# Spark Core01

## Spark编译

参考<http://spark.apache.org/docs/2.4.6/building-spark.html>

整理出编译文档：doc/spark/build\_spark.md

## RDD的五大特性对应五大方法

## Spark-Shell脚本简单使用

1. 备用

# SparkCore02

## Transformations

### map

### mapPartition

### mapPartitionWithIndex

### flatMap

### filter

### glom

把每个分区中的所有元素，整合到一个Array 返回这个RDD[Array]

### sample

### zip

Spark中zip 两边的元素个数必须要对等

### zipWithIndex

### join

### leftOuterJoin

### fullOuterJoin

### cogroup

### union

Union之后的分区数就是2个RDD的分区数相加

### keyBy

### groupByKey

1. SparkCore03

1.Transformations

* 1. reduceByKey
  2. sortBy
  3. sortByKey
  4. Distinct

1. Action
   1. Collect
   2. collectAsMap
   3. First
   4. Take
   5. takeOrdered
   6. Top
   7. Count
   8. Sum
   9. Reduce
   10. Foreach
   11. countByKey
   12. countByValue
   13. Lookup
   14. fold
2. Beiyong