HOMEWORK 4 SOLUTION

```
// Example data:
//2014-03-15:10:10:33,Ronin S2,1a7eca8d-60c9-4d25-8609-
d6cfd1ac80a1,0,24,82,72,enabled,enabled,enabled,41,62,36.49259162,-121.003629078
//2014-03-15:10:10:33/Titanic 2300/d86dbb9d-ff3c-40c6-8685-
01f1fac45d9f/59/83/9/3/28/0/enabled/disabled/enabled/34.3456792864/-117.768326105

// Load the data file
val devstatus = sc.textFile("/loudacre/devicestatus.txt")

// Filter out lines with < 10 characters, use the 20th character as the delimiter, parse the line, and filter
out bad lines
val cleanstatus = devstatus.
   filter(line => line.length>20).
   map(line => line.split(line.charAt(19))).
   filter(values => values.length == 14)

// Create a new RDD containing date, manufacturer, device ID, latitude and longitude
val devicedata = cleanstatus.
   map(values => (values(0), values(1).split(' ')(0), values(2), values(12), values(13)))

// Save to a CSV file as a comma-delimited string (trim parenthesis from tuple toString)
devicedata.
   map(values => values.toString).
   map(s => s.substring(1,s.length-1)).
   saveAsTextFile("/loudacre/devicestatus_etl")
```