**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：**

# 第一部分：大数据平台搭建与运维（10分）

**场景说明**

目前，大数据技术及应用已经深入到各行各业，各高校也在积极建设信息化、智慧化校园，那么大数据中心建设必不可少，它可以服务于学校上游和下游，解决数据孤岛问题。假设现在某大学正在假设大数据中心，首先搭建大数据基础平台，用于数据存储和批量计算以及实时查询，现在请你完成Hadoop、Spark、Hive三个大数据框架的搭建及运维：

**要求：使用不少于三个节点，进行比赛，一个为主节点，其他为从节点**

**要求主节点 主机名为 ahdsj2020amaster，各从节点分别名称为ahdsj2020aslave1，ahdsj2020aslave2**

## 任务一、Hadoop安装（5分）

### 配置 hosts实现主机名和ip地址映射。截图hosts文件配置内容（0.5分）

# cat /etc/hosts

### 配置SSH免密登录（截图成功ssh登录到子节点截图）（0.5分）

使用ssh登陆子节点：

# ssh ahdsj2020aslave1

# ssh ahdsj2020aslave2

### 安装配置JDK（截图java -version命令）（0.5分）

# java -version

### 解压Hadoop安装包并修改配置文件（0.5分）

命令:

截图：

### 配置环境变量（截图profile文件配置的关键位置）（0.5分）

命令：

截图：

### 拷贝hadoop到其他的机器上（截图拷贝命令）（0.5分）

命令：

截图：

### 初始化hadoop集群（截图执行命令）（1分）

命令：

截图：

### 启动Hadoop集群（启动成功后执行jps命令，截图主节点和子节点的进程）(1分)

启动命令：

截图：

验证启动结果：

# jps

主节点进程截图：

两个从节点进程截图：

## 任务二、配置spark集群环境（5分）

### 解压spark、修改Spark的配置文件（3分）

1. 解压命令，进入解压后的目录 执行 ll -li命令 并截图（0.5分）

命令：

截图：

1. spark配置环境变量 配置项截图（1分）

命令：

截图：

1. 编辑spark-env.sh文件，命令以及配置项截图（1分）

命令：

截图：

1. 编辑slaves,添加子节点的主机名，命令以及结果截图 （0.5分）

命令：

截图：

### 启动spark集群（2分）

* 1. 分别在主节点 和从节点执行 jps 命令并进行截图

启动命令:

截图：

#jps

master节点截图 ：

slave1节点截图：

slave2节点截图：

* 1. 在主节点 执行 spark-shell 命令并进行截图

# spark-shell

截图：