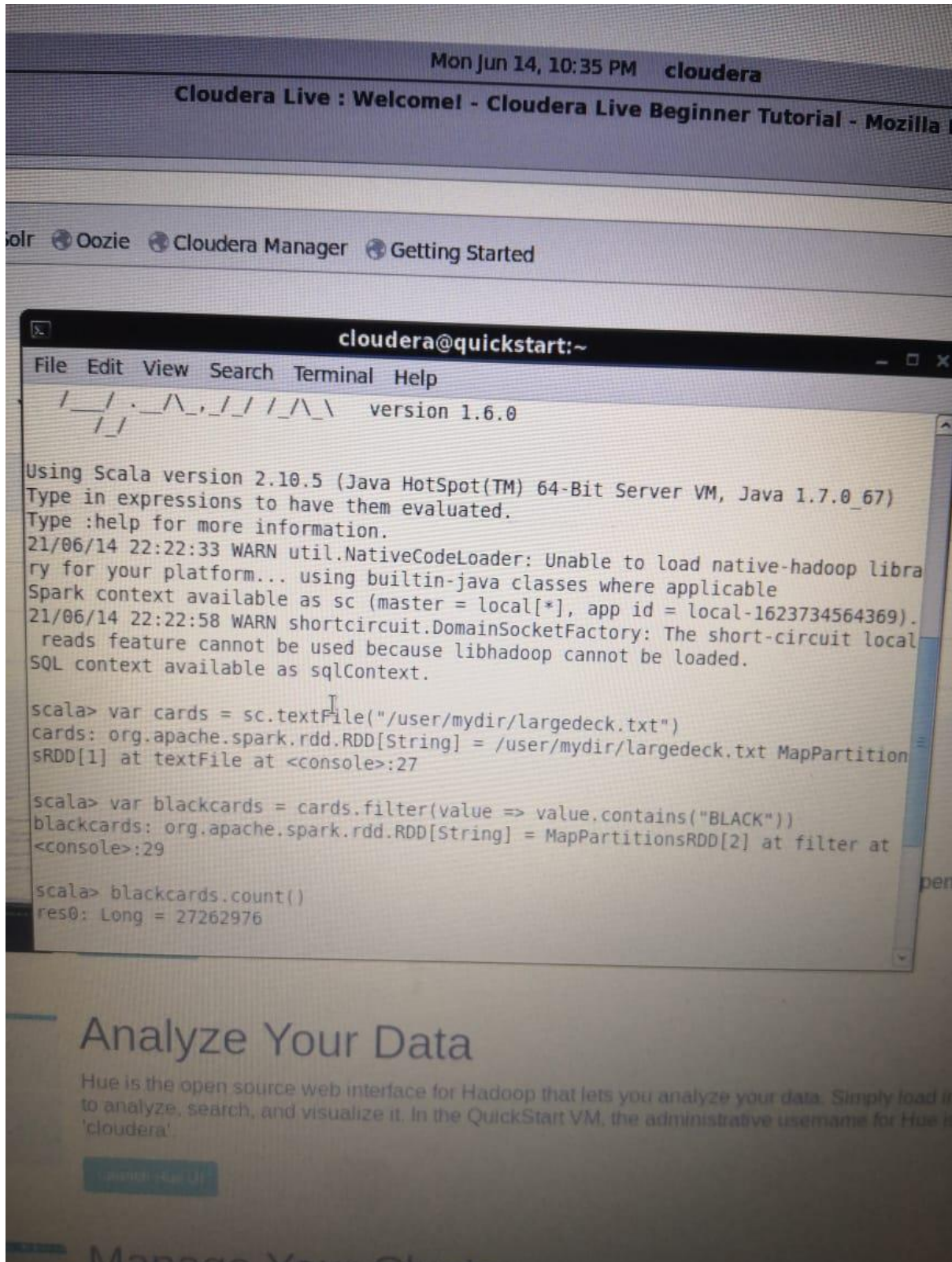


SPARK SCALA TASK for COUNT BLACK QUEENS

Step 1 : open spark shell

Step2 : create variable to store largedeck.txt file

Step3 : count black cards in largedeck.txt file



The screenshot shows a terminal window titled "cloudera@quickstart:~" with a menu bar (File, Edit, View, Search, Terminal, Help). The terminal output includes the Spark version (1.6.0), Scala version (2.10.5), and Java version (1.7.0_67). It also shows warnings about native-hadoop library and short-circuit local reads. The user enters the following commands:

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

scala> blackcards.count()
res0: Long = 27262976
```

Below the terminal window, there is a section titled "Analyze Your Data" with text about Hue and a "Launch Hue UI" button.

SPARK SCALA TASK for COUNT BLACK QUEENS

Step4 : Calculate the count of Black Queens in largedeck.txt file

```
scala> blackcards.count()
res0: Long = 27262976

scala> var blackcards = cards.filter(value => value.contains("Q")).filter(value => value.contains("BLACK"))
<console>:1: error: identifier expected but '(' found.
    var blackcards = cards.filter(value => value.contains("Q")).filter(value => value.contains("BLACK"))
                                     ^

scala> var blackcards = cards.filter(value => value.contains("Q")).filter(value => value.contains("BLACK"))
blackcards: org.apache.spark.rdd.RDD[String] = MapPartitionsRDD[4] at filter at <console>:29

scala> blackcards.count()
res1: Long = 2097152

scala>
```

