# 大数据

## Tips

ETL，是英文Extract-Transform-Load的缩写，用来描述将数据从来源端经过抽取（extract）、转换（transform）、加载（load）至目的端的过程。ETL一词较常用在数据仓库，但其对象并不限于数据仓库。

## Mr

针对每一对<key,value>都会调用一次map()方法（MapTask进程）

）Reduce的输入数据是KV对的形式，同时也是mapper阶段的输出数据，这里的健值对必须跟Map阶段的键值对类型一致，Mapper的输出，就是这里的输入

ReduceTask进程对每一组相同的key调用一次reduce()方



