# 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

# 30-SparkCore03

## 1.Transformations

### reduceByKey

### sortBy

### sortByKey

### distinct

### Coalesce

不发生shuffle 只是做分区的合并

### repartition

## Action

### Collect

### collectAsMap

### First

### Take

### takeOrdered

### Top

### countByKey

### countByValue

### Loopup

### Fold

### Map 和 mapPartitions

### Foreach 和 foreachPartition