Spark Python API Docs!

# pyspark包

## pyspark的子包

[pyspark.sql模块](#_pyspark.sql模块)

[pyspark.streaming模块](#_pyspark.streaming模块)

[pyspark.ml包](#_pyspark.ml包)

[pyspark.mllib包](#_pyspark.mllib包)

## pyspark的公共类

公共类:

[SparkContext](#_SparkContext):

Spark功能的主入口点

[RDD](#_RDD):

弹性分布式数据集,是Spark中的基本抽象(basic abstraction)

[Broadcast](#_Broadcast):

跨任务重复使用的广播变量

[Accumulator](#_Accumulator):

一个共享的变量,任务只能向它添加值(add-only)

[SparkConf](#_SparkConf):

配置Spark

[SparkFiles](#_SparkFiles):

访问作业附带的文件

### SparkConf

class pyspark.SparkConf(lodaDefaults=True, \_jvm=None, \_jconf=None)

Spark应用程序的配置.用于将设置各种Spark参数为键值对

大多数时候,你会用SparkConf()创建一个SparkConf对象.

对于单元测试,可以用SparkConf(false)来跳过加载外部设置并获得相同的配置,无论系统属性是什么.

此类中所有setter方法都支持链接,如:conf.setMaster(‘local’).setAppName(‘My app’).

请注意,一旦SparkConf对象被传递给Spark了,它就被克隆,再也不能被用户修改了

函数：

contains(key):判断这个配置是否包含键key

print conf.contains("spark.master") # 返回True或者False

get(key, default=None):返回配置字典中key对应的value

### SparkContext

### SparkFiles

### RDD

### StorageLevel

### Broadcast

### Accumulator

### AccumulatorParam

### MarshalSerializer

### PickleSerializer

### StatusTracker

### SparkJobInfo

### SparkStageInfo

### Profiler

### BasicProfiler

# pyspark.sql模块

## Module Context

Spark SQL和DataFrames中重要的类:

* pyspark.sql.SparkSession—DataFrame和SQL功能的主入口

## pyspark.sql.types模块

## pyspark.sql.functions模块

## pyspark.sql.streaming模块

# pyspark.streaming模块

## Module contents

## pyspark.streaming.kafa模块

## pyspark.streaming.kinesis模块

## pyspark.streaming.flume模块

# pyspark.ml包

## ML Pipeline APIs

## pyspark.ml.param模块

## pyspark.ml.feature模块

## pyspark.ml.classification模块

## pyspark..ml.clustering模块

## pyspark.ml.linalg模块

## pyspark.ml.recommendation模块

## pyspark.ml.regression模块

## pyspark.ml.tuning模块

## pyspark.ml.evaluation模块

# pyspark.mllib包

## pyspark.mllib.classification模块

## pyspark.mllib.clustering模块

## pyspark.mllib.evaluation模块

## pyspark.mllib.feature模块

## pyspark.mllib.fpm模块

## pyspark.mllib.linalg模块

## pyspark.mllib.linalg.distributed模块

## pyspark.mllib.random模块

## pyspark.mllib.recommendation模块

## pyspark.mllib.stat模块

## pyspark.mllib.tree模块

## pyspark.mllib.util模块