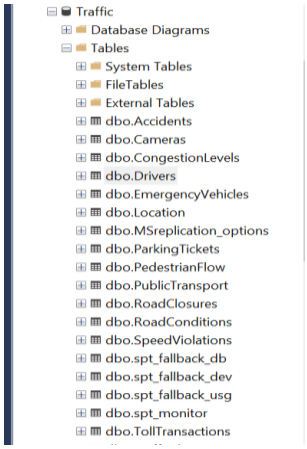
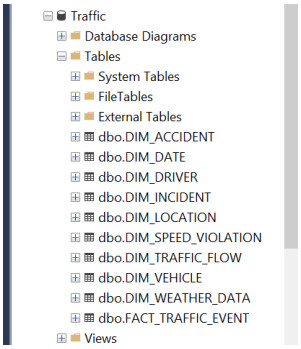
ETL IMPLEMENTATION USING PYTHON AND SQL SERVER MANAGEMENT STUDIO

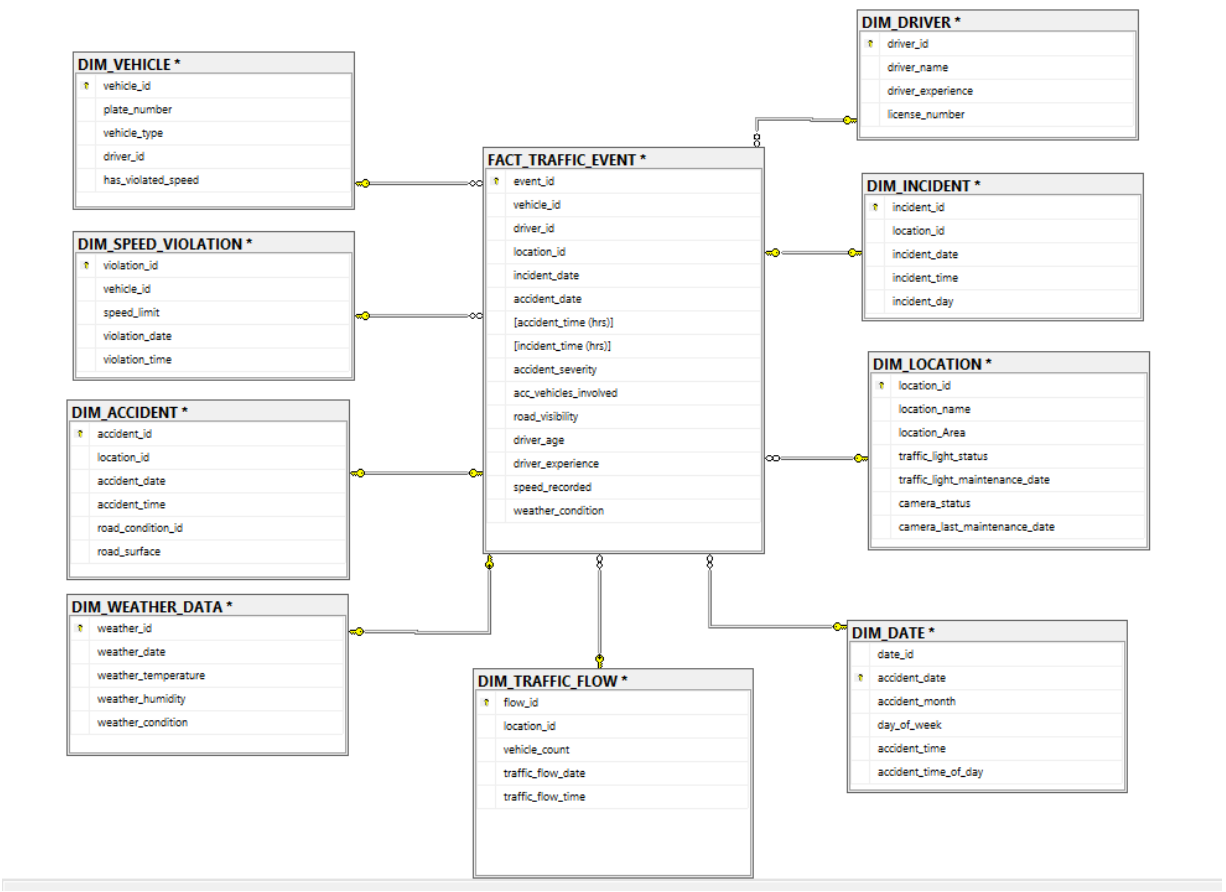




DBMS BEFORE ETL IMPLEMENTATION

RDBMS AFTER ETL IMPLEMENTATION

STAR SCHEMA IMAGE (SSMS)



LIST OF SCREEN SHOTS FOR DIMENSIONS AND FACT TABLE

1. DIM VEHICLE

	vehicle_id	plate_number	vehicle_type	driver_id	has_violated_speed
1	1	92X 5338	Car	64	1
2	2	186-ZJJ	Car	145	0
3	3	835 1WK	Car	140	1
4	4	OYD7711	Bus	70	0
5	5	KFU-0955	Bus	24	0
6	6	41T•740	Motorcycle	43	1
7	7	93P L10	Motorcycle	66	0
8	8	0N H8535	Car	5	0
9	9	Q81 4GN	Truck	166	0
10	10	049-PVL	Truck	142	1

2. DIM SPEEDVIOLATION

	violation_id	vehicle_id	speed_limit	violation_date	violation_time
1	1	179	80	2025-02-12	16:54:07.0000000
2	2	41	80	2025-02-23	02:07:01.0000000
3	3	169	70	2025-03-03	09:24:10.0000000
4	4	6	70	2025-02-03	03:45:09.0000000
5	5	106	70	2025-02-14	06:10:16.0000000
6	6	67	70	2025-01-06	10:42:22.0000000
7	7	160	90	2025-01-24	00:45:49.0000000
8	8	36	50	2025-01-04	00:30:56.0000000
9	9	10	60	2025-02-24	09:56:41.0000000
10	10	43	80	2025-02-19	01:43:15.0000000

3. DIM DRIVER

	driver_id	driver_name	driver_experience	license_number
1	1	Muhammad Ahmed	42	b76c97e7-f
2	2	Ali Hassan	10	ffba8d24-2
3	3	Kamran Siddiqui	42	0f2ecd15-8
4	4	Bilal Iqbal	43	bd1c6771-6
5	5	Usman Malik	41	a3f98c47-1
6	6	Farhan Ahmed	15	7a4a9952-8
7	7	Imran Raza	30	30ce16ac-e
8	8	Asif Iqbal	4	85e4e983-1
9	9	Zubair Sheikh	25	bd555a70-1
10	10	Adnan Khan	2	5c2817ed-7

4. DIM INCIDENT

	incident_id	location_id	incident_date	incident_time	incident_day
1	1	Shahrah-e-Faisal	2025-02-15	14:46:00.0000000	Saturday
2	2	MA Jinnah Road	2025-02-23	18:29:41.0000000	Sunday
3	3	II Chundrigar Road	2025-01-15	23:55:50.0000000	Wednesday
4	4	University Road	2025-02-05	22:04:32.0000000	Wednesday
5	5	Rashid Minhas Road	2025-01-15	19:18:05.0000000	Wednesday
6	6	Shahrah-e-Pakistan	2025-01-20	00:55:24.0000000	Monday
7	7	Korangi Road	2025-01-30	07:56:37.0000000	Thursday
8	8	Shahrah-e-Quaideen	2025-01-31	08:08:58.0000000	Friday
9	9	Shaheed-e-Millat Road	2025-03-03	00:41:46.0000000	Monday
10	10	Shahrah-e-Gulberg	2025-02-17	06:04:21.0000000	Monday

5. DIM LOCATION

	location_id	location_name	location_Area	traffic_light_status	traffic_light_maintenance_date	camera_status	camera_last_maintenance_date
1	1	Shahrah-e-Faisal	Jinnah International Airport to PIDC	Red	2025-02-04 01:50:11.000	Inactive	2025-02-14 00:00:00.000
2	2	MA Jinnah Road	Saddar to Kharadar	Green	2025-01-10 23:37:06.000	Active	2025-02-10 00:00:00.000
3	3	II Chundrigar Road	Tower to Merewether Tower	Red	2025-01-21 11:05:02.000	Inactive	2025-02-11 00:00:00.000
4	4	University Road	NED University to Hassan Square	Red	2025-01-03 00:17:25.000	Active	2025-02-21 00:00:00.000
5	5	Rashid Minhas Road	Gulshan to Gulistan-e-Jauhar	Green	2025-01-24 07:01:55.000	Active	2025-02-04 00:00:00.000
6	6	Shahrah-e-Pakistan	Sohrab Goth to Liaquatabad	Green	2025-02-26 18:28:13.000	Inactive	2025-02-08 00:00:00.000
7	7	Korangi Road	Korangi Crossing to Qayyumabad	Yellow	2025-01-02 06:23:28.000	Inactive	2025-02-04 00:00:00.000
8	8	Shahrah-e-Quaideen	Clifton to NIPA	Green	2025-01-09 19:50:50.000	Active	2025-01-07 00:00:00.000
9	9	Shaheed-e-Millat Road	Nursery to KDA Chowrangi	Green	2025-01-28 12:23:45.000	Active	2025-01-23 00:00:00.000
10	10	Sunset Boulevard	DHA Phase 2	Yellow	2025-02-04 00:19:38.000	Active	2025-02-03 00:00:00.000

6. DIM ACCIDENT

	accident_id	location_id	accident_date	accident_time	road_condition_id	road_surface
1	1	Shahrah-e-Bandar Road	2025-02-20	15:15:34.00...	32	Icy
2	2	Shahrah-e-Gulberg	2025-01-07	13:14:37.00...	79	Icy
3	3	Miran Shah Road	2025-02-03	15:03:38.00...	92	Wet
4	4	Shahrah-e-PECHS	2025-02-19	01:05:37.00...	83	Wet
5	5	Sharae Quaid-e-Azam	2025-02-15	15:15:27.00...	7	Snowy
6	6	Shahrah-e-Manora	2025-01-06	20:48:02.00...	16	Snowy
7	7	Shahrah-e-Saeedabad	2025-02-01	01:52:06.00...	90	Dry
8	8	Shahrah-e-Ibn-e-Sina	2025-02-01	17:34:58.00...	26	Icy
9	9	Shahrah-e-Bahadurabad	2025-01-13	14:28:48.00...	40	Wet
10	10	Khayaban-e-Shamsheer	2025-02-24	02:57:45.00...	73	Snowy

7. DIM DATE

	date_id	accident_date	accident_month	day_of_week	accident_time	accident_time_of_day
1	2025-01-02 00:00:00.000	2025-01-02	January	Thursday	11:15:13.0000000	Morning
2	2025-01-03 00:00:00.000	2025-01-03	January	Friday	12:18:56.0000000	Afternoon
3	2025-01-05 00:00:00.000	2025-01-05	January	Sunday	21:58:43.0000000	Evening
4	2025-01-06 00:00:00.000	2025-01-06	January	Monday	20:48:02.0000000	Evening
5	2025-01-07 00:00:00.000	2025-01-07	January	Tuesday	13:14:37.0000000	Afternoon
6	2025-01-08 00:00:00.000	2025-01-08	January	Wednesday	11:18:38.0000000	Morning
7	2025-01-09 00:00:00.000	2025-01-09	January	Thursday	03:09:53.0000000	Night
8	2025-01-10 00:00:00.000	2025-01-10	January	Friday	18:31:59.0000000	Evening
9	2025-01-13 00:00:00.000	2025-01-13	January	Monday	14:28:48.0000000	Afternoon
10	2025-01-14 00:00:00.000	2025-01-14	January	Tuesday	15:52:30.0000000	Afternoon

8. DIM WEATHER DATA

	weather_id	weather_date	weather_temperature	weather_humidity	weather_condition
1	1	2025-01-08 05:19:23.000	14.7	40	Cloudy
2	2	2025-02-22 11:50:40.000	36.3	48	Cloudy
3	3	2025-02-20 05:23:27.000	9.8	59	Cloudy
4	4	2025-01-09 09:43:52.000	8.5	41	Sunny
5	5	2025-01-29 17:16:29.000	25.2	27	Cloudy
6	6	2025-02-15 12:51:23.000	7.4	57	Sunny
7	7	2025-02-13 23:39:08.000	32.5	31	Stormy
8	8	2025-02-13 06:54:31.000	25.7	56	Rainy
9	9	2025-01-28 22:01:41.000	21.4	82	Cloudy
10	10	2025-01-03 13:59:53.000	24	90	Rainy

9. DIM TRAFFIC FLOW

	flow_id	location_id	vehicle_count	traffic_flow_date	traffic_flow_time
1	1	1	418	2025-02-15	22:08:31.0000000
2	2	2	992	2025-02-05	18:12:09.0000000
3	3	3	793	2025-02-17	07:56:45.0000000
4	4	4	804	2025-02-28	05:24:57.0000000
5	5	5	190	2025-02-26	03:51:13.0000000
6	6	6	836	2025-01-29	13:06:11.0000000
7	7	7	735	2025-01-17	16:34:50.0000000
8	8	8	415	2025-01-28	03:56:05.0000000
9	9	9	510	2025-01-15	04:42:43.0000000
10	10	10	295	2025-01-06	02:52:23.0000000

10. FACT TRAFFIC EVENT

	event_id	vehicle_id	driver_id	location_id	incident_date	accident_date	accident_time (hrs)	incident_time (hrs)	accident_severity	acc_vehicles_involved	road_visibility	driver_age	driver_experience	speed_recorded	weather_condition
1	1	1	1	1	2025-02-15	2025-02-20	15:15:34.0000000	14:46:00.0000000	Minor	3	Clear	56	42	140	Cloudy
2	2	2	2	2	2025-02-23	2025-01-07	13:14:37.0000000	18:29:41.0000000	Minor	2	Foggy	24	10	105	Cloudy
3	3	3	3	3	2025-01-15	2025-02-03	15:03:38.0000000	23:55:50.0000000	Severe	3	Snowy	25	42	92	Cloudy
4	4	4	4	4	2025-02-05	2025-02-19	01:05:37.0000000	22:04:32.0000000	Minor	1	Clear	42	43	120	Sunny
5	5	5	5	5	2025-01-15	2025-02-15	15:15:27.0000000	19:18:05.0000000	Severe	5	Snowy	62	41	113	Cloudy
6	6	6	6	6	2025-01-20	2025-01-06	20:48:02.0000000	00:55:24.0000000	Minor	1	Foggy	60	15	134	Sunny
7	7	7	7	7	2025-01-30	2025-02-01	01:52:06.0000000	07:56:37.0000000	Moderate	5	Clear	58	30	124	Stormy
8	8	8	8	8	2025-01-31	2025-02-01	17:34:58.0000000	08:08:58.0000000	Severe	1	Foggy	59	4	103	Rainy
9	9	9	9	9	2025-03-03	2025-01-13	14:28:48.0000000	00:41:46.0000000	Moderate	2	Clear	26	25	65	Cloudy
10	10	10	10	10	2025-02-17	2025-02-24	02:57:45.0000000	06:04:21.0000000	Minor	4	Foggy	56	2	107	Rainy

SQL QUERY 1. Identify the top 10 location areas with the highest traffic flow between February and March 2025, and analyze the correlation with road conditions (e.g., wet, icy, snowy) and number of accidents.

```sql
SELECT TOP 10
    dl.location_Area,
    SUM(tf.vehicle_count) AS TotalVehicles
FROM
    dbo.dim_traffic_flow tf
JOIN
    dbo.dim_location dl ON tf.location_id = dl.location_id
WHERE
    tf.traffic_flow_date >= '2025-02-01' AND tf.traffic_flow_date < '2025-03-01'
GROUP BY
    dl.location_Area
ORDER BY
    TotalVehicles DESC;
```

90 %

Results | Messages

	location_Area	TotalVehicles
1	Gulshan-e-Iqbal	6138
2	Gulistan-e-Jauhar	5043
3	Saddar	2898
4	Tower to Merewether Tower	2532
5	Jinnah International Airport to PIDC	1752
6	North Karachi	1565
7	PECHS	1305
8	Nazimabad	1223
9	NED University to Hassan Square	1215
10	DHA/Clifton to Gadap	1147

Here we see the areas with highest traffic flow also have high accident rates.

Direct co relation observed between *traffic flow* and *accidents* influenced by *road surface* condition.

```sql
SELECT TOP 10 -- To limit the result to the top 10 locations
    dl.location_Area,da.road_surface,
    COUNT(da.accident_id) AS AccidentCount
FROM
    dbo.dim_accident da
JOIN
    dbo.dim_location dl ON da.location_id = dl.location_name
GROUP BY
    dl.location_Area,da.road_surface
ORDER BY
    AccidentCount DESC;
```

74 %

Results | Messages

	location_Area	road_surface	AccidentCount
1	Gulshan-e-Iqbal	Icy	7
2	Gulshan-e-Iqbal	Wet	5
3	Gulshan-e-Iqbal	Snowy	3
4	Saddar	Snowy	3
5	Malir	Dry	2
6	DHA Phase 6	Dry	2
7	North Nazimabad	Dry	2
8	Korangi	Icy	2
9	DHA Phase 5	Snowy	2
10	Gulistan-e-Jauhar	Snowy	2

SQL QUERY 2. What are the average speed violations recorded for each vehicle type (Car, Bus, Truck, Motorcycle) during peak hours (7-9 AM and 4-6 PM) in February 2025.

```
SELECT
    dv.vehicle_type,
    AVG(CASE
        WHEN DATEPART(hour, dsv.violation_time) BETWEEN 7 AND 9
          OR DATEPART(hour, dsv.violation_time) BETWEEN 16 AND 18
        THEN dsv.speed_limit
        ELSE NULL
    END) AS AvgSpeedViolationsDuringPeakHours
FROM
    dbo.dim_speed_violation dsv
JOIN
    dbo.dim_vehicle dv ON dsv.vehicle_id = dv.vehicle_id
GROUP BY
    dv.vehicle_type
ORDER BY AvgSpeedViolationsDuringPeakHours DESC;
```

90 %

Results | Messages

	vehicle_type	AvgSpeedViolationsDuringPeakHours
1	Motorcycle	78
2	Truck	75
3	Bus	73
4	Car	66

SQL QUERY 3. What is the incident frequency at locations (e.g., Shahrah-e-Faisal, University Road) from January 2025 onwards.

```
SELECT
    da.location_id,
    dl.location_name,
    COUNT(da.incident_id) AS IncidentFrequency
FROM
    dbo.dim_incident da
JOIN
    dbo.dim_location dl ON da.location_id = dl.location_name
WHERE
    dl.location_name IN ('University Road','Shahrah-e-Faisal')
    AND da.incident_date >= '2025-01-01'

GROUP BY
    da.location_id,
    dl.location_name
ORDER BY
    IncidentFrequency DESC;
```

89 %

Results | Messages

	location_id	location_name	IncidentFrequency
1	Shahrah-e-Faisal	Shahrah-e-Faisal	2
2	University Road	University Road	2

SQL QUERY 4. Calculate the percentage of inactive cameras at locations with the highest traffic flow (Top 5) and assess its impact in February 2025.

```
WITH TopLocations AS (
    -- 1. Find Top 5 Location Areas by Traffic Flow in February 2025
    SELECT TOP 10
        dl.location_Area,
        SUM(tf.vehicle_count) AS TotalTraffic
    FROM
        dbo.dim_traffic_flow tf
    JOIN
        dbo.dim_location dl ON tf.location_id = dl.location_id
    WHERE
        tf.traffic_flow_date >= '2025-02-01' AND tf.traffic_flow_date < '2025-03-01'
    GROUP BY
        dl.location_Area
    ORDER BY
        TotalTraffic DESC
),
CameraStatus AS (
    -- 2. Get Camera Status for each Location Area
    SELECT
        tl.location_Area,
        CASE
            WHEN dl.camera_status = 'Inactive' THEN 1
            ELSE 0
        END AS IsInactive
    FROM
        TopLocations tl
    JOIN
        dbo.dim_location dl ON tl.location_Area = dl.location_Area
)
-- 3. Calculate Percentage of Inactive Cameras by Location Area
SELECT
    location_Area,
    (SUM(IsInactive * 1.0) / COUNT(*)) * 100 AS PercentageInactive
FROM
    CameraStatus
GROUP BY
    location_Area
ORDER BY
    PercentageInactive DESC;
```

Results | Messages

	location_Area	PercentageInactive
1	Tower to Merewether Tower	100.000000
2	Jinnah International Airport to PIDC	100.000000
3	PECHS	100.000000
4	DHA/Clifton to Gadap	100.000000
5	Nazimabad	75.000000
6	Gulistan-e-Jauhar	75.000000
7	Gulshan-e-Iqbal	42.857100
8	Saddar	20.000000
9	North Karachi	0.000000
10	NED University to Hassan Square	0.000000

4 highly traffic oriented sites have 100 % of inactive cameras.

SQL QUERY 5. Identify drivers aged under 25 or younger with the highest number of speed violations, and analyze their associated vehicle types (e.g., Motorcycle, Car) to prioritize targeted safety intervention.

```
SELECT
fte.driver_age,dv.vehicle_type,  COUNT(dv.vehicle_id) AS ViolationCount
FROM
    [dbo].[DIM_VEHICLE] dv
JOIN
     [dbo].[FACT_TRAFFIC_EVENT] fte ON fte.vehicle_id = dv.vehicle_id
    WHERE fte.driver_age<25 and dv.has_violated_speed=1

GROUP BY
    fte.driver_age,dv.vehicle_type
ORDER BY
    ViolationCount DESC;
```

89 %

Results | Messages

	driver_age	vehicle_type	ViolationCount
1	22	Motorcycle	2
2	18	Truck	2
3	20	Truck	1
4	19	Bus	1
5	22	Bus	1
6	23	Bus	1
7	19	Car	1
8	23	Car	1

We observe a balanced proportion of associated vehicle type used in speed violations for drivers under the age of 25 with truck and bus leading with 3 violations.