Honda Huang - honda.huang@baruchmail.cuny.edu

Mark Won - mark.won@baruchmail.cuny.edu

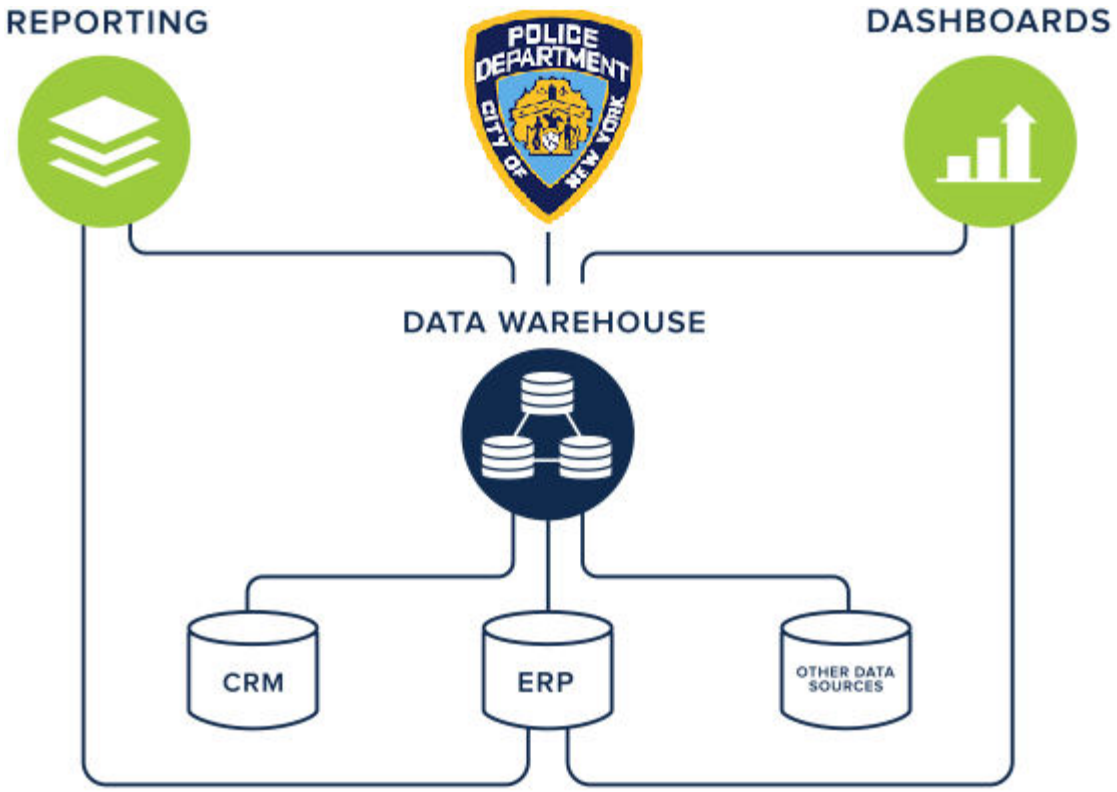
Steve Yun - steve.yun@baruchmail.cuny.edu

Ezra Magder - ezra.magder@baruchmail.cuny.edu

CIS 4400 DMWA

Group 01

NYC Vehicle Collisions Data Warehouse



Introduction

Our project is in direct cooperation with the NYPD to provide a BI Application for analysis on traffic accidents in NYC. Our enterprise business process revolves around helping the NYC government analyze vehicle collision data in conjunction with weather data from 2013 - 2015 to find potential solutions to reduce traffic accidents. Our core business function is to incorporate the data on NYC vehicle collisions with weather and temperature data to create a data warehouse for analytical purposes.

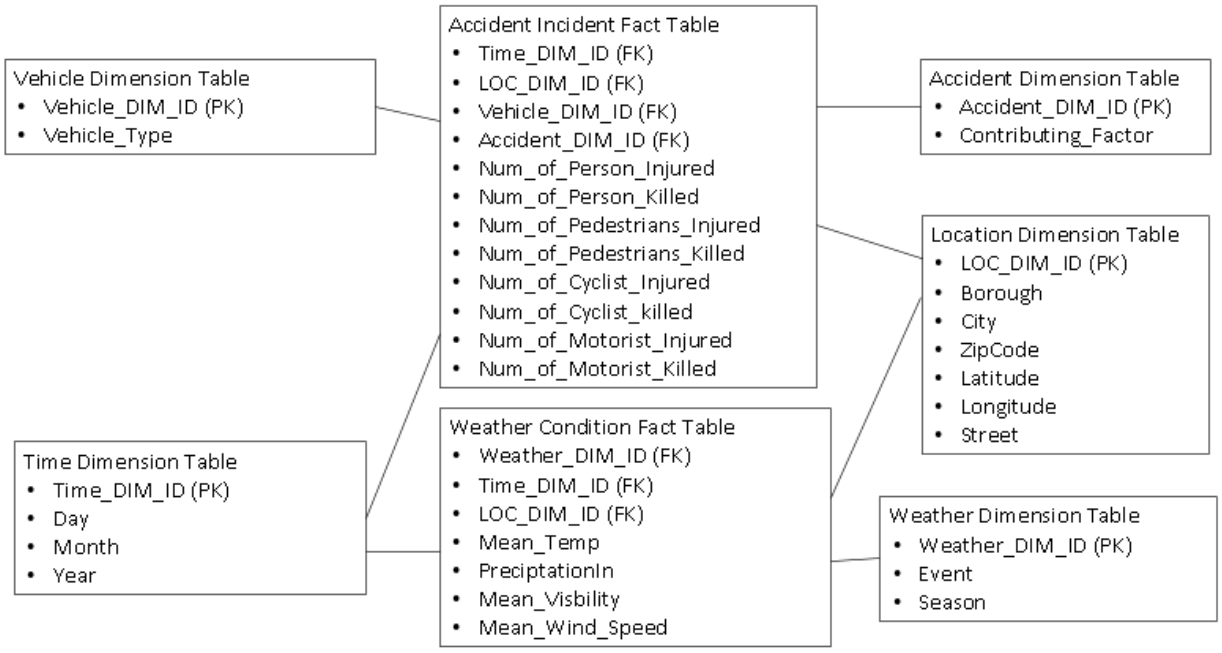
We are using two different data sources: NYPD Motor Vehicle Collisions Data and NYC weather data. The NYPD Motor Vehicle Collisions Data is available online on the NYC open data website and the NYC weather data is available to download from github. The links are provided below. The collisions data provides sufficient information on all police recorded accidents during the two years such as vehicle type and contribution factor. The weather data lists the weather and temperature for each day during the two years for the entire city. We cleaned the datasets to remove any blank cells to avoid bad data and then created a separate ID field for the ETL step. We also fixed up errors such as
 tags inappropriately located in some cells.

Data Sources :

https://data.cityofnewyork.us/Public-Safety/NYPD-Motor-Vehicle-Collisions/h9gi-nx95

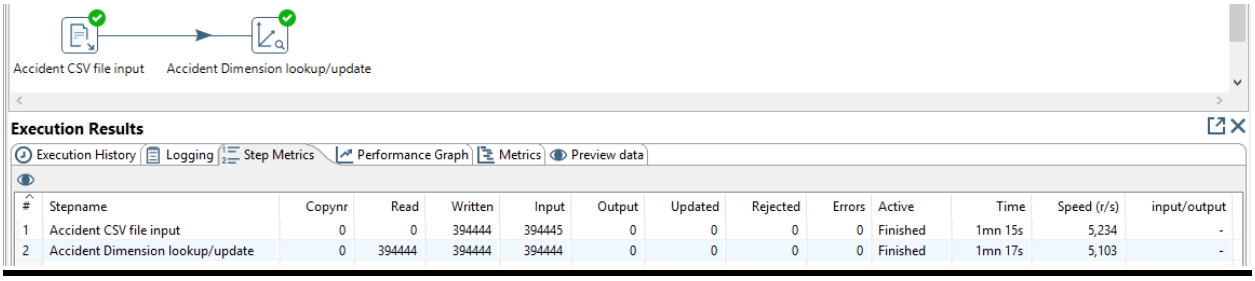
https://github.com/zonination/weather-us/blob/master/nyc.csv

Dimensional Model Diagram



ETL Processes

Accident Dimension



#	Stepname	Copynr	Read	Written	Input	Output	Updated	Rejected	Errors	Active	Time	Speed (r/s)	input/output
1	Accident CSV file input	0	0	394444	394445	0	0	0	0	Finished	1mn 15s	5,234	-
2	Accident Dimension lookup/update	0	394444	394444	394444	0	0	0	0	Finished	1mn 17s	5,103	-

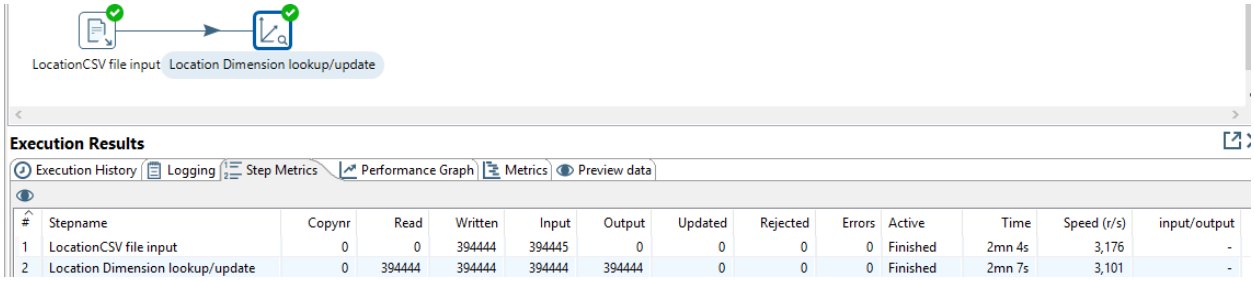
Caption: This is the accident dimension table transformation.

Rows of step: accident_dim (100 rows)

#	ACCIDENT_DIM_ID	VERSION	DATE_FROM	DATE_TO	ACCIDENT_ID	CONTRIBUTING FACTOR VEHICLE 1	CONTRIBUTING FACTOR VEHICLE 2	CONTRIBUTING FACTOR VEHICLE 3	CONTRIBUTING FACTOR VEHICLE 4	CONTRIBUTING FACTOR VEHICLE 5
1	0	1	<null>	<null>	<null>	<null>	<null>	<null>	<null>	<null>
2	1	1	1900/01/01 00:00:00.000000000	2199/12/31 23:59:59.999000000	1	Unspecified	Unspecified	<null>	<null>	<null>
3	2	1	1900/01/01 00:00:00.000000000	2199/12/31 23:59:59.999000000	2	Unspecified	<null>	<null>	<null>	<null>
4	3	1	1900/01/01 00:00:00.000000000	2199/12/31 23:59:59.999000000	3	Fatigued/Drowsy	Unspecified	<null>	<null>	<null>
5	4	1	1900/01/01 00:00:00.000000000	2199/12/31 23:59:59.999000000	4	Unspecified	Unspecified	<null>	<null>	<null>
6	5	1	1900/01/01 00:00:00.000000000	2199/12/31 23:59:59.999000000	5	Alcohol Involvement	Unspecified	<null>	<null>	<null>
7	6	1	1900/01/01 00:00:00.000000000	2199/12/31 23:59:59.999000000	6	Unspecified	Unspecified	<null>	<null>	<null>
8	7	1	1900/01/01 00:00:00.000000000	2199/12/31 23:59:59.999000000	7	Unspecified	Unspecified	<null>	<null>	<null>
9	8	1	1900/01/01 00:00:00.000000000	2199/12/31 23:59:59.999000000	8	Turning Improperly	Unspecified	<null>	<null>	<null>
10	9	1	1900/01/01 00:00:00.000000000	2199/12/31 23:59:59.999000000	9	Fell Asleep	Passing or Lane Usage Improper	<null>	<null>	<null>
11	10	1	1900/01/01 00:00:00.000000000	2199/12/31 23:59:59.999000000	10	Unspecified	Unspecified	<null>	<null>	<null>
12	11	1	1900/01/01 00:00:00.000000000	2199/12/31 23:59:59.999000000	11	Unspecified	Unspecified	<null>	<null>	<null>
13	12	1	1900/01/01 00:00:00.000000000	2199/12/31 23:59:59.999000000	12	Unspecified	<null>	<null>	<null>	<null>
14	13	1	1900/01/01 00:00:00.000000000	2199/12/31 23:59:59.999000000	13	Unspecified	Unspecified	<null>	<null>	<null>
15	14	1	1900/01/01 00:00:00.000000000	2199/12/31 23:59:59.999000000	14	Outside Car Distraction	Fatigued/Drowsy	<null>	<null>	<null>
16	15	1	1900/01/01 00:00:00.000000000	2199/12/31 23:59:59.999000000	15	Unspecified	Unspecified	<null>	<null>	<null>
17	16	1	1900/01/01 00:00:00.000000000	2199/12/31 23:59:59.999000000	16	Unspecified	Unspecified	<null>	<null>	<null>
18	17	1	1900/01/01 00:00:00.000000000	2199/12/31 23:59:59.999000000	17	Unspecified	Unspecified	<null>	<null>	<null>
19	18	1	1900/01/01 00:00:00.000000000	2199/12/31 23:59:59.999000000	18	Driver Inattention/Distraction	Unspecified	<null>	<null>	<null>
20	19	1	1900/01/01 00:00:00.000000000	2199/12/31 23:59:59.999000000	19	Unspecified	Unspecified	<null>	<null>	<null>
21	20	1	1900/01/01 00:00:00.000000000	2199/12/31 23:59:59.999000000	20	Fatigued/Drowsy	Backing Unsafely	<null>	<null>	<null>
22	21	1	1900/01/01 00:00:00.000000000	2199/12/31 23:59:59.999000000	21	Fatigued/Drowsy	Backing Unsafely	<null>	<null>	<null>
23	22	1	1900/01/01 00:00:00.000000000	2199/12/31 23:59:59.999000000	22	Unspecified	Unspecified	<null>	<null>	<null>
24	23	1	1900/01/01 00:00:00.000000000	2199/12/31 23:59:59.999000000	23	Driver Inattention/Distraction	Driver Inattention/Distraction	<null>	<null>	<null>
25	24	1	1900/01/01 00:00:00.000000000	2199/12/31 23:59:59.999000000	24	Driver Inattention/Distraction	Unspecified	<null>	<null>	<null>
26	25	1	1900/01/01 00:00:00.000000000	2199/12/31 23:59:59.999000000	25	Unspecified	Unspecified	Unspecified	<null>	<null>
27	26	1	1900/01/01 00:00:00.000000000	2199/12/31 23:59:59.999000000	26	Unspecified	Unspecified	<null>	<null>	<null>
28	27	1	1900/01/01 00:00:00.000000000	2199/12/31 23:59:59.999000000	27	Other Vehicular	Driver Inattention/Distraction	<null>	<null>	<null>
29	28	1	1900/01/01 00:00:00.000000000	2199/12/31 23:59:59.999000000	28	Unspecified	Unspecified	<null>	<null>	<null>
30	29	1	1900/01/01 00:00:00.000000000	2199/12/31 23:59:59.999000000	29	Other Vehicular	Unspecified	<null>	<null>	<null>
31	30	1	1900/01/01 00:00:00.000000000	2199/12/31 23:59:59.999000000	30	Unspecified	Unspecified	<null>	<null>	<null>
32	31	1	1900/01/01 00:00:00.000000000	2199/12/31 23:59:59.999000000	31	Physical Disability	Physical Disability	Unspecified	<null>	<null>

Caption:This is the accident dimension table preview.

Location Dimension



#	Stepname	Copynr	Read	Written	Input	Output	Updated	Rejected	Errors	Active	Time	Speed (r/s)	input/output
1	LocationCSV file input	0	0	394444	394445	0	0	0	0	Finished	2mn 4s	3,176	-
2	Location Dimension lookup/update	0	394444	394444	394444	394444	0	0	0	Finished	2mn 7s	3,101	-

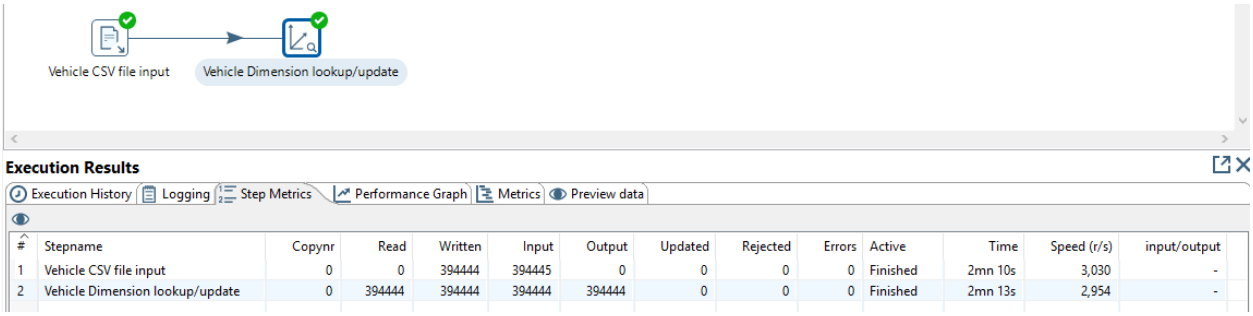
Caption: This is the location dimension table transformation

Rows of step: Location_dim (100 rows)

#	LOCATION_DIM_ID	VERSION	DATE_FROM	DATE_TO	ACCIDENT_ID	BOROUGH	CITY	ZIP CODE	LATITUDE	LONGITUDE	ON STREET NAME	CROSS STREET NAME	OFF STREET NAME
1	0	1	<null>	<null>	<null>	<null>	<null>	<null>	<null>	<null>	<null>	<null>	<null>
2	1	1	1900/01/01 00:00:00.000000000	2199/12/31 23:59:59.999000000	1	MANHATTAN	New York City (USA)	10023	40.8	-74	CENTRAL PARK WEST	WEST 69 STREET	<null>
3	2	1	1900/01/01 00:00:00.000000000	2199/12/31 23:59:59.999000000	2	BROOKLYN	New York City (USA)	11211	40.7	-74	KEAP STREET	LEE AVENUE	<null>
4	3	1	1900/01/01 00:00:00.000000000	2199/12/31 23:59:59.999000000	3	MANHATTAN	New York City (USA)	10012	40.7	-74	KENMARE STREET	MULBERRY STREET	<null>
5	4	1	1900/01/01 00:00:00.000000000	2199/12/31 23:59:59.999000000	4	BROOKLYN	New York City (USA)	11234	40.6	-73.9	EAST 48 STREET	AVENUE K	<null>
6	5	1	1900/01/01 00:00:00.000000000	2199/12/31 23:59:59.999000000	5	QUEENS	New York City (USA)	11373	40.7	-73.9	BROADWAY	45 AVENUE	<null>
7	6	1	1900/01/01 00:00:00.000000000	2199/12/31 23:59:59.999000000	6	MANHATTAN	New York City (USA)	10065	40.8	-74	EAST 64 STREET	MADISON AVENUE	<null>
8	7	1	1900/01/01 00:00:00.000000000	2199/12/31 23:59:59.999000000	7	BROOKLYN	New York City (USA)	11211	40.7	-74	ROEBLING STREET	SOUTH 4 STREET	<null>
9	8	1	1900/01/01 00:00:00.000000000	2199/12/31 23:59:59.999000000	8	MANHATTAN	New York City (USA)	10016	40.7	-74	EAST 40 STREET	3 AVENUE	<null>
10	9	1	1900/01/01 00:00:00.000000000	2199/12/31 23:59:59.999000000	9	QUEENS	New York City (USA)	11362	40.8	-73.7	NORTHERN BOULEVARD	247 STREET	<null>
11	10	1	1900/01/01 00:00:00.000000000	2199/12/31 23:59:59.999000000	10	BRONX	New York City (USA)	10466	40.9	-73.8	PITMAN AVENUE	MONTICELLO AVENUE	<null>
12	11	1	1900/01/01 00:00:00.000000000	2199/12/31 23:59:59.999000000	11	BRONX	New York City (USA)	10470	40.9	-73.9	ONEIDA AVENUE	EAST 238 STREET	<null>
13	12	1	1900/01/01 00:00:00.000000000	2199/12/31 23:59:59.999000000	12	MANHATTAN	New York City (USA)	10002	40.7	-74	BOWERY	RIVINGTON STREET	<null>
14	13	1	1900/01/01 00:00:00.000000000	2199/12/31 23:59:59.999000000	13	BROOKLYN	New York City (USA)	11221	40.7	-73.9	THROOP AVENUE	QUINCY STREET	<null>
15	14	1	1900/01/01 00:00:00.000000000	2199/12/31 23:59:59.999000000	14	QUEENS	New York City (USA)	11357	40.8	-73.8	WILLETS POINT BOULEVARD	15 DRIVE	<null>
16	15	1	1900/01/01 00:00:00.000000000	2199/12/31 23:59:59.999000000	15	BRONX	New York City (USA)	10470	40.9	-73.9	EAST 233 STREET	WEBSTER AVENUE	<null>
17	16	1	1900/01/01 00:00:00.000000000	2199/12/31 23:59:59.999000000	16	BROOKLYN	New York City (USA)	11201	40.7	-74	NAVY STREET	YORK STREET	<null>
18	17	1	1900/01/01 00:00:00.000000000	2199/12/31 23:59:59.999000000	17	BROOKLYN	New York City (USA)	11207	40.7	-73.9	PENNSYLVANIA AVENUE	HEGEMAN AVENUE	<null>
19	18	1	1900/01/01 00:00:00.000000000	2199/12/31 23:59:59.999000000	18	MANHATTAN	New York City (USA)	10065	40.8	-74	EAST 62 STREET	YORK AVENUE	<null>
20	19	1	1900/01/01 00:00:00.000000000	2199/12/31 23:59:59.999000000	19	QUEENS	New York City (USA)	11355	40.8	-73.8	BLOSSOM AVENUE	COLLEGE POINT BOULEVARD	<null>
21	20	1	1900/01/01 00:00:00.000000000	2199/12/31 23:59:59.999000000	20	BROOKLYN	New York City (USA)	11232	40.7	-74	24 STREET	4 AVENUE	<null>
22	21	1	1900/01/01 00:00:00.000000000	2199/12/31 23:59:59.999000000	21	MANHATTAN	New York City (USA)	10001	40.8	-74	WEST 25 STREET	11 AVENUE	<null>
23	22	1	1900/01/01 00:00:00.000000000	2199/12/31 23:59:59.999000000	22	MANHATTAN	New York City (USA)	10003	40.7	-74	EAST 14 STREET	UNION SQUARE WEST	<null>
24	23	1	1900/01/01 00:00:00.000000000	2199/12/31 23:59:59.999000000	23	MANHATTAN	New York City (USA)	10280	40.7	-74	WEST STREET	1 PLACE	<null>
25	24	1	1900/01/01 00:00:00.000000000	2199/12/31 23:59:59.999000000	24	BRONX	New York City (USA)	10463	40.9	-73.9	VANCORTLANDT PARK SOUTH	PUTNAM AVENUE WEST	<null>
26	25	1	1900/01/01 00:00:00.000000000	2199/12/31 23:59:59.999000000	25	QUEENS	New York City (USA)	11106	40.8	-73.9	31 STREET	35 AVENUE	<null>
27	26	1	1900/01/01 00:00:00.000000000	2199/12/31 23:59:59.999000000	26	QUEENS	New York City (USA)	11355	40.7	-73.8	BOOTH MEMORIAL AVENUE	133 STREET	<null>
28	27	1	1900/01/01 00:00:00.000000000	2199/12/31 23:59:59.999000000	27	BROOKLYN	New York City (USA)	11220	40.6	-74	63 STREET	5 AVENUE	<null>
29	28	1	1900/01/01 00:00:00.000000000	2199/12/31 23:59:59.999000000	28	MANHATTAN	New York City (USA)	10029	40.8	-74	EAST 97 STREET	MADISON AVENUE	<null>
30	29	1	1900/01/01 00:00:00.000000000	2199/12/31 23:59:59.999000000	29	BROOKLYN	New York City (USA)	11238	40.7	-74	FLATBUSH AVENUE	PARK PLACE	<null>
31	30	1	1900/01/01 00:00:00.000000000	2199/12/31 23:59:59.999000000	30	MANHATTAN	New York City (USA)	10017	40.8	-74	EAST 49 STREET	LEXINGTON AVENUE	<null>
32	31	1	1900/01/01 00:00:00.000000000	2199/12/31 23:59:59.999000000	31	MANHATTAN	New York City (USA)	10019	40.8	-74	WEST 57 STREET	9 AVENUE	<null>
33	32	1	1900/01/01 00:00:00.000000000	2199/12/31 23:59:59.999000000	32	BROOKLYN	New York City (USA)	11210	40.6	-73.9	FLATBUSH AVENUE	AVENUE H	<null>

Caption:This is the location dimension table preview

Vehicle Dimension



#	Stepname	Copynr	Read	Written	Input	Output	Updated	Rejected	Errors	Active	Time	Speed (r/s)	input/output
1	Vehicle CSV file input	0	0	394444	394445	0	0	0	0	Finished	2mn 10s	3,030	-
2	Vehicle Dimension lookup/update	0	394444	394444	394444	394444	0	0	0	Finished	2mn 13s	2,954	-

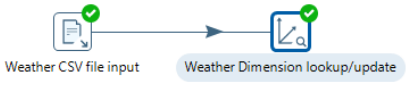
Caption: This is the vehicle dimension table transformation

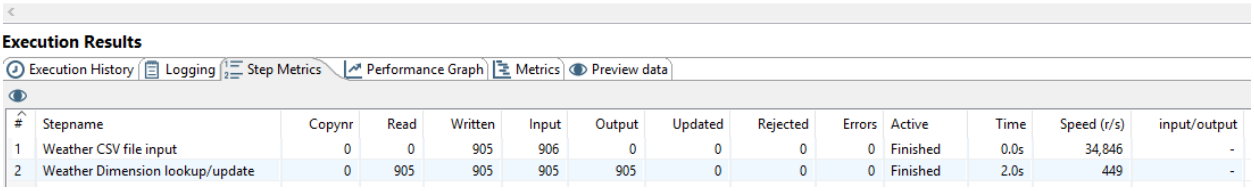
Rows of step: VEHICLE_DIM (100 rows)

#	VEHICLE_DIM_ID	VERSION	DATE_FROM	DATE_TO	ACCIDENT_ID	VEHICLE TYPE CODE 1	VEHICLE TYPE CODE 2	VEHICLE TYPE CODE 3	VEHICLE TYPE CODE 4	VEHICLE TYPE CODE 5
1	0	1	<null>	<null>	<null>	<null>	<null>	<null>	<null>	<null>
2	1	1	1900/01/01 00:00:00.000000000	2199/12/31 23:59:59.999000000	1	PASSENGER VEHICLE	TAXI	<null>	<null>	<null>
3	2	1	1900/01/01 00:00:00.000000000	2199/12/31 23:59:59.999000000	2	PASSENGER VEHICLE	<null>	<null>	<null>	<null>
4	3	1	1900/01/01 00:00:00.000000000	2199/12/31 23:59:59.999000000	3	SPORT UTILITY / STATION WAGON	PASSENGER VEHICLE	<null>	<null>	<null>
5	4	1	1900/01/01 00:00:00.000000000	2199/12/31 23:59:59.999000000	4	SPORT UTILITY / STATION WAGON	UNKNOWN	<null>	<null>	<null>
6	5	1	1900/01/01 00:00:00.000000000	2199/12/31 23:59:59.999000000	5	PASSENGER VEHICLE	PASSENGER VEHICLE	<null>	<null>	<null>
7	6	1	1900/01/01 00:00:00.000000000	2199/12/31 23:59:59.999000000	6	UNKNOWN	TAXI	<null>	<null>	<null>
8	7	1	1900/01/01 00:00:00.000000000	2199/12/31 23:59:59.999000000	7	TAXI	BICYCLE	<null>	<null>	<null>
9	8	1	1900/01/01 00:00:00.000000000	2199/12/31 23:59:59.999000000	8	PASSENGER VEHICLE	PASSENGER VEHICLE	<null>	<null>	<null>
10	9	1	1900/01/01 00:00:00.000000000	2199/12/31 23:59:59.999000000	9	PASSENGER VEHICLE	UNKNOWN	<null>	<null>	<null>
11	10	1	1900/01/01 00:00:00.000000000	2199/12/31 23:59:59.999000000	10	PASSENGER VEHICLE	UNKNOWN	<null>	<null>	<null>
12	11	1	1900/01/01 00:00:00.000000000	2199/12/31 23:59:59.999000000	11	SPORT UTILITY / STATION WAGON	UNKNOWN	<null>	<null>	<null>
13	12	1	1900/01/01 00:00:00.000000000	2199/12/31 23:59:59.999000000	12	VAN	VAN	<null>	<null>	<null>
14	13	1	1900/01/01 00:00:00.000000000	2199/12/31 23:59:59.999000000	13	PASSENGER VEHICLE	BICYCLE	<null>	<null>	<null>
15	14	1	1900/01/01 00:00:00.000000000	2199/12/31 23:59:59.999000000	14	PASSENGER VEHICLE	PASSENGER VEHICLE	<null>	<null>	<null>
16	15	1	1900/01/01 00:00:00.000000000	2199/12/31 23:59:59.999000000	15	PASSENGER VEHICLE	UNKNOWN	<null>	<null>	<null>
17	16	1	1900/01/01 00:00:00.000000000	2199/12/31 23:59:59.999000000	16	PASSENGER VEHICLE	UNKNOWN	<null>	<null>	<null>
18	17	1	1900/01/01 00:00:00.000000000	2199/12/31 23:59:59.999000000	17	PASSENGER VEHICLE	PASSENGER VEHICLE	<null>	<null>	<null>
19	18	1	1900/01/01 00:00:00.000000000	2199/12/31 23:59:59.999000000	18	VAN	LARGE COM VEH(6 OR MORE TIRES)	<null>	<null>	<null>
20	19	1	1900/01/01 00:00:00.000000000	2199/12/31 23:59:59.999000000	19	PASSENGER VEHICLE	PASSENGER VEHICLE	<null>	<null>	<null>
21	20	1	1900/01/01 00:00:00.000000000	2199/12/31 23:59:59.999000000	20	PASSENGER VEHICLE	PASSENGER VEHICLE	<null>	<null>	<null>
22	21	1	1900/01/01 00:00:00.000000000	2199/12/31 23:59:59.999000000	21	PASSENGER VEHICLE	VAN	<null>	<null>	<null>
23	22	1	1900/01/01 00:00:00.000000000	2199/12/31 23:59:59.999000000	22	PASSENGER VEHICLE	PASSENGER VEHICLE	<null>	<null>	<null>
24	23	1	1900/01/01 00:00:00.000000000	2199/12/31 23:59:59.999000000	23	BUS	SPORT UTILITY / STATION WAGON	<null>	<null>	<null>
25	24	1	1900/01/01 00:00:00.000000000	2199/12/31 23:59:59.999000000	24	PASSENGER VEHICLE	LIVERY VEHICLE	<null>	<null>	<null>
26	25	1	1900/01/01 00:00:00.000000000	2199/12/31 23:59:59.999000000	25	SPORT UTILITY / STATION WAGON	SPORT UTILITY / STATION WAGON	SPORT UTILITY / STATION WAGON	<null>	<null>
27	26	1	1900/01/01 00:00:00.000000000	2199/12/31 23:59:59.999000000	26	PASSENGER VEHICLE	PASSENGER VEHICLE	<null>	<null>	<null>
28	27	1	1900/01/01 00:00:00.000000000	2199/12/31 23:59:59.999000000	27	PASSENGER VEHICLE	PASSENGER VEHICLE	<null>	<null>	<null>
29	28	1	1900/01/01 00:00:00.000000000	2199/12/31 23:59:59.999000000	28	BUS	SPORT UTILITY / STATION WAGON	<null>	<null>	<null>
30	29	1	1900/01/01 00:00:00.000000000	2199/12/31 23:59:59.999000000	29	PASSENGER VEHICLE	OTHER	<null>	<null>	<null>
31	30	1	1900/01/01 00:00:00.000000000	2199/12/31 23:59:59.999000000	30	TAXI	UNKNOWN	<null>	<null>	<null>
32	31	1	1900/01/01 00:00:00.000000000	2199/12/31 23:59:59.999000000	31	TAXI	TAXI	SPORT UTILITY / STATION WAGON	<null>	<null>
33	32	1	1900/01/01 00:00:00.000000000	2199/12/31 23:59:59.999000000	32	LARGE COM VEH(6 OR MORE TIRES)	VAN	<null>	<null>	<null>

Caption: This is the vehicle dimension table preview

Weather Dimension





#	Stepname	Copynr	Read	Written	Input	Output	Updated	Rejected	Errors	Active	Time	Speed (r/s)	input/output
1	Weather CSV file input	0	0	905	906	0	0	0	0	Finished	0.0s	34,846	-
2	Weather Dimension lookup/update	0	905	905	905	905	0	0	0	Finished	2.0s	449	-

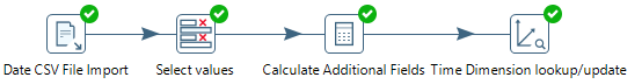
Caption: This is the weather dimension table transformation

Rows of step: Weather_Dim (100 rows)

#	WEATHER_DIM_ID	VERSION	DATE_FROM	DATE_TO	WEATHERID	EVENTS	SEASON
1	0	1	<null>	<null>	<null>	<null>	<null>
2	1	1	2013/01/01 00:00:00.000000000	2015/12/31 23:59:59.999000000	1	Rain-Thunderstorm	Summer
3	2	1	2013/01/01 00:00:00.000000000	2015/12/31 23:59:59.999000000	2	Rain	Summer
4	3	1	2013/01/01 00:00:00.000000000	2015/12/31 23:59:59.999000000	3	Rain	Summer
5	4	1	2013/01/01 00:00:00.000000000	2015/12/31 23:59:59.999000000	4	<null>	Summer
6	5	1	2013/01/01 00:00:00.000000000	2015/12/31 23:59:59.999000000	5	<null>	Summer
7	6	1	2013/01/01 00:00:00.000000000	2015/12/31 23:59:59.999000000	6	<null>	Summer
8	7	1	2013/01/01 00:00:00.000000000	2015/12/31 23:59:59.999000000	7	<null>	Summer
9	8	1	2013/01/01 00:00:00.000000000	2015/12/31 23:59:59.999000000	8	<null>	Summer
10	9	1	2013/01/01 00:00:00.000000000	2015/12/31 23:59:59.999000000	9	<null>	Summer
11	10	1	2013/01/01 00:00:00.000000000	2015/12/31 23:59:59.999000000	10	<null>	Summer
12	11	1	2013/01/01 00:00:00.000000000	2015/12/31 23:59:59.999000000	11	Rain-Thunderstorm	Summer
13	12	1	2013/01/01 00:00:00.000000000	2015/12/31 23:59:59.999000000	12	<null>	Summer
14	13	1	2013/01/01 00:00:00.000000000	2015/12/31 23:59:59.999000000	13	Rain	Summer
15	14	1	2013/01/01 00:00:00.000000000	2015/12/31 23:59:59.999000000	14	Rain-Thunderstorm	Summer
16	15	1	2013/01/01 00:00:00.000000000	2015/12/31 23:59:59.999000000	15	<null>	Summer
17	16	1	2013/01/01 00:00:00.000000000	2015/12/31 23:59:59.999000000	16	Rain	Summer
18	17	1	2013/01/01 00:00:00.000000000	2015/12/31 23:59:59.999000000	17	Rain	Summer
19	18	1	2013/01/01 00:00:00.000000000	2015/12/31 23:59:59.999000000	18	<null>	Summer
20	19	1	2013/01/01 00:00:00.000000000	2015/12/31 23:59:59.999000000	19	Rain	Summer
21	20	1	2013/01/01 00:00:00.000000000	2015/12/31 23:59:59.999000000	20	Rain-Thunderstorm	Summer
22	21	1	2013/01/01 00:00:00.000000000	2015/12/31 23:59:59.999000000	21	<null>	Summer
23	22	1	2013/01/01 00:00:00.000000000	2015/12/31 23:59:59.999000000	22	<null>	Summer
24	23	1	2013/01/01 00:00:00.000000000	2015/12/31 23:59:59.999000000	23	Rain	Summer
25	24	1	2013/01/01 00:00:00.000000000	2015/12/31 23:59:59.999000000	24	<null>	Summer
26	25	1	2013/01/01 00:00:00.000000000	2015/12/31 23:59:59.999000000	25	Rain	Summer
27	26	1	2013/01/01 00:00:00.000000000	2015/12/31 23:59:59.999000000	26	<null>	Summer
28	27	1	2013/01/01 00:00:00.000000000	2015/12/31 23:59:59.999000000	27	<null>	Summer
29	28	1	2013/01/01 00:00:00.000000000	2015/12/31 23:59:59.999000000	28	<null>	Summer
30	29	1	2013/01/01 00:00:00.000000000	2015/12/31 23:59:59.999000000	29	Rain	Summer
31	30	1	2013/01/01 00:00:00.000000000	2015/12/31 23:59:59.999000000	30	Rain	Summer
32	31	1	2013/01/01 00:00:00.000000000	2015/12/31 23:59:59.999000000	31	Rain	Summer
33	32	1	2013/01/01 00:00:00.000000000	2015/12/31 23:59:59.999000000	32	<null>	Summer

Caption: This is the weather dimension table preview

Time Dimension



Execution Results

Execution History | Logging | Step Metrics | Performance Graph | Metrics | Preview data

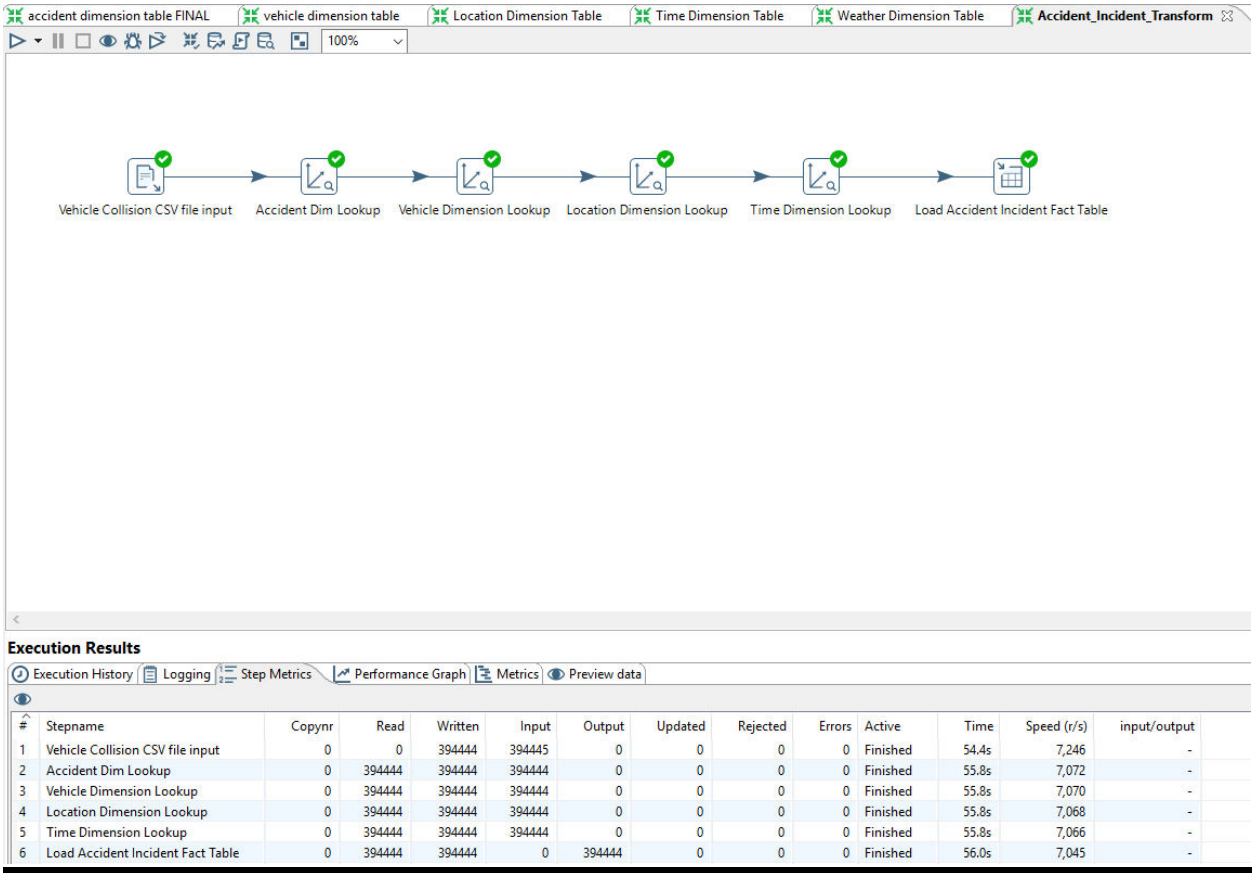
#	Stepname	Copynr	Read	Written	Input	Output	Updated	Rejected	Errors	Active	Time	Speed (r/s)	input/output
1	Date CSV File Import	0	0	905	906	0	0	0	0	Finished	0.1s	17,094	-
2	Select values	0	905	905	0	0	0	0	0	Finished	0.1s	15,603	-
3	Calculate Additional Fields	0	905	905	0	0	0	0	0	Finished	0.1s	7,669	-
4	Time Dimension lookup/update	0	905	905	905	905	0	0	0	Finished	1.1s	838	-

Caption: This is the time dimension table transformation

#	TIME_DIM_ID	VERSION	DATE_FROM	DATE_TO	DATE	DAY_OF_YEAR	MONTH	YEAR	MONTH_NAME	DAY_OF_WEEK_NAME
1	0	1	<null>	<null>	<null>	<null>	<null>	<null>	<null>	<null>
2	1	1	1900/01/01 00:00:00.000000000	2199/12/31 23:59:59.999000000	2013/07/10 00:00:00.000000000	191.0	7.0	2013.0	July	Wednesday
3	2	1	1900/01/01 00:00:00.000000000	2199/12/31 23:59:59.999000000	2013/07/11 00:00:00.000000000	192.0	7.0	2013.0	July	Thursday
4	3	1	1900/01/01 00:00:00.000000000	2199/12/31 23:59:59.999000000	2013/07/12 00:00:00.000000000	193.0	7.0	2013.0	July	Friday
5	4	1	1900/01/01 00:00:00.000000000	2199/12/31 23:59:59.999000000	2013/07/13 00:00:00.000000000	194.0	7.0	2013.0	July	Saturday
6	5	1	1900/01/01 00:00:00.000000000	2199/12/31 23:59:59.999000000	2013/07/14 00:00:00.000000000	195.0	7.0	2013.0	July	Sunday
7	6	1	1900/01/01 00:00:00.000000000	2199/12/31 23:59:59.999000000	2013/07/15 00:00:00.000000000	196.0	7.0	2013.0	July	Monday
8	7	1	1900/01/01 00:00:00.000000000	2199/12/31 23:59:59.999000000	2013/07/16 00:00:00.000000000	197.0	7.0	2013.0	July	Tuesday
9	8	1	1900/01/01 00:00:00.000000000	2199/12/31 23:59:59.999000000	2013/07/17 00:00:00.000000000	198.0	7.0	2013.0	July	Wednesday
10	9	1	1900/01/01 00:00:00.000000000	2199/12/31 23:59:59.999000000	2013/07/18 00:00:00.000000000	199.0	7.0	2013.0	July	Thursday
11	10	1	1900/01/01 00:00:00.000000000	2199/12/31 23:59:59.999000000	2013/07/19 00:00:00.000000000	200.0	7.0	2013.0	July	Friday
12	11	1	1900/01/01 00:00:00.000000000	2199/12/31 23:59:59.999000000	2013/07/20 00:00:00.000000000	201.0	7.0	2013.0	July	Saturday
13	12	1	1900/01/01 00:00:00.000000000	2199/12/31 23:59:59.999000000	2013/07/21 00:00:00.000000000	202.0	7.0	2013.0	July	Sunday
14	13	1	1900/01/01 00:00:00.000000000	2199/12/31 23:59:59.999000000	2013/07/22 00:00:00.000000000	203.0	7.0	2013.0	July	Monday
15	14	1	1900/01/01 00:00:00.000000000	2199/12/31 23:59:59.999000000	2013/07/23 00:00:00.000000000	204.0	7.0	2013.0	July	Tuesday
16	15	1	1900/01/01 00:00:00.000000000	2199/12/31 23:59:59.999000000	2013/07/24 00:00:00.000000000	205.0	7.0	2013.0	July	Wednesday
17	16	1	1900/01/01 00:00:00.000000000	2199/12/31 23:59:59.999000000	2013/07/25 00:00:00.000000000	206.0	7.0	2013.0	July	Thursday
18	17	1	1900/01/01 00:00:00.000000000	2199/12/31 23:59:59.999000000	2013/07/26 00:00:00.000000000	207.0	7.0	2013.0	July	Friday
19	18	1	1900/01/01 00:00:00.000000000	2199/12/31 23:59:59.999000000	2013/07/27 00:00:00.000000000	208.0	7.0	2013.0	July	Saturday
20	19	1	1900/01/01 00:00:00.000000000	2199/12/31 23:59:59.999000000	2013/07/28 00:00:00.000000000	209.0	7.0	2013.0	July	Sunday
21	20	1	1900/01/01 00:00:00.000000000	2199/12/31 23:59:59.999000000	2013/07/29 00:00:00.000000000	210.0	7.0	2013.0	July	Monday
22	21	1	1900/01/01 00:00:00.000000000	2199/12/31 23:59:59.999000000	2013/07/30 00:00:00.000000000	211.0	7.0	2013.0	July	Tuesday
23	22	1	1900/01/01 00:00:00.000000000	2199/12/31 23:59:59.999000000	2013/07/31 00:00:00.000000000	212.0	7.0	2013.0	July	Wednesday
24	23	1	1900/01/01 00:00:00.000000000	2199/12/31 23:59:59.999000000	2013/08/01 00:00:00.000000000	213.0	8.0	2013.0	August	Thursday
25	24	1	1900/01/01 00:00:00.000000000	2199/12/31 23:59:59.999000000	2013/08/02 00:00:00.000000000	214.0	8.0	2013.0	August	Friday
26	25	1	1900/01/01 00:00:00.000000000	2199/12/31 23:59:59.999000000	2013/08/03 00:00:00.000000000	215.0	8.0	2013.0	August	Saturday
27	26	1	1900/01/01 00:00:00.000000000	2199/12/31 23:59:59.999000000	2013/08/04 00:00:00.000000000	216.0	8.0	2013.0	August	Sunday
28	27	1	1900/01/01 00:00:00.000000000	2199/12/31 23:59:59.999000000	2013/08/05 00:00:00.000000000	217.0	8.0	2013.0	August	Monday
29	28	1	1900/01/01 00:00:00.000000000	2199/12/31 23:59:59.999000000	2013/08/06 00:00:00.000000000	218.0	8.0	2013.0	August	Tuesday
30	29	1	1900/01/01 00:00:00.000000000	2199/12/31 23:59:59.999000000	2013/08/07 00:00:00.000000000	219.0	8.0	2013.0	August	Wednesday
31	30	1	1900/01/01 00:00:00.000000000	2199/12/31 23:59:59.999000000	2013/08/08 00:00:00.000000000	220.0	8.0	2013.0	August	Thursday
32	31	1	1900/01/01 00:00:00.000000000	2199/12/31 23:59:59.999000000	2013/08/09 00:00:00.000000000	221.0	8.0	2013.0	August	Friday
33	32	1	1900/01/01 00:00:00.000000000	2199/12/31 23:59:59.999000000	2013/08/10 00:00:00.000000000	222.0	8.0	2013.0	August	Saturday

Caption: This is the time dimension table preview

Accident Incident Fact Table



#	Stepname	Copynr	Read	Written	Input	Output	Updated	Rejected	Errors	Active	Time	Speed (r/s)	input/output
1	Vehicle Collision CSV file input	0	0	394444	394445	0	0	0	0	Finished	54.4s	7,246	-
2	Accident Dim Lookup	0	394444	394444	394444	0	0	0	0	Finished	55.8s	7,072	-
3	Vehicle Dimension Lookup	0	394444	394444	394444	0	0	0	0	Finished	55.8s	7,070	-
4	Location Dimension Lookup	0	394444	394444	394444	0	0	0	0	Finished	55.8s	7,068	-
5	Time Dimension Lookup	0	394444	394444	394444	0	0	0	0	Finished	55.8s	7,066	-
6	Load Accident Incident Fact Table	0	394444	394444	0	394444	0	0	0	Finished	56.0s	7,045	-

Caption: This is the accident incident fact table transformation

#	ACCIDENT_ID	ACCIDENT_DIM_ID	VEHICLE_DIM_ID	LOCATION_DIM_ID	TIME_DIM_ID	NUMBER OF PERSONS INJURED	NUMBER OF PERSONS KILLED	NUMBER OF PEDESTRIANS INJURED	NUMBER OF PEDESTRIANS KILLED	NUMBER OF CYCLIST INJURED	NUMBER OF CYCLIST KILLED	NUMBER
1	1	1	1	1	1	0	0	0	0	0	0	
2	2	2	2	2	1	1	0	1	0	0	0	
3	3	3	3	3	1	0	0	0	0	0	0	
4	4	4	4	4	1	0	0	0	0	0	0	
5	5	5	5	5	1	0	0	0	0	0	0	
6	6	6	6	6	1	0	0	0	0	0	0	
7	7	7	7	7	1	1	0	0	0	1	0	
8	8	8	8	8	1	0	0	0	0	0	0	
9	9	9	9	9	1	0	0	0	0	0	0	
10	10	10	10	10	1	0	0	0	0	0	0	
11	11	11	11	11	1	0	0	0	0	0	0	
12	12	12	12	12	1	0	0	0	0	0	0	
13	13	13	13	13	1	2	0	1	0	0	0	
14	14	14	14	14	1	0	0	0	0	0	0	
15	15	15	15	15	1	0	0	0	0	0	0	
16	16	16	16	16	1	0	0	0	0	0	0	
17	17	17	17	17	1	0	0	0	0	0	0	
18	18	18	18	18	1	0	0	0	0	0	0	
19	19	19	19	19	1	0	0	0	0	0	0	
20	20	20	20	20	1	0	0	0	0	0	0	
21	21	21	21	21	1	0	0	0	0	0	0	
22	22	22	22	22	1	0	0	0	0	0	0	
23	23	23	23	23	1	0	0	0	0	0	0	
24	24	24	24	24	1	5	0	0	0	0	0	
25	25	25	25	25	1	0	0	0	0	0	0	
26	26	26	26	26	1	0	0	0	0	0	0	
27	27	27	27	27	1	0	0	0	0	0	0	
28	28	28	28	28	1	0	0	0	0	0	0	
29	29	29	29	29	1	0	0	0	0	0	0	
30	30	30	30	30	1	0	0	0	0	0	0	
31	31	31	31	31	1	0	0	0	0	0	0	
32	32	32	32	32	1	0	0	0	0	0	0	
33	33	33	33	33	1	0	0	0	0	0	0	
34	34	34	34	34	1	3	0	0	0	0	0	
35	35	35	35	35	1	0	0	0	0	0	0	
36	36	36	36	36	1	0	0	0	0	0	0	
37	37	37	37	37	1	3	0	0	0	0	0	
38	38	38	38	38	1	0	0	0	0	0	0	
39	39	39	39	39	1	0	0	0	0	0	0	
40	40	40	40	40	1	0	0	0	0	0	0	

Caption: This is the accident incident fact table preview

Weather Condition Fact Table

Execution Results

Execution History | Logging | Step Metrics | Performance Graph | Metrics | Preview data

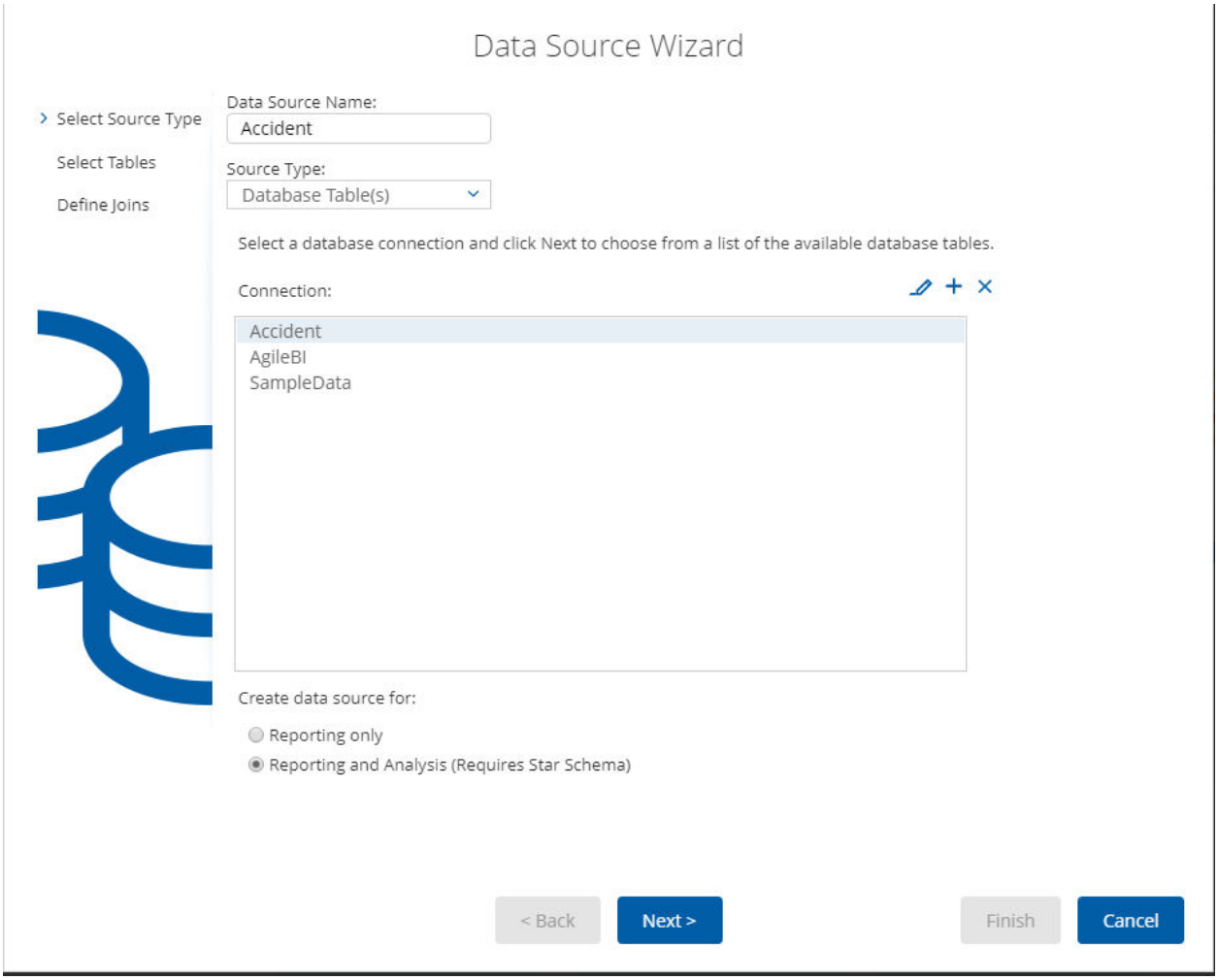
#	Stepname	Copynr	Read	Written	Input	Output	Updated	Rejected	Errors	Active	Time	Speed (r/s)	input/output
1	Weather CSV file input	0	0	905	906	0	0	0	0	Finished	0.0s	181,200	-
2	Time Dimension Lookup	0	905	905	905	0	0	0	0	Finished	0.2s	5,801	-
3	Weather Dimension Lookup	0	905	905	905	0	0	0	0	Finished	0.2s	5,231	-
4	Location Dimension Lookup	0	905	905	905	0	0	0	0	Finished	0.2s	4,840	-
5	Load Weather Condition Fact Table	0	905	905	0	905	0	0	0	Finished	0.3s	2,967	-

Caption: This is the weather condition fact table transformation

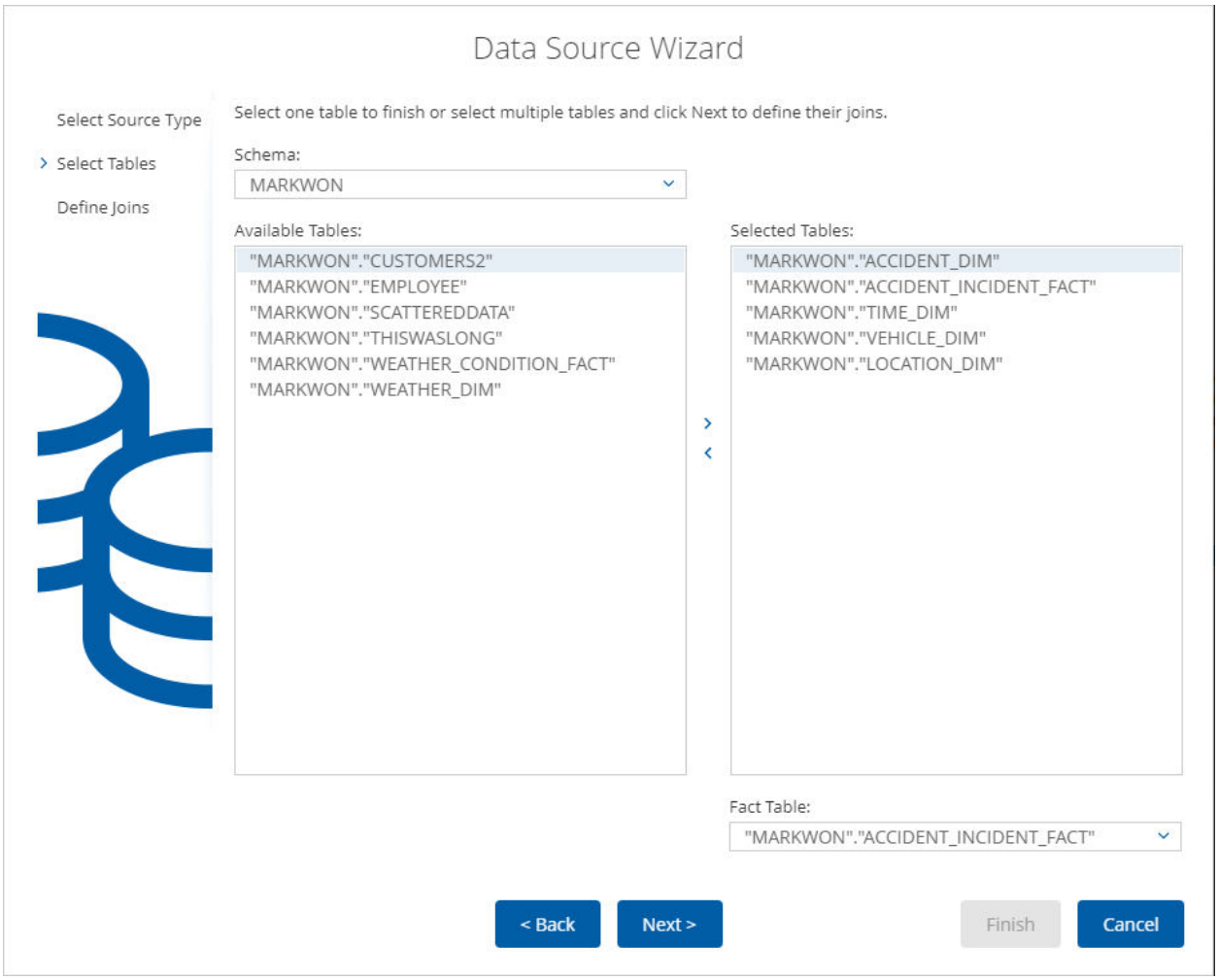
#	WEATHERID	WEATHER_DIM_ID	TIME_DIM_ID	LOCATION_DIM_ID	Mean.TemperatureF	PRECIPITATIONIN	Mean.VisibilityMiles	Mean.Wind.SpeedMPH
1	1	0	1	1	79	0.1	9	9
2	2	0	2	1	81	0	10	7
3	3	0	3	1	75	0.3	8	12
4	4	0	4	1	78	0	8	8
5	5	0	5	1	81	0	9	11
6	6	0	6	1	87	0	10	7
7	7	0	7	1	87	0	10	11
8	8	0	8	1	87	0	10	7
9	9	0	9	1	89	0	10	9
10	10	0	10	1	88	0	8	12
11	11	0	11	1	87	0	9	14
12	12	0	12	1	82	0	10	8
13	13	0	13	1	80	0.1	9	8
14	14	0	14	1	80	0.4	9	11
15	15	0	15	1	79	0	10	11
16	16	0	16	1	67	0	10	13
17	17	0	17	1	74	0	10	11
18	18	0	18	1	75	0	10	8
19	19	0	19	1	73	0.1	9	8
20	20	0	20	1	78	0	10	9
21	21	0	21	1	76	0	10	9
22	22	0	22	1	75	0	10	8
23	23	0	23	1	73	0.7	9	10
24	24	0	24	1	77	0	10	10
25	25	0	25	1	74	0	10	11
26	26	0	26	1	75	0	10	11
27	27	0	27	1	72	0	10	11
28	28	0	28	1	72	0	10	9
29	29	0	29	1	75	0	10	11
30	30	0	30	1	76	0.7	9	14
31	31	0	31	1	78	0	7	13
32	32	0	32	1	78	0	10	10
33	33	0	33	1	74	0	10	8
34	34	0	34	1	75	0.5	9	5
35	35	0	35	1	73	0.7	7	6
36	36	0	36	1	69	0	10	17
37	37	0	37	1	69	0	10	10
38	38	0	38	1	71	0	10	6
39	39	0	39	1	72	0	10	6
40	40	0	40	1	72	0	10	4
41	41	0	41	1	72	0	10	11

Caption: This is the weather condition fact table preview

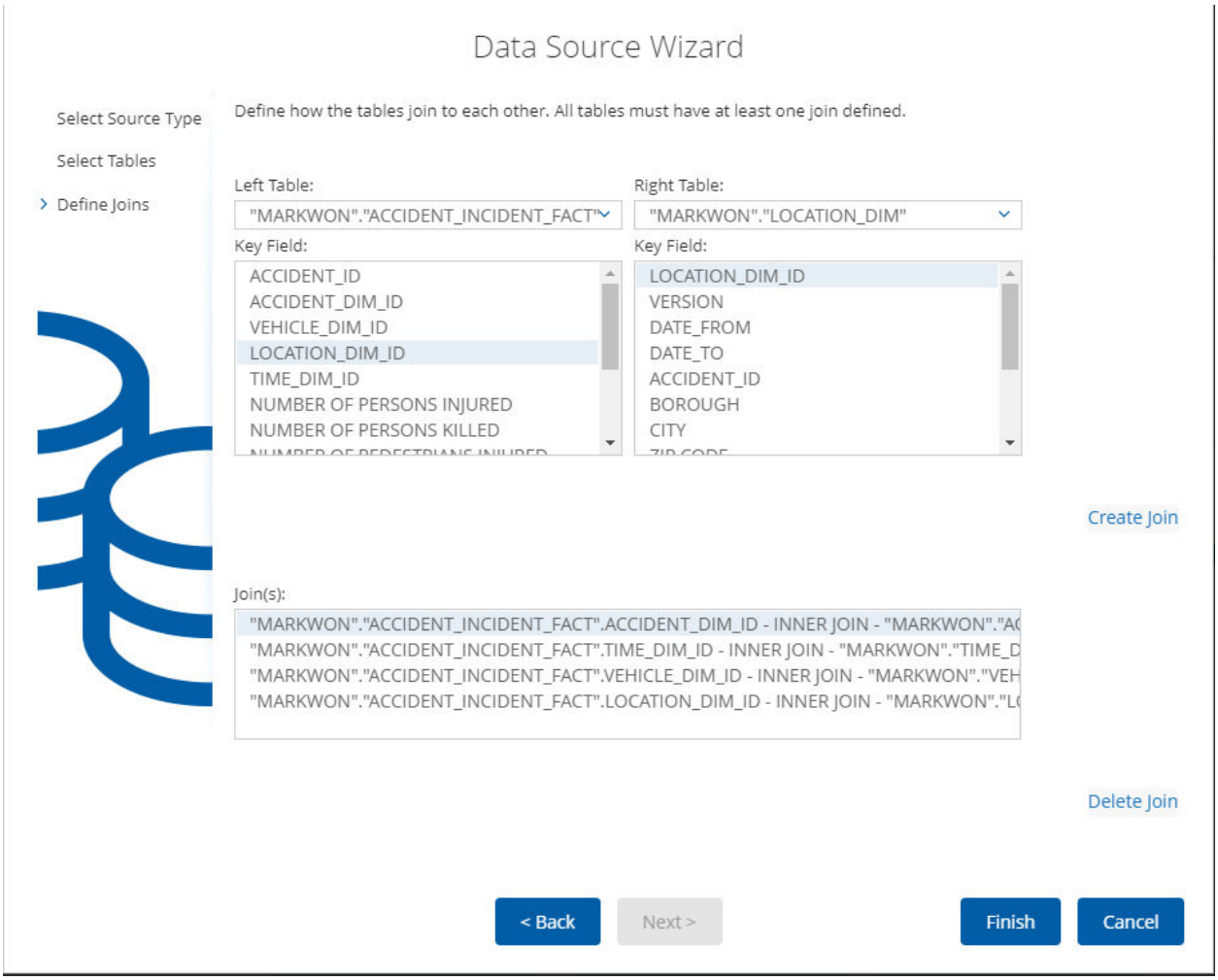
Final Schema



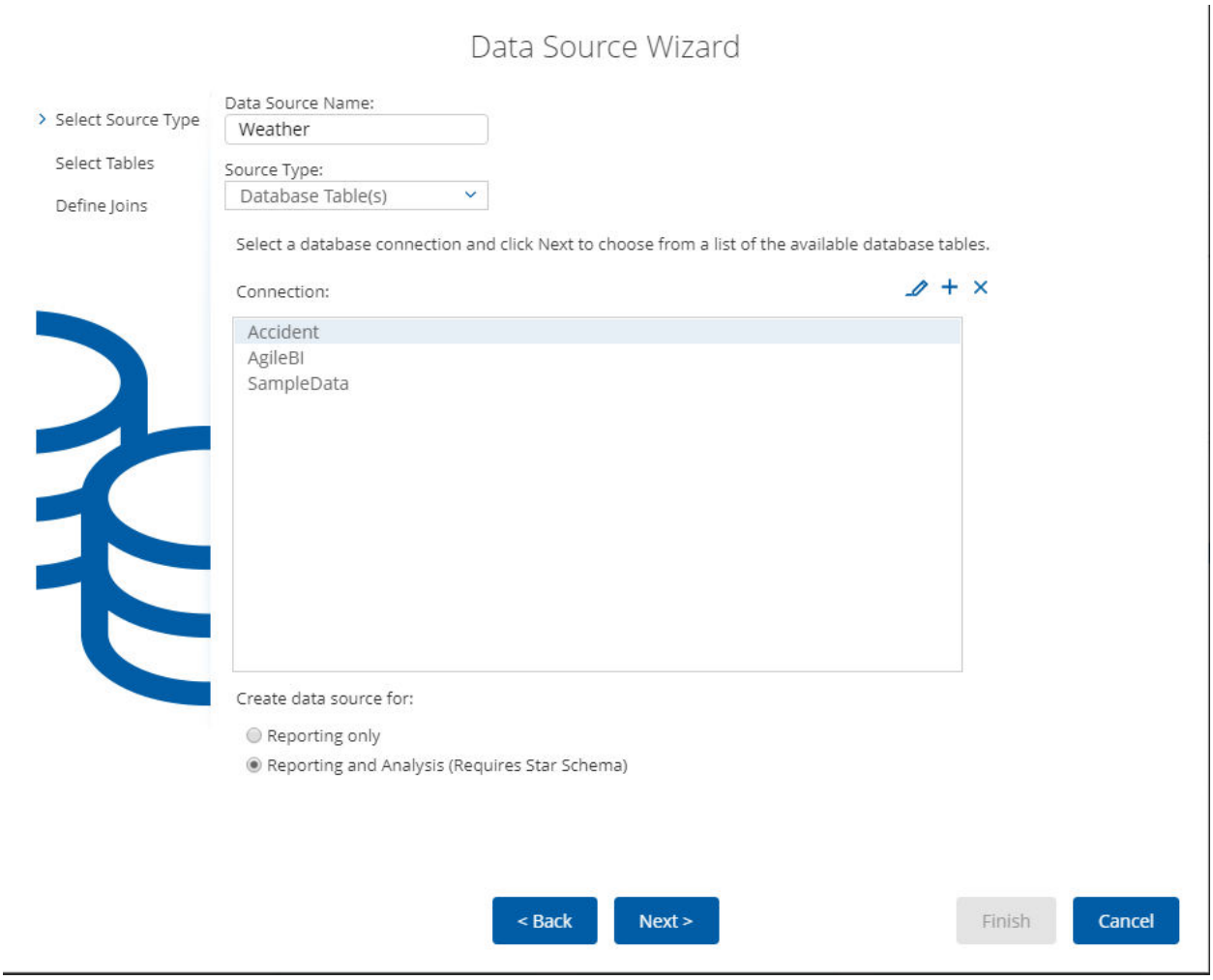
Caption: Starting the creation of the Accident Incident Fact Cube



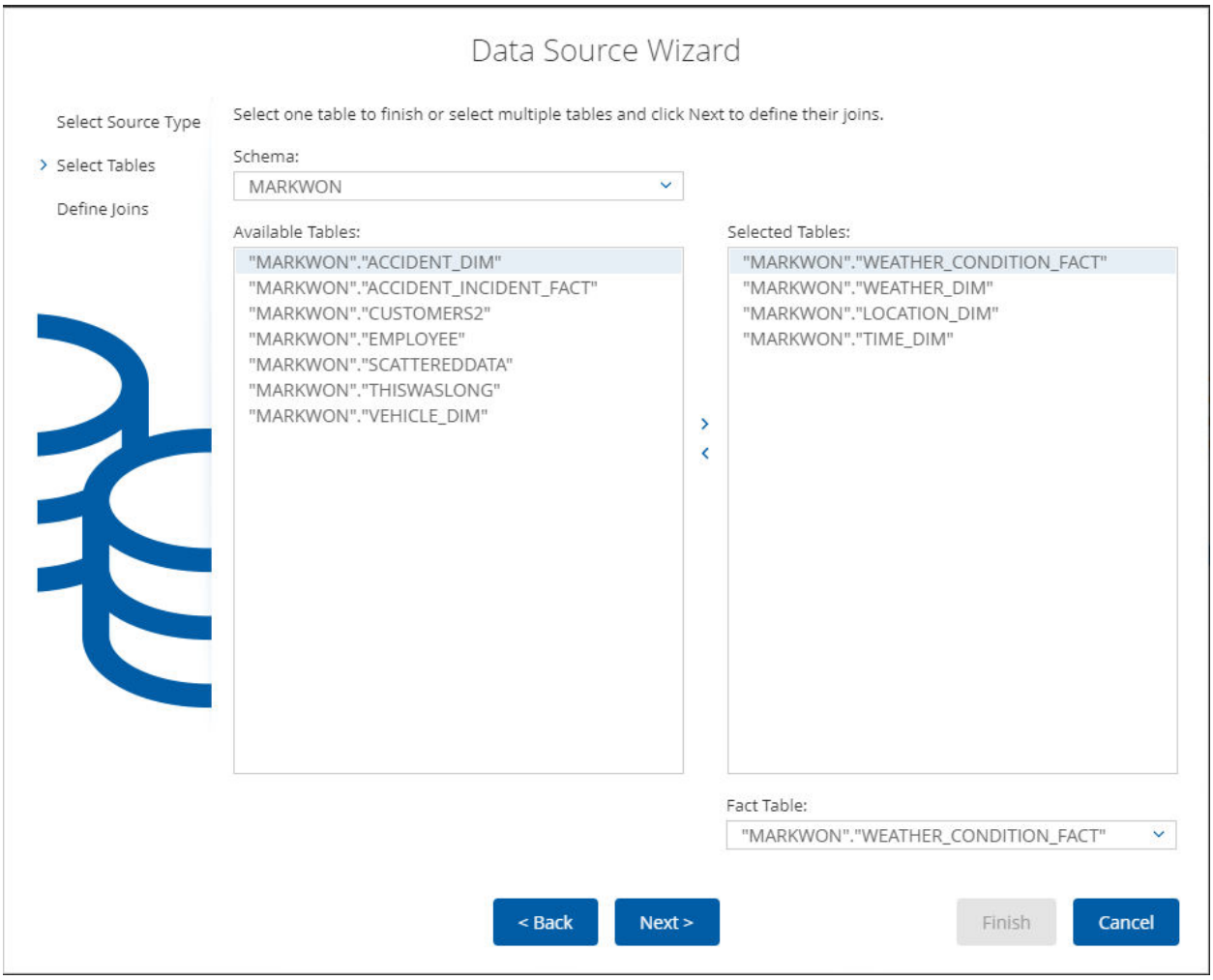
Caption: Selecting Tables Relating to the Accident Incident Fact Cube



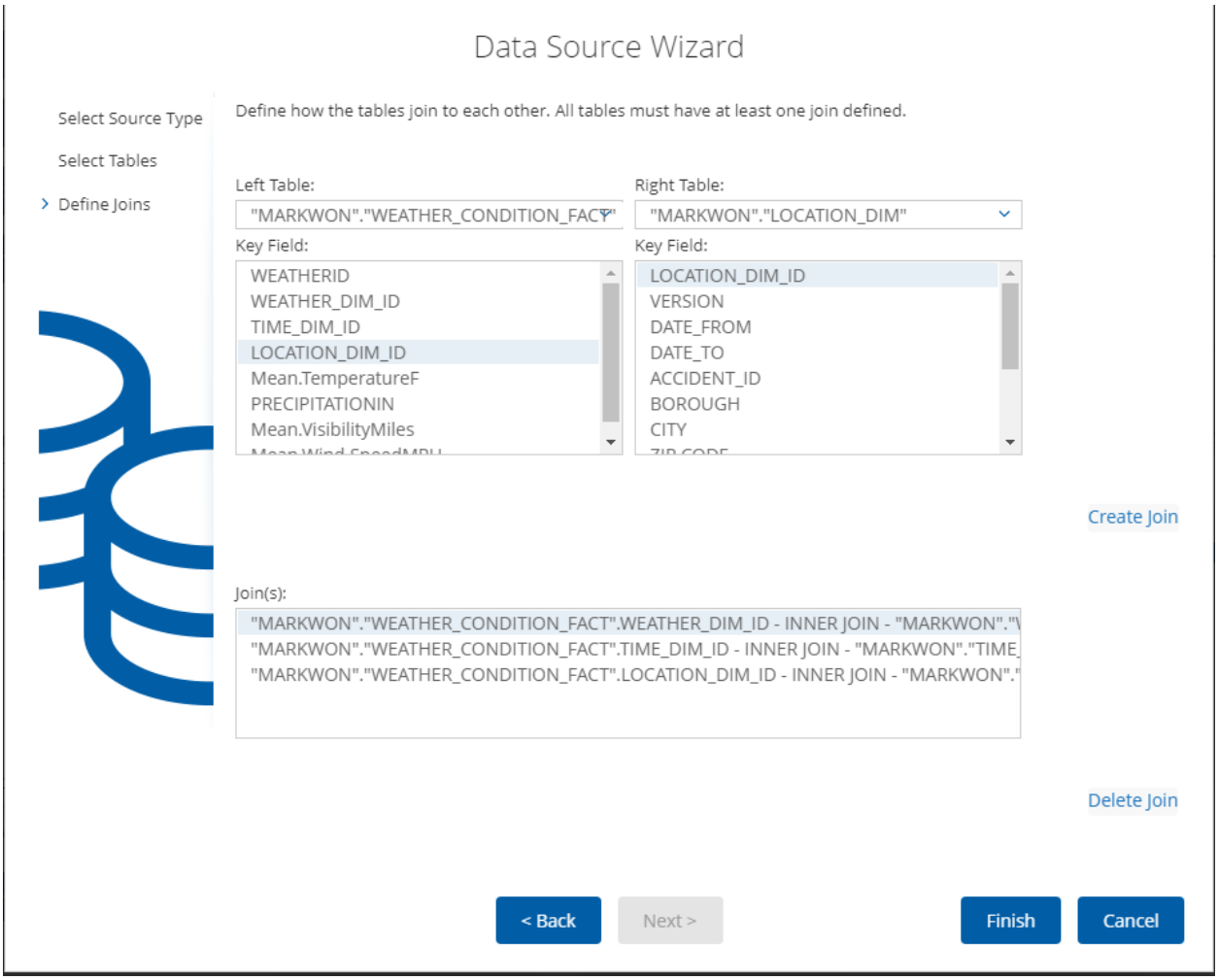
Caption: Creating the Join statements to complete the cube



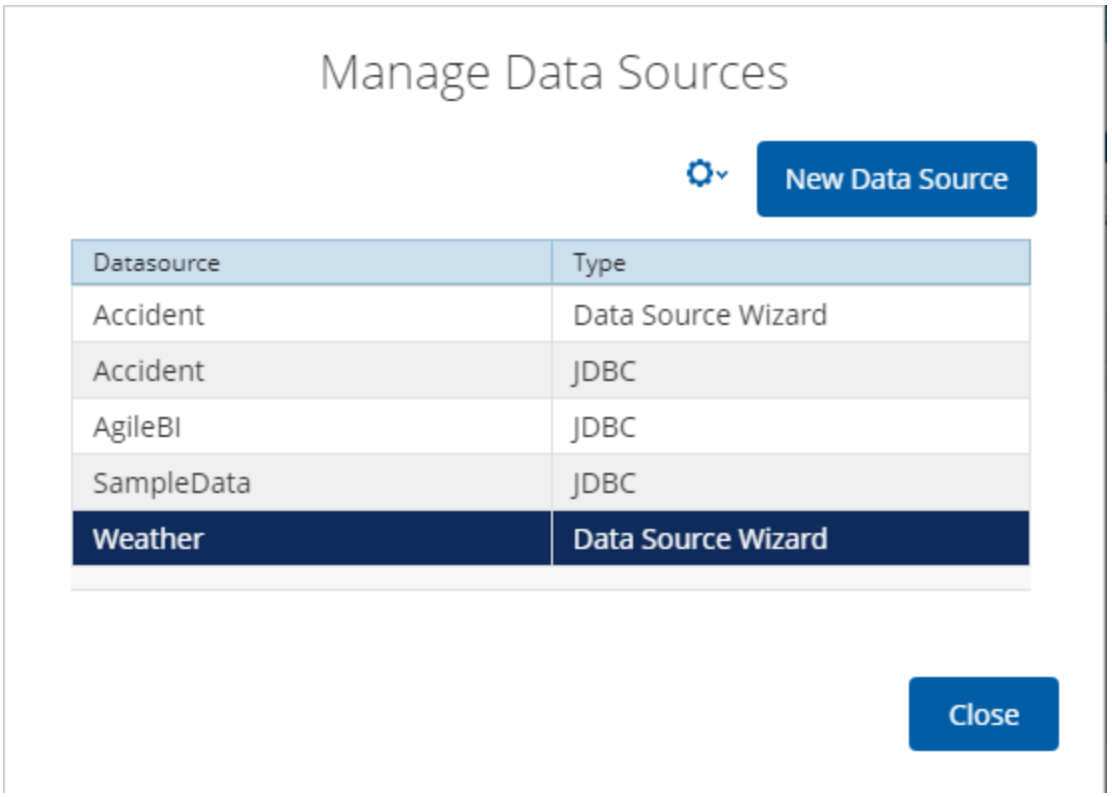
Caption: Starting the creation of the Weather Condition Fact Cube



Caption: Selecting Tables Relating to the Weather Condition Fact Cube



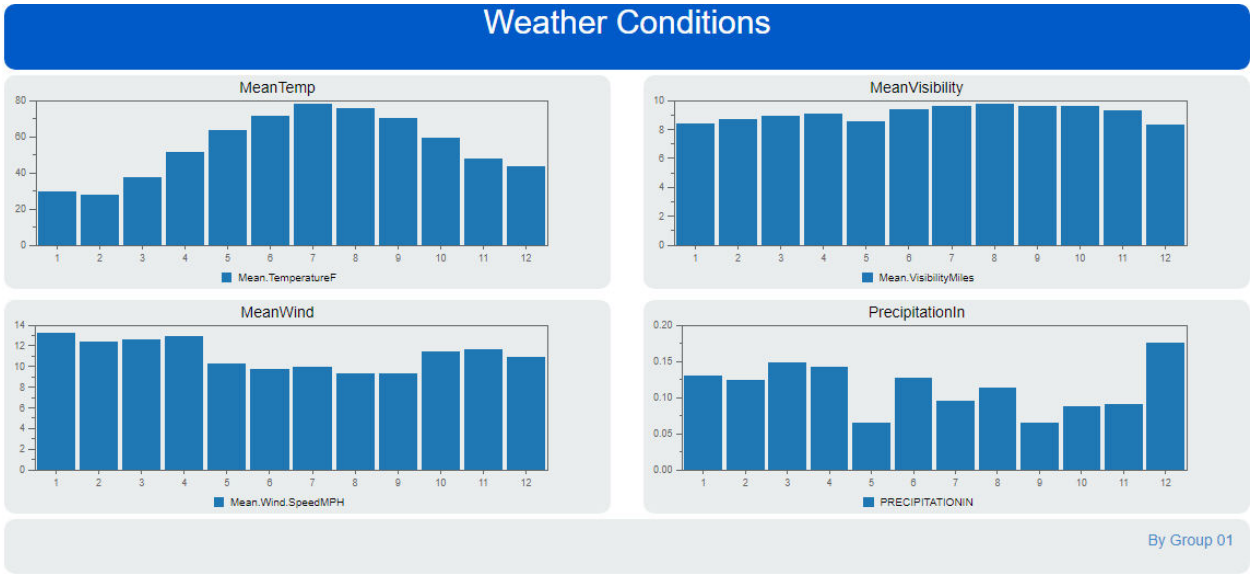
Caption: Creating the Join statements to complete the cube



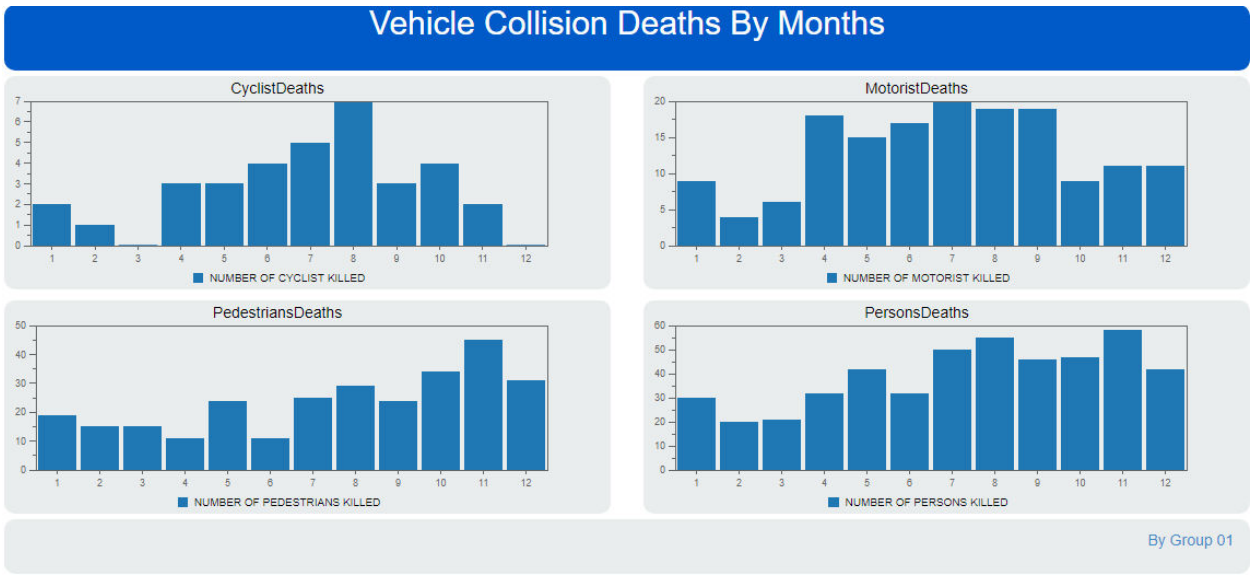
Caption: Weather and Accident Cube created and Accident Connection made

We used the Pentaho CDE to create our cubes and used the oracle connection to connect to the CDE.

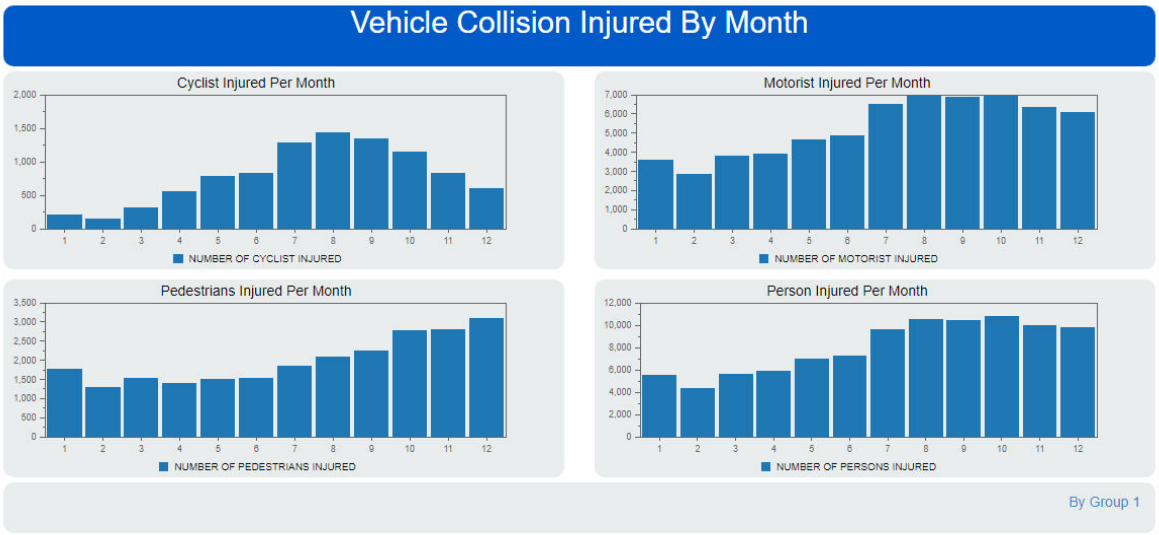
Dashboard Application



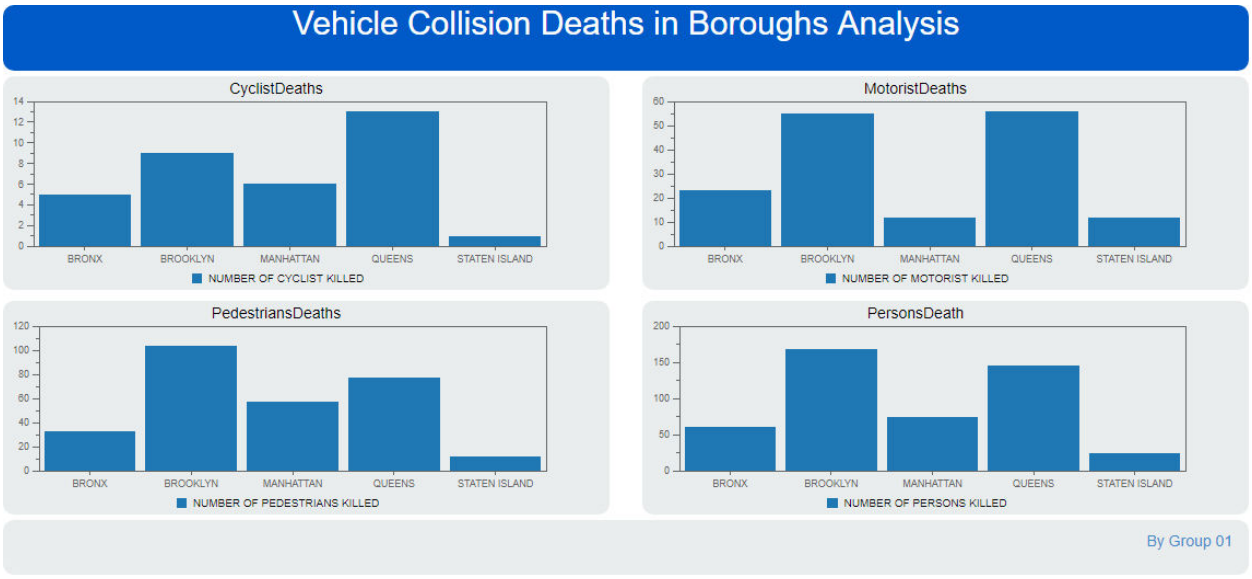
Caption: This shows the average temperature, visibility, wind speed, and precipitation per month from 2013-2015



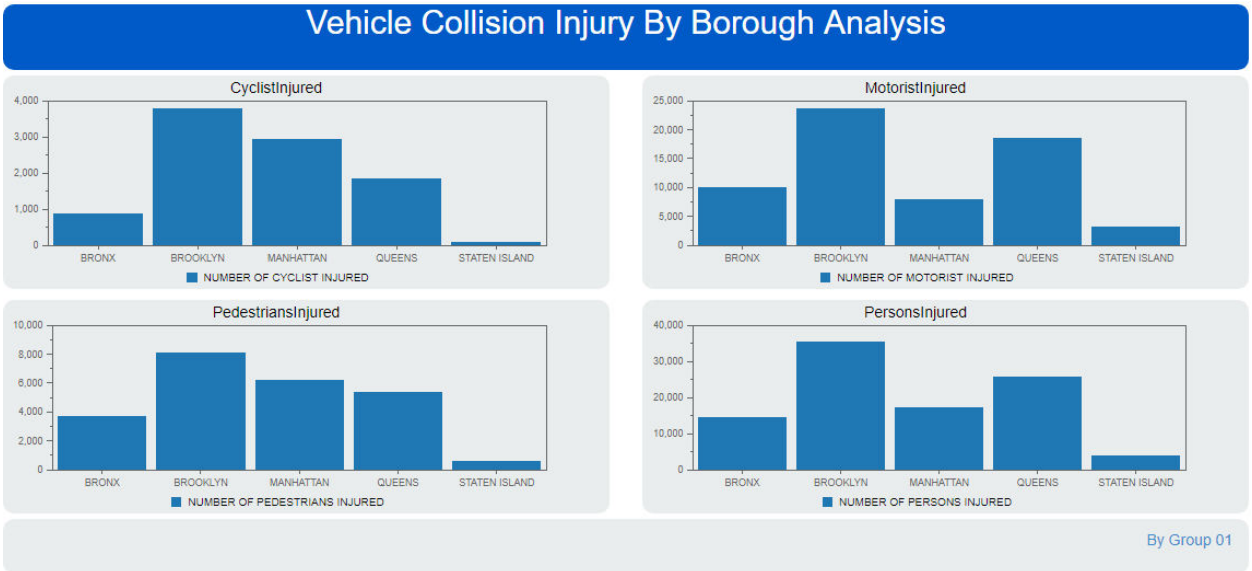
Caption: This shows the sums of the vehicle collisions that resulted in deaths for cyclists, motorists, and pedestrians per month from 2013-2015



Caption: This shows the sums of the vehicle collisions that resulted in injuries for cyclists, motorists, and pedestrians per month from 2013-2015



Caption: This shows the sums of vehicle collisions that resulted in deaths for cyclists, motorists, and pedestrians by boroughs from 2013-2015



Caption: This shows the sums of vehicle collisions that resulted in injuries for cyclists, motorists, and pedestrians by boroughs from 2013-2015

Analysis

By comparing the weather dashboard with the vehicle collision dashboards, we were able to find a specific correlation between the two datasets. For example there is a positive correlation between mean temperature and vehicle collisions that resulted in deaths for both cyclists and motorists (Higher mean temperature resulted in more frequent deaths). Another analysis shows that there is a negative correlation between mean wind speed and vehicle collision that resulted in deaths for motorists. This shows that on days with a high mean wind speed, motorists should ride with caution.

When analyzing the deaths in boroughs, we found that most vehicle collision deaths occured in Brooklyn and Queens while Staten Island had the lowest death occurrence. When picking a place to live in New York City, it may be a good idea for those with a high priority in vehicle safety to live in Staten Island to minimize your chance of death in vehicular accidents.

When analyzing the injuries by boroughs, we found that Brooklyn had the highest number of accidents and Staten Island had the least. As noted in the death analysis, Staten Island appears to be the safest borough in NYC and would require the least government attention while Brooklyn should have a heavier focus on traffic safety and control.