Steps:

a. Load the dataset

b. Determine which delimiter to use - hint: the character at position 19 (the 20th character) is the first use of the delimiter.

c. Filter out any records which do not parse correctly - hint: each record should have exactly 14 values.

d. Extract the date (first field), mfr_model (second field), device ID (third field), and latitude and longitude (13th and 14th fields respectively).

e. The second field (mfr_model) contains the device manufacturer and model name (e.g. "Ronin S2" or "Sorrento F41L") Split this field on the blank(s) to separate the manufacturer from the model (e.g. manufacturer "Ronin", model "S2"), and assign the value extracted for manufacturer to the second field of your output.

f. Save the extracted data to comma delimited text files in the **/loudacre/devstatus/devicestatus_etl** directory on HDFS.

g. Confirm that the data in the file(s) was saved correctly.

Answers to a-f :

```
val logfile = "bdad17/hw4/devicestatus.txt"
val mydata = sc.textFile(logfile)
val mydata_parted = mydata.map(line=> line.split(line.charAt(19)))
val mydata_filtered = mydata_parted.filter(_.length==14)
val mydata_vals =
mydata_filtered.map(line=>(line1(0),line1(1),line1(2),line1(12),line1(13)))
val mydata_final = mydata_vals.map(line=>line._1+","+line._2.split("
")(0)+","+line._3+","+line._4+","+line._5)
mydata_final.saveAsTextFile("bdad17/hw4/loudacre/devstatus/devicestatus_etl")
```

g)
```
2014-03-15:10:49:30,Titanic,609d7e61-faec-488d-b3f0-fb680a1100f6,33.7189909197,-112.127791734
2014-03-15:10:49:30,MeeToo,de6b7b6b-5c01-4f88-948c-5733716ab410,33.1941652647,-116.374666652
2014-03-15:10:49:30,Sorrento,7c06751c-f692-473e-9143-9280594a9740,45.1387902196,-117.739092779
2014-03-15:10:49:30,Sorrento,40e61459-5448-4dc9-bb89-42e73a4e19cf,39.4463417571,-114.736213453
2014-03-15:10:49:30,Titanic,d4f2958f-27a9-49ec-b176-eca99e486b57,36.524950844,-115.056582933
2014-03-15:10:49:30,Ronin,b13ece99-62ab-4c9f-a366-6a06bd5e877f,38.4282665514,-121.25933863
2014-03-15:10:49:30,Titanic,e683c864-2c49-46dd-8aa4-4f10364f3cd2,37.5589514366,-121.212136808
2014-03-15:10:49:30,Sorrento,32af1a0b-ca7f-4906-9772-9eb9435e7e4c,33.7778202246,-108.575470704
2014-03-15:10:49:30,Ronin,a48a5559-d916-481b-84a9-5dce6272cce1,38.2596913494,-122.295712621
2014-03-15:10:49:30,iFruit,d86fbaa6-b71b-435f-a0bf-5304a202a70b,34.2415255221,-118.23526739
[cloudera@quickstart ~]$ hdfs dfs -ls bdad17/hw4/loudacre/devstatus/devicestatus_etl
Found 3 items
-rw-r--r--   1 cloudera cloudera          0 2017-10-14 19:44 bdad17/hw4/loudacre/devstatus/devicestatus_etl/_SUCCESS
-rw-r--r--   1 cloudera cloudera    4616923 2017-10-14 19:44 bdad17/hw4/loudacre/devstatus/devicestatus_etl/part-00000
-rw-r--r--   1 cloudera cloudera    4617183 2017-10-14 19:44 bdad17/hw4/loudacre/devstatus/devicestatus_etl/part-00001
[cloudera@quickstart ~]$
```