



Mapper → <A, zhangsan>, <B, zhangsan>, <D, zhangsan>

<B, lisi>, <C, lisi>, <D, lisi>

<D, wangwu>

Reducer → <A, (zhangsan)>

<B, (zhangsan, lisi)>

<C, (lisi)>

<D, (zhangsan, lisi, wangwu)> 这不是我们想要的, value 并没有排序

解决方法, 用 secondary sort:

Mapper → <<A, zhangsan>, zhangsan>, <<B, zhangsan>, zhangsan>, <<D, zhangsan>, zhangsan>

<<B, lisi>, lisi>, <<C, lisi>, lisi>, <<D, lisi>, lisi>

<<D, wangwu>, wangwu> →

这地方用到 WritableComparator

Shuffle and sort → <<A, zhangsan>, zhangsan> → <<A, zhangsan>, (zhangsan)>

<<B, lisi>, lisi>

<<B, lisi>, (lisi, zhangsan)>

<<B, zhangsan>, zhangsan>

<<C, lisi>, (lisi)>

<<C, lisi>, lisi>

<<D, lisi>, (lisi, wangwu, zhangsan)>

<<D, lisi>, lisi>

<<D, wangwu>, wangwu>

这地方能按要求排好序是因为我们写了 pair 里的 compareTo 方法

<<D, zhangsan>, zhangsan>

→ Reducer → <A, (zhangsan)>

<B, (lisi, zhangsan)>

<C, (lisi)>

<D, (lisi, wangwu, zhangsan)>

所有复合 Key 中, 第一个元素相同的被分到同一个 partitioner 进行计算, 是因为写了 partitioner

