**2020年安徽省大数据与人工智能应用竞赛A卷**

**学校名称： 队伍名称： 队伍编号：**

# 第三部分：大数据分析（20分）

Hive作为基于hadoop的海量数据处理工具之一，在企业中有广泛应用。可以将复杂的MapReduce计算逻辑使用SQL逻辑处理。

数据  

### 1、 创建一个数据库，以你的组名命名，创建成功后使用use命令切换为该库，并执行set hive.cli.print.current.db=true;截图作为答案（2分）

1）命令截图：

2）执行命令结果截图：

### 2、在HIVE中创建需要的表，并写出建表语句，表结构如下：（3分）

(1).学生资料表(student)

|  |  |  |
| --- | --- | --- |
| 字段 | 字段描述 | 字段类型 |
| s\_id | 学生ID | int |
| s\_name | 学生姓名 | String |
| s\_sex | 学生性别 | String |
| s\_age | 学生年龄 | int |
| s\_dept | 所在系 | String |

1. .课程表 (course)

|  |  |  |
| --- | --- | --- |
| 字段 | 字段描述 | 字段类型 |
| c\_id | 课程ID | int |
| c\_name | 课程名称 | String |

1. 考试成绩表(sc)

|  |  |  |
| --- | --- | --- |
| 字段 | 字段描述 | 字段类型 |
| s\_id | 学生id | int |
| c\_id | 课程id | int |
| score | 成绩 | int |

1. 创建hive表 语句截图：

2）创建成功，执行show tables；截图；

### 3、将数据加载到表中，写出加载数据的语句(3分)

数据文件名称分别与表名称对应，（字段分隔符为，），注意，linux系统和hive默认不支持中文，要自己设置成utf-8编码才可以支持中文。

1) 导入命令截图；

2）执行 select \* from table\_name 并截图（其中table\_name 为前面创建的表名）

执行 select \* from student; 截图

执行select \* from sc; 截图

执行select \* from course; 截图

### 4、利用hive进行数据分析查询（12分）。

1）查询各科成绩平均分(1分)

sql语句截图：

sql语句结果截图：

2）查询选修1号课程的学生最高分数(1分)

sql语句截图：

sql语句结果截图：

3）查询每个学生及其选修课程的情况 (1分)

sql语句截图：

sql语句结果截图：

4）查询选修了3门以上（包含3门）的课程的学生学号(4分)

sql语句截图：

sql语句结果截图：

5）查询每门课程成绩前两名的学生信息（5分）

sql语句截图：

sql语句结果截图：