

第十二周作业

在你所在的公司（行业、领域），正在用大数据处理哪些业务？可以用大数据实现哪些价值？

我们公司用大数据，主要有：

1. 做各种取数工作，用来观察业务指标，主要是运行HQL
2. 数据挖掘和分析，针对用户信息来改善营销策略
3. 机器学习，训练模型，做风控，搜索，推荐等

分析如下 HiveQL，生成的 MapReduce 执行程序，map 函数输入是什么？输出是什么，reduce 函数输入是什么？输出是什么？

这个程序有2个mapper，一个reducer

第一个mapper，输入的key是偏移量，value是pv表的值，输出的key是userid，value是表的序号以及pageid

第二个mapper，输入的key是偏移量，value是user表的值，输出的key是userid，value是表的序号以及age

reducer输入的key是userid，value是第一个表的pageid以及第二个表的userid，输出是每个userid对应的pageid以及age的笛卡尔积