PageRank.md 2020/11/23

PageRank作业

ZY2006109-姬轶

一、Scala代码的简要说明讲解

```
// Create RDD (构造RDD函数)
val links = sc.parallelize(List(
   ("A",List("B","C","D")),
   ("B",List("A")),
    ("C",List("A", "B")),
   ("D",List("B","C"))
)).partitionBy(new HashPartitioner(10)).persist()
// 每个页面的排序值初始化为 1.0
var ranks = links.mapValues(v => 1.0)
// For to Compute (十次循环计算更新ranks)
for(i <- 0 until 10){
   val contributions = links.join(ranks).flatMap{
       case(pageId, (links, rank)) =>links.map(link => (link, rank / links.size))
   }
// 每个页面的排序值设为 0.15 + 0.85 * contributionsReceived
   ranks = contributions.reduceByKey((x,y) => x+y).mapValues(v => 0.15 + 0.85*v)
}
// 输出跑出的结果
ranks.collect().foreach(println)
// Save File (保存文件到"ranks"中)
ranks.saveAsTextFlie("ranks")
```

二、Spark运行的截图

PageRank.md 2020/11/23

```
20/11/21 22:31:31 WARN NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable Using Spark's default log4j profile: org/apache/spark/log4j-defaults.properties
Setting default log level to "WARN".

To adjust logging level use sc.setLogLevel(newLevel). For SparkR, use setLogLevel(newLevel).

Spark context Web UI available at http://c04b03.lc.act.buaa.edu.cn:4040

Spark context available as 'sc' (master = local[*], app id = local-1605969101562).

Spark session available as 'spark'.
     /__/.__/\._./_/ /_/\_\ version 3.0.1
Using Scala version 2.12.10 (OpenJDK 64-Bit Server VM, Java 1.8.0_212)
Type in expressions to have them evaluated.
Type :help for more information.
scala> import org.apache.spark.{HashPartitioner, SparkConf, SparkContext}
import org.apache.spark.{HashPartitioner, SparkConf, SparkContext}
 scala> // Create RDD
              )).partitionBy(new HashPartitioner(10)).persist()
                    ache.spark.rdd.RDD[(String, List[String])] = ShuffledRDD[1] at partitionBy at <console>:30
 links: org.ap
scala> var ranks = links.mapValues(v => 1.0)
ranks: org.apache.spark.rdd.RDD[(String, Double)] = MapPartitionsRDD[2] at mapValues at <console>:26
 scala> // For to Compute
                for(i <- 0 until 10){
  val contributions = links.join(ranks).flatMap{</pre>
                      case(pageId, (links, rank)) =>links.map(link => (link, rank / links.size))
                   ranks = contributions.reduceByKey((x,y) => x+y).mapValues(v => 0.15 + 0.85*v)
                     ranks.collect().foreach(println)
 (A,1.4729483816191942)
 (B,1.151795159344013)
 (C,0.8081470492728162)
```

三、ranks文件的截图

ranks.saveAsTextFile("ranks")

(D,0.5671094097639748)

scala> // Save File

PageRank.md 2020/11/23

■ part-00000 - 记事本

文件(F) 编辑(E) 格式(O) 查看(V) 帮助(H)

(A,1.4729483816191942)

[B,1.151795159344013]

[C,0.8081470492728162]

[D,0.5671094097639748]