

Hive练习8 group by操作

编程要求

在右侧编辑器中补充 SQL，计算不同工作年限以及其平均工资并且过滤出平均工资大于 10000 的。（其中库名：db1，表名：table1）

table1 表结构：

INFO	TYPE
eduLevel_name	String
company_name	String
jobName	String
salary	int
city_code	int
responsibility	String
workingExp	String

本地部分文件内容：

本科,北京联通支付有限公司,大数据开发工程师,10000,530,熟练使用hive等,1-3年
专科,北京联科数创科技有限公司,大数据分析师,8000,530,熟练使用MySQL等数据库,1-3年
本科,湖南智湘赢播网络技术有限公司,大数据开发工程师,16000,749,熟练使用spark等,3-5年

代码

```
-----禁止修改-----
create database if not exists db1;
use db1;

create table if not exists table1(
eduLevel_name string comment '学历',
company_name string comment '公司名',
jobName string comment '职位名称',
salary int comment '薪资',
city_code int comment '城市编码',
responsibility string comment '岗位职责',
workingExp string comment '工作经验'
)
row format delimited fields terminated by ','
lines terminated by '\n'
stored as textfile;
```

```
truncate table table1;
load data local inpath '/root/t1.txt' into table table1;
-----禁止修改-----

-----Begin-----
SELECT avg(salary), workingExp FROM table1 WHERE salary > 10000 GROUP BY
workingExp;
-----End-----
```

解析

相关知识

group by

group by 表示按照某些字段的值进行分组，有相同的值放到一起，需要注意的是 select 后面的非聚合列必须出现在 group by 中；

假设存在 st 表：

city	salary	job
长沙	7000	大数据开发
北京	10000	大数据开发
广州	11000	大数据开发
长沙	7000	大数据开发

可以使用 group by 求出不同省份的平均工资：

```
select city,avg(salary) from st group by city;
```

输出：

```
长沙 7000
北京 10000
广州 11000
```