

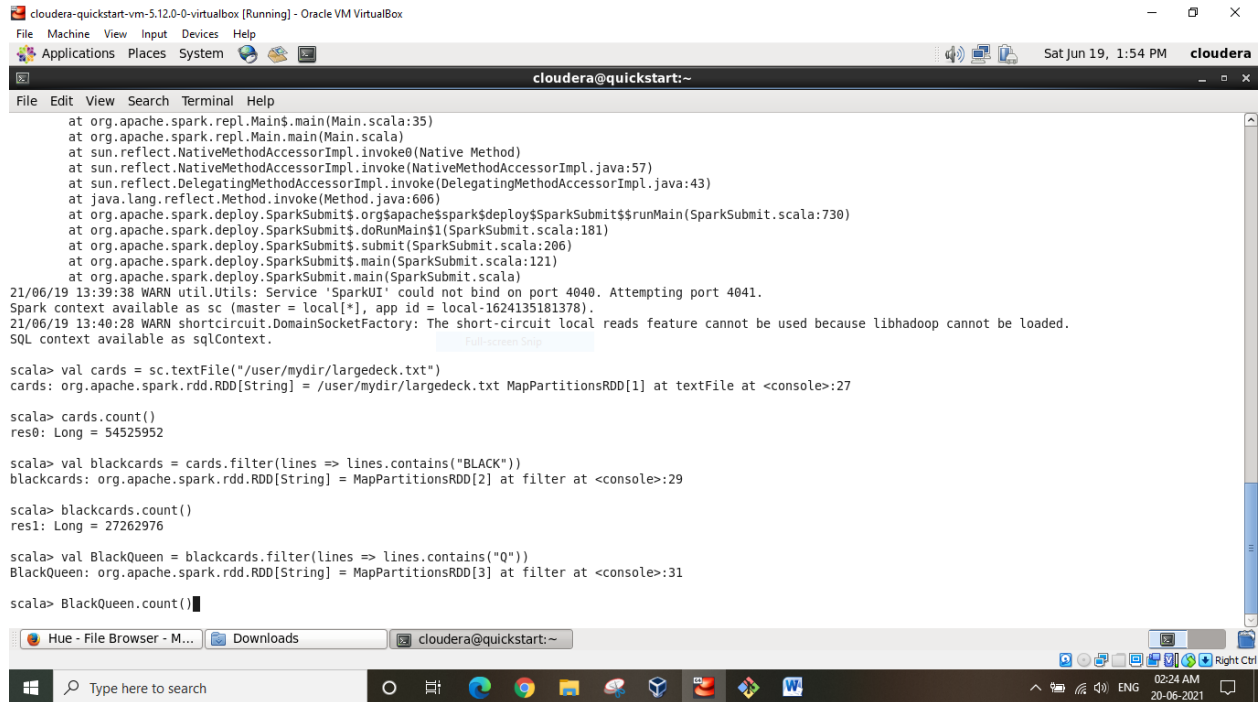
## SPARK TASK

Q: Write a program to count the number of Black Queens from largedeck.txt file.

```
val cards = sc.textFile("/user/mydir/largedeck.txt")

cards.count()

val blackcards = cards.filter(lines => lines.contains("BLACK"))
```



```
cloudera-quickstart-vm-5.12.0-0-virtualbox [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Applications Places System
cloudera@quickstart:~
File Edit View Search Terminal Help
at org.apache.spark.repl.Main$.main(Main.scala:35)
at org.apache.spark.repl.Main.main(Main.scala)
at sun.reflect.NativeMethodAccessorImpl.invoke0(Native Method)
at sun.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAccessorImpl.java:57)
at sun.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMethodAccessorImpl.java:43)
at java.lang.reflect.Method.invoke(Method.java:606)
at org.apache.spark.deploy.SparkSubmit$.org$apache$spark$deploy$SparkSubmit$$runMain(SparkSubmit.scala:730)
at org.apache.spark.deploy.SparkSubmit$.doRunMain$1(SparkSubmit.scala:181)
at org.apache.spark.deploy.SparkSubmit$.submit(SparkSubmit.scala:206)
at org.apache.spark.deploy.SparkSubmit$.main(SparkSubmit.scala:121)
at org.apache.spark.deploy.SparkSubmit.main(SparkSubmit.scala)
21/06/19 13:39:38 WARN util.Utils: Service 'SparkUI' could not bind on port 4040. Attempting port 4041.
Spark context available as sc (master = local[*], app id = local-1624135181378).
21/06/19 13:40:28 WARN shortcircuit.DomainSocketFactory: The short-circuit local reads feature cannot be used because libhadoop cannot be loaded.
SQL context available as sqlContext.

scala> val cards = sc.textFile("/user/mydir/largedeck.txt")
cards: org.apache.spark.rdd.RDD[String] = /user/mydir/largedeck.txt MapPartitionsRDD[1] at textFile at <console>:27

scala> cards.count()
res0: Long = 54525952

scala> val blackcards = cards.filter(lines => lines.contains("BLACK"))
blackcards: org.apache.spark.rdd.RDD[String] = MapPartitionsRDD[2] at filter at <console>:29

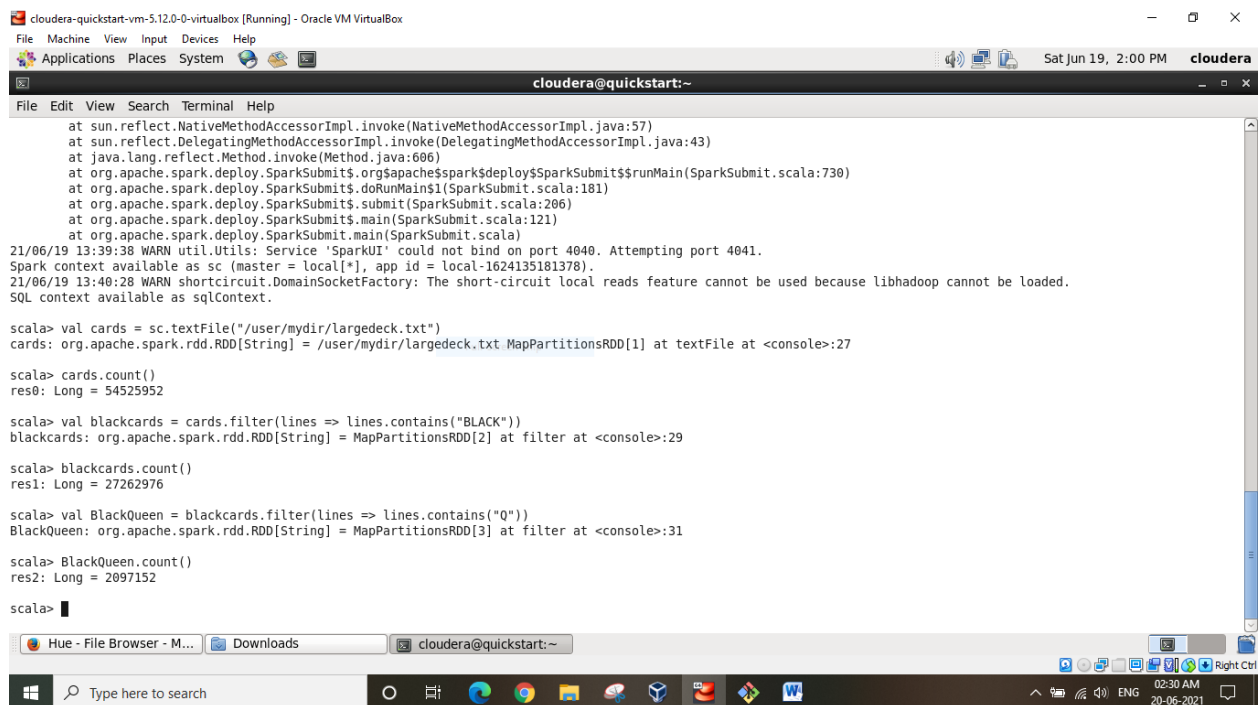
scala> blackcards.count()
res1: Long = 27262976

scala> val BlackQueen = blackcards.filter(lines => lines.contains("Q"))
BlackQueen: org.apache.spark.rdd.RDD[String] = MapPartitionsRDD[3] at filter at <console>:31

scala> BlackQueen.count()
res2: Long = 2897152

scala>
```

```
val BlackQueen = blackcards.filter(lines => lines.contains ("Q"))
```



```
cloudera-quickstart-vm-5.12.0-0-virtualbox [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Applications Places System
cloudera@quickstart:~
File Edit View Search Terminal Help
at sun.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAccessorImpl.java:57)
at sun.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMethodAccessorImpl.java:43)
at java.lang.reflect.Method.invoke(Method.java:606)
at org.apache.spark.deploy.SparkSubmit$.org$apache$spark$deploy$SparkSubmit$$runMain(SparkSubmit.scala:730)
at org.apache.spark.deploy.SparkSubmit$.doRunMain$1(SparkSubmit.scala:181)
at org.apache.spark.deploy.SparkSubmit$.submit(SparkSubmit.scala:206)
at org.apache.spark.deploy.SparkSubmit$.main(SparkSubmit.scala:121)
at org.apache.spark.deploy.SparkSubmit.main(SparkSubmit.scala)
21/06/19 13:39:38 WARN util.Utils: Service 'SparkUI' could not bind on port 4040. Attempting port 4041.
Spark context available as sc (master = local[*], app id = local-1624135181378).
21/06/19 13:40:28 WARN shortcircuit.DomainSocketFactory: The short-circuit local reads feature cannot be used because libhadoop cannot be loaded.
SQL context available as sqlContext.

scala> val cards = sc.textFile("/user/mydir/largedeck.txt")
cards: org.apache.spark.rdd.RDD[String] = /user/mydir/largedeck.txt MapPartitionsRDD[1] at textFile at <console>:27

scala> cards.count()
res0: Long = 54525952

scala> val blackcards = cards.filter(lines => lines.contains("BLACK"))
blackcards: org.apache.spark.rdd.RDD[String] = MapPartitionsRDD[2] at filter at <console>:29

scala> blackcards.count()
res1: Long = 27262976

scala> val BlackQueen = blackcards.filter(lines => lines.contains("Q"))
BlackQueen: org.apache.spark.rdd.RDD[String] = MapPartitionsRDD[3] at filter at <console>:31

scala> BlackQueen.count()
res2: Long = 2897152

scala>
```