Spark函数讲解序列文章

　　本博客近日将对Spark 1.2.1 RDD中所有的函数进行讲解，主要包括函数的解释，实例以及注意事项，每日一篇请关注。以下是将要介绍的函数，按照字母的先后顺序进行介绍，可以点的说明已经发布了。

　　aggregate、aggregateByKey、cache、cartesian、checkpoint、coalesce、cogroup

groupWith
collect, toArray
collectAsMap
combineByKey
compute
context, sparkContext
count
countApprox
countByKey
countByKeyApprox
countByValue
countByValueApprox
countApproxDistinct
countApproxDistinctByKey
dependencies
distinct
first
filter
filterWith
flatMap
flatMapValues
flatMapWith
fold
foldByKey
foreach
foreachPartition
foreachWith
generator, setGenerator
getCheckpointFile
preferredLocations
getStorageLevel
glom
groupBy
groupByKey
histogram
id
intersection

isCheckpointed

iterator

join

keyBy

keys

leftOuterJoin

lookup

map

mapPartitions

mapPartitionsWithContext

mapPartitionsWithIndex

mapPartitionsWithSplit

mapValues

mapWith

max

mean , meanApprox

min

name, setName

partitionBy

partitioner

partitions

persist, cache

pipe

randomSplit

reduce

reduceByKey, reduceByKeyLocally, reduceByKeyToDriver

rightOuterJoin

sample

saveAsHodoopFile, saveAsHadoopDataset, saveAsNewAPIHadoopFile

saveAsObjectFile

saveAsSequenceFile

saveAsTextFile

stats

sortBy

sortByKey

stdev , sampleStdev

subtract

subtractByKey

sum , sumApprox

take

takeOrdered

takeSample

toDebugString

toJavaRDD

top

toString

union, ++
unpersist
values
variance , sampleVariance
zip
zipPartitions
zipWithIndex
zipWithUniquId