

mapreduce过程-例子数据

-----一个的map输入数据

hello java java php java
hello php hadoop mapreduce
hdfs hdfs java 周杰伦 周杰伦 林志玲

-----map分3次处理

第一次拿到：

hello java java php java

-》处理为

hello 1

java 3

php 1

第一次拿到：hello php hadoop mapreduce

-》处理为

hello 1

php 1

hadoop 1

mapreduce 1

第一次拿到：hdfs hdfs java 周杰伦 周杰伦 林志玲

-》处理为

hdfs 2

java 1

周杰伦 2

林志玲 1

-----shuffling (自动完成) 处理后的

hello (1, 1)

java (3, 1)

php (1, 1)

hadoop (1)

周杰伦 (2)

。 。 。

-----reduce (手动完成) 处理后的

hello 2

java 4

php 2

hadoop 1

....