

Hive

```
pi2018_nyu_edu@nyu-dataproc-m:~$ hdfs dfs -put drivers_info.csv hdfs://nyu-dataproc-m/user/pi2018_nyu_edu/
pi2018_nyu_edu@nyu-dataproc-m:~$ hdfs dfs -put timesheet_drivers.csv hdfs://nyu-dataproc-m/user/pi2018_nyu_edu/
pi2018_nyu_edu@nyu-dataproc-m:~$ hdfs dfs -put truck_event.csv hdfs://nyu-dataproc-m/user/pi2018_nyu_edu/
pi2018_nyu_edu@nyu-dataproc-m:~$ ls
HW1 HW2 HW3 HW4 drivers_info.csv timesheet_drivers.csv truck_event.csv
pi2018_nyu_edu@nyu-dataproc-m:~$ hdfs dfs -ls
Found 17 items
drwxr-xr-x - pi2018_nyu_edu pi2018_nyu_edu 0 2023-09-29 01:05 HW2
-rw-r--r-- 1 pi2018_nyu_edu pi2018_nyu_edu 1092 2023-09-21 16:41 MaxTemperatureMapper.java
drwxrwxr-x+ - pi2018_nyu_edu pi2018_nyu_edu 0 2023-09-14 15:42 dirToShareAccess
-rw-r--r-- 1 pi2018_nyu_edu pi2018_nyu_edu 2043 2023-10-24 02:45 drivers_info.csv
-rw-r--r-- 1 pi2018_nyu_edu pi2018_nyu_edu 571 2023-09-29 01:11 input
-rw-r--r-- 1 pi2018_nyu_edu pi2018_nyu_edu 279 2023-09-21 16:40 input.txt
-rw-r--r-- 1 pi2018_nyu_edu pi2018_nyu_edu 571 2023-09-29 01:38 input_hw3
-rw-r--r-- 1 pi2018_nyu_edu pi2018_nyu_edu 7077973 2023-10-04 18:13 input_hw4
drwxr-xr-x - pi2018_nyu_edu pi2018_nyu_edu 0 2023-09-22 16:58 new_output
drwxr-xr-x - pi2018_nyu_edu pi2018_nyu_edu 0 2023-09-29 01:48 output_hw3
drwxr-xr-x - pi2018_nyu_edu pi2018_nyu_edu 0 2023-10-04 18:46 output_hw4
drwxr-xr-x - pi2018_nyu_edu pi2018_nyu_edu 0 2023-10-06 17:19 output_hw4A
drwxr-xr-x - pi2018_nyu_edu pi2018_nyu_edu 0 2023-10-06 17:29 output_hw4B
drwxr-xr-x - pi2018_nyu_edu pi2018_nyu_edu 0 2023-10-06 17:37 output_hw4C
drwxr-xr-x - pi2018_nyu_edu pi2018_nyu_edu 0 2023-10-06 17:43 output_hw4D
-rw-r--r-- 1 pi2018_nyu_edu pi2018_nyu_edu 26205 2023-10-24 02:45 timesheet_drivers.csv
-rw-r--r-- 1 pi2018_nyu_edu pi2018_nyu_edu 2272077 2023-10-24 02:45 truck_event.csv
pi2018_nyu_edu@nyu-dataproc-m:~$
```

```
pi2018_nyu_edu@nyu-dataproc-m:~$ beeline -u jdbc:hive2://localhost:10000
Connecting to jdbc:hive2://localhost:10000
Connected to: Apache Hive (version 3.1.2)
Driver: Hive JDBC (version 3.1.2)
Transaction isolation: TRANSACTION_REPEATABLE_READ
Beeline version 3.1.2 by Apache Hive
0: jdbc:hive2://localhost:10000> use pi2018_nyu_edu;
No rows affected (0.124 seconds)
0: jdbc:hive2://localhost:10000>
```

```
0: jdbc:hive2://localhost:10000> CREATE EXTERNAL TABLE drivers (
. . . . .> id INT,
. . . . .> name STRING,
. . . . .> ssn INT,
. . . . .> location STRING,
. . . . .> certified STRING,
. . . . .> wage_plan STRING
. . . . .> )
. . . . .> ROW FORMAT DELIMITED
. . . . .> FIELDS TERMINATED BY ',';
No rows affected (0.115 seconds)
0: jdbc:hive2://localhost:10000> show tables;
+-----+
| tab_name |
+-----+
| drivers  |
+-----+
1 row selected (0.138 seconds)
```

```
0: jdbc:hive2://localhost:10000> load data inpath 'hdfs://nyu-dataproc-m/user/pi2018_nyu_edu/drivers_info.csv' overwrite into table drivers;
No rows affected (1.055 seconds)
0: jdbc:hive2://localhost:10000>
0: jdbc:hive2://localhost:10000> show tables;
```

```
+-----+
| tab_name |
+-----+
| drivers  |
+-----+
1 row selected (0.244 seconds)
0: jdbc:hive2://localhost:10000> select * from drivers;
```

drivers.id	drivers.name	drivers.ssn	drivers.location	drivers.certified	drivers.wage_plan
NULL	name	NULL	location	certified	wage-plan
10	George Vetticaden	621011971	244-4532 Nulla Rd.	N	miles
11	Jamie Engesser	262112338	366-4125 Ac Street	N	miles
12	Paul Coddin	198041975	Ap #622-957 Risus. Street	Y	hours
13	Joe Niemiec	139907145	2071 Hendrerit. Ave	Y	hours
14	Adis Cesir	820812209	Ap #810-1228 In St.	Y	hours
15	Rohit Bakshi	239005227	648-5681 Dui- Rd.	Y	hours
16	Tom McCuch	363303105	P.O. Box 313- 962 Parturient Rd.	Y	hours
17	Eric Mizell	123808238	P.O. Box 579- 2191 Gravida. Street	Y	hours
18	Grant Liu	171010151	Ap #928-3159 Vestibulum Av.	Y	hours
19	Ajay Singh	160005158	592-9430 Nonummy Avenue	Y	hours
20	Chris Harris	921812303	883-2691 Proin Avenue	Y	hours
21	Jeff Markham	209408086	Ap #852-7966 Facilisis St.	Y	hours
22	Nadeem Asghar	783204269	154-9147 Aliquam Ave	Y	hours
23	Adam Diaz	928312208	P.O. Box 260- 6127 Vitae Road	Y	hours
24	Don Hilborn	254412152	4361 Ac Road	Y	hours
25	Jean-Philippe Playe	913310051	P.O. Box 812- 6238 Ac Rd.	Y	hours
26	Michael Aube	124705141	P.O. Box 213- 8948 Nec Ave	Y	hours
27	Mark Lochbihler	392603159	8355 Ipsum St.	Y	hours
28	Olivier Renault	959908181	P.O. Box 243- 6509 Erat. Avenue	Y	hours

```
0: jdbc:hive2://localhost:10000> CREATE EXTERNAL TABLE timesheet (
. . . . .> id INT,
. . . . .> week INT,
. . . . .> hours_logged INT,
. . . . .> miles_logged INT
. . . . .> )
. . . . .> ROW FORMAT DELIMITED
. . . . .> FIELDS TERMINATED BY ',';
```

```
No rows affected (0.147 seconds)
0: jdbc:hive2://localhost:10000> load data inpath 'hdfs://nyu-dataproc-m/user/pi2018_nyu_edu/timesheet_drivers.csv' overwrite into table timesheet;
```

```
No rows affected (0.703 seconds)
0: jdbc:hive2://localhost:10000> select * from timesheet;
```

timesheet.id	timesheet.week	timesheet.hours_logged	timesheet.miles_logged
NULL	NULL	NULL	NULL
10	1	70	3300
10	2	70	3300
10	3	60	2800
10	4	70	3100
10	5	70	3200
10	6	70	3300
10	7	70	3000
10	8	70	3300
10	9	70	3200
10	10	50	2500
10	11	70	2900
10	12	70	3100
10	13	70	3300
10	14	70	3300
10	15	70	3300
10	16	70	3400
10	17	70	3300
10	18	70	3300
10	19	70	3300
10	20	30	1200
10	21	50	2500
10	22	70	3300
10	23	70	3300
10	24	70	3300
10	25	70	3300
10	26	60	2600
10	27	66	2700
10	28	33	1200
10	29	70	3300

```

0: jdbc:hive2://localhost:10000> CREATE EXTERNAL TABLE event (
. . . . .> driver_id INT,
. . . . .> truck_id INT,
. . . . .> event_time STRING,
. . . . .> event_type STRING,
. . . . .> longitude INT,
. . . . .> latitude INT,
. . . . .> event_key STRING,
. . . . .> corr_id STRING,
. . . . .> driver_name STRING,
. . . . .> route_id INT,
. . . . .> route_name STRING,
. . . . .> event_date STRING
. . . . .> )
. . . . .> ROW FORMAT DELIMITED
. . . . .> FIELDS TERMINATED BY ',';
No rows affected (0.402 seconds)
0: jdbc:hive2://localhost:10000> load data inpath 'hdfs://nyu-dataproc-m/user/pi2018_nyu_edu/truck_event.csv' overwrite into table event;
No rows affected (0.801 seconds)
0: jdbc:hive2://localhost:10000> select * from event;
-----+-----+-----+-----+-----+-----+-----+
| event.driver_id | event.truck_id | event.event_time | event.event_type | event.longitude | event.latitude | event.event_key |
| event.corr_id | event.driver_name | event.route_id | event.route_name | event.event_date |
-----+-----+-----+-----+-----+-----+-----+
| NULL | NULL | eventTime | eventType | NULL | NULL | eventKey |
| 14 | CorrelationId | 25 | driverName | 59:21.4 | NULL | Normal | routeName | -94 | 37 | 2016-05-27-22 | 14|25|92233705724648 |
14373 | 3.66E+18 | 16 | Adis Cesir | 160405074 | Joplin to Kansas City Route 2 | 2016-05-27-22 | 37 | 25 | 92233705724648 |
18 | 18 | 16 | 59:21.7 | Normal | -89 | 39 | 18|16|92233705724648 |
14089 | 3.66E+18 | 105 | Grant Liu | 1565885487 | Springfield to KC Via Hanibal | 2016-05-27-22 | 38 | 105 | 92233705724648 |
27 | 27 | 105 | 59:21.7 | Normal | -90 | 38 | 27|105|92233705724648 |
814070 | 3.66E+18 | 74 | Mark Lochbihler | 1325562373 | Springfield to KC Via Columbia Route 2 | 2016-05-27-22 | 38 | 74 | 92233705724648 |
11 | 11 | 74 | 59:21.7 | Normal | -90 | 38 | 11|74|92233705724648 |
14123 | 3.66E+18 | 87 | Jamie Engesser | 1567254452 | Saint Louis to Memphis Route2 | 2016-05-27-22 | 35 | 87 | 92233705724648 |
22 | 22 | 87 | 59:21.7 | Normal | -90 | 35 | 22|87|92233705724648 |
14101 | 3.66E+18 | 87 | Nadeem Asghar | 1198242881 | Saint Louis to Chicago Route2 | 2016-05-27-22 | 35 | 87 | 92233705724648 |
22 | 22 | 87 | 59:22.3 | Normal | -90 | 35 | 22|87|92233705724648 |
13486 | 3.66E+18 | 68 | Nadeem Asghar | 1198242881 | Saint Louis to Chicago Route2 | 2016-05-27-22 | 40 | 68 | 92233705724648 |
23 | 23 | 68 | 59:22.4 | Normal | -89 | 40 | 23|68|92233705724648 |
13450 | 3.66E+18 | Adam Diaz | 160405074 | Joplin to Kansas City Route 2 | 2016-05-27-22 |

```

```

0: jdbc:hive2://localhost:10000> DESCRIBE drivers;
-----+-----+-----+
| col_name | data_type | comment |
-----+-----+-----+
| id | int | |
| name | string | |
| ssn | int | |
| location | string | |
| certified | string | |
| wage_plan | string | |
-----+-----+-----+
6 rows selected (0.078 seconds)
0: jdbc:hive2://localhost:10000> DESCRIBE timesheet;
-----+-----+-----+
| col_name | data_type | comment |
-----+-----+-----+
| id | int | |
| week | int | |
| hours_logged | int | |
| miles_logged | int | |
-----+-----+-----+
4 rows selected (0.073 seconds)
0: jdbc:hive2://localhost:10000>

```

```
0: jdbc:hive2://localhost:10000> SELECT a.*, b.*
.....> FROM drivers a
.....> JOIN timesheet b
.....> ON a.id = b.id;
```

a.id	a.name	a.ssn	a.location	a.certified	a.wage_plan	b.id	b.week	b.hours_logged	b.miles_logged
38	Scott Shaw	386411175	276 Lobortis Road	Y	hours	38	1	55	2634
38	Scott Shaw	386411175	276 Lobortis Road	Y	hours	38	2	56	2599
38	Scott Shaw	386411175	276 Lobortis Road	Y	hours	38	3	47	2791
38	Scott Shaw	386411175	276 Lobortis Road	Y	hours	38	4	45	2733
38	Scott Shaw	386411175	276 Lobortis Road	Y	hours	38	5	52	2686
38	Scott Shaw	386411175	276 Lobortis Road	Y	hours	38	6	54	2727
38	Scott Shaw	386411175	276 Lobortis Road	Y	hours	38	7	58	2516
38	Scott Shaw	386411175	276 Lobortis Road	Y	hours	38	8	47	2524
38	Scott Shaw	386411175	276 Lobortis Road	Y	hours	38	9	46	2755
38	Scott Shaw	386411175	276 Lobortis Road	Y	hours	38	10	46	2620
38	Scott Shaw	386411175	276 Lobortis Road	Y	hours	38	11	59	2510
38	Scott Shaw	386411175	276 Lobortis Road	Y	hours	38	12	50	2508
38	Scott Shaw	386411175	276 Lobortis Road	Y	hours	38	13	53	2747
38	Scott Shaw	386411175	276 Lobortis Road	Y	hours	38	14	55	2697
38	Scott Shaw	386411175	276 Lobortis Road	Y	hours	38	15	60	2511
38	Scott Shaw	386411175	276 Lobortis Road	Y	hours	38	16	47	2785
38	Scott Shaw	386411175	276 Lobortis Road	Y	hours	38	17	52	2599
38	Scott Shaw	386411175	276 Lobortis Road	Y	hours	38	18	60	2527
38	Scott Shaw	386411175	276 Lobortis Road	Y	hours	38	19	46	2720
38	Scott Shaw	386411175	276 Lobortis Road	Y	hours	38	20	56	2634
38	Scott Shaw	386411175	276 Lobortis Road	Y	hours	38	21	55	2742
38	Scott Shaw	386411175	276 Lobortis Road	Y	hours	38	22	49	2569
38	Scott Shaw	386411175	276 Lobortis Road	Y	hours	38	23	60	2537
38	Scott Shaw	386411175	276 Lobortis Road	Y	hours	38	24	56	2791
38	Scott Shaw	386411175	276 Lobortis Road	Y	hours	38	25	60	2735
38	Scott Shaw	386411175	276 Lobortis Road	Y	hours	38	26	54	2644
38	Scott Shaw	386411175	276 Lobortis Road	Y	hours	38	27	60	2694
38	Scott Shaw	386411175	276 Lobortis Road	Y	hours	38	28	55	2515
38	Scott Shaw	386411175	276 Lobortis Road	Y	hours	38	29	59	2661
38	Scott Shaw	386411175	276 Lobortis Road	Y	hours	38	30	50	2607
38	Scott Shaw	386411175	276 Lobortis Road	Y	hours	38	31	58	2628
38	Scott Shaw	386411175	276 Lobortis Road	Y	hours	38	32	57	2530
38	Scott Shaw	386411175	276 Lobortis Road	Y	hours	38	33	52	2508

```
0: jdbc:hive2://localhost:10000> show tables;
```

```
+-----+
| tab_name |
+-----+
| drivers  |
| event    |
| timesheet |
+-----+
```

```
3 rows selected (0.258 seconds)
```

```
0: jdbc:hive2://localhost:10000> DROP TABLE event;
```

```
No rows affected (0.198 seconds)
```

```
0: jdbc:hive2://localhost:10000> show tables;
```

```
+-----+
| tab_name |
+-----+
| drivers  |
| timesheet |
+-----+
```

```
2 rows selected (0.047 seconds)
```

```
0: jdbc:hive2://localhost:10000> 
```

 SSH-in-browser

```
0: jdbc:hive2://localhost:10000>
```

```
pi2018_nyu_edu@nyu-dataproc-m:~$ presto
```

```
presto> show catalogs;  
Catalog
```

```
-----  
bigquery  
bigquery_public_data  
hive  
memory  
system  
tpcds  
tpch  
(7 rows)
```

```
Query 20231024_032614_00000_g5q2c, FINISHED, 2 nodes  
Splits: 36 total, 36 done (100.00%)  
0.24 [0 rows, 0B] [0 rows/s, 0B/s]
```

```
presto> show schemas from hive;  
Schema
```

```
-----  
aa10350_nyu_edu  
ab10465_nyu_edu  
ab10535_nyu_edu  
adm209_nyu_edu  
ag7432_nyu_edu  
ag8172_nyu_edu  
ag8766_nyu_edu  
ag9489_nyu_edu  
ag9544_nyu_edu  
ahg342_nyu_edu  
aj3437_nyu_edu  
ajk8815_nyu_edu  
ak10514_nyu_edu  
ak8827_nyu_edu  
al8291_nyu_edu  
am10865_nyu_edu  
am11351_nyu_edu  
am13801_nyu_edu  
an3731_nyu_edu  
an4207_nyu_edu  
ap7986_nyu_edu  
ap7989_nyu_edu  
ap8246_nyu_edu  
apk9563_nyu_edu  
ar7165_nyu_edu  
ar7996_nyu_edu  
ar8012_nyu_edu  
as14661_nyu_edu
```

```

pi2018_nyu_edu@nyu-dataproc-m:~$ presto
presto> show tables from hive.pi2018_nyu_edu;
  Table
  -----
  drivers
  timesheet
  weather
(3 rows)

Query 20231024_033031_00005_g5q2c, FINISHED, 3 nodes
Splits: 36 total, 36 done (100.00%)
0.75 [3 rows, 95B] [3 rows/s, 126B/s]

presto> DESCRIBE hive.pi2018_nyu_edu.weather;
  Column  | Type   | Extra | Comment
  -----+-----+-----+-----
  column_one | integer |      | 
  year      | integer |      | 
  column_three | varchar |      | 
  temperature | integer |      | 
  column_five | varchar |      | 
  column_six  | varchar |      | 
(6 rows)

Query 20231024_033036_00006_g5q2c, FINISHED, 3 nodes
Splits: 36 total, 36 done (100.00%)
0.41 [6 rows, 460B] [14 rows/s, 1.1KB/s]

presto> █

```

```

presto> SELECT year, MAX(temperature) AS max_temperature
        -> FROM hive.pi2018_nyu_edu.weather
        -> GROUP BY year;
  year | max_temperature
  -----+-----
  1950 |                22
  1949 |                111
(2 rows)

Query 20231024_033129_00008_g5q2c, FINISHED, 2 nodes
Splits: 81 total, 81 done (100.00%)
0.38 [5 rows, 555B] [13 rows/s, 1.41KB/s]

presto> █

```