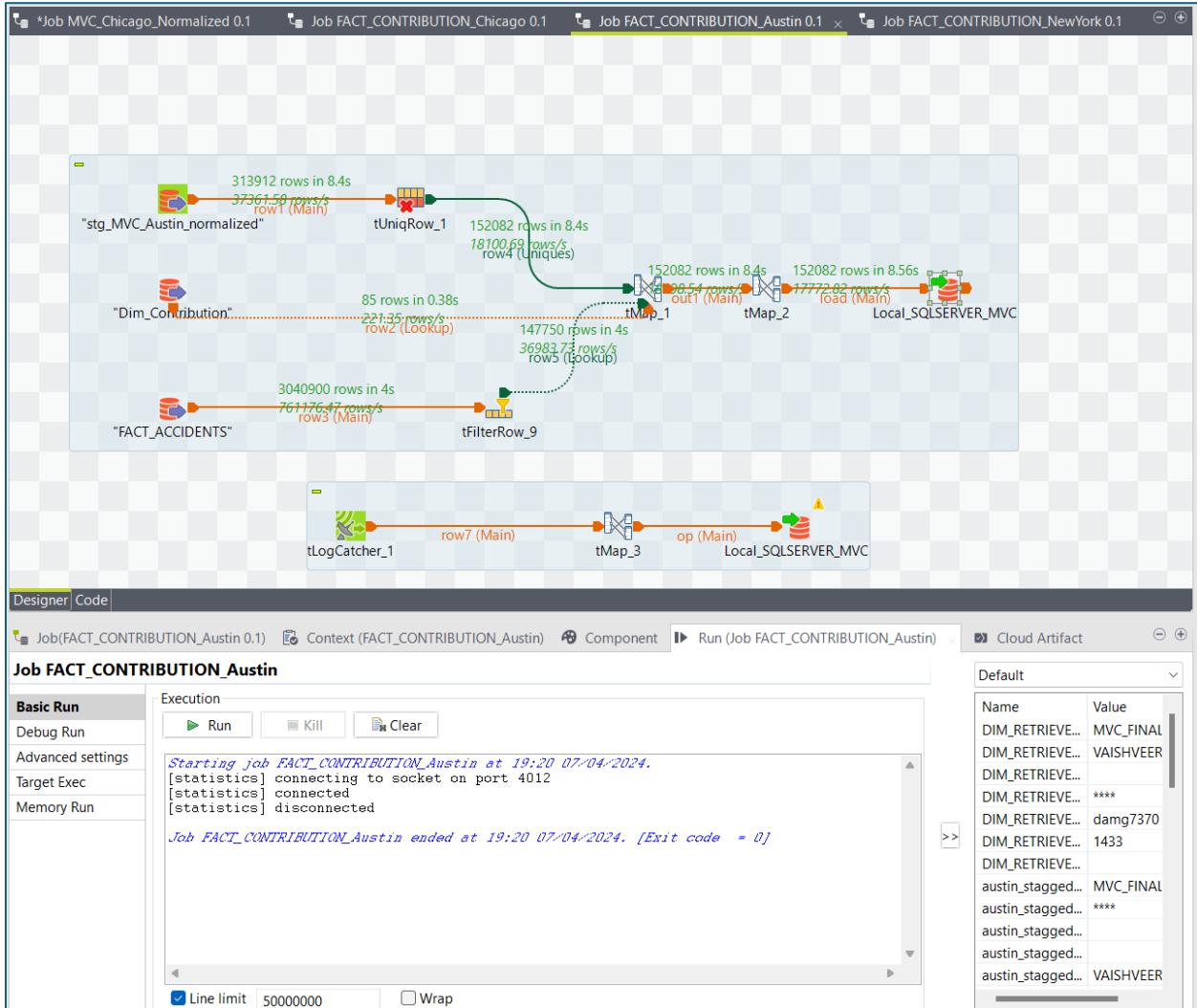
2:34 AM (35 minutes ago)

Hello,

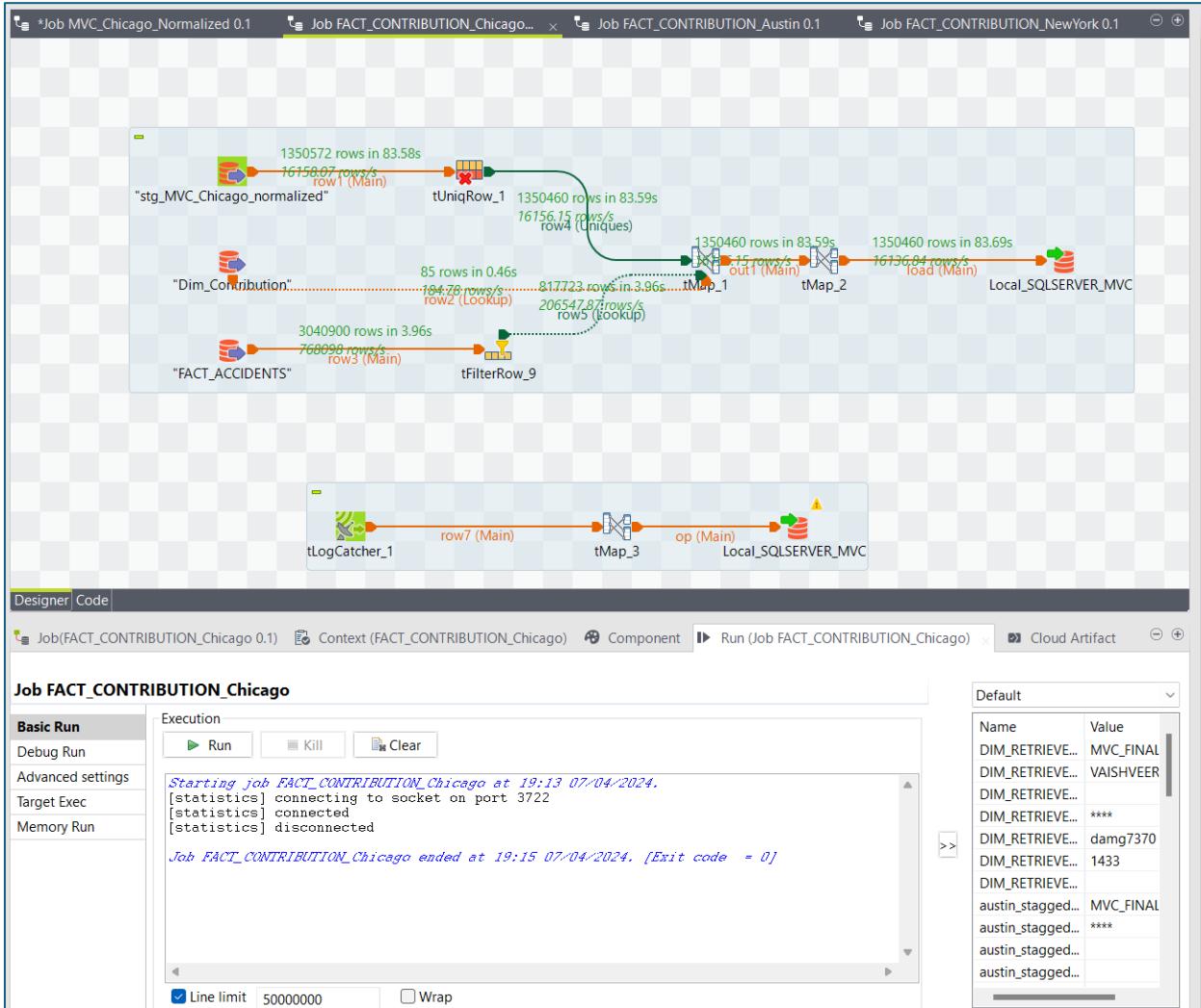
FACT\_VEHICLE HAS BEEN LOADED SUCCESSFULLY!

**For FACT\_CONTRIBUTION** 3 individual workflows were created for each city and then all 3 jobs were connected in an orchestration job where the FACT CONTRIBUTION was loaded successfully, and an email notification was sent out upon completion.

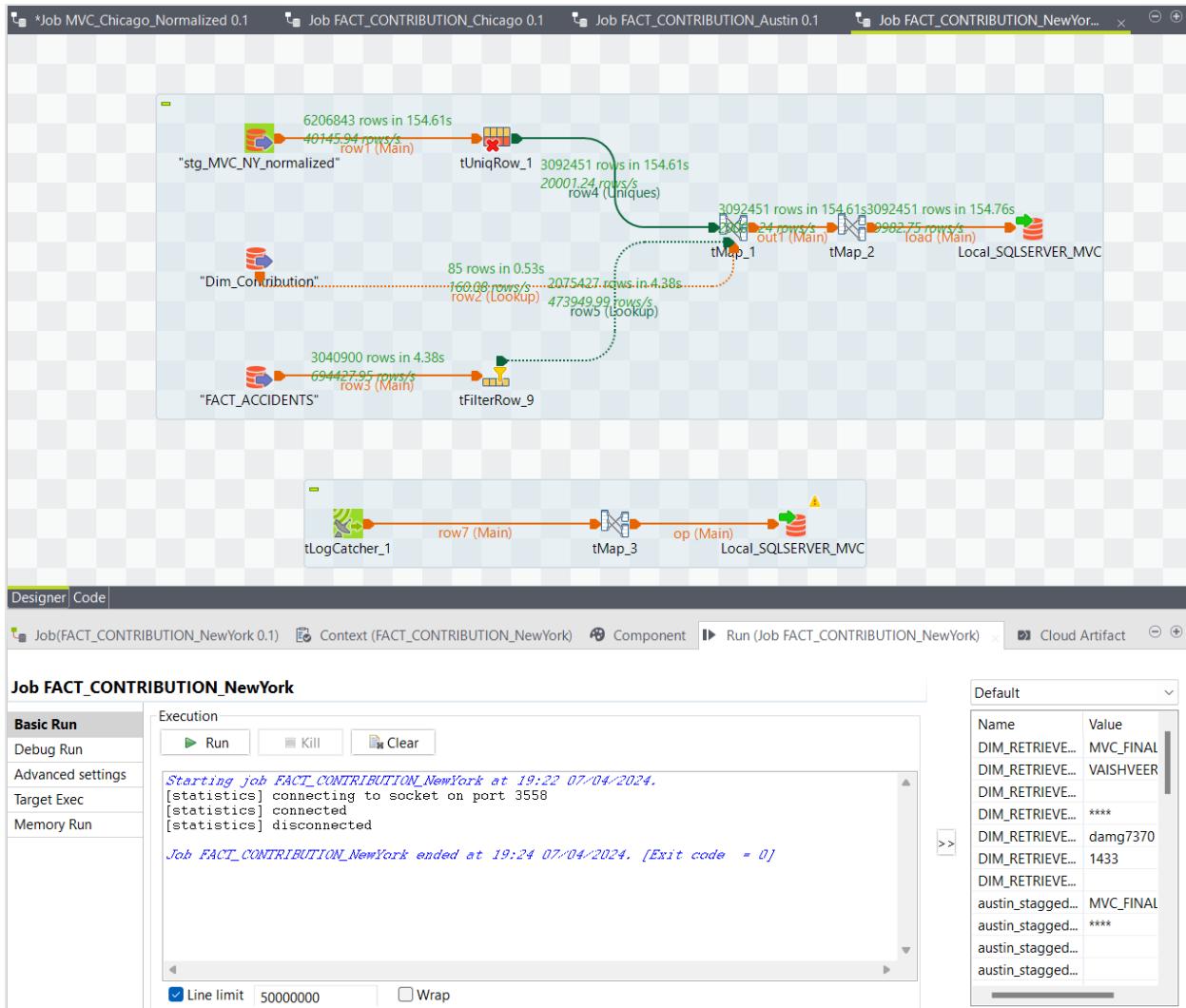
### **FACT CONTRIBUTION- AUSTIN INDIVIDUAL**



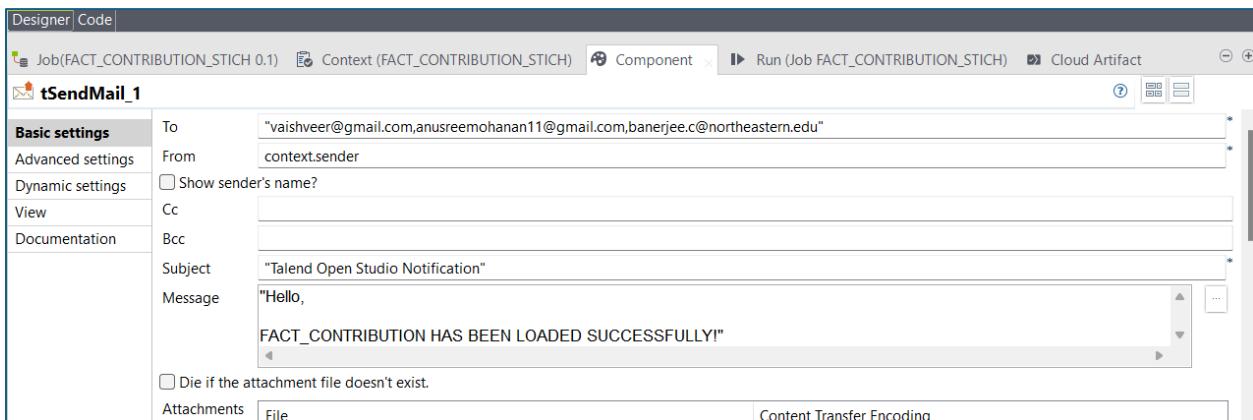
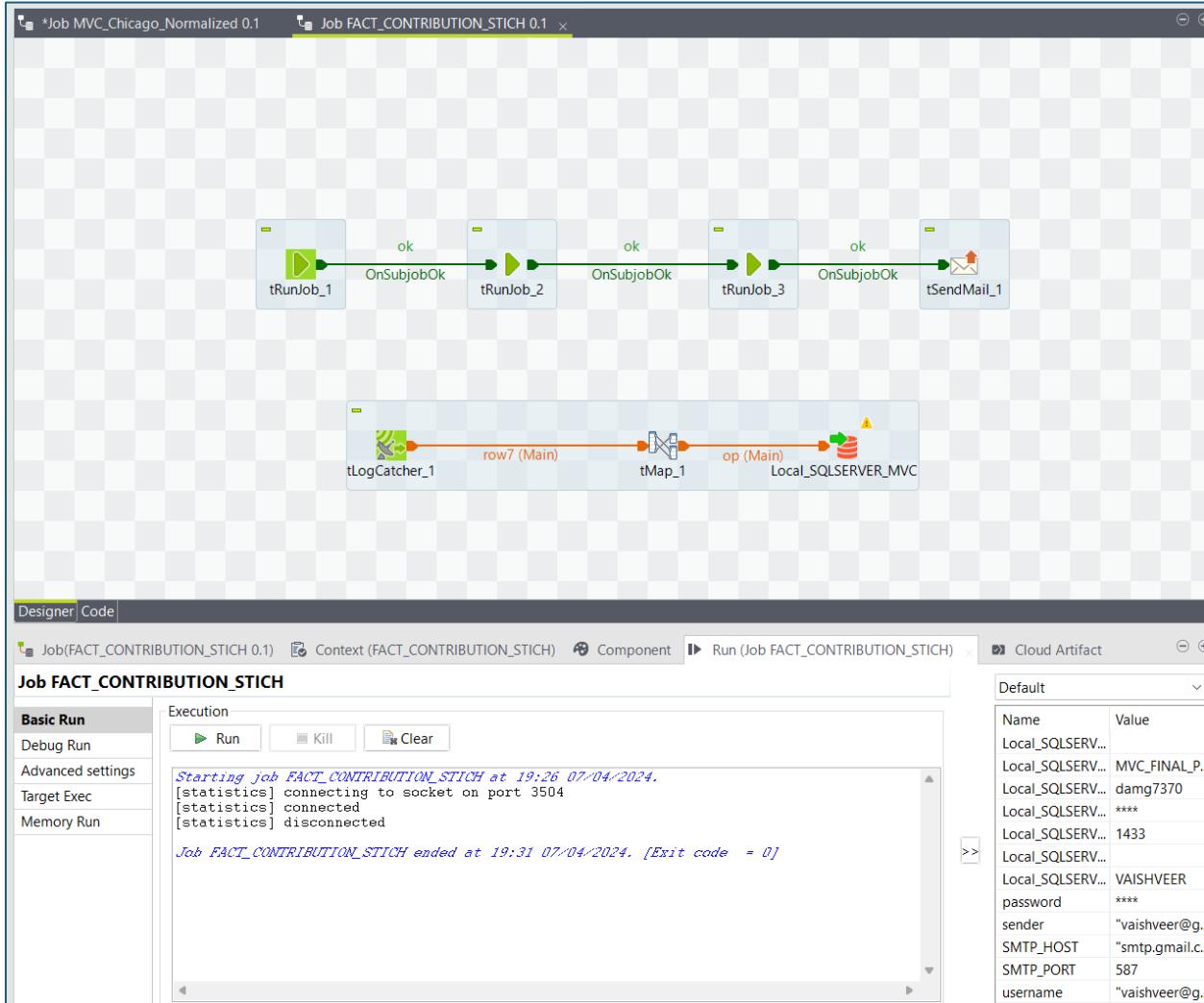
## FACT CONTRIBUTION- CHICAGO INDIVIDUAL



## FACT CONTRIBUTION- NEW YORK INDIVIDUAL



## FACT CONTRIBUTION- FINAL STICH JOB



vaishveer@gmail.com  
to me, anusreemohanani11, banerjee.c

2:53 AM (16 minutes ago) ⚡ 😊 ← ⏮

Hello,

FACT\_CONTRIBUTION HAS BEEN LOADED SUCCESSFULLY!

## VALIDATION OF THE DIMENSIONAL MODEL USING SQL QUERIES

### AUSTIN DATASET

The screenshot shows a SQL Server Management Studio window with several queries run against a database named 'SQL QUERIES'. The queries validate the dimensional model by joining facts and dimensions.

```
-- TEST QUERIES
--AUSTIN

SELECT * FROM stg_MVC_Austin_normalized where crash_id=13790694
SELECT * FROM FACT_ACCIDENTS where Crash_ID=13790694

SELECT * FROM DIM_Source WHERE SK_Source=2
SELECT * FROM DIM_Location WHERE SK_Location=65
SELECT * FROM DIM_DATE WHERE Date_SK=20140330
SELECT * FROM DIM_TIME WHERE Time_SK=3

SELECT * FROM stg_MVC_Austin_normalized where crash_id=13790694
SELECT * FROM FACT_CONTRIBUTION WHERE SK_FACT_ACCIDENTS=2893215
SELECT * FROM Dim_Contribution WHERE SK_Contribution IN (7,9)

SELECT * FROM FACT_VEHICLE WHERE SK_FACT_ACCIDENTS=2893215

SELECT * FROM DIM_Vehicle_Type WHERE SK_Vehicle_Type IN(16,126,34)
SELECT * FROM stg_MVC_Austin_normalized where crash_id=13790694
```

The results pane displays the output of the last query, which lists 15 rows of accident data from the 'stg\_MVC\_Austin\_normalized' table. The columns include crash\_id, Source, crash\_fatal\_flg, crash\_date, crash\_date\_only, crash\_time, crash\_hour, case\_id, rpt\_latitude, rpt\_longitude, rpt\_block\_num, rpt\_street\_pfx, rpt\_street\_name, rpt\_street\_sfx, crash\_speed\_limit, road\_co, SK\_FACT\_ACCIDENTS, CRASH\_ID, SK\_Source, SK\_Location, SK\_Date, SK\_Time, INJURIES\_TOTAL, IS\_PEDESTRIAN, MOTORIST\_INJURY\_COUNT, MOTORIST\_KILLED\_COUNT, PEDESTRIAN\_INJURY\_COUNT, and PEDESTRIAN\_KILLED\_COUNT.

crash_id	Source	crash_fatal_flg	crash_date	crash_date_only	crash_time	crash_hour	case_id	rpt_latitude	rpt_longitude	rpt_block_num	rpt_street_pfx	rpt_street_name	rpt_street_sfx	crash_speed_limit	road_co	
1	13790694	AUSTIN	N	2014-03-30 02:45:00.000	2014-03-30 00:00:00.000	02:45:00	2	140890422	-1	-1	1088	NA	PARK PLZ	NA	-1	N
2	13790694	AUSTIN	N	2014-03-30 02:45:00.000	2014-03-30 00:00:00.000	02:45:00	2	140890422	-1	-1	1088	NA	PARK PLZ	NA	-1	N
3	13790694	AUSTIN	N	2014-03-30 02:45:00.000	2014-03-30 00:00:00.000	02:45:00	2	140890422	-1	-1	1088	NA	PARK PLZ	NA	-1	N
4	13790694	AUSTIN	N	2014-03-30 02:45:00.000	2014-03-30 00:00:00.000	02:45:00	2	140890422	-1	-1	1088	NA	PARK PLZ	NA	-1	N
5	13790694	AUSTIN	N	2014-03-30 02:45:00.000	2014-03-30 00:00:00.000	02:45:00	2	140890422	-1	-1	1088	NA	PARK PLZ	NA	-1	N
6	13790694	AUSTIN	N	2014-03-30 02:45:00.000	2014-03-30 00:00:00.000	02:45:00	2	140890422	-1	-1	1088	NA	PARK PLZ	NA	-1	N
7	13790694	AUSTIN	N	2014-03-30 02:45:00.000	2014-03-30 00:00:00.000	02:45:00	2	140890422	-1	-1	1088	NA	PARK PLZ	NA	-1	N
8	13790694	AUSTIN	N	2014-03-30 02:45:00.000	2014-03-30 00:00:00.000	02:45:00	2	140890422	-1	-1	1088	NA	PARK PLZ	NA	-1	N

Query executed successfully.

## NEW YORK DATASET

SQL QUERIES.sql ...ISHVEER\vaish (57) X SQLQuery1.sql - VA...SHVEER\vaish (54)\*

```

SELECT * FROM stg_MVC_Chicago_normalized WHERE crash_id=100

-- NEW YORK

SELECT * FROM stg_MVC_NY_normalized WHERE COLLISION_ID=4455765
SELECT * FROM FACT_ACCIDENTS WHERE CRASH_ID=4455765

SELECT * FROM DIM_Source WHERE SK_Source=3
SELECT * FROM DIM_Location WHERE SK_Location=398545
SELECT * FROM DIM_DATE WHERE Date_SK=20210911
SELECT * FROM DIM_TIME WHERE Time_SK=3
SELECT * FROM DIM_CONTRIBUTION WHERE SK_Contribution IN(32,1)

SELECT * FROM stg_MVC_NY_normalized WHERE COLLISION_ID=4455765
SELECT * FROM FACT_CONTRIBUTION WHERE SK_FACT_ACCIDENTS=929992

SELECT * FROM FACT_VEHICLE WHERE SK_FACT_ACCIDENTS=929992
SELECT * FROM DIM_Vehicle_Type WHERE SK_Vehicle_Type IN(3)
SELECT * FROM stg_MVC_NY_normalized WHERE COLLISION_ID=4455765

```

110 % ▾

Results Messages

	COLLISION_ID	CRASH_DATE	Source	CRASH_TIME	CRASH_HOUR	BOROUGH	ZIP_CODE	LATITUDE	LONGITUDE	LOCATION	ON_STREET_NAME	CROSS_STREET_NAME	OFF_STREET_NAME	NUMBER_OF_PERSONS
1	4455765	2021-09-11 00:00:00.000	NEW YORK	2:39:00	2	NA	-1	-1	-1	WHitestone Expressway	20 Avenue	NA	2	
2	4455765	2021-09-11 00:00:00.000	NEW YORK	2:39:00	2	NA	-1	-1	-1	Whitestone Expressway	20 Avenue	NA	2	
3	4455765	2021-09-11 00:00:00.000	NEW YORK	2:39:00	2	NA	-1	-1	-1	Whitestone Expressway	20 Avenue	NA	2	
4	4455765	2021-09-11 00:00:00.000	NEW YORK	2:39:00	2	NA	-1	-1	-1	Whitestone Expressway	20 Avenue	NA	2	

	SK_FACT_ACCIDENTS	CRASH_ID	SK_Source	SK_Location	SK_Date	SK_Time	INJURIES_TOTAL	IS_PEDESTRIAN	MOTORIST_INJURY_COUNT	MOTORIST_KILLED_COUNT	PEDESTRIAN_INJURY_COUNT	PEDESTRIAN_KILLED_COUNT
1	929992	4455765	3	398545	20210911	3	2	0	2	0	0	0

	SK_Source	Source	DL_CreateDate	DL_JobID	DL_WorkflowName
1	3	NEW YORK	2024-04-07 00:26:23.970	b73A	Load_Dim_Source

	SK_Location	Street_Name	Latitude	Longitude	DL_CreateDate	DL_JobID	DL_WorkflowFileName
1	398545	Whitestone Expressway	-1	-1	2024-04-07 15:01:14.820	s5roa9	Load_Dim_Location

	Date_SK	Date	Day	Month	Year	Season	DL_CreateDate	DL_JobID	DL_WorkflowName
1	20210911	2021-09-11 00:00:00.000	11	9	2021	Fall	2024-04-07 00:24:05.127	LNxL1	Load_Dim_Date

	Time_SK	Hour_24	Time_Period	DL_CreateDate	DL_JobID	DL_WorkflowName
1	3	2	Late Night	2024-04-07 09:48:28.677	kSYUwb	Load_Dim_Time

Query executed successfully.

VAISHVEER (16.0 RTM) | VAISHVEER\vaish (57) | MVC\_FINAL\_PROJECT | 00:00:00 | 9 rows

## CHICAGO DATASET

SQL VALIDATION QU...HVEER(vaish (57)) + X SQLQuery1.sql - VA...SHVEER(vaish (54))

```

SELECT * FROM stg_MVC_Chicago_normalized WHERE crash_id=100
SELECT * FROM FACT_ACCIDENTS WHERE Crash_ID=100 AND SK_Source=1

SELECT * FROM DIM_Source WHERE SK_Source=1
SELECT * FROM DIM_Location WHERE SK_Location=97124
SELECT * FROM DIM_DATE WHERE Date_SK=20170615
SELECT * FROM DIM_TIME WHERE Time_SK=13
SELECT * FROM DIM_CONTRIBUTION WHERE SK_Contribution IN(70)

SELECT * FROM stg_MVC_Chicago_normalized WHERE CONTRIBUTORY_CAUSE IS NULL
SELECT * FROM FACT_CONTRIBUTION WHERE SK_FACT_ACCIDENTS=2075428

SELECT COUNT(*) FROM FACT_CONTRIBUTION WHERE SK_Contribution=1

SELECT * FROM stg_MVC_Chicago WHERE CRASH_RECORD_ID='71248bae100b1360c1a2c419a317d5a87f8878cf14163acd879b5b315aeaba1c3d91ed98976bef0e029fb59acc827c3b8d07ec8575937df4fcf7fee6'
SELECT * FROM FACT_CONTRIBUTION WHERE SK_FACT_ACCIDENTS=2075428

```

110 % 4 Messages

CRASH_ID	SOURCE	CRASH_RECORD_ID	CRASH_DATE_EST_J	CRASH_DATE	CRASH_DATE_ONLY	CRASH_TIME	POSTED_SPEED_LIMIT	TRAFFIC_CONTROL_DEVICE	DEVICE_CONDITION	WEATHER_CONDITION	LIGHTING_CONDITION
100	CHICAGO	711d632dd484b30cd03cf068edf788a561d784588cd5e0...	NA	2017-06-15 20:00:00.000	2017-06-15 00:00:00.000	12:00:00.000	35	TRAFFIC SIGNAL	FUNCTIONING PROPERLY	UNKNOWN	UNKNC

SK_FACT_ACCIDENTS	CRASH_ID	SK_Source	SK_Location	SK_Date	SK_Time	INJURIES_TOTAL	IS_PEDESTRIAN	MOTORIST_INJURY_COUNT	MOTORIST_KILLED_COUNT	PEDESTRIAN_INJURY_COUNT	PEDESTRIAN_KILLED_COUNT
2075428	100	1	97124	20170615	13	2	0	0	0	0	0

SK_Source	Source	DL_CreateDate	DL_JobID	DL_WorkflowName
1	CHICAGO	2024-04-07 00:26:23.967	6y73A	Load_Dim_Source

SK_Location	Street_Name	Latitude	Longitude	DL_CreateDate	DL_JobID	DL_WorkflowFileName
97124	CALIFORNIA AVE	41.7062	-87.6916	2024-04-07 15:00:51.277	s9r0q9	Load_Dim_Location

Date_SK	Date	Day	Month	Year	Season	DL_CreateDate	DL_JobID	DL_WorkflowName
20170615	2017-06-15 00:00:00.000	15	6	2017	Summer	2024-04-07 00:24:05.100	LNXLxV1	Load_Dim_Date

Time_SK	Hour_24	Time_Period	DL_CreateDate	DL_JobID	DL_WorkflowName
13	12	Early Afternoon	2024-04-07 09:48:28.677	kSYIwB	Load_Dim_Time

SK_Contribution	DK_Contribution	Contributing_Factor_code	Contributing_Factor_Description	scd_start	scd_end	scd_version	scd_active	DL_JobID	DL_WorkFlow_Name	DL_Created_Date
70	132	UNABLE TO DETERMINE		2024-04-07 15:04:05.287	9999-01-01 12:00:00.000	1	1	QBaSFz	Load_Dim_Contribution	2024-04-07

Query executed successfully.

VAISHVEER (16.0 RTM) | VAISHVEER(vaish (57)) | MVC\_FINAL\_PROJECT | 00:00:00 | 7 rows

## SCHEMA'S FOR THE STAGGED TABLES

The screenshot shows a SQL Server Management Studio (SSMS) window. The query pane at the top contains the following T-SQL code:

```
SELECT TABLE_SCHEMA, TABLE_NAME
FROM MVC_FINAL_PROJECT.INFORMATION_SCHEMA.TABLES
WHERE TABLE_TYPE = 'BASE TABLE'
ORDER BY TABLE_SCHEMA, TABLE_NAME;
```

The results pane below displays a table with two columns: TABLE\_SCHEMA and TABLE\_NAME. The data consists of 32 rows, each showing a schema name followed by a table name. The schemas are primarily 'dbo' with some exceptions like 'stg\_MVC\_Austin' and 'stg\_MVC\_NY'. The table names are descriptive, such as 'Dim\_Contribution', 'FACT\_ACCIDENTS', and 'Log\_Catcher' variants for Austin, Chicago, and New York.

TABLE_SCHEMA	TABLE_NAME
dbo	Dim_Contribution
dbo	Dim_Contribution_Test
dbo	DIM_DATE
dbo	DIM_Location
dbo	DIM_Source
dbo	DIM_TIME
dbo	DIM_Vehicle_Type
dbo	FACT_ACCIDENTS
dbo	FACT_ACCIDENTS_AUSTIN
dbo	FACT_ACCIDENTS_Chicago
dbo	FACT_ACCIDENTS_New_York
dbo	FACT_ACCIDENTS_TEST
dbo	FACT_CONTRIBUTION
dbo	FACT_CONTRIBUTION_Austin
dbo	FACT_CONTRIBUTION_Chicago
dbo	FACT_CONTRIBUTION_NY
dbo	FACT_VEHICLE
dbo	FACT_VEHICLE_Austin
dbo	FACT_VEHICLE_Chicago
dbo	FACT_VEHICLE_NY
dbo	Log_Catcher_Austin_Normalized
dbo	Log_Catcher_Chicago_Cleaned
dbo	Log_Catcher_Fact_Vehicle_Au..
dbo	Log_Catcher_NY_normalized
dbo	Log_Catcher_test
dbo	stg_MVC_Austin
dbo	stg_MVC_Austin_normalized
dbo	stg_MVC_Chicago
dbo	stg_MVC_Chicago_cleaned
dbo	stg_MVC_Chicago_normalized
dbo	stg_MVC_NY
dbo	stg_MVC_NY_normalized

At the bottom of the results grid, a message states "Query executed successfully." To the right of the grid, status information is displayed: VAISHVEER (16.0 RTM) | VAISHVEER(vaish (52)) | MVC\_FINAL\_PROJECT | 00:00:00 | 32 rows.