

Spark函数讲解：cartesian

从名字就可以看出这是笛卡儿的意思，就是对给的两个RDD进行笛卡儿计算。官方文档说明：

Return the Cartesian product of this RDD and another one, that is, the RDD of all pairs of elements (a, b) where a is in `this` and b is in `other`.

函数原型

```
def cartesian[U: ClassTag](other: RDD[U]): RDD[(T, U)]
```

该函数返回的是Pair类型的RDD，计算结果是当前RDD和other RDD中每个元素进行笛卡儿计算的结果。最后返回的是CartesianRDD。

实例

```
/**
 * User: 过往记忆
 * Date: 15-03-07
 * Time: 上午06:30
 * bolg:
 * 本文地址：/archives/1277
 * 过往记忆博客，专注于hadoop、hive、spark、shark、flume的技术博客，大量的干货
 * 过往记忆博客微信公共帐号：iteblog_hadoop
 */
scala> val a = sc.parallelize(List(1,2,3))
a: org.apache.spark.rdd.RDD[Int] = ParallelCollectionRDD[62] at parallelize at <console>:12

scala> val b = sc.parallelize(List(4,5,6))
b: org.apache.spark.rdd.RDD[Int] = ParallelCollectionRDD[63] at parallelize at <console>:12

scala> val result = a.cartesian(b)
result: org.apache.spark.rdd.RDD[(Int, Int)] = CartesianRDD[64] at cartesian at <console>:16

scala> result.collect
res78: Array[(Int, Int)] = Array((1,4), (1,5), (1,6), (2,4),
    (2,5), (2,6), (3,4), (3,5), (3,6))
```

注意

笛卡儿计算是很恐怖的，它会迅速消耗大量的内存，所以在使用这个函数的时候请小心。

本博客文章除特别声明，全部都是原创！
原创文章版权归过往记忆大数据（[过往记忆](#)）所有，未经许可不得转载。
本文链接: **【】**（）