**实验三 RDD转换、持久化与分区**

**【实验目的】**

**1．掌握RDD的创建与五种转换操作；**

**2．掌握RDD的持久化与分区。**

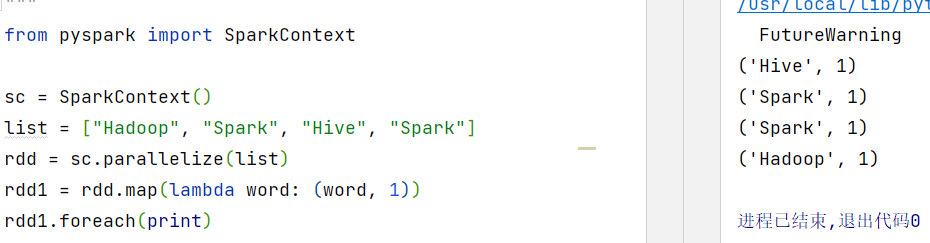
**【实验内容】**

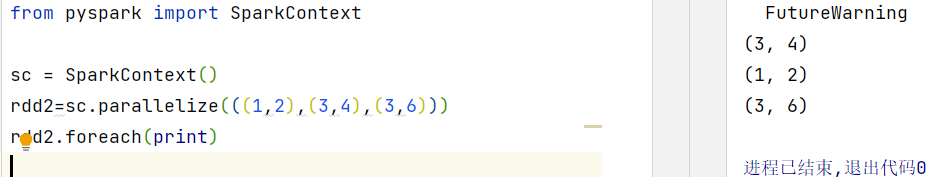
**一、spark之RDD的创建与转换操作。**

**1、使用三种键值对RDD的创建方式创建rdd0，rdd1，rdd2。**

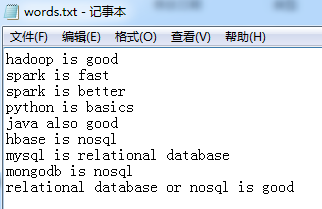
**内容自已定义**

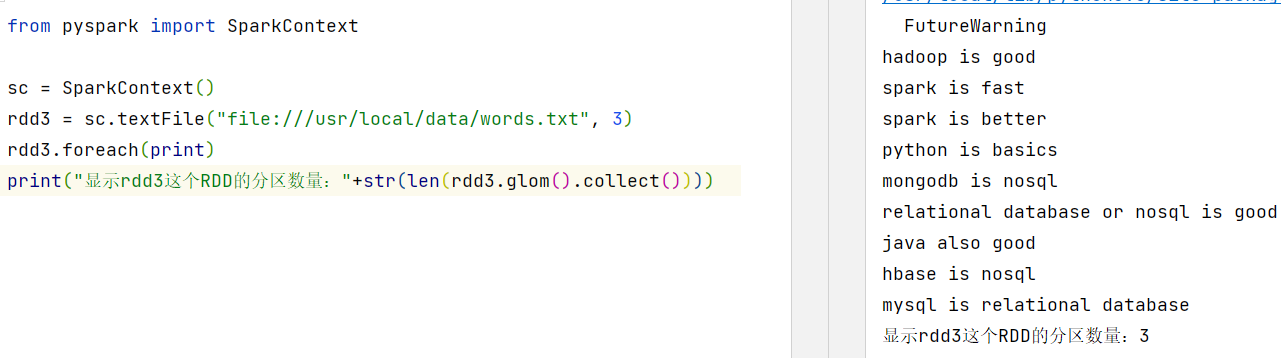






**2、将word.txt文档转换成RDD，并设置为3个分区，命名为rdd3，并分别显示rdd3与分区个数。**

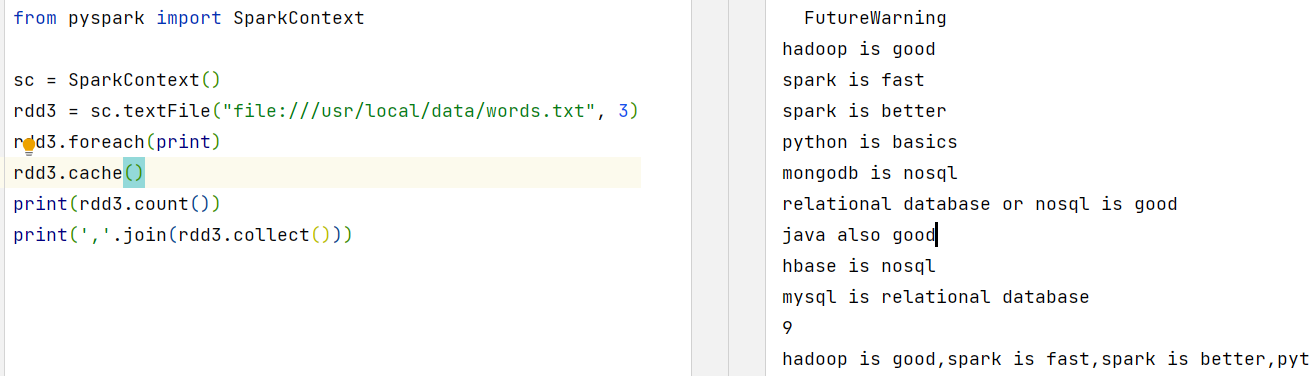
****



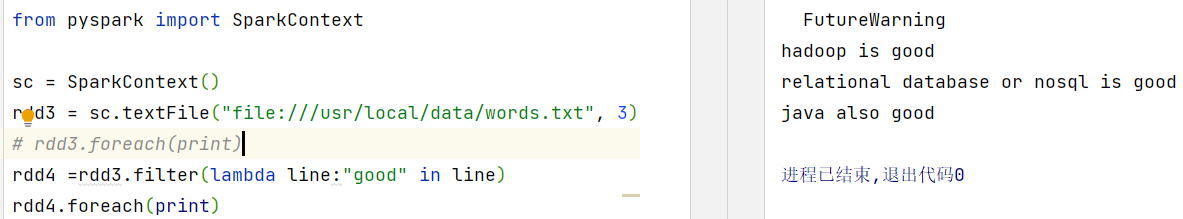
1. **重新设置分区个数为2，并显示。**



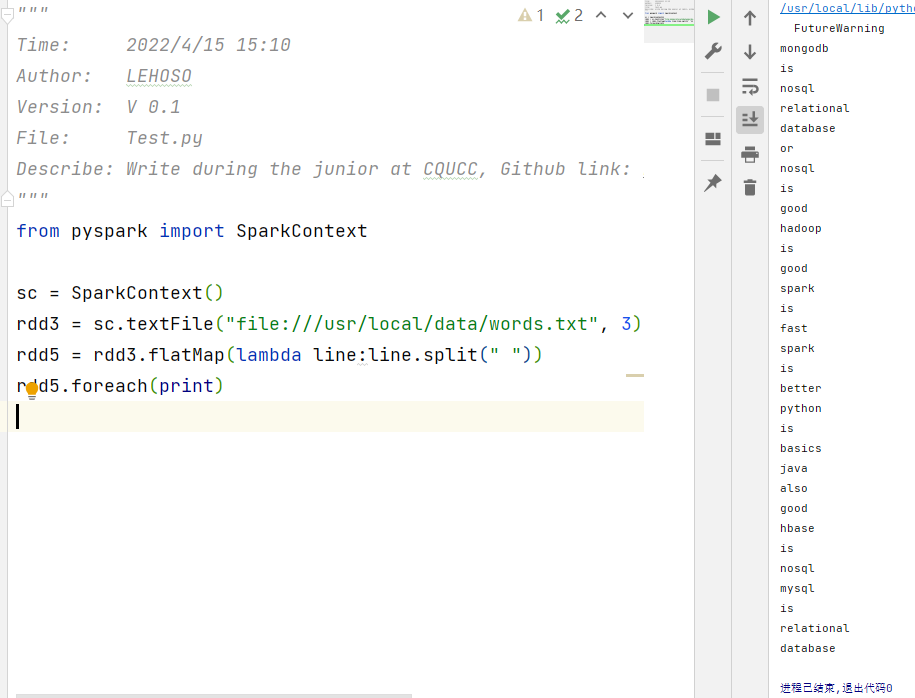
1. **将rdd3标记为持久化；**



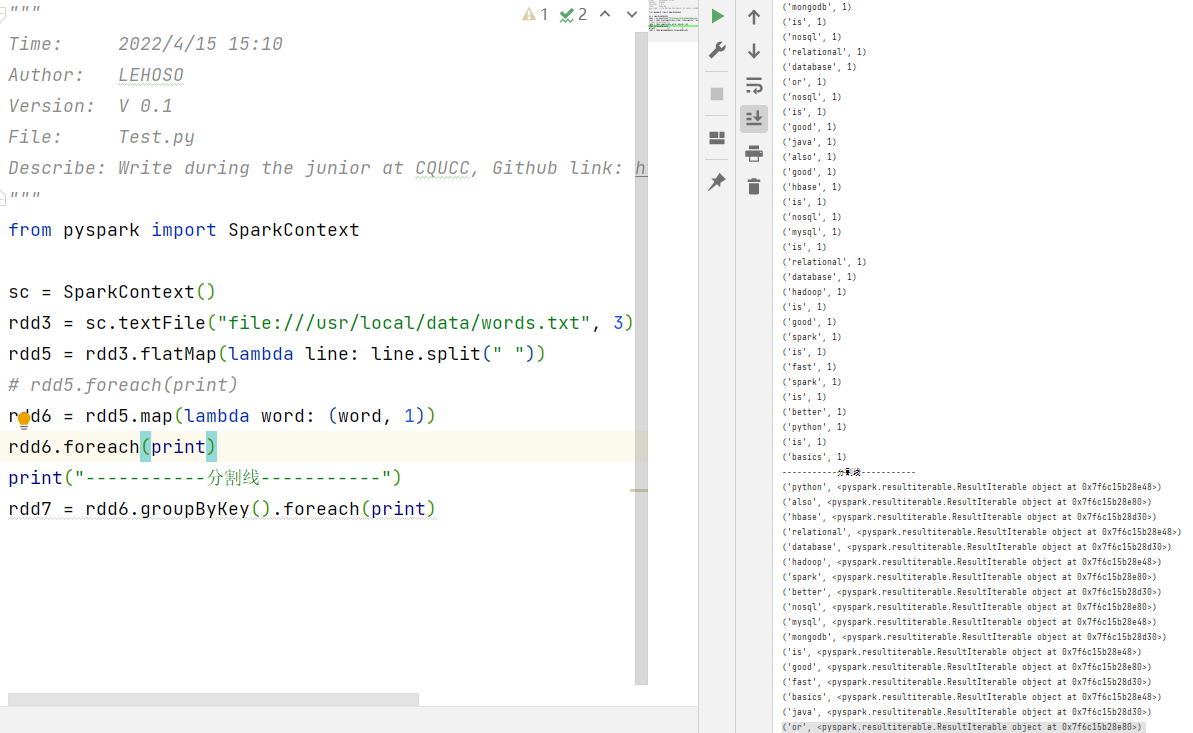
1. **在rdd3中筛选出包含“good”的字符串保存到新的rdd4中，并打印显示；**



1. **将rdd3使用flatMap转换，使用空格分离转换，保存成rdd5并查看其结果；**



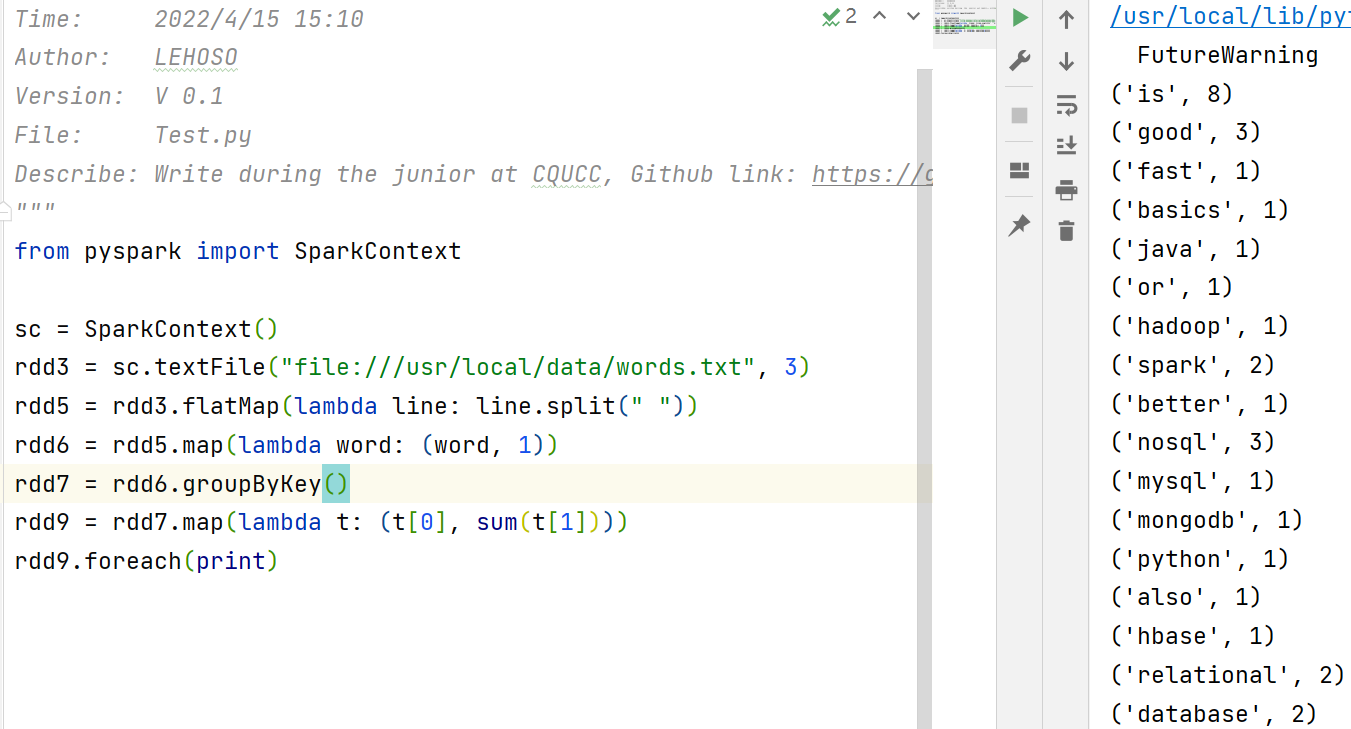
1. **将rdd5使用map转换生成键值对，保存成rdd6，再使用groupByKey转换保存为rdd7，并查看其结果；**



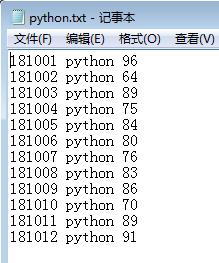
1. **将rdd6，使用reduceByKey转换，计算词频，保存成rdd8并查看其结果；**



1. **将rdd7计算词频，保存成rdd9并查看其结果；**



**10、将python.txt文件转换为rdd10，并通过RDD转换操作计算python课程的总分（尝试有几种方法）。**

****

