**2020年安徽省大数据与人工智能应用初赛**

**学校名称： 队伍名称： 队伍编号：**

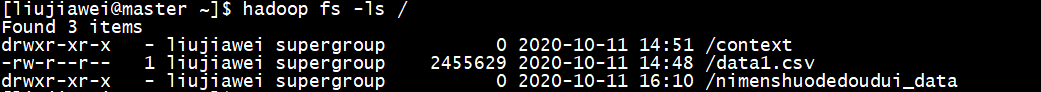
# 第三部分：大数据分析（20分）

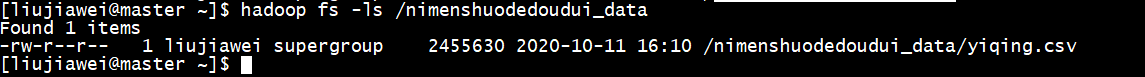
疫情期间各类政府媒体及社交网站，均发布了相关疫情每日统计数据，下面基于数据仓库工具Hive请你统计分析相关疫情数据。



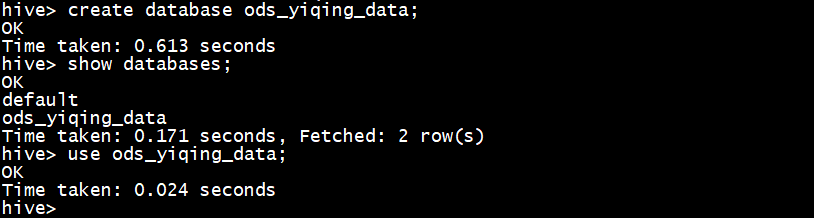
数据字段为：日期、省份、城市、新增确诊、新增出院、新增死亡、消息来源、来源1，来源2，来源3

### 请在hdfs根目录下中创建一个目录，以你的小组名称英文全拼命名\_data，并将以上疫情数据上传到这个目录中，截图成功后结果（2分）。

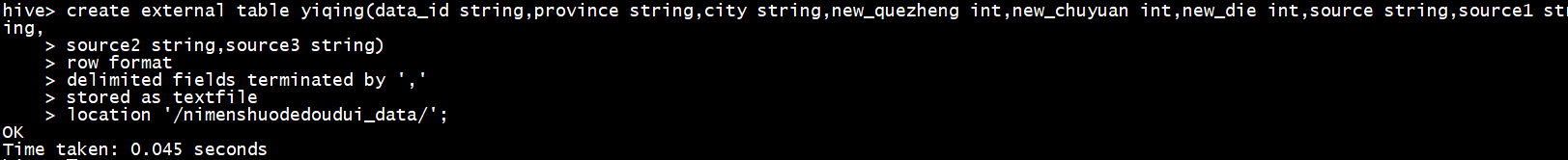




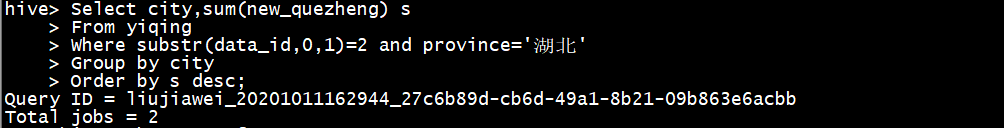
1. 请你在hive中创建数据库名为: ods\_yiqing\_data，并切换使用这个数据库，截图成功后结果（2分）。

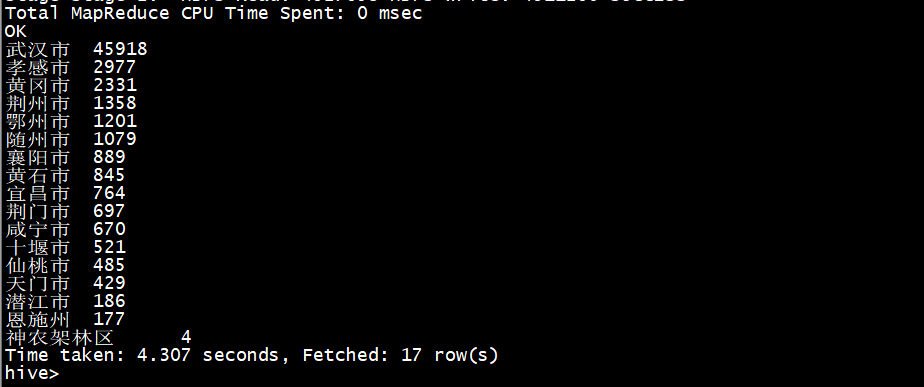


1. 创建一个hive外部表，字段为以上疫情数据全部字段（自行命名字段），数据存储位置为第1小题创建的hdfs目录，请粘贴建表语句及运行成功截图（3分）。

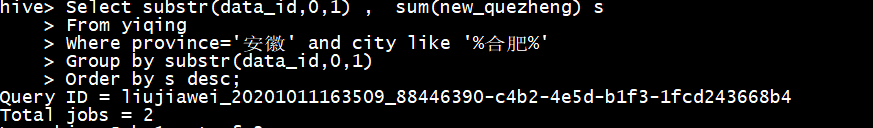


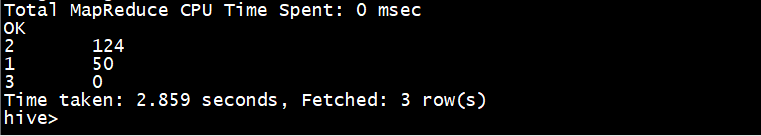
1. 统计湖北省各市2月新增确诊病例总数，按照总数降序排列，请提供SQL语句及运行结果截图（4分）





1. 统计文件中安徽省合肥市每月新增确诊病例总数，按照降序排列，请提供SQL语句及运行结果截图（4分）





1. 统计文件中湖北每月新增出院病例总数最多的前2个城市，请提供SQL语句及运行结果截图（5分）

