1. **HYPERLOGLOG应用**
2. 业务应用

统计注册 IP 数

统计每日访问 IP 数

统计页面实时 UV 数

统计在线用户数

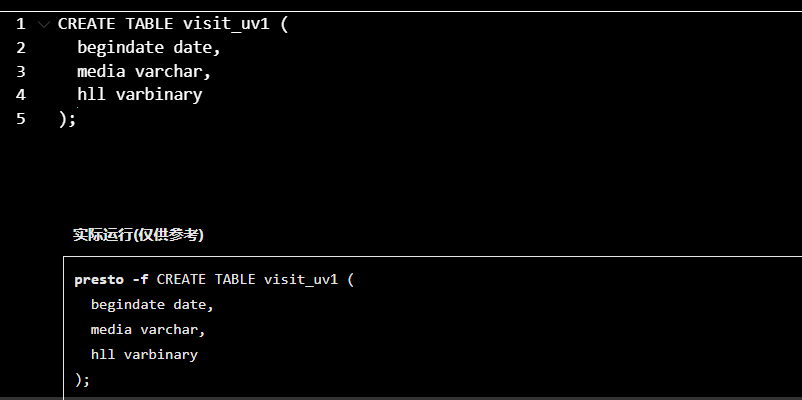
统计用户每天搜索不同词条的个数

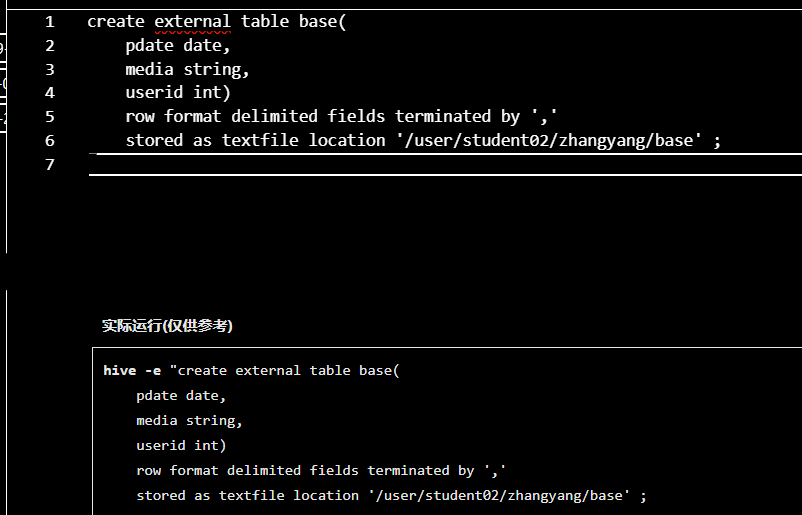
1. 集成hll的组件

spark-alchemy的预先聚合为distinct counts提升性能

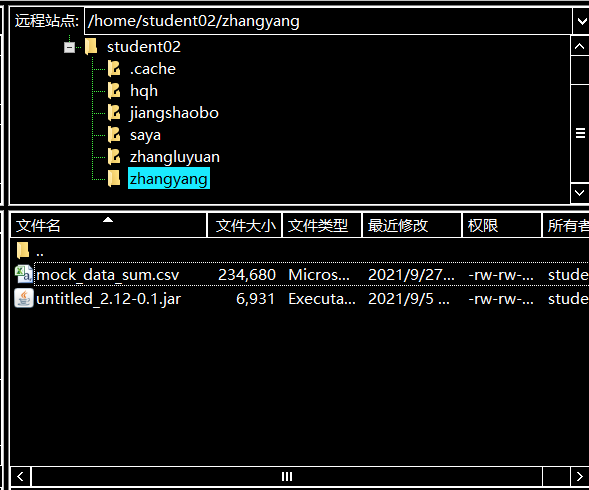
elasticsearch的聚合方法中的基数聚合

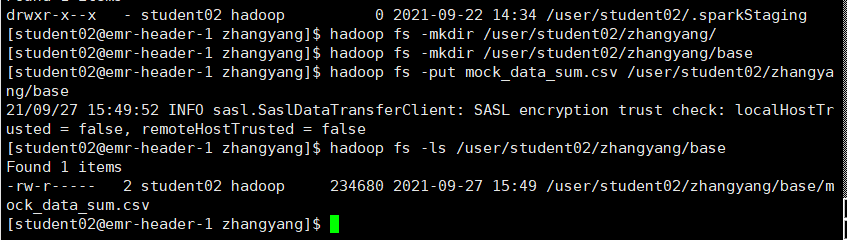
1. **近似计算**
2. 在EMR环境中创建UV用户表visit\_uv1与用户访问事实表base，其中base表为外部表



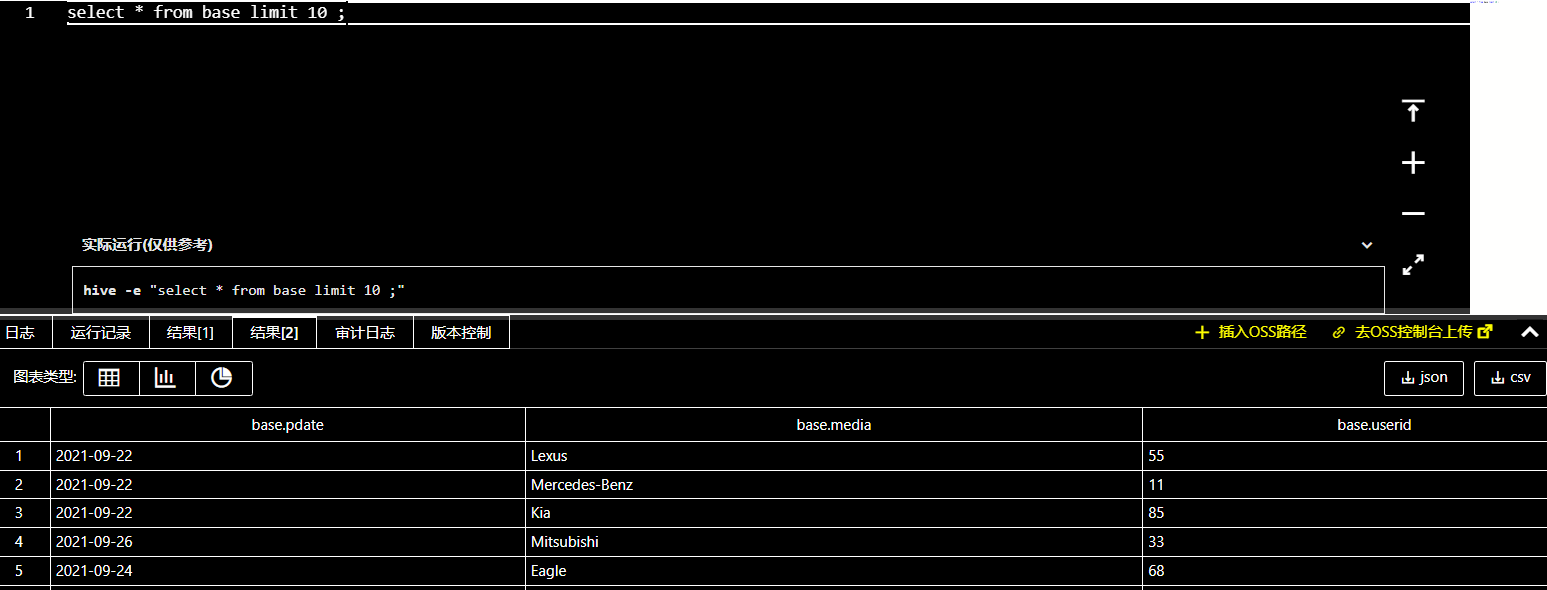


1. 向事实表中插入数据（FILEZILLA传输文件至服务器，再PUT至HDFS）





1. 在EMR环境HIVESQL中查看数据可以正常访问



1. 将base表的数据插入新建的中间表，其中需要求uv的字段用函数approx\_set计算



1. 利用函数，将预估基数值解析出来，得到uv

