

Hive练习9 join操作

编程要求

在右侧编辑器中补充 SQL，求出表 `table2` 中所有城市名的平均工资。（其中库名：`db1`，表名：`table1`，表名：`table2`）

表 `table1` 结构：

INFO	TYPE
<code>eduLevel_name</code>	<code>String</code>
<code>company_name</code>	<code>String</code>
<code>jobName</code>	<code>String</code>
<code>salary</code>	<code>int</code>
<code>city_code</code>	<code>int</code>
<code>responsibility</code>	<code>String</code>
<code>workingExp</code>	<code>String</code>

`table1` 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

表 `table2` 结构：

INFO	TYPE
<code>city_code</code>	<code>int</code>
<code>city_name</code>	<code>String</code>

`table2` 本地部分文件内容：

538,上海
653,杭州
749,长沙
763,广州

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/t2.txt' into table table1;

create table if not exists table2(
city_code int comment '城市编码',
city_name string comment '城市名'
)
row format delimited fields terminated by ','
lines terminated by '\n'
stored as textfile;
truncate table table2;
load data local inpath '/root/t22.txt' into table table2;
-----禁止修改-----

-----Begin-----
SELECT  avg(table1.salary), table2.city_name FROM  table1 RIGHT OUTER JOIN
table2 ON table1.city_code = table2.city_code GROUP BY city_name;
-----End-----

```

解析

相关知识

Hive 只支持等值连接，即 ON 子句中只能使用等号连接

假设存在表 a：

id	name
1	bob
2	lily
3	herry

表 b：

cid	score
1	80
2	90
5	60

内连接(JOIN)

内连接指的是把符合两边连接条件的数据查询出来：

```
select a.name,b.score from a join b on a.id=b.cid;
```

输出结果：

```
bob 80
lily 90
```

左外连接 (LEFT OUTER JOIN)

左表全部查询出来，右表不符合连接条件的显示为空：

```
select a.name,b.score from a left outer join b on a.id=b.cid;
```

输出结果：

```
bob    80
lily    90
herry  null
```

右外连接 (RIGHT OUTER JOIN)

右表全部查询出来，左表不符合连接条件的显示为空：

```
select a.name,b.score from a right outer join b on a.id=b.cid;
```

输出结果：

```
bob    80
lily    90
null    60
```

全外连接 (FULL OUTER JOIN)

左右表符合连接条件和不符合连接条件的都查出来，不符合的显示空：

```
select a.name,b.score from a full outer join b on a.id=b.cid;
```

输出结果：

bob	80
lily	90
herry	null
null	60

左半开连接 (LEFT SEMI JOIN)

查询出满足连接条件的左边表记录，需要注意的是 `select` 和 `where` 语句中都不能使用右表的字段。

Hive 不支持右半开连接：

```
select a.name from a LEFT SEMI JOIN b on a.id=b.cid;
```

输出结果：

bob
lily