Problem Set C Submission Form

Overview

Your Name	Hendi Kushta
Your SU Email	hkushta@syr.edu

Instructions

Put your name and SU email at the top. Answer these questions all from the lab. When asked to include screenshots, please follow the screen shot guidelines from the first homework.

Remember as you complete the homework it is not only about getting it right / correct. We will discuss the answers in class so it’s important to articulate anything you would like to contribute to the discussion in your answer:

- If you feel the question is vague, include any assumptions you've made.
- If you feel the answer requires interpretation or justification provide it.
- If you do not know the answer to the question, articulate what you tried and how you are stuck.
- Highlight any doubts or questions you would like me to review.

This how you receive credit for answering questions which might not be correct. In addition, you must complete the reflection portion of the homework assignment for full credit. Since most answers will be similar this is an important part of your individual submission.

Complete Part II of this document first, then go back and complete the Reflection in Part I.

Part I - Reflection

Use this section to reflect on your learning. To achieve the highest grade on the assignment you must be as descriptive and personal as possible with your reflection.

1. As you completed this assignment, identify what you learned.

   Learned how to load data in HDFS from the command line, create internal nad external Hive tables, create a Spark session configured to integrate with Hive.

2. What barriers or challenges did you encounter while completing this assignment?

   Question number 7. It looks like the version of Hive is not compatible in Pyspark to overwrite the data as CSV file in HDFS system.

3. How prepared were you to complete this assignment? What can you do to be better prepared?

4. Rate your comfort level with this week’s material. Use the rubric provided.

4 ==> I understand this material and can explain it to others.
3 ==> I understand this material.
2 ==> I somewhat understand the material but sometimes need guidance from others.
1 ==> I understand very little of this material and need extra help.

Part II – Questions

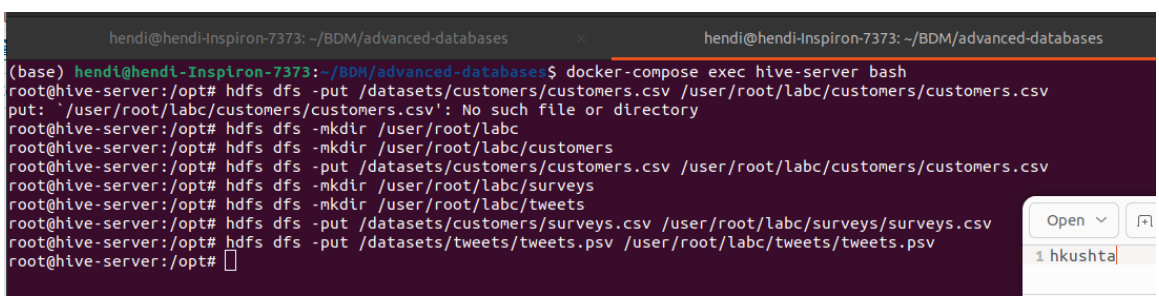
Paste your answers to the Exercises found in the lab document. Make sure to include your netid in any screenshots you provide. If the question asks for commands, only include those commands which are necessary to complete the tasks. Number each answer.

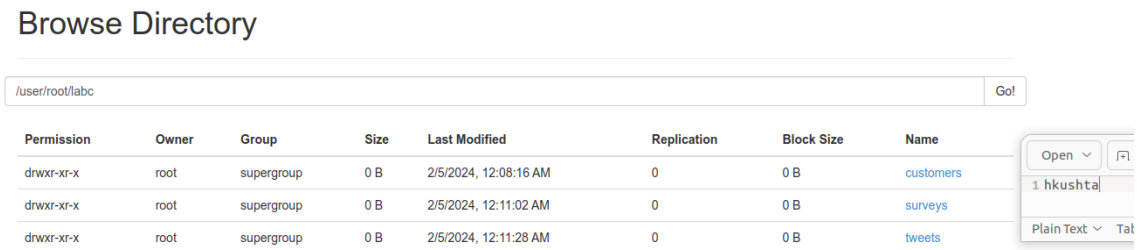
1. Connect to the Linux shell on the **hive-server** (this is where the Hadoop client has been installed for you.) On this server you will see the **/datasets** folder is mounted. Load the:
   a. customers/customers.csv,
   b. customers/surveys.csv, and
   c. tweets/tweets.psv into HDFS.

   Specifically:

Source	HDFS Location
customers/customers.csv	/user/root/labc/customers/customers.csv
customers/surveys.csv	/user/root/labc/surveys/surveys.csv
tweets/tweets.psv	/user/root/labc/tweets/tweets.psv

   Record the Hadoop commands you entered to complete this task. provide a screenshot of evidence these files are in HDFS. The screen shot can use the Hadoop client output or the HDFS website.





2. Create a Hive database called **labc**. In the **labc** database create an external hive table for the **tweets**. Your external table will point to the existing location on HDFS.
   NOTE: You will need to view the tweets.psv file to see the format of the file before you can create the table schema correctly.

   After you create the table write a SELECT query to display all of the tweets for a user a single user of your choice. Please include the HQL code you wrote to create and query the **tweets** table. Along with screenshots of a **describe tweets** command output along with your SELECT query output.

```
0: jdbc:hive2://hive-server:10000/default> create database labc;
No rows affected (0.079 seconds)
0: jdbc:hive2://hive-server:10000/default> show databases;
+----------------+
| database_name  |
+----------------+
| default        |
| demo           |
| labc           |
+----------------+
3 rows selected (0.079 seconds)
0: jdbc:hive2://hive-server:10000/default>
```

```
Open    *Un..
1 hkushta
Plain Text    Tab Width: 8
```

```
0: jdbc:hive2://hive-server:10000/default> create external table tweets (
. . . . . . . . . . . . . . . . . . . . .> tweet_id bigint,
. . . . . . . . . . . . . . . . . . . . .> tweet_decimal decimal,
. . . . . . . . . . . . . . . . . . . . .> tweet_date string,
. . . . . . . . . . . . . . . . . . . . .> tweet_user string,
. . . . . . . . . . . . . . . . . . . . .> tweet_text string)
. . . . . . . . . . . . . . . . . . . . .> row format delimited
. . . . . . . . . . . . . . . . . . . . .> fields terminated by '|'
. . . . . . . . . . . . . . . . . . . . .> location '/user/root/labc/tweets/';
No rows affected (0.839 seconds)
```

```
Open
1 hkushta
Plain Text
```

```
0: jdbc:hive2://hive-server:10000/default> select tweet_text from tweets where tweet_user = 'sladd';
+--------------------------------------------------+
|                    tweet_text                    |
+--------------------------------------------------+
| Just placed an order today. thx @fudgemart       |
| Worst purchase ever: digital downloads at @fudgemart. |
| I love you, @fudgemart                           |
| I make my first order from fudgemart last week.  |
| This is fudgemart. The e-commerce platform.      |
| Happy with my service from @fudgemart            |
| Fudgemart just is.                               |
| There are new digital downloads on #fudgemart    |
+--------------------------------------------------+
8 rows selected (0.386 seconds)
```

```
Open    *Un...    Save
1 hkushta
Plain Text    Tab Width: 8    Ln 1, Col 8
```

```
0: jdbc:hive2://hive-server:10000/default> describe tweets;
+----------------+----------------+----------+
|    col_name    |   data_type    | comment  |
+----------------+----------------+----------+
| tweet_id       | bigint         |          |
| tweet_decimal  | decimal(10,0)  |          |
| tweet_date     | string         |          |
| tweet_user     | string         |          |
| tweet_text     | string         |          |
+----------------+----------------+----------+
5 rows selected (0.092 seconds)
```

3. In the **labc** database, let's create an internal hive table for **customers**. After you create the table, use the LOAD command to move the data from the current HDFS location into the Hive data warehouse.

NOTE 1: if you screw up you will need to drop table and reload the file back into HDFS from step 1.

NOTE 2: there is a header row in this file, you might need to search the Hive docs on the web for how to exclude this first row.

When you have created the table and imported the data, provide the HQL code you entered to complete the task and provide screenshots of the **describe customers** command, a SELECT output to show data is there, and a screenshot on Web HDFS to show the data is located in **/user/hive/warehouse**.

```
jdbc:hive2://hive-server:10000/default> create table customers (
 . . . . . . . . . . . . . . . . . . .> cfirst string, clast string, cemail string,
 . . . . . . . . . . . . . . . . . . .> cgender string, cip string, ccity string, cstate string,
 . . . . . . . . . . . . . . . . . . .> ctotalorders int, ctotalpurchased int, cmonthscustomer int)
 . . . . . . . . . . . . . . . . . . .> row format delimited fields terminated by ',';
rows affected (0.319 seconds)
```

```
0: jdbc:hive2://hive-server:10000/default> describe customers;
+-------------------+------------+----------+
|     col_name      | data_type  | comment  |
+-------------------+------------+----------+
| cfirst            | string     |          |
| clast             | string     |          |
| cemail            | string     |          |
| cgender           | string     |          |
| cip               | string     |          |
| ccity             | string     |          |
| cstate            | string     |          |
| ctotalorders      | int        |          |
| ctotalpurchased   | int        |          |
| cmonthscustomer   | int        |          |
+-------------------+------------+----------+
10 rows selected (0.104 seconds)
```

```
0: jdbc:hive2://hive-server:10000/default> load data local inpath '/datasets/customers/customers.csv' overwrite into table customers;
No rows affected (1.125 seconds)
0: jdbc:hive2://hive-server:10000/default> alter table customers set tblproperties ("skip.header.line.count"="1");
No rows affected (0.187 seconds)
0: jdbc:hive2://hive-server:10000/default> select * from customers limit 2;
+------------------+-----------------+-------------------+------------------+---------------+----------------+-----------------+-----------------------+-----------------
---------+--------------------------+
| customers.cfirst  | customers.clast  |  customers.cemail   | customers.cgender  | customers.cip  | customers.ccity  | customers.cstate  | customers.ctotalorders  | customers.ctotalp
urchased  | customers.cmonthscustomer  |
+------------------+-----------------+-------------------+------------------+---------------+----------------+-----------------+-----------------------+-----------------
---------+--------------------------+
| Al               | Fresco          | afresco@dayrep.com  | M                | 74.111.18.161  | Syracuse        | NY              | 1                     | 45
        | 1                        |
| Abby             | Kuss            | akuss@rhyta.com     | F                | 23.80.125.101  | Phoenix         | AZ              | 1                     | 25
        | 2                        |
+------------------+-----------------+-------------------+------------------+---------------+----------------+-----------------+-----------------------+-----------------
---------+--------------------------+
2 rows selected (0.189 seconds)
```

Browse Directory

/user/hive/warehouse/customers

Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
-rwxrwxr-x	root	supergroup	2.05 KB	2/5/2024, 11:52:35 PM	3	128 MB	customers.csv

4. Like the previous step, import the surveys.csv into a Hive internal table in the **labc** database called **surveys**. When you have created the table and imported the data, provide all the commands you entered to complete the task, a screenshots of the table description, the select statement output, and Web HDFS location.

```
0: jdbc:hive2://hive-server:10000/default> create table surveys (
. . . . . . . . . . . . . . . . . . . . .> semail string, stwitter string,
. . . . . . . . . . . . . . . . . . . . .> smstatus string, sincome int, sownhome string,
. . . . . . . . . . . . . . . . . . . . .> seducation string, sfavdept string )
. . . . . . . . . . . . . . . . . . . . .> row format delimited fields terminated by ','
. . . . . . . . . . . . . . . . . . . . .> tblproperties ("skip.header.line.count"="1");
No rows affected (0.081 seconds)
```

```
0: jdbc:hive2://hive-server:10000/default> describe surveys;
+-------------+------------+----------+
|  col_name   | data_type  | comment  |
+-------------+------------+----------+
| semail      | string     |          |
| stwitter    | string     |          |
| smstatus    | string     |          |
| sincome     | int        |          |
| sownhome    | string     |          |
| seducation  | string     |          |
| sfavdept    | string     |          |
+-------------+------------+----------+
7 rows selected (0.083 seconds)
```

```
0: jdbc:hive2://hive-server:10000/default> load data local inpath '/datasets/customers/surveys.csv' overwrite into table surveys;

No rows affected (0.356 seconds)
0: jdbc:hive2://hive-server:10000/default> select * from surveys limit 2;
+-------------------------+-------------------+-------------------+-------------------+-------------------+----------------------+-
------------------+
|     surveys.semail      | surveys.stwitter  | surveys.smstatus  | surveys.sincome   | surveys.sownhome  | surveys.seducation   |
surveys.sfavdept  |
+-------------------------+-------------------+-------------------+-------------------+-------------------+----------------------+-
------------------+
| ojouglad@einrot.com     | ojouglad          | Married           | 65000             | No                | 4 Year Degree        |
Electronics       |
| lkarfurless@dayrep.com  | lkarfurless       | Single            | 143000            | Yes               | Graduate Degree      |
Apparel           |
+-------------------------+-------------------+-------------------+-------------------+-------------------+----------------------+-
------------------+
2 rows selected (0.193 seconds)
```

Open *Un... Save
1 hkushta

Open *Un... Save
1 hkushta
Plain Text Tab Width: 8 Ln 1, Col 8 INS

Browse Directory

/user/hive/warehouse/surveys Go!

Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
-rwxrwxr-x	root	supergroup	1.95 KB	2/6/2024, 12:12:47 AM	3	128 MB	surveys.csv

5. Open Jupyter Lab. Create a new notebook called **labc.** Copy over the code from an example to create a spark session connected to Hive.

   In a separate cell, write Spark SQL code to join customers to surveys on email address. Include all rows and columns and show output in the notebook. Provide a screenshot of the notebook cell with a reasonable amount of output (doesn't need to be the entire set of rows and columns as that will be too large).

```
[8]: query = '''
select c.*, s.*
    from labc.customers c
    join labc.surveys s
      on c.cemail = s.semail
    where cemail != 'Email';
'''

spark.sql(query).show()
```

Open *Un... Save
1 hkushta
Plain Text Tab Width: 8 Ln 1, Col 8 INS

cfirst	clast	cemail	cgender	cip	ccity	cstate	ctotalorders	ctotalpurchased	cmonthscustomer	semail	stwitter	smstatus	sincome	sownhome	seducation	sfavdept
Al	Fresco	afresco@dayrep.com	M	74.111.18.161	Syracuse	NY	1	45	1	afresco@dayrep.com	afresco	Married	45000	No	High School	Apparel
Abby	Kuss	akuss@rhyta.com	F	23.80.125.101	Phoenix	AZ	1	25	2	akuss@rhyta.com	akuss	Single	22500	No	High School	None
Bette	Alott	balott@rhyta.com	F	56.216.127.219	Raleigh	NC	6	560	18	balott@rhyta.com	balott	Married	null	Prefer not to Answer	Graduate Degree	Apparel
Barb	Barion	bbarion@superrito...	F	38.68.15.223	Dallas	TX	4	1590	1	bbarion@superrito...	bbarion	Single	74000	No	4 Year Degree	Jewelry
Barry	DeHatchett	bdehatchett@dayre...	M	23.192.215.78	Boston	MA	1	15	6	bdehatchett@dayre...	bdehatchett	Prefer not to Answer	67000	Yes	4 Year Degree	Electronics
Bill	Melator	bmelator@einrot.com	M	24.11.125.10	Orem	UT	9	6090	35	bmelator@einrot.com	bmelator	Single	13000	No	2 Year Degree	Digital Downloads
Candi	Cayne	ccayne@rhyta.com	F	24.39.14.15	Portland	ME	1	620	2	ccayne@rhyta.com	ccayne	Married	62000	Yes	4 Year Degree	None
Cam	Rha	crha@einrot.com	M	24.1.25.140	Chicago	IL	0	0	1	crha@einrot.com	crha	Married	34000	Prefer not to Answer	4 Year Degree	Books
Dan	Delyons	ddelyons@dayrep.com	M	24.38.224.161	Greenwich	CT	2	2570	10	ddelyons@dayrep.com	ddelyons	Single	105000	Yes	High School	Electronics
Erin	Detyers	edetyers@dayrep.com	F	70.209.14.54	Tampa	FL	5	1105	38	edetyers@dayrep.com	edetyers	Single	null	No	Prefer not to Answer	Apparel
Euron	Tasomthin	etasomthin@superr...	M	68.199.40.156	Hempstead	NY	13	4630	28	etasomthin@superr...	etasomthin	Married	39000	No	2 Year Degree	Prefer not to Answer
Jean	Poole	jpoole@dayrep.com	F	23.182.25.40	Kingston	NY	7	185	12	jpoole@dayrep.com	jpoole	Married	null	Yes	Some College	Books
Lee	Hvmeehom	lhvmeehom@einrot.com	F	215.82.23.2	Columbus	OH	9	207	18	lhvmeehom@einrot.com	lhvmeehom	Prefer not to Answer	75000	Yes	4 Year Degree	Prefer not to Answer
Lisa	Karfurless	lkarfurless@dayre...	F	172.189.252.8	Fairfax	VA	6	250	27	lkarfurless@dayre...	lkarfurless	Single	143000	Yes	Graduate Degree	Apparel
Mary	Melator	mmelator@rhyta.com	F	23.88.15.5	Los Angeles	CA	8	4275	40	mmelator@rhyta.com	mmelator	Prefer not to Answer	42000	No	4 Year Degree	Books
Mike	Rofone	mrofone@dayrep.com	M	23.224.160.4	Cheyenne	WY	0	0	0	mrofone@dayrep.com	mrofone	Single	121000	Yes	Prefer not to Answer	Jewelry
Oren	Jouglad	ojouglad@einrot.com	M	128.122.140.238	New York	NY	12	4500	36	ojouglad@einrot.com	ojouglad	Married	65000	No	4 Year Degree	Electronics
Rowan	Deboat	rdeboat@dayrep.com	M	23.84.32.22	Topeka	KS	1	3500	42	rdeboat@dayrep.com	rdeboat	Single	69000	Yes	Graduate Degree	Electronics
Ray	Ovlight	rovlight@dayrep.com	M	74.111.18.59	Syracuse	NY	6	125	42	rovlight@dayrep.com	rovlight	Prefer not to Answer	28000	No	2 Year Degree	Digital Downloads
Sara	Bellum	sbellum@superrito...	F	74.111.6.173	Alexandria	VA	2	189	2	sbellum@superrito...	sbellum	Married	100000	Yes	Graduate Degree	Jewelry

only showing top 20 rows

6. The marketing department would like a dataset of customers / surveys for analysis. In a separate cell in the **labc** Juypter Notebook, write a Spark SQL query to create a hive table called **marketing** in **AVRO** file format from a SELECT query that once again joins customers and surveys on email addresses. Include the following columns in the new table: **Household Income**, **Education**, **Marital Status**, **Gender**, **City** and **State**.

   Provide a screenshot of the Jupyter cell and output that creates the new table, and another of the cell and output of executing a SELECT on the table.

```
query = '''
select c.cgender, c.ccity, c.cstate, s.smstatus, s.seducation, s.sincome
    from labc.customers c
    join labc.surveys s
        on c.cemail = s.semail
    where cemail != 'Email';
'''

spark.sql(query).show()
```

Open *Un... Save

1 hkushta

Plain Text Tab Width: 8 Ln 1, Col 8 INS

```
+-------+-----------+------+--------------------+--------------------+-------+
|cgender|      ccity|cstate|            smstatus|          seducation|sincome|
+-------+-----------+------+--------------------+--------------------+-------+
|      M|   Syracuse|    NY|             Married|         High School|  45000|
|      F|    Phoenix|    AZ|              Single|         High School|  22500|
|      F|    Raleigh|    NC|             Married|     Graduate Degree|   null|
|      F|     Dallas|    TX|              Single|       4 Year Degree|  74000|
|      M|     Boston|    MA|Prefer not to Answer|       4 Year Degree|  67000|
|      M|       Orem|    UT|              Single|       2 Year Degree|  13000|
|      F|   Portland|    ME|             Married|       4 Year Degree|  62000|
|      M|    Chicago|    IL|             Married|       4 Year Degree|  34000|
|      M|  Greenwich|    CT|              Single|         High School| 105000|
|      F|      Tampa|    FL|              Single|Prefer not to Answer|   null|
|      M|  Hempstead|    NY|             Married|       2 Year Degree|  39000|
|      F|   Kingston|    NY|             Married|        Some College|   null|
|      F|   Columbus|    OH|Prefer not to Answer|       4 Year Degree|  75000|
|      F|    Fairfax|    VA|              Single|     Graduate Degree| 143000|
|      F|Los Angeles|    CA|Prefer not to Answer|       4 Year Degree|  42000|
|      M|   Cheyenne|    WY|              Single|Prefer not to Answer| 121000|
|      M|   New York|    NY|             Married|       4 Year Degree|  65000|
|      M|     Topeka|    KS|              Single|     Graduate Degree|  69000|
|      M|   Syracuse|    NY|Prefer not to Answer|       2 Year Degree|  28000|
|      F| Alexandria|    VA|             Married|     Graduate Degree| 100000|
+-------+-----------+------+--------------------+--------------------+-------+
only showing top 20 rows
```

```
query = '''
create table labc.marketing stored as avro as
select c.cgender, c.ccity, c.cstate, s.smstatus, s.seducation, s.sincome
    from labc.customers c
    join labc.surveys s
        on c.cemail = s.semail
    where cemail != 'Email';
'''

spark.sql(query).show()
```

Open *Un... Save

1 hkushta

Plain Text Tab Width: 8 Ln 1, Col 8 INS

```
++
||
++
++
```

```
query = '''
select * from labc.marketing;
'''

spark.sql(query).show()
```

Open *Un... Save

1 hkushta

Plain Text Tab Width: 8 Ln 1, Col 8 INS

cgender	ccity	cstate	smstatus	seducation	sincome
M	Syracuse	NY	Married	High School	45000
F	Phoenix	AZ	Single	High School	22500
F	Raleigh	NC	Married	Graduate Degree	null
F	Dallas	TX	Single	4 Year Degree	74000
M	Boston	MA	Prefer not to Answer	4 Year Degree	67000
M	Orem	UT	Single	2 Year Degree	13000
F	Portland	ME	Married	4 Year Degree	62000
M	Chicago	IL	Married	4 Year Degree	34000
M	Greenwich	CT	Single	High School	105000
F	Tampa	FL	Single	Prefer not to Answer	null
M	Hempstead	NY	Married	2 Year Degree	39000
F	Kingston	NY	Married	Some College	null
F	Columbus	OH	Prefer not to Answer	4 Year Degree	75000
F	Fairfax	VA	Single	Graduate Degree	143000
F	Los Angeles	CA	Prefer not to Answer	4 Year Degree	42000
M	Cheyenne	WY	Single	Prefer not to Answer	121000
M	New York	NY	Married	4 Year Degree	65000
M	Topeka	KS	Single	Graduate Degree	69000
M	Syracuse	NY	Prefer not to Answer	2 Year Degree	28000
F	Alexandria	VA	Married	Graduate Degree	100000

only showing top 20 rows

7. Stupid marketing doesn't know what they want! Now they would like the same query in the previous step, only output as a Comma-Delimited file instead of a Hive table. In a new Jupyter Lab cell, write Spark SQL to execute the Hive query but save the output back to HDFS in the folder **/user/root/marketing**.

   Provide a screenshot of the Spark code cell and its output, as well as a screenshot of the file on Web HDFS.

   Question number 7, I tried to write the Spark code, but it didn't run properly as shown in the screenshot below.

```
: query = '''
insert overwrite directory '/user/root/marketing'
row format delimited fields terminated by ','
select c.cgender, c.ccity, c.cstate, s.smstatus, s.seducation, s.sincome
    from labc.customers c
    join labc.surveys s
        on c.cemail = s.semail
    where cemail != 'Email';;
'''
spark.sql(query).show()
```

Open *Un... Save

1 hkushta

Plain Text Tab Width: 8 Ln 1, Col 8 INS

```
chgrp: changing ownership of                                    33-06_785_820199
of '/tmp/my_staging_dir_hive_                                   Operation not pe
---------------------------------------------------------------------------
Py4JJavaError                             Traceback (most recent call last)
/tmp/ipykernel_614/2694115020.py in <module>
      8     where cemail != 'Email';;
      9 '''
---> 10 spark.sql(query).show()

/usr/local/spark/python/pyspark/sql/session.py in sql(self, sqlQuery)
    721         [Row(f1=1, f2='row1'), Row(f1=2, f2='row2'), Row(f1=3, f2='row3')]
    722         """
--> 723         return DataFrame(self._jsparkSession.sql(sqlQuery), self._wrapped)
    724
    725     def table(self, tableName):

/usr/local/spark/python/lib/py4j-0.10.9-src.zip/py4j/java_gateway.py in __call__(self, *args
   1302
```

   So to finish the exercise, I used Hive to create the marketing file and save the file in HDFS as csv.

Browse Directory

/user/root/marketing Go!

Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
-rwxrwxrwx	root	supergroup	1.09 KB	2/6/2024, 1:44:32 AM	3	128 MB	000000_0

Open

1 hkushta

```
0: jdbc:hive2://hive-server:10000/default> insert overwrite directory '/user/root/marketing/'
. . . . . .> row format delimited fields terminated by ','
. . . . . .> select * from labc.marketing;
WARNING: Hive-on-MR is deprecated in Hive 2 and may not be available in the future versions. Consider using a
different execution engine (i.e. spark, tez) or using Hive 1.X releases.
No rows affected (1.853 seconds)
0: jdbc:hive2://hive-server:10000/default>
```

Open

1 hkushta

Plain Text

```
root@hive-server:/opt# hadoop fs -cat /user/root/marketing/*
M,Syracuse,NY,Married,High School,45000
F,Phoenix,AZ,Single,High School,22500
F,Raleigh,NC,Married,Graduate Degree,\N
F,Dallas,TX,Single,4 Year Degree,74000
M,Boston,MA,Prefer not to Answer,4 Year Degree,67000
M,Orem,UT,Single,2 Year Degree,13000
F,Portland,ME,Married,4 Year Degree,62000
M,Chicago,IL,Married,4 Year Degree,34000
M,Greenwich,CT,Single,High School,105000
F,Tampa,FL,Single,Prefer not to Answer,\N
M,Hempstead,NY,Married,2 Year Degree,39000
F,Kingston,NY,Married,Some College,\N
F,Columbus,OH,Prefer not to Answer,4 Year Degree,75000
F,Fairfax,VA,Single,Graduate Degree,143000
F,Los Angeles,CA,Prefer not to Answer,4 Year Degree,42000
M,Cheyenne,WY,Single,Prefer not to Answer,121000
M,New York,NY,Married,4 Year Degree,65000
M,Topeka,KS,Single,Graduate Degree,69000
M,Syracuse,NY,Prefer not to Answer,2 Year Degree,28000
F,Alexandria,VA,Married,Graduate Degree,100000
M,Rochester,NY,Married,2 Year Degree,52000
M,Cleveland,OH,Single,4 Year Degree,50000
M,San Jose,CA,Single,2 Year Degree,26000
M,Buffalo,NY,Prefer not to Answer,Graduate Degree,89000
M,Green Bay,WI,Prefer not to Answer,High School,17500
root@hive-server:/opt#
```

Open

1 hkushta

Plain Text