

# Apache Spark: Split a pair RDD into multiple RDDs by key

This drove me crazy but I finally found a solution.

If you want to Split a pair RDD of type (A, Iterable(B)) by key, so the result is several RDDs of type B, then here how you go:

```
[crayon-6a05e5042a322497064597/]
```

The trick is twofold (1) get the list of all the keys, (2) iterate through the list of keys, and for each key, create a new RDD by filtering the original pair RDD to get only values attached to that key. Note the use of Java 8's Lambda expression in line 17.

**Remark:** bare in mind that the action `[crayon-6a05e5042a329306340171-i/]` gets you the keys of the *entire* distributed RDD into the driver machine. Generally, keys are integers or small strings, so collecting them into one machine wouldn't be problematic. Otherwise, this wouldn't be the best way to go. If you have a suggestion, you are more than welcome to leave a comment below... spread the word, help the world!