



# New York City Parking Tickets

# Data Source

From NYC OpenData

Fiscal Year 2014 :

<https://data.cityofnewyork.us/City-Government/Parking-Violations-Issued-Fiscal-Year-2014/jt7v-77mi>

Fiscal Year 2015:

<https://data.cityofnewyork.us/City-Government/Parking-Violations-Issued-Fiscal-Year-2015/c284-tqph>

Fiscal Year 2016:

<https://data.cityofnewyork.us/City-Government/Parking-Violations-Issued-Fiscal-Year-2016/kiv2-tbuis>

Fiscal Year 2017:

<https://data.cityofnewyork.us/City-Government/Parking-Violations-Issued-Fiscal-Year-2017/2bnn-yakx>

# Hardware Specifications - Google Dataproc

## **-Master node**

Standard (1 master, N workers)

## **-Machine type**

n1-standard-4 (4 vCPU, 15.0 GB memory)

## **-Primary disk type**

pd-standard

## **-Primary disk size**

500 GB

## **-Worker nodes**

2

## **-Machine type**

n1-standard-4 (4 vCPU, 15.0 GB memory)

## **-Primary disk type**

pd-standard

## **-Primary disk size**

500 GB

## Hardware Specifications cont.

**Image Version:**1.3.45-debian9

**Applications:** Hadoop 2.9.2, Spark 2.3.4, Hive 2.3.5, Pig 0.17.0, Tez 0.9.2

# Ticket Summary

```
CREATE EXTERNAL TABLE IF NOT EXISTS tickets_summary  
(summons_number INT, plate_id STRING, vehicle_make STRING, vehicle_body_type STRING, vehicle_color  
STRING, violation_code STRING, violation_location STRING)  
ROW FORMAT DELIMITED FIELDS TERMINATED BY ','  
STORED AS TEXTFILE LOCATION '/output/tickets_summary'  
TBLPROPERTIES ('skip.header.line.count'='1');
```

```
INSERT OVERWRITE TABLE tickets_summary SELECT summons_number, plate_id, vehicle_make,  
vehicle_body_type, vehicle_color, violation_code, violation_location FROM tickets;
```

```
select * from tickets_summary LIMIT 10;
```

# Results

```
0: jdbc:hive2://localhost:10000/default> CREATE EXTERNAL TABLE IF NOT EXISTS tickets_summary
. . . . .> (summons_number INT, plate_id STRING, vehicle_make STRING, vehicle_body_type STRING, vehicle_color STRING, violation_code STRING, violation_location STRING)
. . . . .> ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
. . . . .> STORED AS TEXTFILE LOCATION '/output/tickets_summary'
. . . . .> TBLPROPERTIES ('skip.header.line.count'='1');
No rows affected (0.128 seconds)
0: jdbc:hive2://localhost:10000/default>
0: jdbc:hive2://localhost:10000/default> INSERT OVERWRITE TABLE tickets_summary SELECT summons_number, plate_id, vehicle_make, vehicle_body_type, vehicle_color, violation_code, violation_location FROM tickets;
No rows affected (187.126 seconds)
0: jdbc:hive2://localhost:10000/default> select * from tickets_summary LIMIT 10;
```

tickets_summary.summons_number	tickets_summary.plate_id	tickets_summary.vehicle_make	tickets_summary.vehicle_body_type	tickets_summary.vehicle_color	tickets_summary.violation_code	tickets_summary.violation_location
1366962000	63540MC	FRUEH	DELV	BRN	46	0068
1342296187	GCY4187	FORD	VAN	BLUE	21	104
1342296199	95V6675	GMC		SILVR	21	104
1342296217	FYMS117	NISSA	SUBN	WHITE	21	104
1356906515	GFM1421	MAZDA	SDN	BLK	40	0033
1337077380	18972BB	INTER	BUS	YELLOW	14	115
1364523796	WNJ4730	TOYOT	SDN	BLK	14	0073
1359914924	68091JZ	TOYOT	DELV	WH	46	0010
1355498326	EWV4127	ACURA	SUBN	GREY	21	0068
1361272259	P4Z0671	BMW	SDN	BK	21	0014

10 rows selected (2.429 seconds)

# How many tickets were given to Toyotas

```
SELECT * FROM tickets_summary WHERE vehicle_make LIKE 'TOYOT' LIMIT 10;
```

```
0: jdbc:hive2://localhost:10000/default> SELECT * FROM tickets_summary WHERE vehicle_make LIKE 'TOYOT' LIMIT 10;
```

tickets_summary.summons_number	tickets_summary.plate_id	tickets_summary.vehicle_make	tickets_summary.vehicle_body_type	tickets_summary.vehicle_color	tickets_summary.violation_code	tickets_summary.violation_location
1364523796	WNJ4730	TOYOT	SDN	BLK	14	0073
1359914924	68091JZ	TOYOT	DELV	WH	46	0010
1365454538	FWF7605	TOYOT	VAN	BLK	40	114
1357082800	END3168	TOYOT	SDN	GRAY	98	110
1353764930	BAW1554	TOYOT	SDN	WH	21	0073
1358207150	GDX5773	TOYOT	SDN	GOLD	71	105
1358235636	ERP5947	TOYOT	SDN	GREEN	71	105
1359127665	EDC1820	TOYOT	SDN	RED	98	106
1359744400	FZA1990	TOYOT	SDN	YELLOW	99	108
1369757682	T615651C	TOYOT	SDN	BLACK	98	0083

```
10 rows selected (0.155 seconds)
```

# Top 10 Registration State by number

```
0: jdbc:hive2://localhost:10000/default> Select Count (*) as count, registration_state
- - - - - .> From tickets
- - - - - .> Group by registration_state
- - - - - .> Order by count desc
- - - - - .> limit 10;
```

count	registration_state
9081366	NY
1125070	NJ
313659	PA
177766	CT
155466	FL
102787	MA
85164	VA
67692	MD
64714	NC
61546	IN



# Bottom 15 Registration State by number

```
0: jdbc:hive2://localhost:10000/default> Select Count (*) as count, registration_state
- - - - -> From tickets
- - - - -> Group by registration_state
- - - - -> Having Count (*) > 549
- - - - -> Order by count
- - - - -> limit 15;

+-----+-----+
| count | registration_state |
+-----+-----+
| 724   | WY                 |
| 746   | BC                 |
| 866   | NS                 |
| 974   | HI                 |
| 1150  | ND                 |
| 1632  | MT                 |
| 2246  | AK                 |
| 2368  | SD                 |
| 2404  | UT                 |
| 3254  | KS                 |
| 4136  | NE                 |
| 4552  | WV                 |
| 4656  | GV                 |
| 5166  | NV                 |
| 6724  | AR                 |
+-----+-----+
15 rows selected (154.377 seconds)
```

# Top 15 Violation times

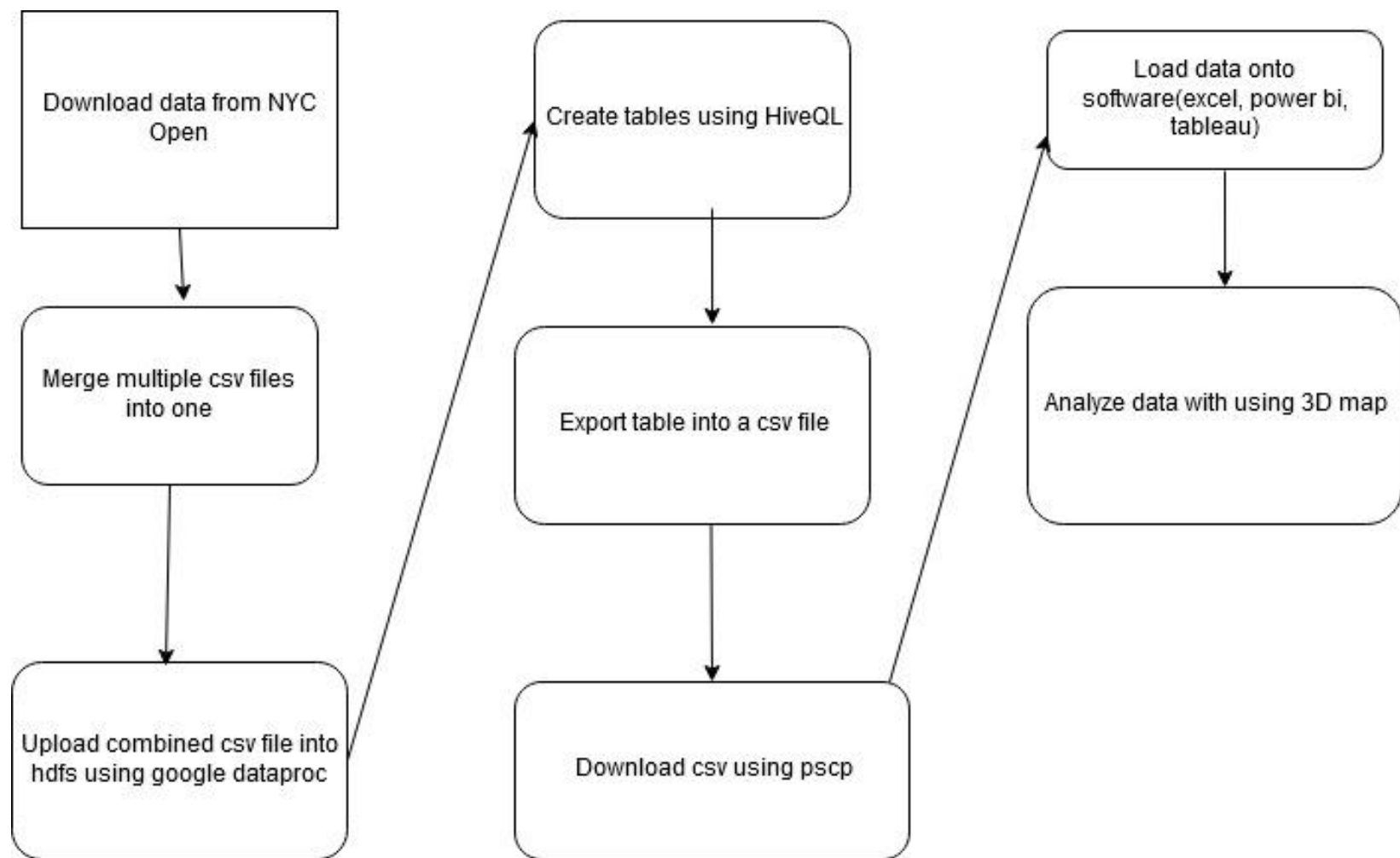
```
0: jdbc:hive2://localhost:10000/default> Select Count (*) as count, violation_time
- - - - -> From tickets
- - - - -> Group by violation_time
- - - - -> Order by count desc
- - - - -> limit 15;
```

count	violation_time
32230	0836A
28632	1136A
27829	1140A
27596	0936A
27409	0806A
26922	0840A
26557	0906A
26505	0940A
25956	1145A
25033	0945A
24937	1139A
24709	1137A
24683	1138A
24672	0838A
24630	0837A

## Bottom 15 Violation time

```
O: jdbc:hive2://localhost:10000/default> Select Count (*) as count, violation_time
- - - - -> From tickets
- - - - -> Group by violation_time
- - - - -> Having Count(*) > 412
- - - - -> Order by count
- - - - -> limit 15;

+-----+-----+
| count | violation_time |
+-----+-----+
| 413    | 0004A          |
| 437    | 0002A          |
| 441    | 0003A          |
| 458    | 0429A          |
| 468    | 0501A          |
| 489    | 0504A          |
| 490    | 0419A          |
| 491    | 0434A          |
| 491    | 0416A          |
| 496    | 0439A          |
| 500    | 0359A          |
| 500    | 0449A          |
| 504    | 0421A          |
| 507    | 0441A          |
| 509    | 0424A          |
+-----+-----+
15 rows selected (133.781 seconds)
```



# Visualization

