

1.



cdacuser5235@ip-172-31-16-2 Hue - File Browser Nuvepro BigData.Exam - Google Docs

npapc.cloudloka.com/shell/

For quick access, place your bookmarks here on the bookmarks bar: [import bookmarks now...](#)

txnrccsbycat6
txnrccsbystate
zerostops
Time taken: 0.133 seconds, Fetched: 20 row(s)
hive (cdac_rohit)> desc airport;
OK
airport_id int
name string
city string
country string
iata string
icao string
latitude double
longitude double
altitude int
timezone double
dst string
tz string
Time taken: 0.048 seconds, Fetched: 12 row(s)
hive (cdac_rohit)> desc routes;
OK
airline_iata string
airline_id int
src_airport_iata string
src_airport_id int
dest_airport_iata string
dest_airport_id int
codeshare string
stops int
equipment string
Time taken: 0.045 seconds, Fetched: 9 row(s)
hive (cdac_rohit)> create table routes_partitioned(airline_iata string,airline_id int,src_airport_iata string,src_airport_id int,dest_airport_iata string,codeshare string,stops int)
> partitioned by (destination_airport_id int) row format
> delimited fields terminated by ',' stored as textfile;
OK
Time taken: 0.135 seconds
hive (cdac_rohit)>

32°C Smoke Search ENG IN 16:59 21-11-2024

Subscription Details | Nuvepro

cdacuser5235@ip-172-31-16-205

Hue - File Browser

Nuvepro Web FTP

BigData.Exam - Google Docs

< > ↺

Not secure npacc.cloudkloka.com:8132/hue/filebrowser/view=%2Fuser%2Fcdacuser5235#/user/hive/warehouse/cdac_rohit.db

VPN

All Bookmarks

For quick access, place your bookmarks here on the bookmarks bar: [Import bookmarks now...](#)

Hue

Search data and saved documents...

Jobs

Files

Filter...

Simplelearn

hadoop

hbase

home

opt

system

tmp

user

Search for file name

Actions Delete forever

Upload New

<input type="checkbox"/>	custbuck	cdacuser5235	hive	drwxr-xr-x	November 20, 2024 11:25 PM
<input type="checkbox"/>	custbyage	cdacuser5235	hive	drwxr-xr-x	November 20, 2024 11:51 PM
<input type="checkbox"/>	customer	cdacuser5235	hive	drwxr-xr-x	November 06, 2024 08:18 PM
<input type="checkbox"/>	emp100	cdacuser5235	hive	drwxr-xr-x	November 09, 2024 05:15 AM
<input type="checkbox"/>	employee	cdacuser5235	hive	drwxr-xr-x	November 11, 2024 04:00 AM
<input type="checkbox"/>	employee_header	cdacuser5235	hive	drwxr-xr-x	November 09, 2024 05:13 AM
<input type="checkbox"/>	nyc_parking_issued	cdacuser5235	hive	drwxr-xr-x	November 13, 2024 04:44 AM
<input type="checkbox"/>	nyse	cdacuser5235	hive	drwxr-xr-x	November 05, 2024 10:42 PM
<input type="checkbox"/>	routes	cdacuser5235	hive	drwxr-xr-x	November 07, 2024 12:50 AM
<input type="checkbox"/>	routes_partitioned	cdacuser5235	hive	drwxr-xr-x	November 21, 2024 03:28 AM
<input type="checkbox"/>	sales	cdacuser5235	hive	drwxr-xr-x	November 09, 2024 05:09 AM
<input type="checkbox"/>	txn_orc	cdacuser5235	hive	drwxr-xr-x	November 09, 2024 02:21 AM
<input type="checkbox"/>	txn_parquet	cdacuser5235	hive	drwxr-xr-x	November 09, 2024 03:09 AM
<input type="checkbox"/>	txnrecsbycat	cdacuser5235	hive	drwxr-xr-x	November 09, 2024 03:57 AM
<input type="checkbox"/>	txnrecsbycat2	cdacuser5235	hive	drwxr-xr-x	November 09, 2024 04:11 AM
<input type="checkbox"/>	txnrecsbycat3	cdacuser5235	hive	drwxr-xr-x	November 11, 2024 01:26 AM
<input type="checkbox"/>	txnrecsbycat4	cdacuser5235	hive	drwxr-xr-x	November 11, 2024 01:12 AM
<input type="checkbox"/>	txnrecsbycat5	cdacuser5235	hive	drwxr-xr-x	November 11, 2024 01:54 AM

2.

```
Insert overwrite table routes_partitioned
partition(dest_airport_iata) select
airline_iata,airline_id,dest_airport_iata from routes_partitioned;
```

```
cdacuser5235@ip-172-31-16-205:~$ hive
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/opt/hive-3.1.1/lib/log4j-slf4j-impl-2.10.0.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/opt/hadoop/share/hadoop/common/lib/slf4j-log4j12-1.7.25.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.apache.logging.slf4j.Log4jLoggerFactory]
Hive Session ID = 0aafbb54-bc95-4ca8-aa71-077d634a6e98

Logging initialized using configuration in jar:file:/opt/hive-3.1.1/lib/hive-common-3.1.1.jar!/hive-log4j2.properties Async: true
Hive Session ID = 894beb89-9154-4b51-8e1a-a09673d21217
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.
hive> create table routes_partitioned(airline_iata string,airline_id int,src_airport_iata string,src_airport_id int,dest_airport_iata string,codeshare string,stops int)

> set hive.cli.current.db=true;
FAILED: ParseException line 2:0 missing EOF at 'set' near ''''
hive> set hive.cli.print.current.db=true;
hive (default)> use cdac_rohit;
OK
Time taken: 0.081 seconds
hive (cdac_rohit)> Insert overwrite table routes_partitioned partition(dest_airport_iata) select airline_iata,airline_id,dest_airport_iata from routes_partitioned;
FAILED: ValidationFailureSemanticException cdac_rohit.routes_partitioned: Partition spec {dest_airport_iata=null} contains non-partition columns
hive (cdac_rohit)> Insert overwrite table routes_partitioned partition(stops) select airline_iata,airline_id,stops from routes_partitioned;
FAILED: ValidationFailureSemanticException cdac_rohit.routes_partitioned: Partition spec {stops=null} contains non-partition columns
hive (cdac_rohit)> Insert overwrite table routes_partitioned partition(codeshare) select airline_iata,airline_id,codeshare from routes_partitioned;
FAILED: ValidationFailureSemanticException cdac_rohit.routes_partitioned: Partition spec {codeshare=null} contains non-partition columns
hive (cdac_rohit)> []
```

4.

```
Display routes_partitioned();
```

SPARK

3.

```
Subscription Details | Nuvepro Hue - File Browser cdcuser5235@ip-172-31-16-2 Nuvepro Web FTP BigData.Exam - Google Docs
npapc.cloudloka.com/shell/
For quick access, place your bookmarks here on the bookmarks bar. import bookmarks now... All Bookmarks

[GCC 11.2.0] :: Anaconda, Inc. on linux
Type "help", "copyright", "credits" or "license" for more information.
24/11/21 12:19:15 WARN NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Setting default log level to "WARN".
To adjust logging level use sc.setLogLevel(newLevel). For SparkR, use setLogLevel(newLevel).
24/11/21 12:19:16 WARN Utils: Service 'SparkUI' could not bind on port 4040. Attempting port 4041.
24/11/21 12:19:16 WARN Utils: Service 'SparkUI' could not bind on port 4041. Attempting port 4042.
24/11/21 12:19:16 WARN Utils: Service 'SparkUI' could not bind on port 4042. Attempting port 4043.
24/11/21 12:19:16 WARN Utils: Service 'SparkUI' could not bind on port 4043. Attempting port 4044.
24/11/21 12:19:16 WARN Utils: Service 'SparkUI' could not bind on port 4044. Attempting port 4045.
24/11/21 12:19:16 WARN Utils: Service 'SparkUI' could not bind on port 4045. Attempting port 4046.
24/11/21 12:19:16 WARN Utils: Service 'SparkUI' could not bind on port 4046. Attempting port 4047.
24/11/21 12:19:16 WARN Utils: Service 'SparkUI' could not bind on port 4047. Attempting port 4048.
24/11/21 12:19:16 WARN Utils: Service 'SparkUI' could not bind on port 4048. Attempting port 4049.
24/11/21 12:19:17 WARN Client: Neither spark.yarn.jars nor spark.yarn.archive is set, falling back to uploading libraries under SPARK_HOME.
Welcome to

      ____
     / ___/
    / __/
   /___/
  version 3.1.2

Using Python version 3.9.13 (main, Aug 25 2022 23:26:10)
Spark context Web UI available at http://ip-172-31-16-205.ap-south-1.compute.internal:4049
Spark context available as 'sc' (master = yarn, app id = application_1732089968849_3040).
SparkSession available as 'spark'.
>>> Df=spark.read.format("csv").option("header",True).option("inferSchema",True).load("/user/cdcuser5235/seat/seats.csv")
>>> Df.groupBy("Quarter").avg("Avg_rev_per_seat").show()
+-----+
|Quarter|avg(Avg_rev_per_seat)|
+-----+
|1|330.61238095238093|
|3|327.5557142857143|
|4|328.48285714285714|
|2|332.33904761904756|
+-----+

>>>
```

4.

```
Subscription Details | Nuvepro Hue - File Browser cdcuser5235@ip-172-31-16-2 Nuvepro Web FTP BigData.Exam - Google Docs
npapc.cloudloka.com/shell/
For quick access, place your bookmarks here on the bookmarks bar. import bookmarks now... All Bookmarks

>>> ut.groupby("Quarter").avg("Avg_rev_per_seat").show()
+-----+
|Quarter|avg(Avg_rev_per_seat)|
+-----+
|1|330.61238095238093|
|3|327.5557142857143|
|4|328.48285714285714|
|2|332.33904761904756|
+-----+

>>> from pyspark.sql.functions import *
>>> Df.groupBy("Year").avg("Avg_rev_per_seat").show()
+-----+
|Year|avg(Avg_rev_per_seat)|
+-----+
|2003|315.4675|
|2007|325.14|
|2015|377.1275|
|2006|328.3|
|2013|382.0025|
|1997|287.155|
|2014|391.7|
|2004|305.875|
|1996|276.8925|
|1998|309.285|
|2012|374.675|
|2009|310.61|
|1995|292.2475|
|2001|319.7975|
|2005|307.185|
|2000|339.0325|
|2010|335.8325|
|2011|363.63250000000005|
|2008|346.1575|
|1999|324.0575|
+-----+
only showing top 20 rows

>>>
```

```
Subscription Details | Nuvepro Hue - File Browser cdcuser5235@ip-172-31-16-2 Nuvepro Web FTP BigData.Exam - Google Docs
npapc.cloudloka.com/shell/
For quick access, place your bookmarks here on the bookmarks bar. import bookmarks now... All Bookmarks

|2015|          377.1275|
|2006|          328.3 |
|2013|          382.0025|
|1997|          287.155|
|2014|          391.7 |
|2004|          305.875|
|1996|          276.8925|
|1998|          309.285|
|2012|          374.675|
|2009|          310.61 |
|1995|          292.2475|
|2001|          319.7975|
|2005|          307.185|
|2000|          339.0325|
|2010|          335.8325|
|2011| 363.6325000000005|
|2008|          346.1575|
|1999|          324.0575|
+-----+
only showing top 20 rows

>>> Df.groupBy("Year").avg("Avg_rev_per_seat<290").show()
Traceback (most recent call last):
  File "<stdin>", line 1, in <module>
  File "/opt/spark-3.1.2/python/pyspark/sql/group.py", line 41, in _api
    jdf = getattr(self._jgd, name)(_to_seq(self.sql_ctx._sc, cols))
  File "/opt/spark-3.1.2/python/lib/py4j-0.10.9-src.zip/py4j/java_gateway.py", line 1304, in _call_
  File "/opt/spark-3.1.2/python/pyspark/sql/utlis.py", line 117, in deco
    raise converted from None
pyspark.sql.utils.AnalysisException: Cannot resolve column name "Avg_rev_per_seat<290" among (Year, Quarter, Avg_rev_per_seat, booked_seats)
>>> Df.groupby("Year","Quarter").agg(sum("booked_seats").alias("Booked_seats")).orderBy("Booked_seats",ascending=False).limit(1).show()
+-----+-----+
|Year|Quarter|Booked_seats|
+-----+-----+
|2010|      1|      49678|
+-----+-----+

>>> |
```

1.

Average

```
Subscription Details | Nuvepro Hue - File Browser cdcuser5235@ip-172-31-16-2 Nuvepro Web FTP BigData.Exam - Google Docs
npapc.cloudloka.com/shell/
For quick access, place your bookmarks here on the bookmarks bar. import bookmarks now... All Bookmarks

|2015| 4| 1.548602528E7|
|2008| 4| 1.5459962940000001E7|
|2012| 2| 1.5447962530000001E7|
|2006| 3| 1.533935592E7|
|2001| 1| 1.524724957E7|
|2010| 4| 1.515749928E7|
|2011| 3| 1.512474598E7|
+-----+
only showing top 20 rows

>>> Df.groupBy("Year").agg(sum("Avg_rev_per_seat").alias("Avg_rev_per_Year")).show()
+-----+-----+
|Year| Avg_rev_per_Year|
+-----+-----+
|2003|      1261.87|
|2007|      1300.56|
|2015|      1508.51|
|2006|      1313.2 |
|2013|      1528.01|
|1997|      1148.62|
|2014|      1566.8 |
|2004|      1223.5 |
|1996|      1107.57|
|1998|      1237.14|
|2012|      1498.7 |
|2009|      1242.44|
|1995|      1168.99|
|2001|      1279.19|
|2005|      1228.74|
|2000|      1356.13|
|2010|      1343.33|
|2011| 1454.5300000000002|
|2008|      1384.63|
|1999|      1296.23|
+-----+
only showing top 20 rows

>>> |
```

Maximum

```
Subscription Details | Nuvepro Hue - File Browser cdcuser5235@ip-172-31-16-2 Nuvepro Web FTP BigData.Exam - Google Docs +
npapc.cloudloka.com/shell/
For quick access, place your bookmarks here on the bookmarks bar. import bookmarks now... All Bookmarks

|Year| Avg_rev_per_Year|
+-----+
|2003| 1261.87|
|2007| 1300.56|
|2015| 1508.51|
|2006| 1313.2|
|2013| 1528.01|
|1997| 1148.62|
|2014| 1566.8|
|2004| 1223.5|
|1996| 1107.57|
|1998| 1237.14|
|2012| 1498.7|
|2009| 1242.44|
|1995| 1168.99|
|2001| 1279.19|
|2005| 1228.74|
|2000| 1356.13|
|2010| 1343.33|
|2011| 1454.5300000000002|
|2008| 1384.63|
|1999| 1296.23|
+-----+
only showing top 20 rows

>>> Df.groupby("Year").agg(sum("Avg_rev_per_seat").alias("Avg_rev_per_Year")).orderBy("Avg_rev_per_Year",ascending=False).limit(1).show()
File "<stdin>", line 1
Df.groupby("Year").agg(sum("Avg_rev_per_seat").alias("Avg_rev_per_Year")).orderBy("Avg_rev_per_Year",ascending=False).limit(1).show()
^
SyntaxError: invalid syntax
>>> Df.groupby("Year").agg(sum("Avg_rev_per_seat").alias("Avg_rev_per_Year")).orderBy("Avg_rev_per_Year",ascending=False).limit(1).show()
+-----+
|Year|Avg_rev_per_Year|
+-----+
|2014| 1566.8|
+-----+

>>>
```

Minimum

```
Subscription Details | Nuvepro Hue - File Browser cdcuser5235@ip-172-31-16-2 Nuvepro Web FTP BigData.Exam - Google Docs +
npapc.cloudloka.com/shell/
For quick access, place your bookmarks here on the bookmarks bar. import bookmarks now... All Bookmarks

|1997| 1148.62|
|2014| 1566.8|
|2004| 1223.5|
|1996| 1107.57|
|1998| 1237.14|
|2012| 1498.7|
|2009| 1242.44|
|1995| 1168.99|
|2001| 1279.19|
|2005| 1228.74|
|2000| 1356.13|
|2010| 1343.33|
|2011| 1454.5300000000002|
|2008| 1384.63|
|1999| 1296.23|
+-----+
only showing top 20 rows

>>> Df.groupby("Year").agg(sum("Avg_rev_per_seat").alias("Avg_rev_per_Year")).orderBy("Avg_rev_per_Year",ascending=False).limit(1).show()
File "<stdin>", line 1
Df.groupby("Year").agg(sum("Avg_rev_per_seat").alias("Avg_rev_per_Year")).orderBy("Avg_rev_per_Year",ascending=False).limit(1).show()
^
SyntaxError: invalid syntax
>>> Df.groupby("Year").agg(sum("Avg_rev_per_seat").alias("Avg_rev_per_Year")).orderBy("Avg_rev_per_Year",ascending=False).limit(1).show()
+-----+
|Year|Avg_rev_per_Year|
+-----+
|2014| 1566.8|
+-----+

>>> Df.groupby("Year").agg(sum("Avg_rev_per_seat").alias("Avg_rev_per_Year")).orderBy("Avg_rev_per_Year",ascending=True).limit(1).show()
+-----+
|Year|Avg_rev_per_Year|
+-----+
|1996| 1107.57|
+-----+

>>>
```

