











Group TOPN sort

Read each line of data

2, generate pairs K, V key-value pairs

Enter a line of data output KEY value is the name, Value is the score

3, grouped by name

4, sorting after the group

Enter groupdata, where KEY is the name of the group, and VALUE is the set of scores

Output KEY: the group name after the grouping, VALUE: is a set of scores after the ranking to take five values

Sort Algorithm RangePartitioner Insider Decryption

RangePartitioner is mainly dependent on the RDD data is divided into different areas, the key is that different areas are ordered.

In addition to being the cornerstone of a well-ordered result, the most important thing about RangePartitioner is to try to ensure that the amount of data in each Partition is uniform! The company is located in:

The algorithm for sampling the ponds is used