Truck Accident Study using Big Data Analytics

Rohan Sharma

The University of Texas at Dallas

Big Data Analytics

Task 1: Creating the set of geographic data needed for the remaining exercises.

Downloaded from the projects folder on eLearning.

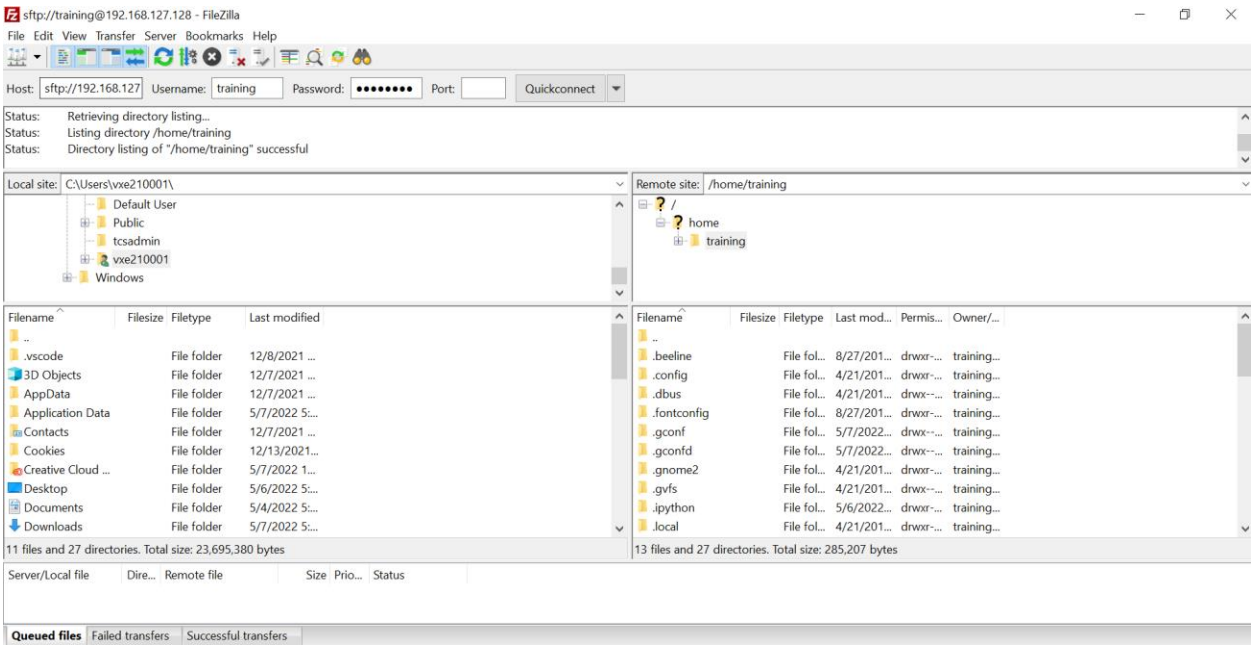
Task 2: Loading Data into Hadoop File System (HDFS)

Checked the IP Address using ifconfig

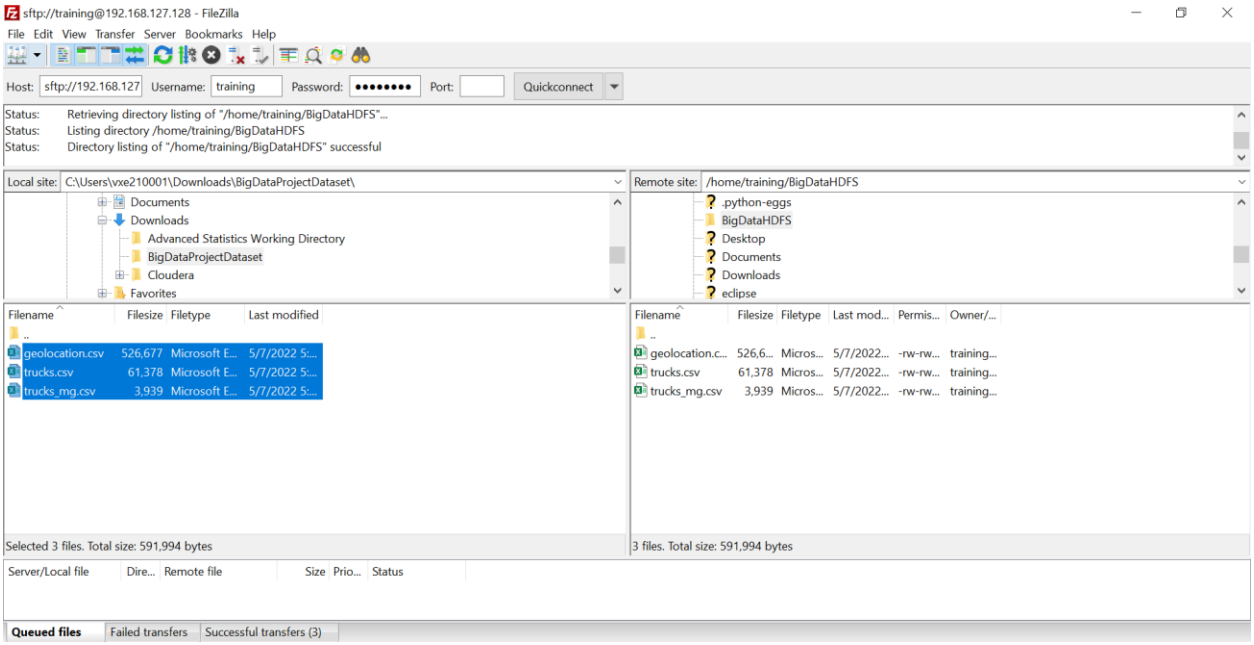
```
[training@localhost ~]$ ifconfig
eth0      Link encap:Ethernet  HWaddr 00:0C:29:59:27:2E
          inet addr:192.168.127.128  Bcast:192.168.127.255  Mask:255.255.255.0
          UP BROADCAST RUNNING MULTICAST  MTU:1500  Metric:1
          RX packets:83 errors:0 dropped:0 overruns:0 frame:0
          TX packets:62 errors:0 dropped:0 overruns:0 carrier:0
          collisions:0 txqueuelen:1000
          RX bytes:7949 (7.7 KiB)  TX bytes:6032 (5.8 KiB)

lo        Link encap:Local Loopback
          inet addr:127.0.0.1  Mask:255.0.0.0
          UP LOOPBACK RUNNING  MTU:65536  Metric:1
          RX packets:7249 errors:0 dropped:0 overruns:0 frame:0
          TX packets:7249 errors:0 dropped:0 overruns:0 carrier:0
          collisions:0 txqueuelen:0
          RX bytes:2784104 (2.6 MiB)  TX bytes:2784104 (2.6 MiB)
```

IP Address - 192.168.127.128



Copied files from local to HDFS using sftp. (Used FileZilla)



Viewed the files in HDFS.

```
[training@localhost ~]$ cd BigDataHDFS/
[training@localhost BigDataHDFS]$ ls -l
total 580
-rw-rw-r-- 1 training training 526677 May  7 16:09 geolocation.csv
-rw-rw-r-- 1 training training  61378 May  7 16:09 trucks.csv
-rw-rw-r-- 1 training training   3939 May  7 16:09 trucks_mg.csv
```

Task 3: Hive Table Creation

Created a Hive Table for geolocation tab and load data from Geolocation file.

```
Impala
File  Edit  View  Search  Terminal  Help
[localhost.localdomain:21000] > CREATE TABLE geolocation_stage
                              > (
                              > truckid string,
                              > driverid string,
                              > event string,
                              > latitude DOUBLE,
                              > longitude DOUBLE,
                              > city string,
                              > state string,
                              > velocity BIGINT,
                              > event_ind BIGINT,
                              > idling_ind BIGINT
                              > )
                              > ROW FORMAT DELIMITED
                              > FIELDS TERMINATED BY ',' STORED AS TEXTFILE
                              > TBLPROPERTIES ("skip.header.line.count"="1");
Query: create TABLE geolocation_stage
(
truckid string,
driverid string,
event string,
latitude DOUBLE,
longitude DOUBLE,
city string,
state string,
velocity BIGINT,
event_ind BIGINT,
idling_ind BIGINT
)
ROW FORMAT DELIMITED
FIELDS TERMINATED BY ',' STORED AS TEXTFILE
TBLPROPERTIES ("skip.header.line.count"="1")

Fetched 0 row(s) in 1.36s
```

```
[training@localhost ~]$ hdfs dfs -ls /
Found 6 items
drwxrwxrwx   - training supergroup          0 2022-05-06 18:17 /dev
drwxrwxrwx   - hbase    supergroup          0 2015-08-27 05:37 /hbase
drwxrwxrwx   - solr     supergroup          0 2015-08-27 05:32 /solr
drwxrwxrwt   - hdfs     supergroup          0 2015-08-27 05:38 /tmp
drwxrwxrwx   - hdfs     supergroup          0 2015-08-27 05:32 /user
drwxrwxrwx   - hdfs     supergroup          0 2015-08-27 05:31 /var
[training@localhost ~]$ hdfs dfs -mkdir /BigDataHDFS

[training@localhost BigDataHDFS]$ hdfs dfs -put /home/training/BigDataHDFS /
[training@localhost BigDataHDFS]$ hdfs dfs -ls /
Found 7 items
drwxrwxrwx   - training supergroup          0 2022-05-07 17:05 /BigDataHDFS
drwxrwxrwx   - training supergroup          0 2022-05-06 18:17 /dev
drwxrwxrwx   - hbase    supergroup          0 2015-08-27 05:37 /hbase
drwxrwxrwx   - solr     supergroup          0 2015-08-27 05:32 /solr
drwxrwxrwt   - hdfs     supergroup          0 2015-08-27 05:38 /tmp
drwxrwxrwx   - hdfs     supergroup          0 2015-08-27 05:32 /user
drwxrwxrwx   - hdfs     supergroup          0 2015-08-27 05:31 /var

[training@localhost BigDataHDFS]$ hdfs dfs -ls /BigDataHDFS/
Found 3 items
-rw-rw-rw-   1 training supergroup     526677 2022-05-07 17:05 /BigDataHDFS/geolocation.csv
-rw-rw-rw-   1 training supergroup      61378 2022-05-07 17:05 /BigDataHDFS/trucks.csv
-rw-rw-rw-   1 training supergroup       3939 2022-05-07 17:05 /BigDataHDFS/trucks_mg.csv
```

```
[localhost.localdomain:21000] > LOAD DATA INPATH '/BigDataHDFS/geolocation.csv' INTO TABLE geolocation_stage;
Query: load DATA INPATH '/BigDataHDFS/geolocation.csv' INTO TABLE geolocation_stage
+-----------------------------------------------------------+
| summary                                                   |
+-----------------------------------------------------------+
| Loaded 1 file(s). Total files in destination location: 1 |
+-----------------------------------------------------------+
Fetched 1 row(s) in 0.34s
```

Data sample for geolocation_stage

View in Metastore Browser

	truckid	driverid	event	latitude	longitude	city	state	velocity	event_ind	idling_ind
0	truckid	driverid	event	NULL	NULL	city	state	NULL	NULL	NULL
1	A54	A54	normal	38.440467	-122.714431	Santa Rosa	California	17	0	0
2	A20	A20	normal	36.977173	-121.899402	Aptos	California	27	0	0
3	A40	A40	overspeed	37.957702	-121.29078	Stockton	California	77	1	0
4	A31	A31	normal	39.409608	-123.355566	Willits	California	22	0	0
5	A71	A71	normal	33.683947	-117.794694	Irvine	California	43	0	0
6	A50	A50	normal	38.40765	-122.947713	Occidental	California	0	0	1
7	A51	A51	normal	37.639097	-120.996878	Modesto	California	0	0	1
8	A19	A19	normal	37.962146	-122.345526	San Pablo	California	0	0	1
9	A77	A77	normal	37.962146	-122.345526	San Pablo	California	25	0	0
10	A92	A92	normal	37.484938	-119.966284	Mariposa	California	0	0	1
11	A89	A89	normal	39.017396	-122.057748	Arbuckle	California	38	0	0
12	A86	A86	normal	32.715329	-117.157255	San Diego	California	45	0	0

Ok

```
Query: create TABLE trucks(driverid string, truckid string, model string, jun13_miles bigint, jun13_gas bigint,
may13_miles bigint, may13_gas bigint, apr13_miles bigint, apr13_gas bigint, mar13_miles bigint,
mar13_gas bigint, feb13_miles bigint, feb13_gas bigint, jan13_miles bigint, jan13_gas bigint, dec12_miles
bigint, dec12_gas bigint, nov12_miles bigint, nov12_gas bigint, oct12_miles bigint, oct12_gas bigint,
sep12_miles bigint, sep12_gas bigint, aug12_miles bigint, aug12_gas bigint, jul12_miles bigint, jul12_gas
bigint, jun12_miles bigint, jun12_gas bigint,may12_miles bigint, may12_gas bigint, apr12_miles bigint,
apr12_gas bigint, mar12_miles bigint, mar12_gas bigint, feb12_miles bigint, feb12_gas bigint, jan12_miles
bigint, jan12_gas bigint, dec11_miles bigint, dec11_gas bigint, nov11_miles bigint, nov11_gas bigint,
oct11_miles bigint, oct11_gas bigint, sep11_miles bigint, sep11_gas bigint, aug11_miles bigint, aug11_gas
bigint, jul11_miles bigint, jul11_gas bigint, jun11_miles bigint, jun11_gas bigint, may11_miles bigint,
may11_gas bigint, apr11_miles bigint, apr11_gas bigint, mar11_miles bigint, mar11_gas bigint,
feb11_miles bigint, feb11_gas bigint, jan11_miles bigint, jan11_gas bigint, dec10_miles bigint, dec10_gas
bigint, nov10_miles bigint, nov10_gas bigint, oct10_miles bigint, oct10_gas bigint, sep10_miles bigint,
sep10_gas bigint, aug10_miles bigint, aug10_gas bigint, jul10_miles bigint, jul10_gas bigint, jun10_miles
bigint, jun10_gas bigint, may10_miles bigint, may10_gas bigint, apr10_miles bigint, apr10_gas bigint,
mar10_miles bigint, mar10_gas bigint, feb10_miles bigint, feb10_gas bigint, jan10_miles bigint, jan10_gas
bigint, dec09_miles bigint, dec09_gas bigint, nov09_miles bigint, nov09_gas bigint, oct09_miles bigint,
oct09_gas bigint, sep09_miles bigint, sep09_gas bigint, aug09_miles bigint, aug09_gas bigint, jul09_miles
bigint, jul09_gas bigint, jun09_miles bigint, jun09_gas bigint, may09_miles bigint, may09_gas bigint,
apr09_miles bigint, apr09_gas bigint, mar09_miles bigint, mar09_gas bigint, feb09_miles bigint,
feb09_gas bigint, jan09_miles bigint, jan09_gas bigint)
ROW FORMAT DELIMITED
FIELDS TERMINATED BY ','
STORED AS TEXTFILE
TBLPROPERTIES ("skip.header.line.count"="1")

Fetched 0 row(s) in 0.25s
```

```
[localhost.localdomain:21000] > LOAD DATA INPATH '/BigDataHDFS/trucks.csv' INTO TABLE trucks;
Query: load DATA INPATH '/BigDataHDFS/trucks.csv' INTO TABLE trucks
+-----------------------------------------------------------+
| summary                                                   |
+-----------------------------------------------------------+
| Loaded 1 file(s). Total files in destination location: 1 |
+-----------------------------------------------------------+
Fetched 1 row(s) in 5.19s
```

Data sample for trucks

View in Metastore Browser

	driverid	truckid	model	jun13_miles	jun13_gas	may13_miles	may13_gas	apr13_miles	apr13_gas	mar13_miles	mar13_gas	feb13_miles	feb13_gas	jan13_miles	jan13_gas	dec12_miles	dec12_gas	nov12_miles	nov12_ga
0	driverid	truckid	model	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL
1	A1	A1	Freightliner	9217	1914	8769	1892	14234	3008	11519	2262	8676	1596	10025	1878	12647	2331	10214	2054
2	A2	A2	Ford	12058	2335	14314	2648	11050	2323	14114	3157	13583	2346	15362	3353	13608	2607	11236	2597
3	A3	A3	Ford	13652	2899	12075	2603	12277	2792	9642	1864	8606	1796	13089	2796	10762	2099	10250	1857
4	A4	A4	Kenworth	12687	2439	10680	2083	11071	2599	12302	2361	8845	1816	14130	2929	12316	2702	11999	2377
5	A5	A5	Hino	10233	1825	14634	3450	9281	2028	13547	2790	12990	3056	13769	2713	11423	2083	12570	2547
6	A6	A6	Caterpillar	14488	2883	13317	3091	10927	2161	10972	1866	14322	2983	12956	2435	14293	2596	12593	2384
7	A7	A7	Ford	10938	2231	12080	2142	13539	2346	13453	2413	12754	2444	11147	2083	11749	2442	15816	2728
8	A8	A8	Navistar	11392	2280	8922	1834	11422	2202	13142	2478	14133	2813	10184	1819	9660	1690	12201	2359
9	A9	A9	Volvo	12601	2515	9707	1869	11914	2479	11531	2207	12913	2342	9648	1663	12035	2497	8453	1746
10	A10	A10	Peterbilt	13699	2583	11583	2077	12833	2473	13168	2436	12751	2622	9391	2104	11971	2472	14276	2762
11	A11	A11	Peterbilt	12447	2596	15726	3145	11461	2512	14566	3360	11557	2524	15319	3025	15166	2951	12052	2725

Ok

```
hive>
hive> CREATE TABLE truck_mileage AS
    > SELECT
    >   truckid,
    >   driverid,
    >   rdate,
    >   miles,
    >   gas,
    >   miles/gas as mpg
    > FROM
    >   trucks LATERAL VIEW stack(
    >     54,
    >     'jun13',
    >     jun13_miles,
    >     jun13_gas,
    >     'may13',
    >     may13_miles,
    >     may13_gas,
    >     'apr13',
    >     apr13_miles,
    >     apr13_gas,
    >     'mar13',
    >     mar13_miles,
    >     mar13_gas,
    >     'feb13',
    >     feb13_miles,
    >     feb13_gas,
    >     'jan13',
    >     jan13_miles,
    >     jan13_gas,
    >     'dec12',
    >     dec12_miles,
    >     dec12_gas,
    >     'nov12',
    >     nov12_miles,
    >     nov12_gas,
    >     'oct12',
    >     oct12_miles,
    >     oct12_gas,
    >     'sep12',
    >     sep12_miles,
    >     sep12_gas,
    >     'aug12',
    >     aug12_miles,
    >     aug12_gas,
```

```
    >     aug09_gas,
    >     'jul09',
    >     jul09_miles,
    >     jul09_gas,
    >     'jun09',
    >     jun09_miles,
    >     jun09_gas,
    >     'may09',
    >     may09_miles,
    >     may09_gas,
    >     'apr09',
    >     apr09_miles,
    >     apr09_gas,
    >     'mar09',
    >     mar09_miles,
    >     mar09_gas,
    >     'feb09',
    >     feb09_miles,
    >     feb09_gas,
    >     'jan09',
    >     jan09_miles,
    >     jan09_gas
    >   ) dummyalias AS rdate,
    >   miles,
    >   gas;
Query ID = training_20220508185757_f796f8bb-3339-4825-b51e-60af1482a723
Total jobs = 3
Launching Job 1 out of 3
Number of reduce tasks is set to 0 since there's no reduce operator
Starting Job = job_1652050965891_0001, Tracking URL = http://localhost:8088/proxy/application_1652050965891_0001/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1652050965891_0001
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 0
2022-05-08 18:58:39,109 Stage-1 map = 0%,  reduce = 0%
2022-05-08 18:59:06,662 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 5.93 sec
MapReduce Total cumulative CPU time: 5 seconds 930 msec
Ended Job = job_1652050965891_0001
Stage-4 is selected by condition resolver.
Stage-3 is filtered out by condition resolver.
Stage-5 is filtered out by condition resolver.
Moving data to: hdfs://localhost:8020/user/hive/warehouse/.hive-staging_hive_2022-05-08_18-57-55_609_5069494975774970920-1/-ext-10001
Moving data to: hdfs://localhost:8020/user/hive/warehouse/truck_mileage
Table default.truck_mileage stats: [numFiles=1, numRows=5400, totalSize=229744, rawDataSize=224344]
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1   Cumulative CPU: 5.93 sec   HDFS Read: 87992 HDFS Write: 229827 SUCCESS
Total MapReduce CPU Time Spent: 5 seconds 930 msec
OK
Time taken: 74.456 seconds
```

Data sample for truck_mileage

View in Metastore Browser

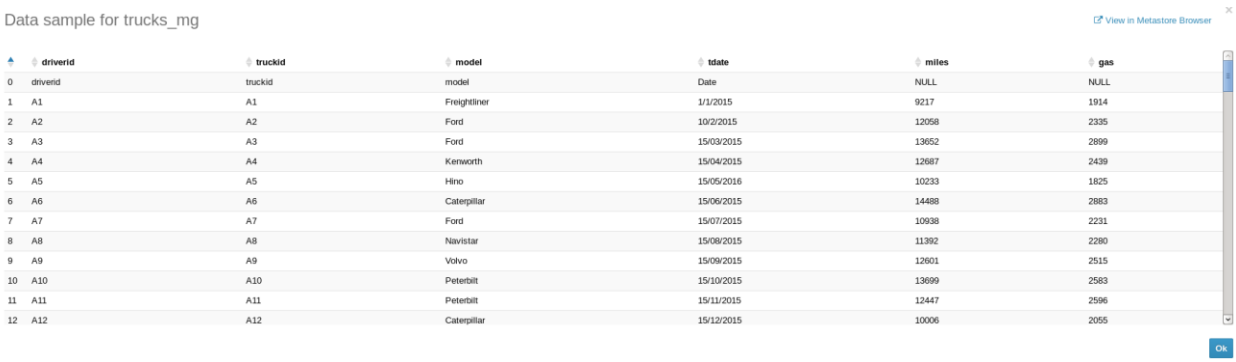
	truckid	driverid	rdate	miles	gas	mpg
0	A1	A1	jun13	9217	1914	4.81556948798
1	A1	A1	may13	8769	1892	4.63477801268
2	A1	A1	apr13	14234	3008	4.73204787234
3	A1	A1	mar13	11519	2262	5.09239610964
4	A1	A1	feb13	8676	1596	5.43609022556
5	A1	A1	jan13	10025	1878	5.3381256656
6	A1	A1	dec12	12647	2331	5.42556842557
7	A1	A1	nov12	10214	2054	4.97273612463
8	A1	A1	oct12	10807	2134	5.06419868791
9	A1	A1	sep12	11127	2191	5.07850296668
10	A1	A1	aug12	9754	1967	4.95882053889
11	A1	A1	jul12	12925	2578	5.01357641583
12	A1	A1	jun12	15792	3313	4.76667672804

Ok

```
hive> CREATE TABLE avg_mileage
    > AS
    > SELECT truckid, avg(mpg) avgmpg
    > FROM truck_mileage
    > GROUP BY truckid;
Query ID = training_20220508190303_d6e17101-9d35-481b-adbd-2215998032bd
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks not specified. Estimated from input data size: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1652050965891_0002, Tracking URL = http://localhost:8088/proxy/application_1652050965891_0002/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1652050965891_0002
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2022-05-08 19:03:37,455 Stage-1 map = 0%,  reduce = 0%
2022-05-08 19:04:06,336 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 6.37 sec
2022-05-08 19:04:28,635 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 10.06 sec
MapReduce Total cumulative CPU time: 10 seconds 60 msec
Ended Job = job_1652050965891_0002
Moving data to: hdfs://localhost:8020/user/hive/warehouse/avg_mileage
Table default.avg_mileage stats: [numFiles=1, numRows=100, totalSize=2189, rawDataSize=2089]
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 10.06 sec   HDFS Read: 237077 HDFS Write: 2267 SUCCESS
Total MapReduce CPU Time Spent: 10 seconds 60 msec
OK
Time taken: 82.515 seconds
```

```
hive> CREATE TABLE DriverMileage
    > AS
    > SELECT driverid, sum(miles) totmiles
    > FROM truck_mileage
    > GROUP BY driverid;
Query ID = training_20220508190505_3748f6d0-a6ba-45d6-8984-065c9045424d
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks not specified. Estimated from input data size: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1652050965891_0003, Tracking URL = http://localhost:8088/proxy/application_1652050965891_0003/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1652050965891_0003
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2022-05-08 19:05:27,279 Stage-1 map = 0%,  reduce = 0%
2022-05-08 19:05:46,972 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 3.91 sec
2022-05-08 19:06:09,696 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 7.67 sec
MapReduce Total cumulative CPU time: 7 seconds 670 msec
Ended Job = job_1652050965891_0003
Moving data to: hdfs://localhost:8020/user/hive/warehouse/drivermileage
Table default.drivermileage stats: [numFiles=1, numRows=100, totalSize=1092, rawDataSize=992]
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 7.67 sec   HDFS Read: 236738 HDFS Write: 1171 SUCCESS
Total MapReduce CPU Time Spent: 7 seconds 670 msec
OK
Time taken: 69.946 seconds
```

```
hive> CREATE TABLE trucks_mg(driverid string, truckid string, model string, Tdate
    > string, miles bigint, gas bigint )
    > ROW FORMAT DELIMITED
    > FIELDS TERMINATED BY ','
    > STORED AS TEXTFILE
    > TBLPROPERTIES ("skip.header.line.count"="1");
OK
Time taken: 0.412 seconds
```

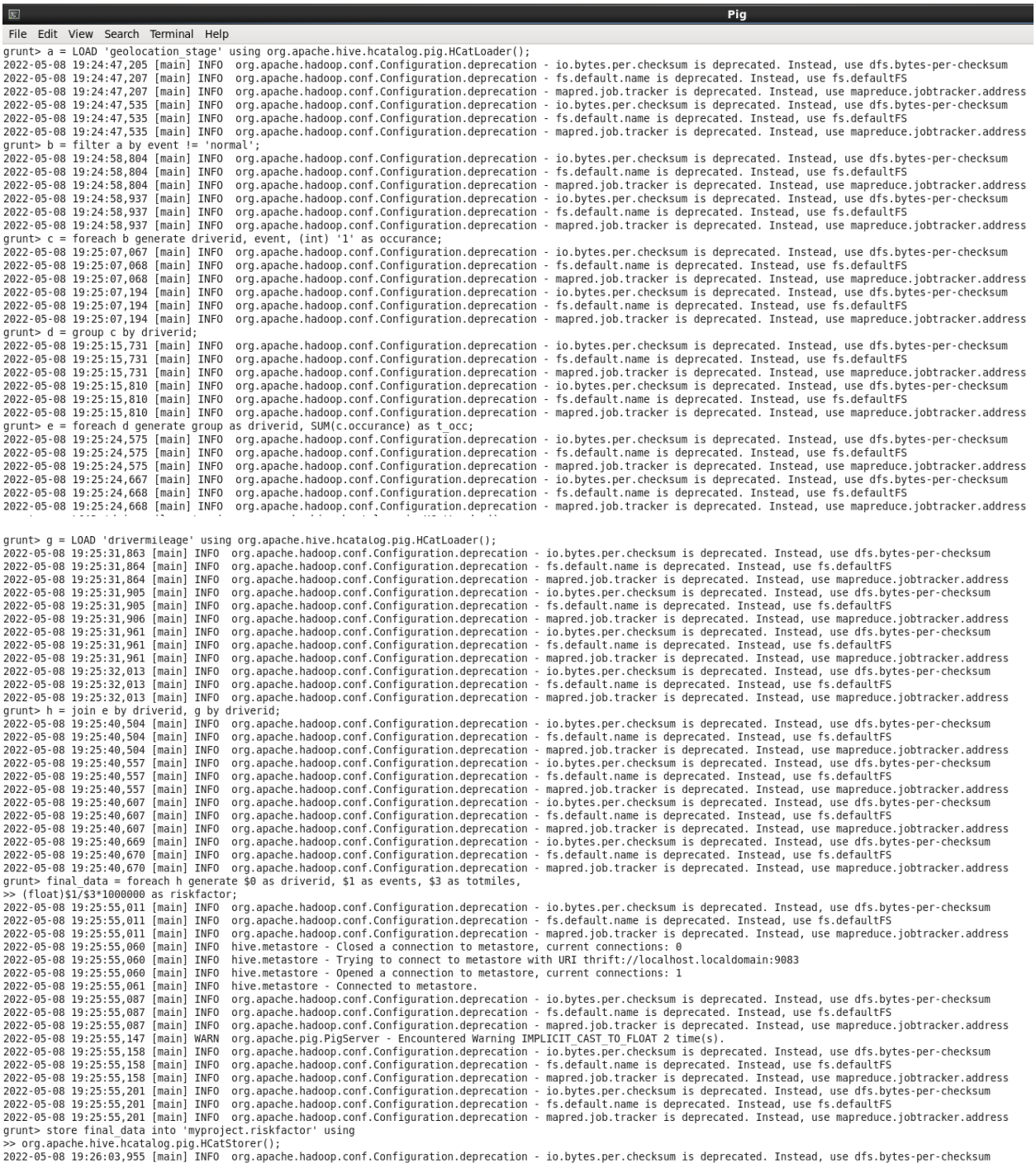


	driverId	truckid	model	tdate	miles	gas
0	driverid	truckid	model	Date	NULL	NULL
1	A1	A1	Freightliner	1/1/2015	9217	1914
2	A2	A2	Ford	10/2/2015	12058	2335
3	A3	A3	Ford	15/03/2015	13652	2899
4	A4	A4	Kenworth	15/04/2015	12687	2439
5	A5	A5	Hino	15/05/2016	10233	1825
6	A6	A6	Caterpillar	15/06/2015	14488	2883
7	A7	A7	Ford	15/07/2015	10938	2231
8	A8	A8	Navistar	15/08/2015	11392	2280
9	A9	A9	Volvo	15/09/2015	12601	2515
10	A10	A10	Peterbilt	15/10/2015	13699	2583
11	A11	A11	Peterbilt	15/11/2015	12447	2596
12	A12	A12	Caterpillar	15/12/2015	10006	2055

```
hive> CREATE TABLE riskfactor
    > (driverid string,
    > events bigint,
    > totmiles bigint,
    > riskfactor float)
    > ;
OK
Time taken: 0.038 seconds
```

Task 4: Loading Risk Factor table using Pig

```
[training@localhost ~]$ pig -x local
log4j:WARN No appenders could be found for logger (org.apache.hadoop.util.Shell).
log4j:WARN Please initialize the log4j system properly.
log4j:WARN See http://logging.apache.org/log4j/1.2/faq.html#noconfig for more info.
2022-05-08 19:22:47,346 [main] INFO  org.apache.pig.Main - Apache Pig version 0.12.0-cdh5.4.3 (rexported) compiled Jun 24 2015, 19:36:38
2022-05-08 19:22:47,346 [main] INFO  org.apache.pig.Main - Logging error messages to: /home/training/pig_1652062967332.log
2022-05-08 19:22:47,414 [main] INFO  org.apache.pig.impl.util.Utils - Default bootup file /home/training/.pigbootup not found
2022-05-08 19:22:47,710 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2022-05-08 19:22:47,714 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2022-05-08 19:22:47,715 [main] INFO  org.apache.pig.backend.hadoop.executionengine.HExecutionEngine - Connecting to hadoop file system at: file:///
2022-05-08 19:22:48,288 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2022-05-08 19:22:48,292 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2022-05-08 19:22:48,292 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2022-05-08 19:22:48,409 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2022-05-08 19:22:48,410 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2022-05-08 19:22:48,414 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2022-05-08 19:22:48,571 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2022-05-08 19:22:48,574 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2022-05-08 19:22:48,575 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2022-05-08 19:22:48,675 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2022-05-08 19:22:48,681 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2022-05-08 19:22:48,681 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2022-05-08 19:22:48,766 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2022-05-08 19:22:48,766 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2022-05-08 19:22:48,774 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2022-05-08 19:22:48,847 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2022-05-08 19:22:48,851 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2022-05-08 19:22:48,854 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2022-05-08 19:22:48,927 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2022-05-08 19:22:48,928 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2022-05-08 19:22:48,928 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2022-05-08 19:22:49,004 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2022-05-08 19:22:49,009 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2022-05-08 19:22:49,009 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2022-05-08 19:22:49,080 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2022-05-08 19:22:49,083 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2022-05-08 19:22:49,084 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
```


Pig
File Edit View Search Terminal Help
grunt> a = LOAD 'geolocation_stage' using org.apache.hive.hcatalog.pig.HCatLoader();
2022-05-08 19:24:47,205 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2022-05-08 19:24:47,207 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2022-05-08 19:24:47,207 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2022-05-08 19:24:47,535 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2022-05-08 19:24:47,535 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2022-05-08 19:24:47,535 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
grunt> b = filter a by event != 'normal';
2022-05-08 19:24:58,804 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2022-05-08 19:24:58,804 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2022-05-08 19:24:58,804 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2022-05-08 19:24:58,937 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2022-05-08 19:24:58,937 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2022-05-08 19:24:58,937 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
grunt> c = foreach b generate driverid, event, (int) '1' as occurance;
2022-05-08 19:25:07,067 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2022-05-08 19:25:07,068 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2022-05-08 19:25:07,068 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2022-05-08 19:25:07,194 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2022-05-08 19:25:07,194 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2022-05-08 19:25:07,194 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
grunt> d = group c by driverid;
2022-05-08 19:25:15,731 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2022-05-08 19:25:15,731 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2022-05-08 19:25:15,731 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2022-05-08 19:25:15,810 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2022-05-08 19:25:15,810 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2022-05-08 19:25:15,810 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
grunt> e = foreach d generate group as driverid, SUM(c.occurance) as t_occ;
2022-05-08 19:25:24,575 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2022-05-08 19:25:24,575 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2022-05-08 19:25:24,575 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2022-05-08 19:25:24,667 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2022-05-08 19:25:24,668 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2022-05-08 19:25:24,668 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
grunt> g = LOAD 'drivermileage' using org.apache.hive.hcatalog.pig.HCatLoader();
2022-05-08 19:25:31,863 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2022-05-08 19:25:31,864 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2022-05-08 19:25:31,864 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2022-05-08 19:25:31,905 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2022-05-08 19:25:31,905 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2022-05-08 19:25:31,906 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2022-05-08 19:25:31,961 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2022-05-08 19:25:31,961 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2022-05-08 19:25:31,961 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2022-05-08 19:25:32,013 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2022-05-08 19:25:32,013 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2022-05-08 19:25:32,013 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
grunt> h = join e by driverid, g by driverid;
2022-05-08 19:25:40,504 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2022-05-08 19:25:40,504 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2022-05-08 19:25:40,504 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2022-05-08 19:25:40,557 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2022-05-08 19:25:40,557 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2022-05-08 19:25:40,557 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2022-05-08 19:25:40,607 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2022-05-08 19:25:40,607 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2022-05-08 19:25:40,607 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2022-05-08 19:25:40,669 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2022-05-08 19:25:40,670 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2022-05-08 19:25:40,670 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
grunt> final_data = foreach h generate $0 as driverid, $1 as events, $3 as totmiles,
>> (float)$1/$3*1000000 as riskfactor;
2022-05-08 19:25:55,011 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2022-05-08 19:25:55,011 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2022-05-08 19:25:55,011 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2022-05-08 19:25:55,060 [main] INFO hive.metastore - Closed a connection to metastore, current connections: 0
2022-05-08 19:25:55,060 [main] INFO hive.metastore - Trying to connect to metastore with URI thrift://localhost.localdomain:9083
2022-05-08 19:25:55,060 [main] INFO hive.metastore - Opened a connection to metastore, current connections: 1
2022-05-08 19:25:55,061 [main] INFO hive.metastore - Connected to metastore.
2022-05-08 19:25:55,087 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2022-05-08 19:25:55,087 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2022-05-08 19:25:55,087 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2022-05-08 19:25:55,147 [main] WARN org.apache.pig.PigServer - Encountered Warning IMPLICIT_CAST_TO_FLOAT 2 time(s).
2022-05-08 19:25:55,158 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2022-05-08 19:25:55,158 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2022-05-08 19:25:55,158 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2022-05-08 19:25:55,201 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2022-05-08 19:25:55,201 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2022-05-08 19:25:55,201 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
grunt> store final_data into 'myproject.riskfactor' using
>> org.apache.hive.hcatalog.pig.HCatStorer();
2022-05-08 19:26:03,955 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum

```
2022-05-08 19:25:40,670 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
grunt> final_data = foreach h generate $0 as driverid, $1 as events, $3 as totmiles,
>> (float)$1/$3*1000000 as riskfactor;
2022-05-08 19:25:55,011 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2022-05-08 19:25:55,011 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2022-05-08 19:25:55,011 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2022-05-08 19:25:55,060 [main] INFO  hive.metastore - Closed a connection to metastore, current connections: 0
2022-05-08 19:25:55,060 [main] INFO  hive.metastore - Trying to connect to metastore with URI thrift://localhost.localdomain:9083
2022-05-08 19:25:55,060 [main] INFO  hive.metastore - Opened a connection to metastore, current connections: 1
2022-05-08 19:25:55,061 [main] INFO  hive.metastore - Connected to metastore.
2022-05-08 19:25:55,087 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2022-05-08 19:25:55,087 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2022-05-08 19:25:55,087 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2022-05-08 19:25:55,147 [main] WARN  org.apache.pig.PigServer - Encountered Warning IMPLICIT_CAST_TO_FLOAT 2 time(s).
2022-05-08 19:25:55,158 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2022-05-08 19:25:55,158 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2022-05-08 19:25:55,158 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2022-05-08 19:25:55,201 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2022-05-08 19:25:55,201 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2022-05-08 19:25:55,201 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
grunt> store final_data into 'myproject.riskfactor' using
>> org.apache.hive.hcatalog.pig.HCatStorer();
2022-05-08 19:26:03,955 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2022-05-08 19:26:03,955 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2022-05-08 19:26:03,955 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2022-05-08 19:26:04,013 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2022-05-08 19:26:04,013 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2022-05-08 19:26:04,013 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2022-05-08 19:26:04,059 [main] WARN  org.apache.pig.PigServer - Encountered Warning IMPLICIT_CAST_TO_FLOAT 2 time(s).
2022-05-08 19:26:04,083 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2022-05-08 19:26:04,083 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2022-05-08 19:26:04,083 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2022-05-08 19:26:04,139 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2022-05-08 19:26:04,139 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2022-05-08 19:26:04,140 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2022-05-08 19:26:04,208 [main] WARN  org.apache.pig.PigServer - Encountered Warning IMPLICIT_CAST_TO_FLOAT 2 time(s).
2022-05-08 19:26:04,215 [main] INFO  org.apache.pig.tools.pigstats.ScriptState - Pig features used in the script: HASH_JOIN,GROUP_BY,FILTER
2022-05-08 19:26:04,252 [main] INFO  org.apache.pig.newplan.logical.optimizer.LogicalPlanOptimizer - {RULES_ENABLED=[AddForEach, ColumnMapKeyPrune, DuplicateForEachColumnRewrite, GroupByConstParallelSetter, ImplicitSplitInserter, LimitOp
timizer, LoadTypeCastInserter, MergeFilter, MergeForEach, NewPartitionFilterOptimizer, PushDownForEachFlatten, PushUpFilter, SplitFilter, StreamTypeCastInserter], RULES_DISABLED=[FilterLogicExpressionSimplifier, PartitionFilterOptimizer
]
2022-05-08 19:26:04,267 [main] INFO  org.apache.pig.newplan.logical.rules.ColumnPruneVisitor - Columns pruned for a: $0, $3, $4, $5, $6, $7, $8, $9
2022-05-08 19:26:04,295 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2022-05-08 19:26:04,304 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2022-05-08 19:26:04,304 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2022-05-08 19:26:04,392 [main] ERROR org.apache.pig.tools.grunt.Grunt - ERROR 1115: org.apache.hive.hcatalog.common.HCatException : 2001 : Error setting output information. Cause : NoSuchObjectException(message:myproject.riskfactor table
 not found)
Details at logfile: /home/training/pig_1652062967332.log
```

```
Pig
File  Edit  View  Search  Terminal  Help
2022-05-08 20:00:35,632 [pool-12-thread-1] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.PigMapReduce$Reduce - Aliases being processed per job phase (AliasName[line,offset]): M: h[59,4],g[41,4],h[59,4] C:  R: final_d
ata[60,13]
2022-05-08 20:00:36,752 [pool-12-thread-1] INFO  org.apache.hadoop.mapred.Task - Task:attempt_local94333845_0002_r_000000_0 is done. And is in the process of committing
2022-05-08 20:00:36,754 [pool-12-thread-1] INFO  org.apache.hadoop.mapred.LocalJobRunner - 2 / 2 copied.
2022-05-08 20:00:36,754 [pool-12-thread-1] INFO  org.apache.hadoop.mapred.Task - Task attempt_local94333845_0002_r_000000_0 is allowed to commit now
2022-05-08 20:00:36,755 [pool-12-thread-1] INFO  org.apache.hadoop.mapreduce.lib.output.FileOutputCommitter - File Output Committer Algorithm version is 1
2022-05-08 20:00:36,761 [pool-12-thread-1] INFO  org.apache.hadoop.mapreduce.lib.output.FileOutputCommitter - Saved output of task 'attempt_local94333845_0002_r_000000_0' to hdfs://localhost:8020/user/hive/warehouse/riskfactor/_SCRATCH0.
3378446339465223/_temporary/0/task_local94333845_0002_r_000000
2022-05-08 20:00:36,763 [pool-12-thread-1] INFO  org.apache.hadoop.mapred.LocalJobRunner - reduce > reduce
2022-05-08 20:00:36,763 [pool-12-thread-1] INFO  org.apache.hadoop.mapred.Task - Task 'attempt_local94333845_0002_r_000000_0' done.
2022-05-08 20:00:36,763 [pool-12-thread-1] INFO  org.apache.hadoop.mapred.LocalJobRunner - Finishing task: attempt_local94333845_0002_r_000000_0
2022-05-08 20:00:36,763 [Thread-22] INFO  org.apache.hadoop.mapred.LocalJobRunner - reduce task executor complete.
2022-05-08 20:00:36,798 [Thread-22] INFO  org.apache.hive.hcatalog.mapreduce.FileOutputCommitterContainer - Cancelling delegation token for the job.
2022-05-08 20:00:36,841 [Thread-22] INFO  org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2022-05-08 20:00:36,841 [Thread-22] INFO  org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2022-05-08 20:00:36,841 [Thread-22] INFO  org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2022-05-08 20:00:36,841 [Thread-22] INFO  hive.metastore - Trying to connect to metastore with URI thrift://localhost.localdomain:9083
2022-05-08 20:00:36,842 [Thread-22] INFO  hive.metastore - Opened a connection to metastore, current connections: 3
2022-05-08 20:00:36,843 [Thread-22] INFO  hive.metastore - Connected to metastore.
2022-05-08 20:00:37,238 [main] WARN  org.apache.pig.tools.pigstats.PigStatsUtil - Failed to get RunningJob for job job_local94333845_0002
2022-05-08 20:00:37,239 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 100% complete
2022-05-08 20:00:37,241 [main] INFO  org.apache.pig.tools.pigstats.SimplePigStats - Detected Local mode. Stats reported below may be incomplete
2022-05-08 20:00:37,245 [main] INFO  org.apache.pig.tools.pigstats.SimplePigStats - Script Statistics:

HadoopVersion   PigVersion      UserId  StartedAt       FinishedAt      Features
2.6.0-cdh5.4.3  0.12.0-cdh5.4.3 training        2022-05-08 20:00:29     2022-05-08 20:00:37     HASH_JOIN,GROUP_BY,FILTER

Success!

Job Stats (time in seconds):
JobId   Alias   Feature Outputs
job_local408986930_0001 a,b,c,d,e       GROUP_BY,COMBINER
job_local94333845_0002  final_data,g,h  HASH_JOIN       riskfactor,

Input(s):
Successfully read records from: "geolocation_stage"
Successfully read records from: "drivermileage"

Output(s):
Successfully stored records in: "riskfactor"

Job DAG:
job_local408986930_0001 ->      job_local94333845_0002,
job_local94333845_0002


2022-05-08 20:00:37,245 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
```

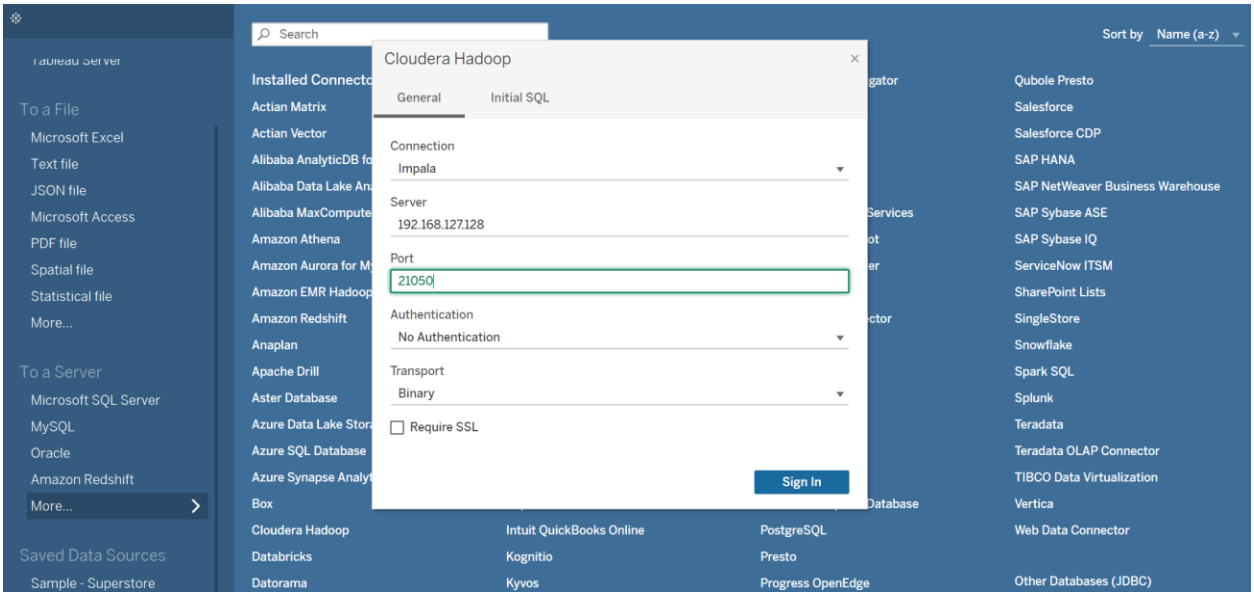
Data sample for riskfactor

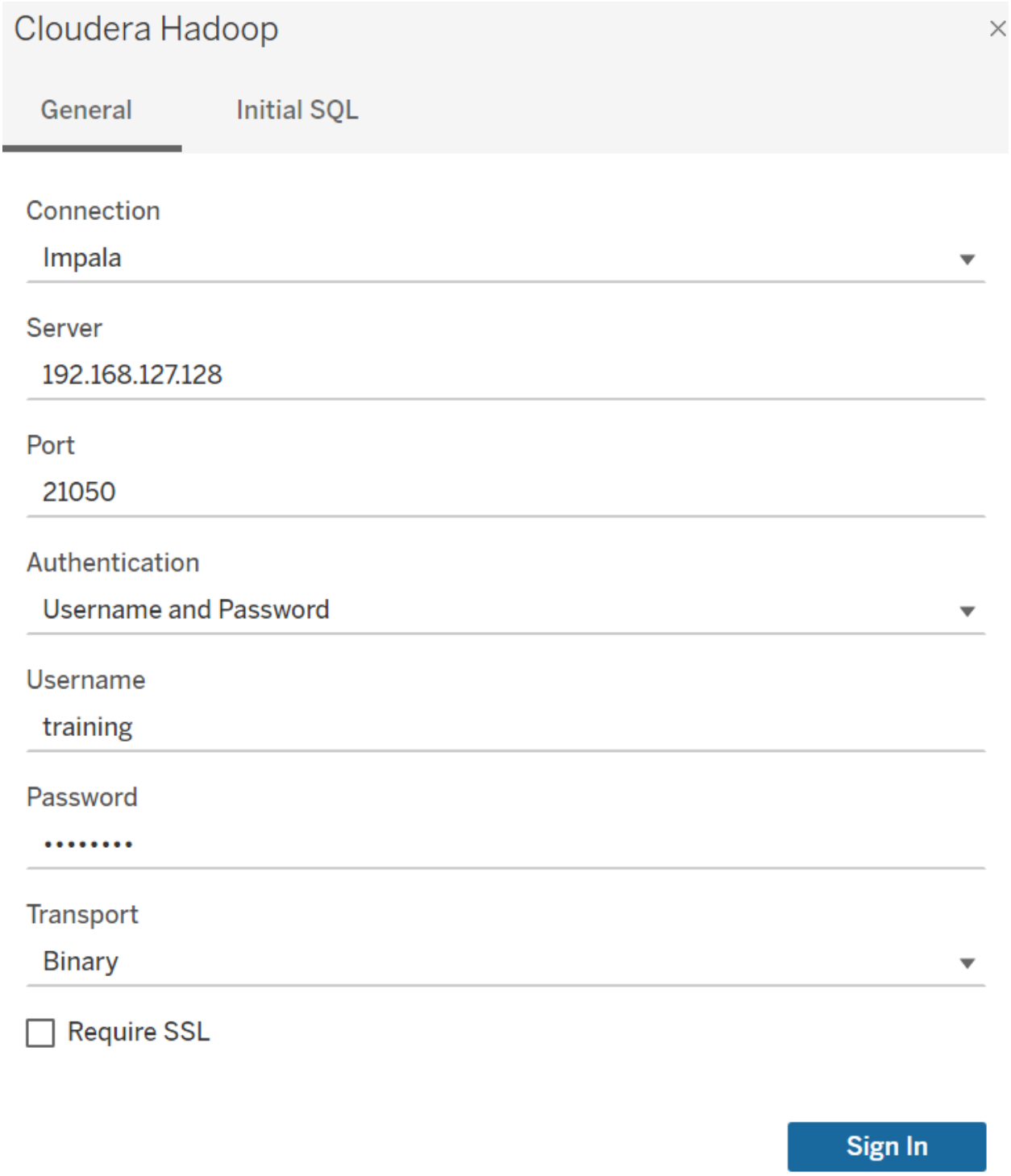
View in Metastore Browser

	driverid	events	totmiles	riskfactor
0	A1	3	628507	4.77321624756
1	A2	1	664543	1.50479352474
2	A3	6	639584	9.38109779358
3	A4	6	663289	9.04583072662
4	A5	9	676574	13.3023138046
5	A6	1	648479	1.54207003117
6	A7	5	653787	7.64775085449
7	A8	2	653991	3.05814623833
8	A9	6	665456	9.01637363434
9	A10	6	675377	8.88392734528
10	A11	5	652452	7.66339874268
11	A12	5	668241	7.4823307991
12	A13	4	654319	6.11322593689

Ok

Task 5: Integrating HDFS with Tableau

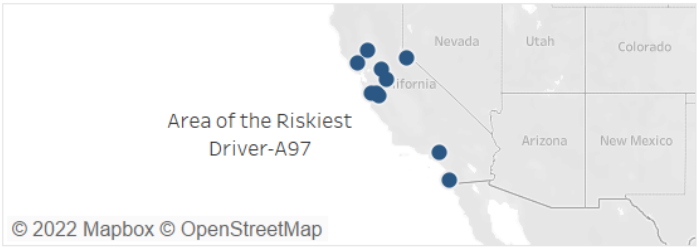

Search
Sort by Name (a-z)
To a File
Microsoft Excel
Text file
JSON file
Microsoft Access
PDF file
Spatial file
Statistical file
More...
To a Server
Microsoft SQL Server
MySQL
Oracle
Amazon Redshift
More...
Saved Data Sources
Sample - Superstore
Cloudera Hadoop
General
Initial SQL
Connection
Impala
Server
192.168.127.128
Port
21050
Authentication
No Authentication
Transport
Binary
Require SSL
Sign In
Installed Connecto
Actian Matrix
Actian Vector
Alibaba AnalyticDB fo
Alibaba Data Lake An
Alibaba MaxCompute
Amazon Athena
Amazon Aurora for M
Amazon EMR Hadoop
Amazon Redshift
Anaplan
Apache Drill
Aster Database
Azure Data Lake Stor
Azure SQL Database
Azure Synapse Analyt
Box
Cloudera Hadoop
Databricks
Datorama
Intuit QuickBooks Online
Kognitio
Kyvos
PostgreSQL
Presto
Progress OpenEdge
Qubole Presto
Salesforce
Salesforce CDP
SAP HANA
SAP NetWeaver Business Warehouse
SAP Sybase ASE
SAP Sybase IQ
ServiceNow ITSM
SharePoint Lists
SingleStore
Snowflake
Spark SQL
Splunk
Teradata
Teradata OLAP Connector
TIBCO Data Virtualization
Vertica
Web Data Connector
Other Databases (JDBC)



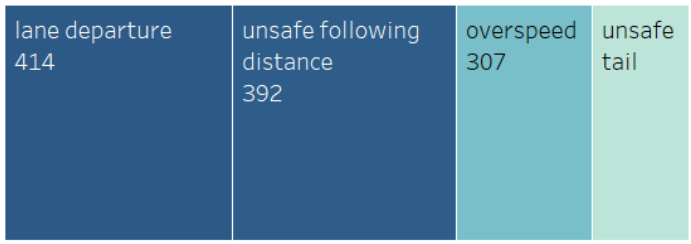
Task 6: Reports and Dashboards

Attached in the PowerPoint presentation.

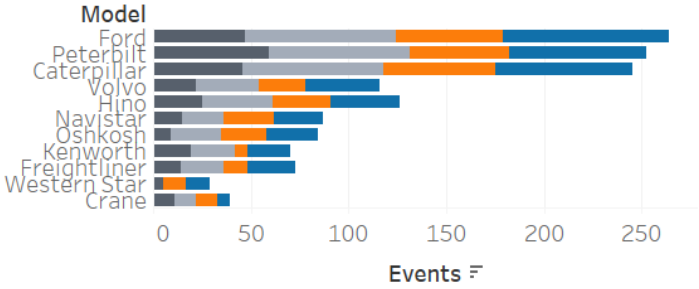
Top 1 Risky Driver (A97) Area



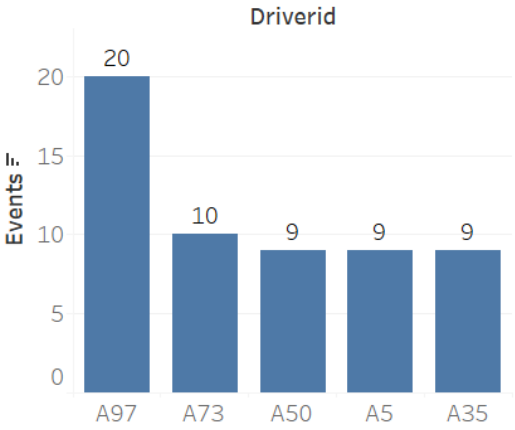
Event Distribution



Event by Truck Model



Top 5 Risky Drivers



Top 5 Cities with most events