
HOMEWORK 5 Solutions

```
hdfs dfs -put /home/cloudera/Desktop/Week5/activations /user/cloudera/loudacre/activations
```

```
val activations = “/user/cloudera/loudacre/activations”
```

```
val mydata = sc.wholeTextFiles(activations)
```

```
import scala.xml._
```

```
def getactivations(xmlstring: String): Iterator[Node] = {  
  | val nodes = XML.loadString(xmlstring) \ \ "activation"  
  | nodes.toIterator  
  | }
```

```
def getmodel(activation: Node): String = {  
  | (activation \ "model").text  
  | }
```

```
def getaccount(activation: Node): String = {  
  | (activation \ "account-number").text  
  | }
```

```
val xmlStrings = mydata.map(tup => tup._2)
```

```
val xmlRDD = xmlStrings.flatMap(line => XML.loadString(line) \ "activation")
```

```
val newRDD = xmlRDD.map(line => getaccount(line) + ":" + getmodel(line))  
newRDD.saveAsTextFile(“loudacre/activations/account-models”)
```