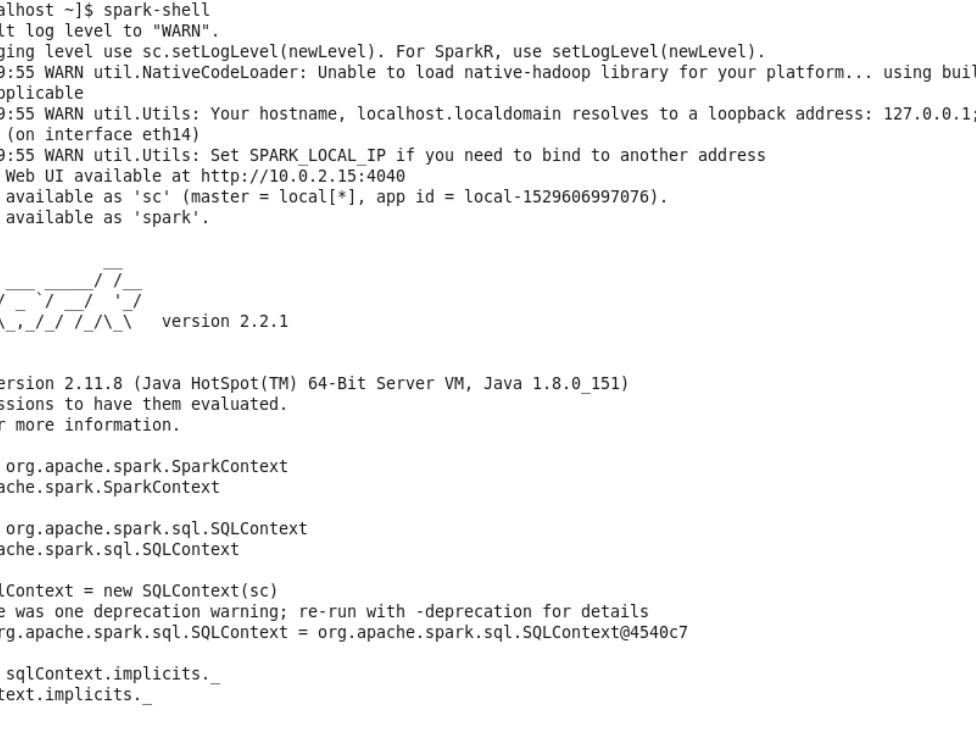
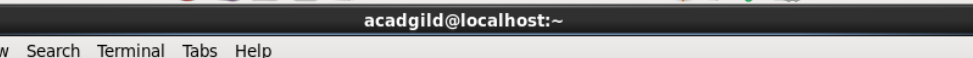


Creating SQL context and importing implicits:



```
acadgild@localhost:~  
File Edit View Search Terminal Help  
[acadgild@localhost ~]$ spark-shell  
Setting default log level to "WARN".  
To adjust logging level use sc.setLogLevel(newLevel). For SparkR, use setLogLevel(newLevel).  
18/06/22 00:19:55 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable  
18/06/22 00:19:55 WARN util.Utils: Your hostname, localhost.localdomain resolves to a loopback address: 127.0.0.1; using 10.0.2.15 instead (on interface eth14)  
18/06/22 00:19:55 WARN util.Utils: Set SPARK_LOCAL_IP if you need to bind to another address  
Spark context Web UI available at http://10.0.2.15:4040  
Spark context available as 'sc' (master = local[*], app id = local-1529606997076).  
Spark session available as 'spark'.  
Welcome to  
  
      ____  
     /   /  
    /___/  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@4540c7  
  
scala> import sqlContext.implicits._  
import sqlContext.implicits._  
  
scala> █
```

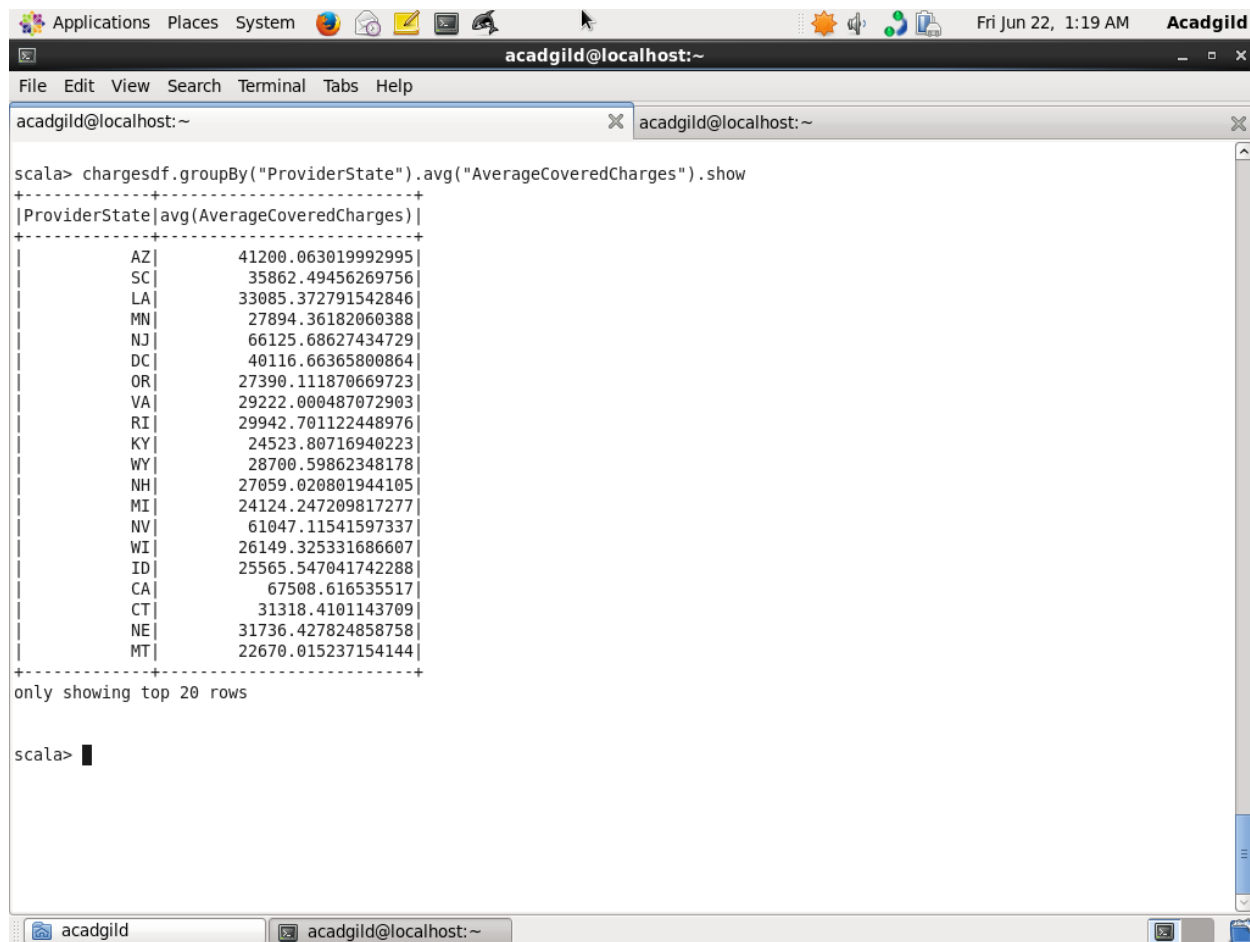
Loading csv file:



The screenshot shows a terminal window with the title bar "Acadgild" and the prompt "acadgild@localhost:~". The terminal contains the following text:

```
scala>
scala>
scala> val chargesdf = sqlContext.read.option("header","true").option("inferSchema","true").csv("file:///home/acadgild/inpatientCharges.csv")
chargesdf: org.apache.spark.sql.DataFrame = [DRGDefinition: string, ProviderId: int ... 10 more fields]
scala> chargesdf.createTempView("TD_HOSPITAL_CHARGES")
```

Average amount of AverageCoveredCharges per state :

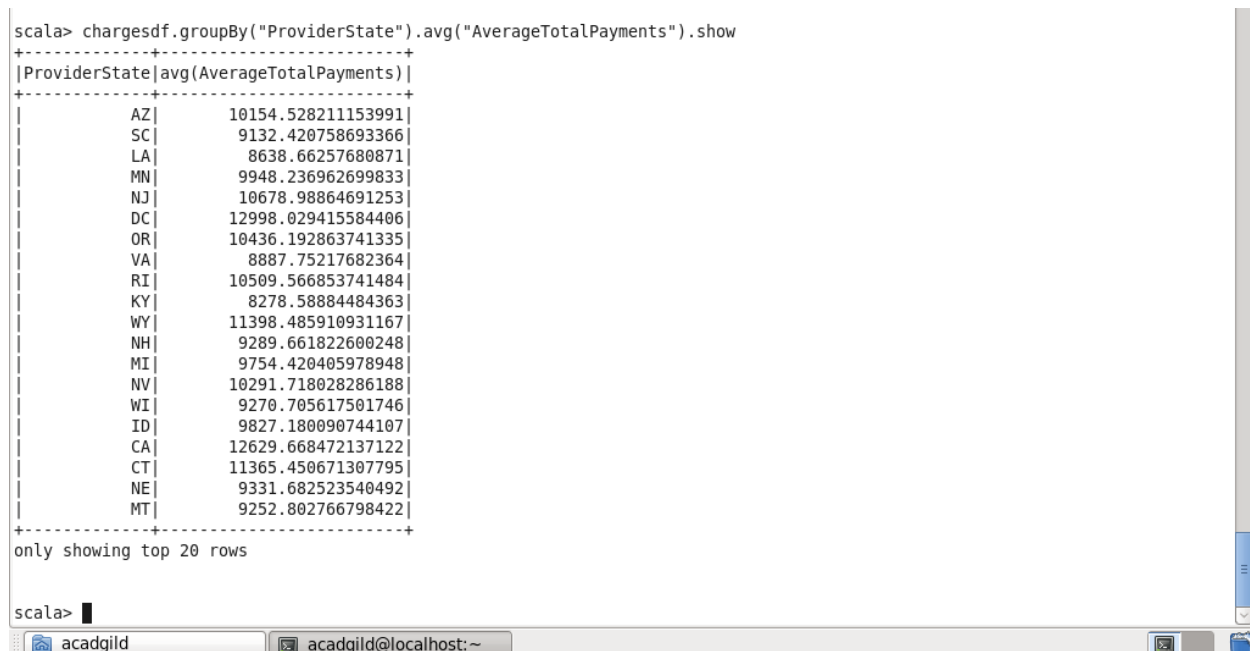


The screenshot shows a Scala REPL window titled 'acadgild@localhost:~'. The command executed is `chargesdf.groupBy("ProviderState").avg("AverageCoveredCharges").show`. The output is a table with two columns: 'ProviderState' and 'avg(AverageCoveredCharges)'. The table lists 20 states with their corresponding average values. The window also shows a taskbar at the bottom with icons for 'acadgild' and 'acadgild@localhost:~'.

```
scala> chargesdf.groupBy("ProviderState").avg("AverageCoveredCharges").show
+-----+
|ProviderState|avg(AverageCoveredCharges)|
+-----+
|AZ|41200.063019992995|
|SC|35862.49456269756|
|LA|33085.372791542846|
|MN|27894.36182060388|
|NJ|66125.68627434729|
|DC|40116.66365800864|
|OR|27390.111870669723|
|VA|29222.000487072903|
|RI|29942.701122448976|
|KY|24523.80716940223|
|WY|28700.59862348178|
|NH|27059.020801944105|
|MI|24124.247209817277|
|NV|61047.11541597337|
|WI|26149.325331686607|
|ID|25565.547041742288|
|CA|67508.616535517|
|CT|31318.4101143709|
|NE|31736.427824858758|
|MT|22670.015237154144|
+-----+
only showing top 20 rows

scala>
```

AverageTotalPayments charges per state:



The screenshot shows a Scala REPL window titled 'acadgild@localhost:~'. The command executed is `chargesdf.groupBy("ProviderState").avg("AverageTotalPayments").show`. The output is a table with two columns: 'ProviderState' and 'avg(AverageTotalPayments)'. The table lists 20 states with their corresponding average values. The window also shows a taskbar at the bottom with icons for 'acadgild' and 'acadgild@localhost:~'.

```
scala> chargesdf.groupBy("ProviderState").avg("AverageTotalPayments").show
+-----+
|ProviderState|avg(AverageTotalPayments)|
+-----+
|AZ|10154.528211153991|
|SC|9132.420758693366|
|LA|8638.66257680871|
|MN|9948.236962699833|
|NJ|10678.98864691253|
|DC|12998.029415584406|
|OR|10436.192863741335|
|VA|8887.75217682364|
|RI|10509.566853741484|
|KY|8278.58884484363|
|WY|11398.485910931167|
|NH|9289.661822600248|
|MI|9754.420405978948|
|NV|10291.718028286188|
|WI|9270.705617501746|
|ID|9827.180090744107|
|CA|12629.668472137122|
|CT|11365.450671307795|
|NE|9331.682523540492|
|MT|9252.802766798422|
+-----+
only showing top 20 rows

scala>
```

Average Medicare payments per state

only showing top 20 rows

```
scala> chargesdf.groupBy("ProviderState").avg("AverageMedicarePayments").show
```

| ProviderState | avg(AverageMedicarePayments) |
|---------------|------------------------------|
| AZ | 8825.717239565045 |
| SC | 7876.33152441167 |
| LA | 7387.704625041281 |
| MN | 8619.214982238007 |
| NJ | 9586.940055946912 |
| DC | 11811.967705627709 |
| OR | 9035.259961508847 |
| VA | 7538.847006001846 |
| RI | 9317.939115646255 |
| KY | 7185.227810467647 |
| WY | 9539.392024291496 |
| NH | 8124.506852976913 |
| MI | 8662.157756043543 |
| NV | 8747.602828618963 |
| WI | 8002.597911079731 |
| ID | 8461.977513611617 |
| CA | 11494.381677893474 |
| CT | 10104.592943809059 |
| NE | 7992.6272504707995 |
| MT | 7981.088063241104 |

only showing top 20 rows

```
scala>
```

acadgild acadgild@localhost: ~

total number of Discharges per state and for each disease :

```
scala> chargesdf.groupBy("ProviderState","DRGDefinition").sum("TotalDischarges").show
```

18/06/22 01:25:14 WARN executor.Executor: Managed memory leak detected; size = 17039360 bytes, TID = 509

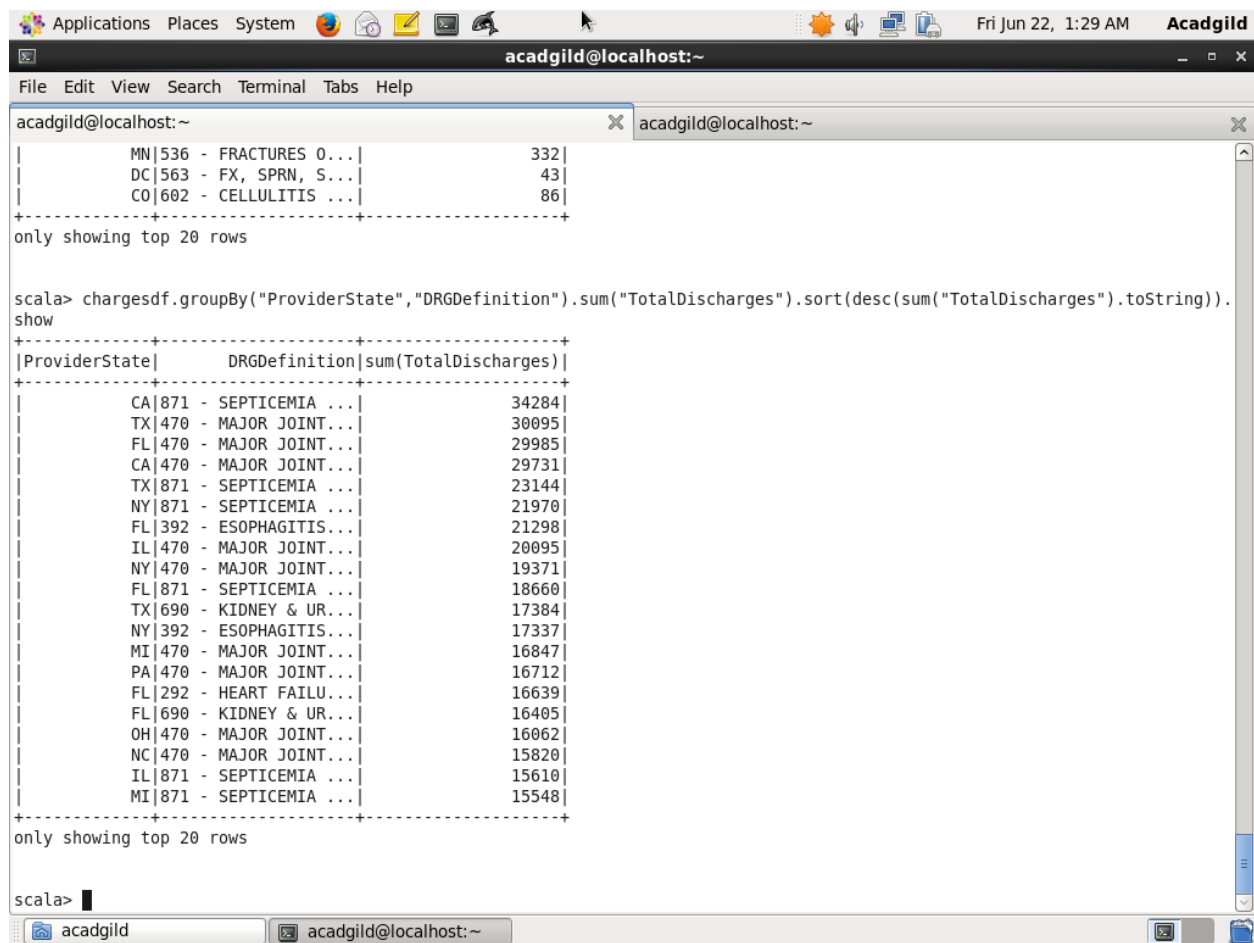
| ProviderState | DRGDefinition | sum(TotalDischarges) |
|---------------|------------------|----------------------|
| KY 065 | - INTRACRANIA... | 1937 |
| NY 101 | - SEIZURES W/... | 4503 |
| IN 149 | - DYSEQUILIBRIUM | 700 |
| IA 178 | - RESPIRATORY... | 540 |
| WI 202 | - BRONCHITIS ... | 338 |
| MO 208 | - RESPIRATORY... | 1840 |
| WI 251 | - PERC CARDIO... | 417 |
| AR 281 | - ACUTE MYOCA... | 413 |
| AZ 292 | - HEART FAILU... | 2643 |
| NY 292 | - HEART FAILU... | 13289 |
| NV 293 | - HEART FAILU... | 519 |
| SD 303 | - ATHEROSCLER... | 53 |
| TN 305 | - HYPERTENSIO... | 730 |
| ME 308 | - CARDIAC ARR... | 312 |
| NV 372 | - MAJOR GASTR... | 126 |
| WA 392 | - ESOPHAGITIS... | 3148 |
| WI 439 | - DISORDERS O... | 215 |
| MN 536 | - FRACTURES O... | 332 |
| DC 563 | - FX, SPRN, S... | 43 |
| CO 602 | - CELLULITIS ... | 86 |

only showing top 20 rows

```
scala>
```

acadgild acadgild@localhost: ~

Sort the output in descending order of totalDischarges :



```
acadmild@localhost:~  
File Edit View Search Terminal Tabs Help  
acadmild@localhost:~  
| MN|536 - FRACTURES 0...| 332|  
| DC|563 - FX, SPRN, S...| 43|  
| CO|602 - CELLULITIS ...| 86|  
+-----+  
only showing top 20 rows  
  
scala> chargesdf.groupBy("ProviderState","DRGDefinition").sum("TotalDischarges").sort(desc(sum("TotalDischarges").toString)).  
show  
+-----+  
|ProviderState| DRGDefinition|sum(TotalDischarges)|  
+-----+  
| CA|871 - SEPTICEMIA ...| 34284|  
| TX|470 - MAJOR JOINT...| 30095|  
| FL|470 - MAJOR JOINT...| 29985|  
| CA|470 - MAJOR JOINT...| 29731|  
| TX|871 - SEPTICEMIA ...| 23144|  
| NY|871 - SEPTICEMIA ...| 21970|  
| FL|392 - ESOPHAGITIS...| 21298|  
| IL|470 - MAJOR JOINT...| 20095|  
| NY|470 - MAJOR JOINT...| 19371|  
| FL|871 - SEPTICEMIA ...| 18660|  
| TX|690 - KIDNEY & UR...| 17384|  
| NY|392 - ESOPHAGITIS...| 17337|  
| MI|470 - MAJOR JOINT...| 16847|  
| PA|470 - MAJOR JOINT...| 16712|  
| FL|292 - HEART FAILU...| 16639|  
| FL|690 - KIDNEY & UR...| 16405|  
| OH|470 - MAJOR JOINT...| 16062|  
| NC|470 - MAJOR JOINT...| 15820|  
| IL|871 - SEPTICEMIA ...| 15610|  
| MI|871 - SEPTICEMIA ...| 15548|  
+-----+  
only showing top 20 rows  
  
scala> |
```