

cloudera-training-capspark-student-rev_cdh5.4.3a - VMware Workstation 17 Player (Non-commercial use only)

Player ▾ | Applications Places System | Thu Jul 20, 3:52 PM

Hue - Hive Editor - Query - Mozilla Firefox

Hue - Hive Editor - Query x | localhost:8888/beeswax/#query

Most Visited ▾ | Cloudera | Hue | YARN RM | Spark UI (local) | Spark Doc | Solr Admin UI | Kite SDK Doc

HUE | Query Editors ▾ | Data Browsers ▾ | Workflows ▾ | Search | File Browser | Job Browser | training ▾ |

Hive Editor | Query Editor | My Queries | Saved Queries | History

Assist Settings

DATABASE default

Table name...

accounts_avro
accounts_by_area
avg_mileage
device
drivemileage
geolocation2
geolocation_stage
geolocation_stage1
load
riskfactor
truck_mileage
trucks
trucks_mg

1 Example: `SELECT * FROM tablename, or press CTRL + space`

Execute Save as... Explain or create a New query

Recent queries Query Log Columns Results Chart

Time Query

07/20/2023 12:28:32 AM CREATE TABLE DriverMileage AS SELECT driverid, sum(miles) totmiles FROM truck_mileage GROUP BY driverid;

07/20/2023 12:28:06 AM drop table drivermileage ;

07/20/2023 12:26:51 AM CREATE TABLE sum_mileage AS SELECT truckid, sum(miles) totmiles FROM truck_mileage GROUP BY truckid;

The installed version of VMware Tools is not up to date. Log in to the guest operating system and click "Update Tools."

Update Tools Remind Me Later Never Remind Me

88°F Clear 5:52 PM 7/20/2023

cloudera-training-capspark-student-rev_cdh5.4.3a - VMware Workstation 17 Player (Non-commercial use only)

Player ▾ | Applications Places System | Thu Jul 20, 6:43 PM

Hue - Hive Editor - Query - Mozilla Firefox

Hue - Hive Editor - Query x | localhost:8888/beeswax/#query

Most Visited ▾ | Cloudera | Hue | YARN RM | Spark UI (local) | Spark Doc | Solr Admin UI | Kite SDK Doc

HUE | Query Editors ▾ | Data Browsers ▾ | Workflows ▾ | Search | File Browser | Job Browser | training ▾ |

Hive Editor | Query Editor | My Queries | Saved Queries | History

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	6.11322583689

[training@localhost:~] Hue - Hive Editor - Qu... training@localhost:~/t...

The installed version of VMware Tools is not up to date. Log in to the guest operating system and click "Update Tools."

Update Tools Remind Me Later Never Remind Me

97°F Sunny 8:43 PM 7/20/2023

cloudera-training-capspark-student-rev_cdh5.4.3a - VMware Workstation 17 Player (Non-commercial use only)

Player ▾ | Applications Places System | Thu Jul 20, 6:43 PM

training@localhost:~/training_materials/dev1/scripts

```
File Edit View Search Terminal Help
[training@localhost ~]$ cd $DEV1
[training@localhost dev1]$ cd scripts
[training@localhost scripts]$ pig
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 in
fo.
2023-07-20 18:38:51,507 [main] INFO org.apache.pig.Main - Apache Pig version 0.
12.0-cdh5.4.3 (rexpoted) compiled Jun 24 2015, 19:36:38
2023-07-20 18:38:51,508 [main] INFO org.apache.pig.Main - Logging error message
s to: /home/training/training_materials/dev1/scripts/pig_1689903531485.log
2023-07-20 18:38:51,536 [main] INFO org.apache.pig.impl.util.Utils - Default bo
otup file /home/training/.pigbootup not found
2023-07-20 18:38:52,546 [main] INFO org.apache.hadoop.conf.Configuration.deprec
ation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.addr
ess
2023-07-20 18:38:52,546 [main] INFO org.apache.hadoop.conf.Configuration.deprec
ation - fs.default.name is deprecated. Instead, use fs.defaultFS
2023-07-20 18:38:52,547 [main] INFO org.apache.pig.backend.hadoop.executionengi
ne.HExecutionEngine - Connecting to hadoop file system at: hdfs://localhost:8020
2023-07-20 18:38:54,818 [main] INFO org.apache.hadoop.conf.Configuration.deprec
ation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.addr
ess
2023-07-20 18:38:54,818 [main] INFO org.apache.pig.backend.hadoop.executionengi
ne.HExecutionEngine - Connecting to map-reduce job tracker at: localhost:8021
2023-07-20 18:38:54,820 [main] INFO org.apache.hadoop.conf.Configuration.deprec
ation - fs.default.name is deprecated. Instead, use fs.defaultFS
2023-07-20 18:38:54,988 [main] INFO org.apache.hadoop.conf.Configuration.deprec
ation - fs.default.name is deprecated. Instead, use fs.defaultFS
```

[training@localhost:~] | Hue - Hive Editor - Qu... | training@localhost:~/t...

The installed version of VMware Tools is not up to date. Log in to the guest operating system and click "Update Tools."

Update Tools | Remind Me Later | Never Remind Me

97°F | ENG IN | 8:43 PM 7/20/2023

cloudera-training-capspark-student-rev_cdh5.4.3a - VMware Workstation 17 Player (Non-commercial use only)

Player ▾ | Applications Places System | Thu Jul 20, 6:44 PM

training@localhost:~/training_materials/dev1/scripts

```
File Edit View Search Terminal Help
grunt> c = foreach b generate driverid, event, (int) '1' as occurrence;
2023-07-20 18:39:02,708 [main] ERROR org.apache.pig.tools.grunt.Grunt - ERROR 12
00: Pig script failed to parse:
<line 1, column 12> Undefined alias: b
Details at logfile: /home/training/training_materials/dev1/scripts/pig_168990353
1485.log
grunt> d = group c by driverid;
2023-07-20 18:39:02,719 [main] ERROR org.apache.pig.tools.grunt.Grunt - ERROR 12
00: Pig script failed to parse:
<line 1, column 10> Undefined alias: c
Details at logfile: /home/training/training_materials/dev1/scripts/pig_168990353
1485.log
grunt> e = foreach d generate group as driverid, SUM(c.occurrence) as t_occ;
2023-07-20 18:39:02,732 [main] ERROR org.apache.pig.tools.grunt.Grunt - ERROR 12
00: Pig script failed to parse:
<line 1, column 12> Undefined alias: d
Details at logfile: /home/training/training_materials/dev1/scripts/pig_168990353
1485.log
grunt> g = LOAD 'drivermileage' using org.apache.hive.hcatalog.pig.HCatLoader();
2023-07-20 18:39:02,743 [main] INFO org.apache.hadoop.conf.Configuration.deprec
ation - fs.default.name is deprecated. Instead, use fs.defaultFS
2023-07-20 18:39:02,744 [main] INFO org.apache.hadoop.conf.Configuration.deprec
ation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.addr
ess
2023-07-20 18:39:03,416 [main] INFO org.apache.hadoop.conf.Configuration.deprec
ation - fs.default.name is deprecated. Instead, use fs.defaultFS
2023-07-20 18:39:03,416 [main] INFO org.apache.hadoop.conf.Configuration.deprec
ation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.addr
ess
grunt> h = join e by driverid, g by driverid;
```

[training@localhost:~] | Hue - Hive Editor - Qu... | training@localhost:~/t...

The installed version of VMware Tools is not up to date. Log in to the guest operating system and click "Update Tools."

Update Tools | Remind Me Later | Never Remind Me

97°F | ENG IN | 8:44 PM 7/20/2023

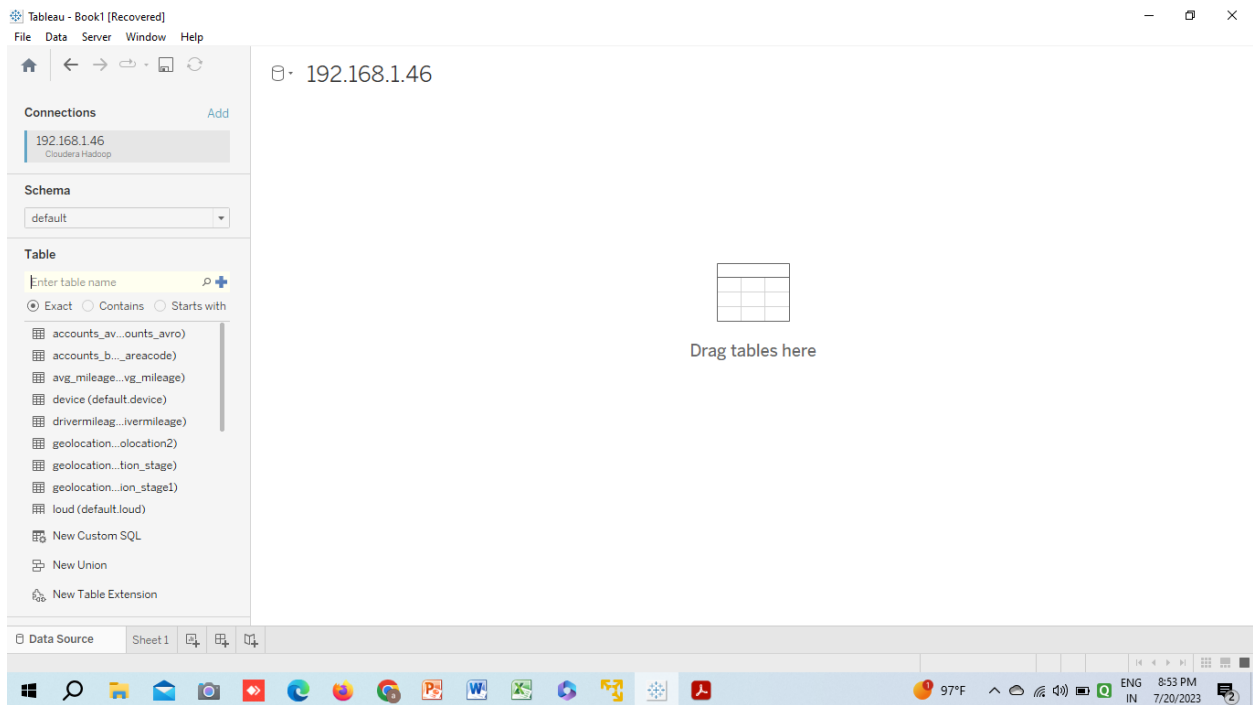
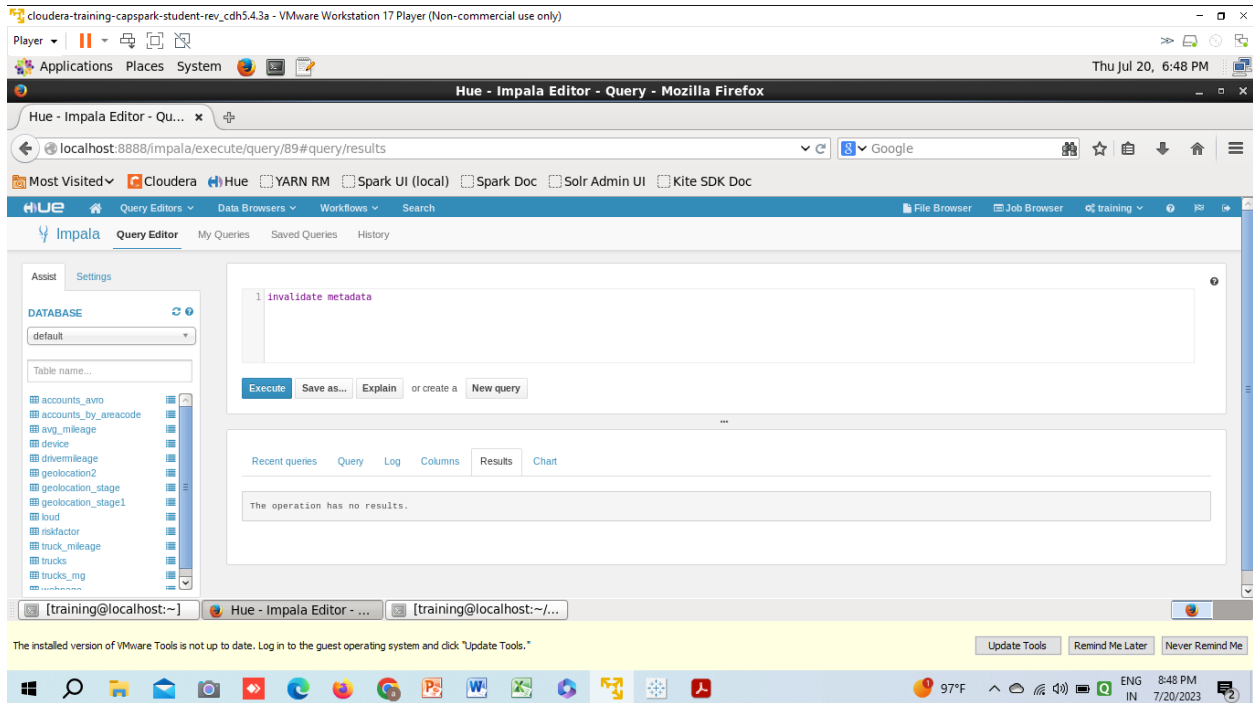


Tableau - Book1 [Recovered]

File Data Server Window Help

192.168.1.46

Connections

192.168.1.46

Schema

default

Table

Enter table name

Exact Contains Starts with

geolocation...ocation2)

geolocation...tion_stage)

geolocation...ion_stage1)

loud (default.loud)

riskfactor (d...t.riskfactor)

truck_milea...ck_mileage)

trucks (default.trucks)

trucks_mg (...trucks_mg)

webpage (de...it.webpage)

New Custom SQL

New Union

New Table Extension

192.168.1.46

Connection

Live Extract

Filters

0 Add

geolocation2

avg_mileage

drivermileage

riskfactor

truck_mileage

riskfactor

5 fields 99 rows

99 rows

Name	Driverid (Riskfactor)	Events	Totmiles	Riskfactor	Riskfactor_calculated
A97		20	631.199	31.6857	20.3493
A73		10	642.222	15.5709	10.0000
A35		9	637.069	14.1272	9.0728
A50		9	640.069	14.0610	9.0303
A5		9	676.574	13.3023	8.5431

Fields

Type	Field Name	Physical Table	Remot...
Abc	Driverid (Riskfactor)	riskfactor	driverid (...)

Data Source

Sheet1

98°F

9:22 PM

7/20/2023

Tableau - Book1 [Recovered]

File Data Server Window Help

192.168.1.46

Connections

192.168.1.46

Schema

default

Table

Enter table name

Exact Contains Starts with

geolocation...ocation2)

geolocation...tion_stage)

geolocation...ion_stage1)

loud (default.loud)

riskfactor (d...t.riskfactor)

truck_milea...ck_mileage)

trucks (default.trucks)

trucks_mg (...trucks_mg)

webpage (de...it.webpage)

New Custom SQL

New Union

New Table Extension

192.168.1.46

Connection

Live Extract Edit Refresh

Filters

0 Add

Extract will contain all data.

geolocation2

avg_mileage

drivermileage

riskfactor

truck_mileage

riskfactor

rows

Name	Driverid (Riskfactor)	Events	Totmiles	Riskfactor	Riskfactor_calculated
A97		20	631.199	31.6857	20.3493
A73		10	642.222	15.5709	10.0000
A35		9	637.069	14.1272	9.0728
A50		9	640.069	14.0610	9.0303
A5		9	676.574	13.3023	8.5431

Fields

Type	Field Name	Physical Table	Remot...
Abc	Driverid (Riskfactor)	riskfactor	driverid (...)

Update Now

Update Automatically

Data Source

Sheet1

98°F

9:24 PM

7/20/2023

Save the data after the above step

Tableau - myfirstprojectbigdatahadoop

Save As

Location: This PC > Documents

File name: myfirstprojectbigdatahadoop2

Save as type: Tableau Packaged Workbook (*.twbx)

Connection: Extract (selected) | Edit | Refresh | Filters: 0 | Add

Extract contains all data: 7/20/2023 9:26:59 PM

Driverid (Riskfactor)	Events	Totmiles	Riskfactor	Riskfactor_calculated
A97	20	631.199	31.6857	20.3493
A73	10	642.222	15.5709	10.0000
A35	9	637.069	14.1272	9.0728
A50	9	640.069	14.0610	9.0303
A5	9	676.574	13.3023	8.5431

Data Source | Sheet 1

Tableau - myfirstprojectbigdatahadoop2

File | Data | Worksheet | Dashboard | Story | Analysis | Map | Format | Server | Window | Help

192.168.1.46

Search

Tables

- avg_mileage
- drivermileage
- geolocation2
- riskfactor
 - Driverid (Riskfactor)
 - Events
 - Riskfactor
 - Riskfactor_calculated
 - Totmiles
 - riskfactor (Count)
- truck_mileage
- trucks
 - Driverid (Trucks)
 - Model
 - Truckid (Trucks)
 - Apr09 Gas
 - Apr09 Miles
 - Apr10 Gas
 - Apr10 Miles
 - Apr11 Gas
 - Apr11 Miles

Columns: Truckid (Trucks)

Rows: SUM(Riskfactor_calculated)

Sheet 1

Truckid (Trucks)

192 marks | 1 row by 101 columns | SUM(Riskfactor_calculated): 453.22

96°F

9:34 PM 7/20/2023

