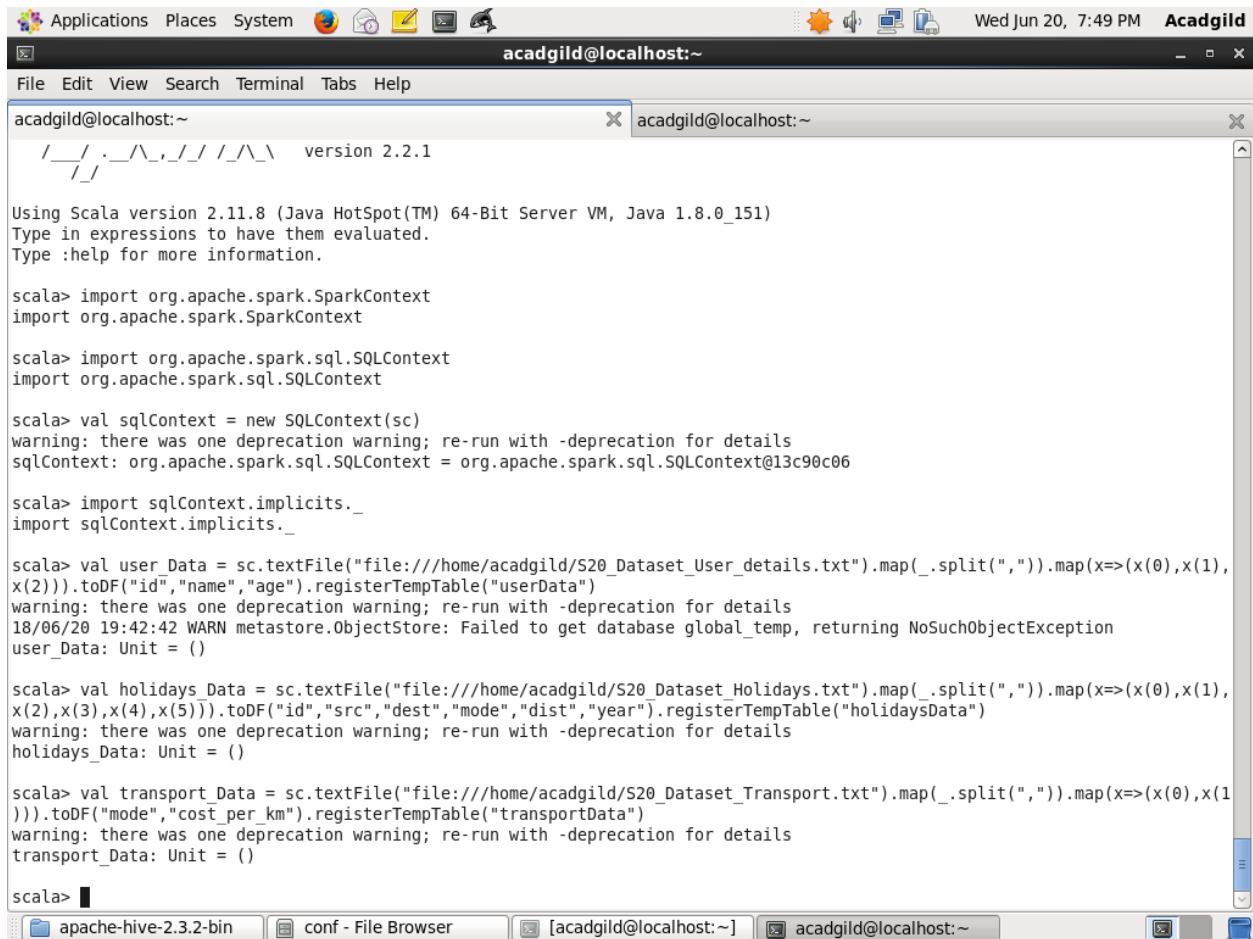


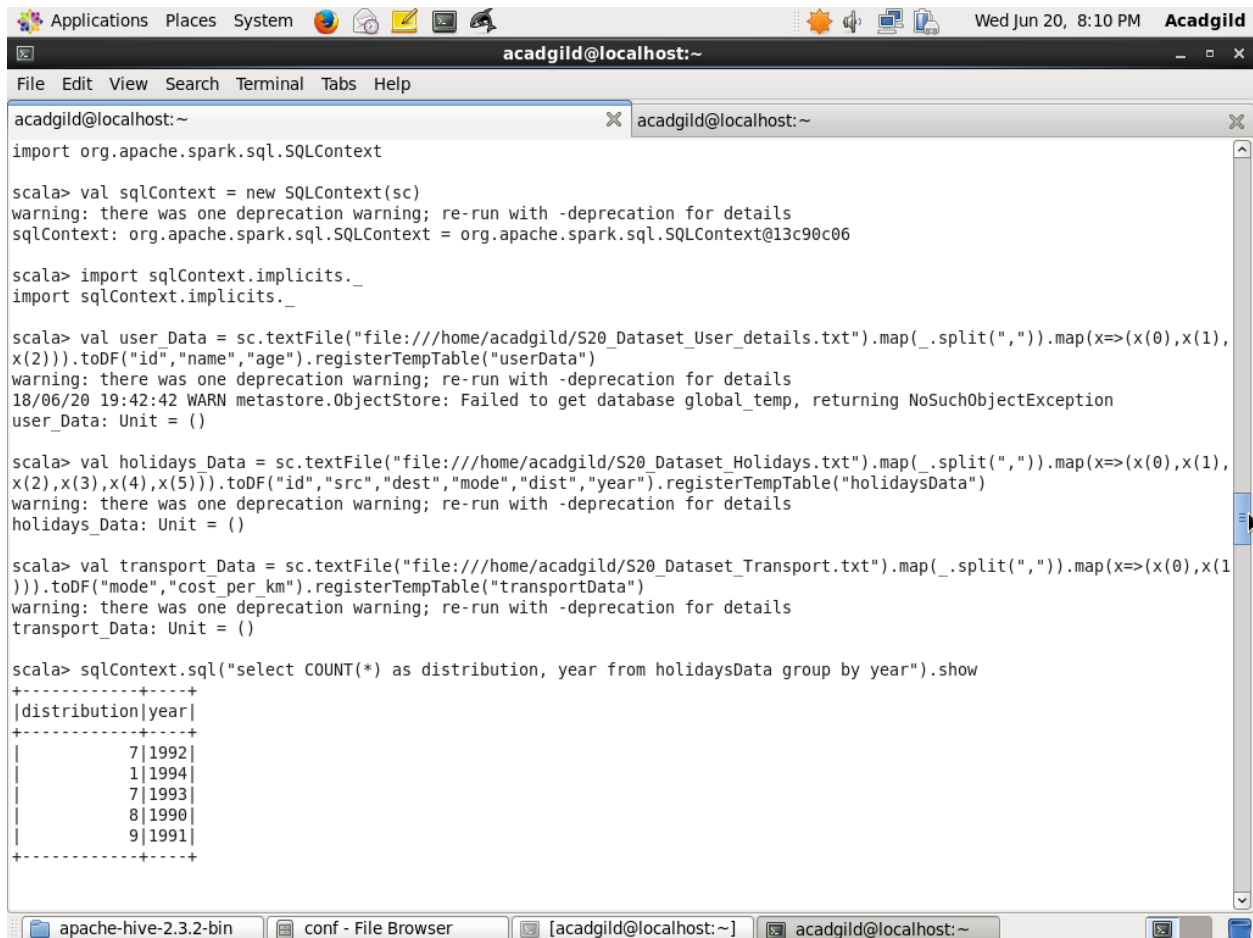
[illegible]

Loading data as dataframe with column name and registering as temp table



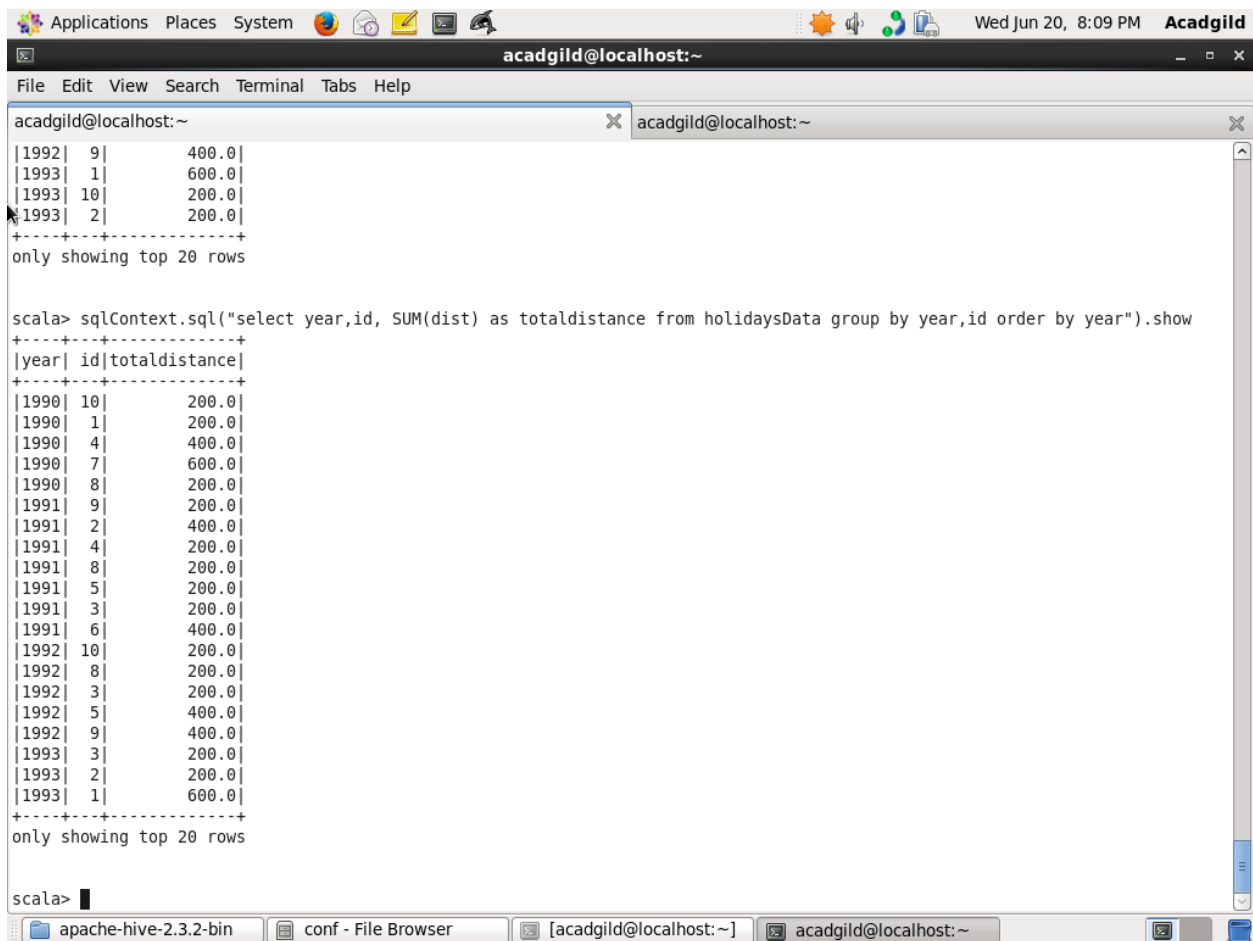
```
acadmild@localhost:~  
File Edit View Search Terminal Tabs Help  
acadmild@localhost:~  
/___/ . _/\_ , _/_ / _/\_ \ version 2.2.1  
/___/  
Using Scala version 2.11.8 (Java HotSpot(TM) 64-Bit Server VM, Java 1.8.0_151)  
Type in expressions to have them evaluated.  
Type :help for more information.  
  
scala> import org.apache.spark.SparkContext  
import org.apache.spark.SparkContext  
  
scala> import org.apache.spark.sql.SQLContext  
import org.apache.spark.sql.SQLContext  
  
scala> val sqlContext = new SQLContext(sc)  
warning: there was one deprecation warning; re-run with -deprecation for details  
sqlContext: org.apache.spark.sql.SQLContext = org.apache.spark.sql.SQLContext@13c90c06  
  
scala> import sqlContext.implicits._  
import sqlContext.implicits._  
  
scala> val user_Data = sc.textFile("file:///home/acadmild/S20_Dataset_User_details.txt").map(_.split(",")).map(x=>(x(0),x(1),  
x(2))).toDF("id","name","age").registerTempTable("userData")  
warning: there was one deprecation warning; re-run with -deprecation for details  
18/06/20 19:42:42 WARN metastore.ObjectStore: Failed to get database global_temp, returning NoSuchObjectException  
user_Data: Unit = ()  
  
scala> val holidays_Data = sc.textFile("file:///home/acadmild/S20_Dataset_Holidays.txt").map(_.split(",")).map(x=>(x(0),x(1),  
x(2),x(3),x(4),x(5))).toDF("id","src","dest","mode","dist","year").registerTempTable("holidaysData")  
warning: there was one deprecation warning; re-run with -deprecation for details  
holidays_Data: Unit = ()  
  
scala> val transport_Data = sc.textFile("file:///home/acadmild/S20_Dataset_Transport.txt").map(_.split(",")).map(x=>(x(0),x(1),  
x(2),x(3),x(4),x(5))).toDF("mode","cost_per_km").registerTempTable("transportData")  
warning: there was one deprecation warning; re-run with -deprecation for details  
transport_Data: Unit = ()  
  
scala> 
```

Finding distribution of total number of air travellers per year



```
acadgild@localhost:~  
File Edit View Search Terminal Tabs Help  
acadgild@localhost:~ acadgild@localhost:~  
import org.apache.spark.sql.SQLContext  
  
scala> val sqlContext = new SQLContext(sc)  
warning: there was one deprecation warning; re-run with -deprecation for details  
sqlContext: org.apache.spark.sql.SQLContext = org.apache.spark.sql.SQLContext@13c90c06  
  
scala> import sqlContext.implicits._  
import sqlContext.implicits._  
  
scala> val user_Data = sc.textFile("file:///home/acadgild/S20_Dataset_User_details.txt").map(_.split(",")).map(x=>(x(0),x(1),  
x(2))).toDF("id","name","age").registerTempTable("userData")  
warning: there was one deprecation warning; re-run with -deprecation for details  
18/06/20 19:42:42 WARN metastore.ObjectStore: Failed to get database global_temp, returning NoSuchObjectException  
user_Data: Unit = ()  
  
scala> val holidays_Data = sc.textFile("file:///home/acadgild/S20_Dataset_Holidays.txt").map(_.split(",")).map(x=>(x(0),x(1),  
x(2),x(3),x(4),x(5))).toDF("id","src","dest","mode","dist","year").registerTempTable("holidaysData")  
warning: there was one deprecation warning; re-run with -deprecation for details  
holidays_Data: Unit = ()  
  
scala> val transport_Data = sc.textFile("file:///home/acadgild/S20_Dataset_Transport.txt").map(_.split(",")).map(x=>(x(0),x(1),  
x(2))).toDF("mode","cost_per_km").registerTempTable("transportData")  
warning: there was one deprecation warning; re-run with -deprecation for details  
transport_Data: Unit = ()  
  
scala> sqlContext.sql("select COUNT(*) as distribution, year from holidaysData group by year").show  
+-----+  
|distribution|year|  
+-----+  
|          7|1992|  
|          1|1994|  
|          7|1993|  
|          8|1990|  
|          9|1991|  
+-----+
```

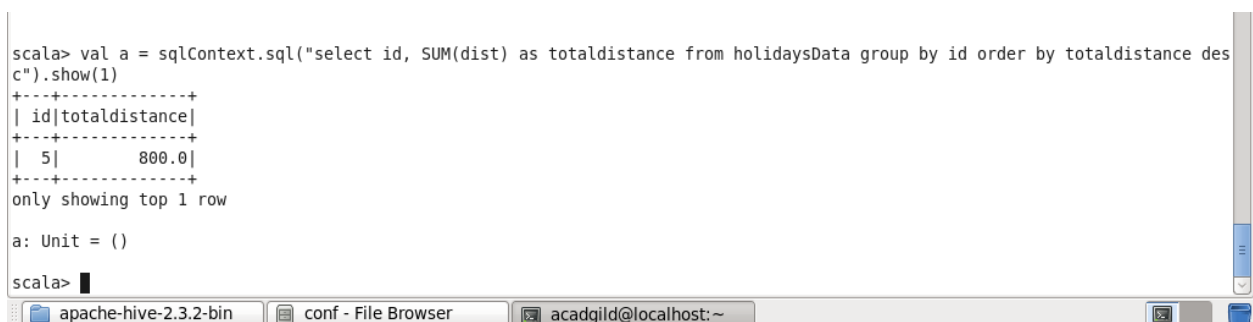
Total distance covered by each user every year



The terminal window shows the execution of a SQL query to calculate the total distance covered by each user every year. The query is: `select year, id, SUM(dist) as totaldistance from holidaysData group by year, id order by year`. The results are displayed in a table format, showing the year, user ID, and total distance. The results are grouped by year and ordered by year. The first 20 rows are shown, with a message indicating that only the top 20 rows are displayed.

```
acadmild@localhost:~  
|1992| 9| 400.0|  
|1993| 1| 600.0|  
|1993| 10| 200.0|  
|1993| 2| 200.0|  
+-----+  
only showing top 20 rows  
  
scala> sqlContext.sql("select year, id, SUM(dist) as totaldistance from holidaysData group by year, id order by year").show  
+-----+  
|year| id|totaldistance|  
+-----+  
|1990| 10| 200.0|  
|1990| 1| 200.0|  
|1990| 4| 400.0|  
|1990| 7| 600.0|  
|1990| 8| 200.0|  
|1991| 9| 200.0|  
|1991| 2| 400.0|  
|1991| 4| 200.0|  
|1991| 8| 200.0|  
|1991| 5| 200.0|  
|1991| 3| 200.0|  
|1991| 6| 400.0|  
|1992| 10| 200.0|  
|1992| 8| 200.0|  
|1992| 3| 200.0|  
|1992| 5| 400.0|  
|1992| 9| 400.0|  
|1993| 3| 200.0|  
|1993| 2| 200.0|  
|1993| 1| 600.0|  
+-----+  
only showing top 20 rows  
  
scala>
```

Finding user who has travelled the most distance :



The terminal window shows the execution of a SQL query to find the user who has travelled the most distance. The query is: `select id, SUM(dist) as totaldistance from holidaysData group by id order by totaldistance desc`. The results are displayed in a table format, showing the user ID and total distance. The results are grouped by user ID and ordered by total distance in descending order. The first row is shown, with a message indicating that only the top 1 row is displayed.

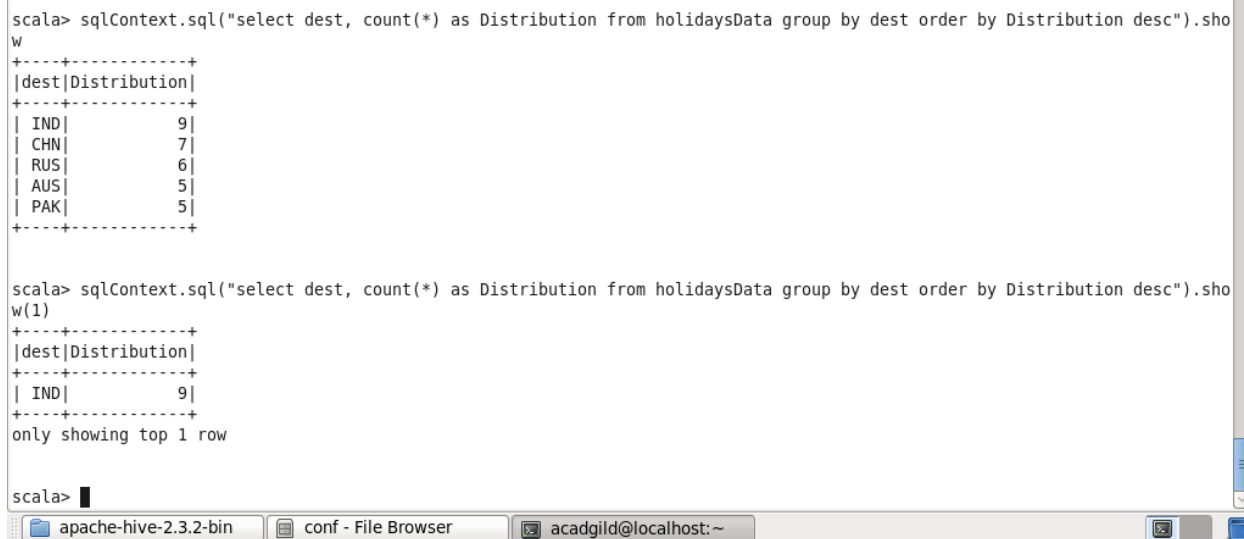
```
scala> val a = sqlContext.sql("select id, SUM(dist) as totaldistance from holidaysData group by id order by totaldistance desc").show(1)  
+-----+  
| id|totaldistance|  
+-----+  
| 5| 800.0|  
+-----+  
only showing top 1 row  
  
a: Unit = ()  
  
scala>
```

Finding most preferred distance for all users:

```
scala> sqlContext.sql("select dest, count(*) as Distribution from holidaysData group by dest order by Distribution desc").show
w
+-----+-----+
|dest|Distribution|
+-----+-----+
| IND|          9|
| CHN|          7|
| RUS|          6|
| AUS|          5|
| PAK|          5|
+-----+-----+

scala> sqlContext.sql("select dest, count(*) as Distribution from holidaysData group by dest order by Distribution desc").show(1)
+-----+-----+
|dest|Distribution|
+-----+-----+
| IND|          9|
+-----+-----+
only showing top 1 row

scala> █
```



Route generating most revenue per year:

Applications Places System Wed Jun 20, 8:52 PM Acadgild

acadgild@localhost:~

File Edit View Search Terminal Tabs Help

acadgild@localhost:~

acadgild@localhost:~

1990	CHN	RUS	1
1991	IND	PAK	1
1991	AUS	PAK	1
+-----+
only showing top 20 rows

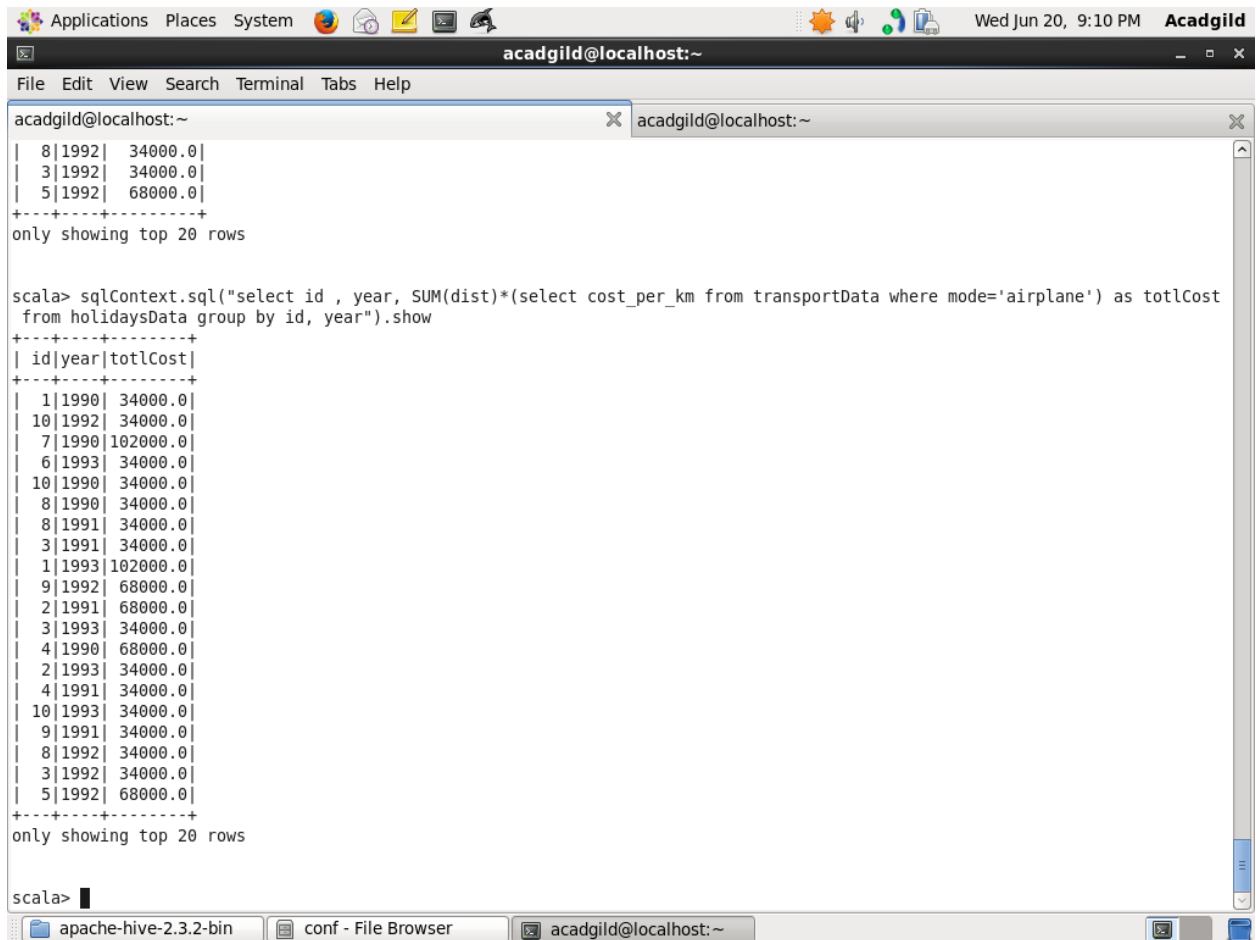
scala> sqlContext.sql("select year, src, dest , COUNT(*) as distribution from holidaysData group by year, src, dest order by distribution desc").show

+-----+
|year|src|dest|distribution|
+-----+
1991	IND	AUS	2
1993	CHN	IND	2
1992	CHN	RUS	2
1993	AUS	CHN	2
1991	IND	RUS	2
1992	RUS	IND	2
1990	CHN	IND	2
1990	RUS	AUS	1
1994	CHN	PAK	1
1990	RUS	IND	1
1992	AUS	IND	1
1991	CHN	PAK	1
1992	RUS	CHN	1
1990	CHN	PAK	1
1991	IND	CHN	1
1993	PAK	AUS	1
1990	AUS	CHN	1
1990	CHN	RUS	1
1991	IND	PAK	1
1991	AUS	PAK	1
+-----+
only showing top 20 rows

scala>

apache-hive-2.3.2-bin conf - File Browser acadgild@localhost:~

Amount spent by every user on air travel per year :



The screenshot shows a terminal window titled 'acadgild@localhost:~' with a menu bar (File, Edit, View, Search, Terminal, Tabs, Help). The terminal displays the following content:

```
acacgild@localhost:~
| 8|1992| 34000.0|
| 3|1992| 34000.0|
| 5|1992| 68000.0|
+-----+
only showing top 20 rows

scala> sqlContext.sql("select id , year, SUM(dist)*(select cost_per_km from transportData where mode='airplane') as totlCost
from holidaysData group by id, year").show
+-----+
| id|year|totlCost|
+-----+
| 1|1990| 34000.0|
| 10|1992| 34000.0|
| 7|1990|102000.0|
| 6|1993| 34000.0|
| 10|1990| 34000.0|
| 8|1990| 34000.0|
| 8|1991| 34000.0|
| 3|1991| 34000.0|
| 1|1993|102000.0|
| 9|1992| 68000.0|
| 2|1991| 68000.0|
| 3|1993| 34000.0|
| 4|1990| 68000.0|
| 2|1993| 34000.0|
| 4|1991| 34000.0|
| 10|1993| 34000.0|
| 9|1991| 34000.0|
| 8|1992| 34000.0|
| 3|1992| 34000.0|
| 5|1992| 68000.0|
+-----+
only showing top 20 rows

scala> 
```

The taskbar at the bottom shows three open applications: 'apache-hive-2.3.2-bin', 'conf - File Browser', and 'acacgild@localhost:~'.