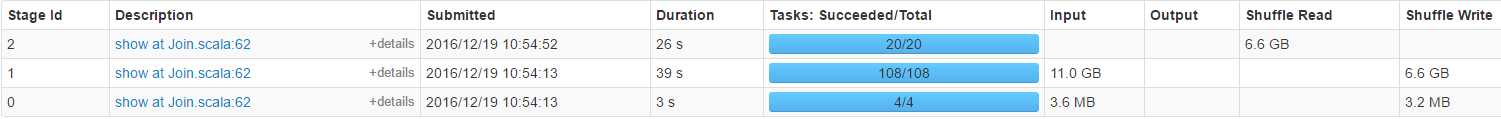
Partition 问题

1. spark.sql.shuffle.partition 参数可以改变shuffle时的partition数目，进而控制task的个数

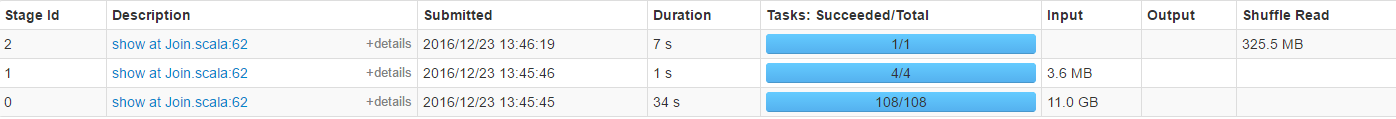
效果：



1. join时的问题

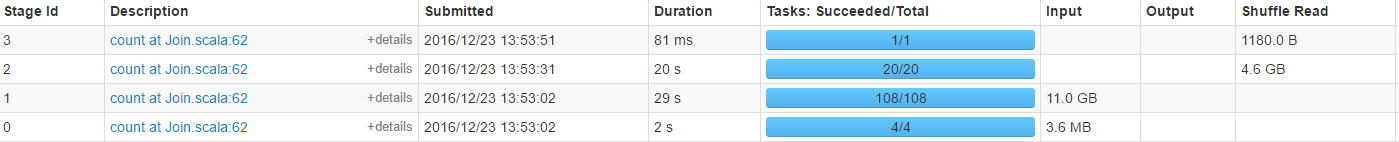
sqltext1 = **"select pagerank,sourceipaddr,url,adrevenue "** +  
 **"from rankings "** +  
 **"INNER JOIN uservisits "** +  
 **"ON url=destinationURL "**

**result.show()**



Sqltext1 = **"select pagerank,sourceipaddr,url,adrevenue "** +  
 **"from rankings "** +  
 **"INNER JOIN uservisits "** +  
 **"ON url=destinationURL "**

**result.count()**



Sqltext2 = **"select pagerank,sourceipaddr,url,adrevenue "** +  
 **"from uservisits "** +  
 **"INNER JOIN rankings "** +  
 **"ON url=destinationURL "** +  
 **"order by adrevenue desc limit 100"**

**result.show()**

