**2020年全国高校计算机能力挑战赛**

**大数据应用赛答题卡第一部分**

**队长姓名： 刘嘉伟 队长学校： 合肥工业大学**

# 第一部分：大数据平台搭建（20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 dsj2020master，各从节点分别名称为dsj2020slave1，dsj2020slave2**

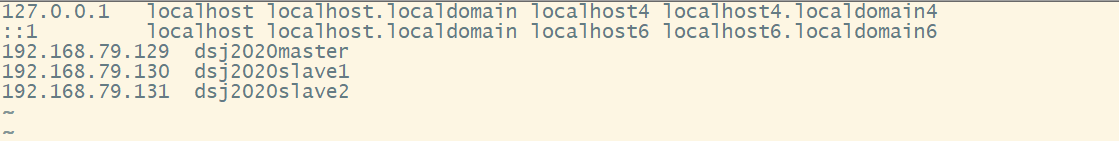
## 任务一、Hadoop集群搭建（5分）

### 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）

命令截图：



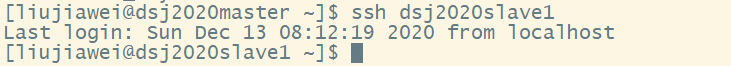
结果截图：



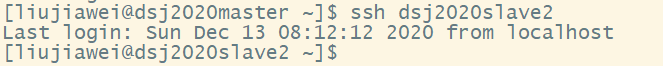
### 配置SSH免密登录（截图成功ssh登录到子节点截图）（0.5分）

使用ssh登陆子节点：

在主节点 执行ssh dsj2020slave1 命令截图

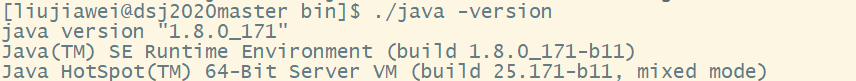


在主节点 执行 ssh dsj2020slave2 命令截图



### 安装配置JDK（截图java -version命令）（0.5分）

执行**java -version命令 截图:**

****

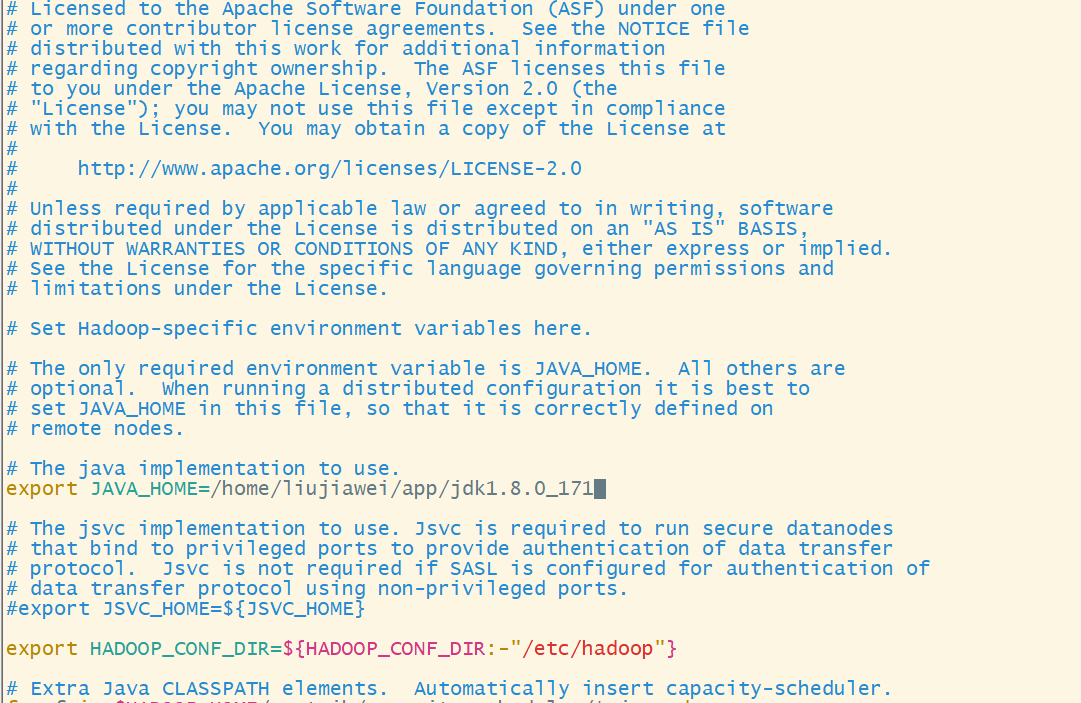
### 解压Hadoop安装包并修改配置文件（1分）

解压命令截图：

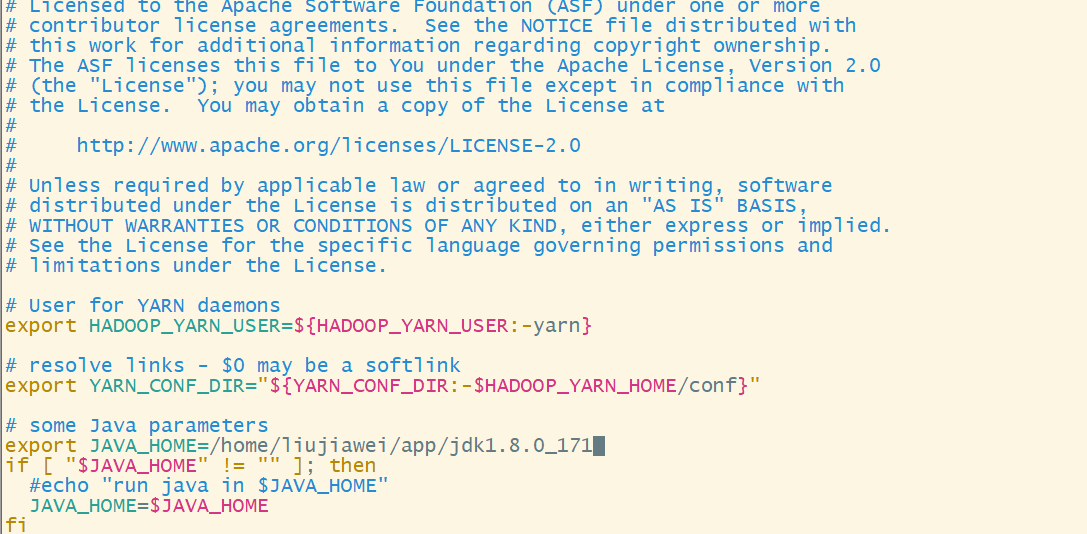


修改配置文件截图：

Hadoop-env.sh



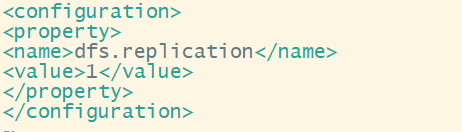
Yarn-env.sh



Core-site.xml



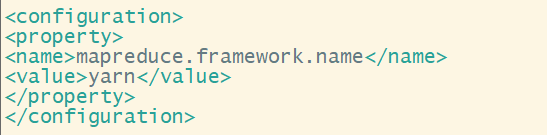
Hdfs-site.xml



Yarn-site.xml



Mapred-site.xml

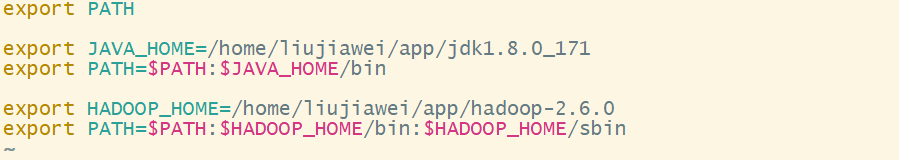


Slaves



### 配置环境变量（截图profile文件配置的关键位置）（0.5分）

**profile文件配置 的配置位置截图：**

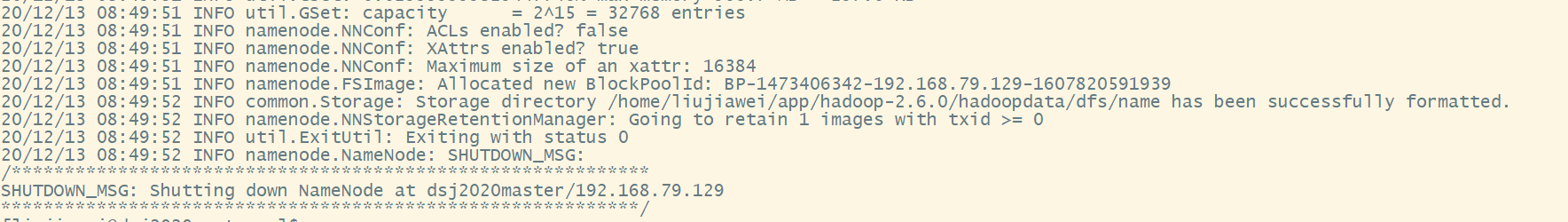
****

### 初始化hadoop集群（截图执行命令）（1分）

初始化命令截图：



结果截图:



### 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)

1）启动命令:（0.5分）

启动命令截图：

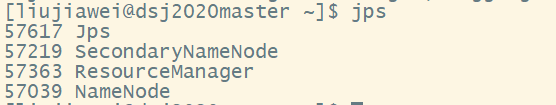




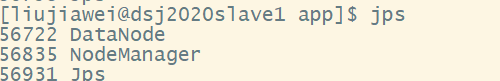
2）分别在主节点 和从节点执行 jps 命令并进行截图（0.5分）

#jps

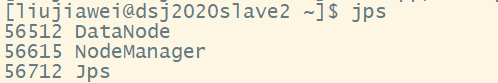
master节点执行jps 截图：



slave1节点执行jps 截图：



slave2节点执行jps 截图：



## 任务二、spark集群搭建（5分）

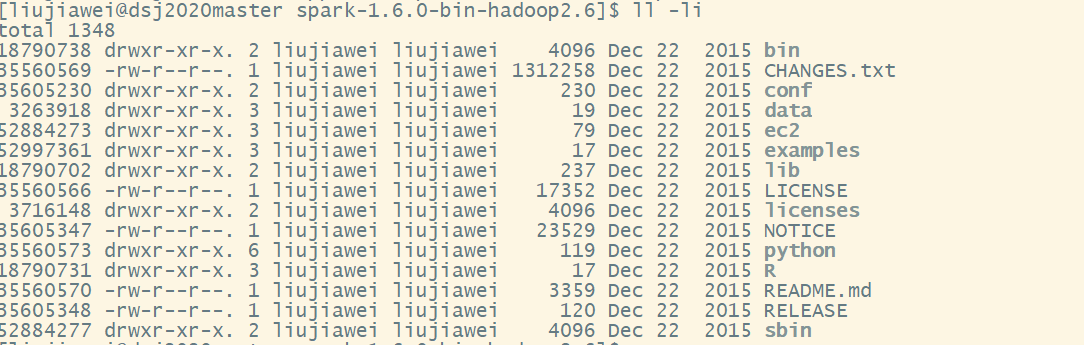
### 解压spark、修改Spark的配置文件（3分）

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 ：



结果截图：

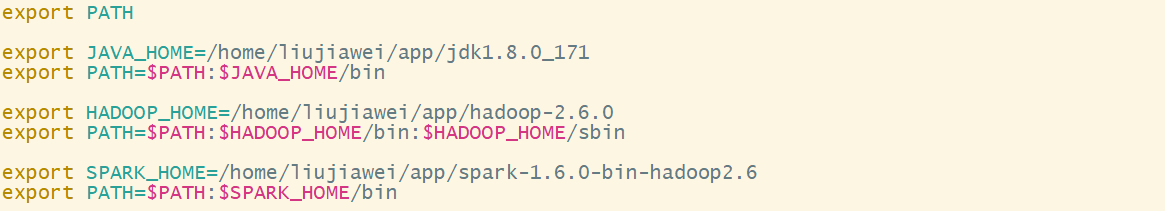


1. spark配置环境变量 配置项截图（1分）

编辑命令截图：



关键项截图：

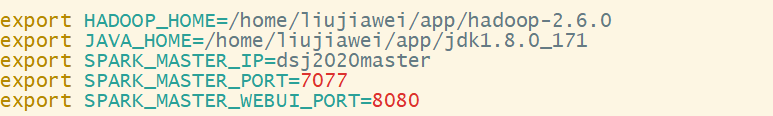


1. 编辑spark-env.sh文件，命令以及配置项截图（1分）

编辑命令截图：



关键项截图：

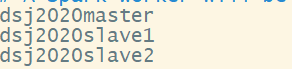


1. 编辑slaves,添加子节点的主机名，命令以及结果截图 （0.5分）

编辑命令截图：



关键项截图：

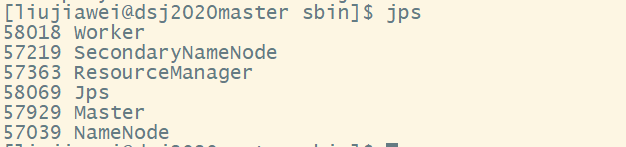


### 启动spark集群（2分）

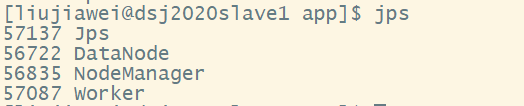
* 1. 分别在主节点 和从节点执行 jps 命令并进行截图（1分）

#jps

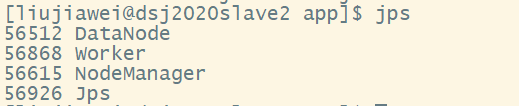
master节点执行jps 截图：



slave1节点执行jps 截图：



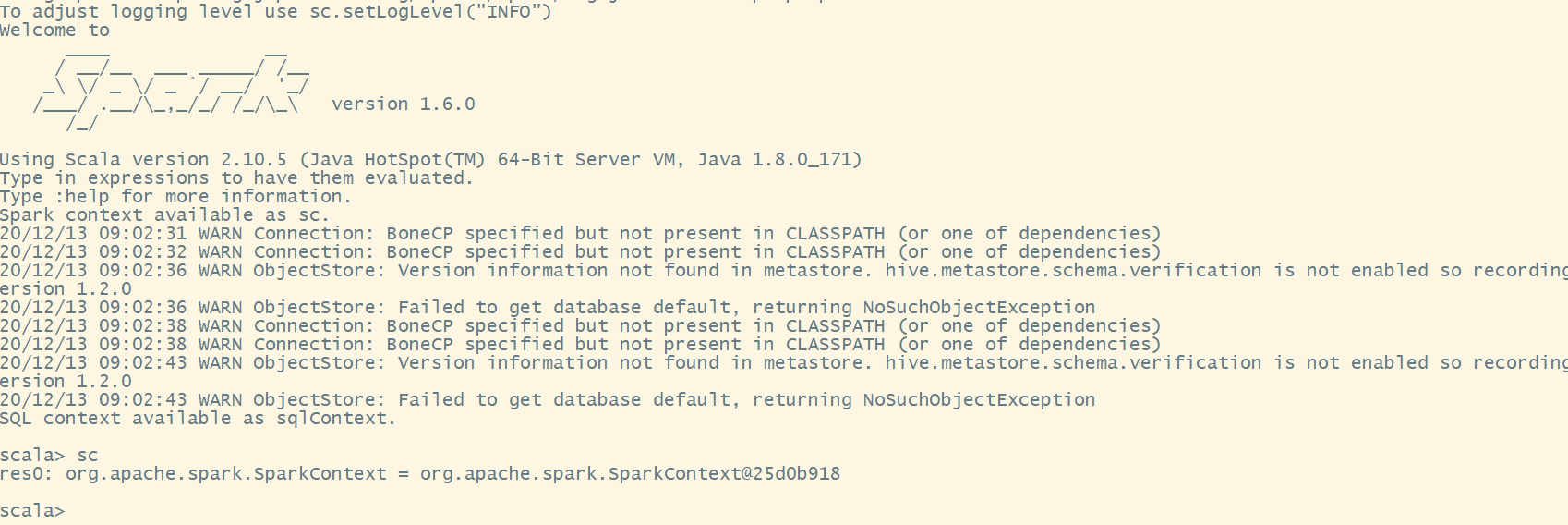
slave2节点执行jps 截图：



* 1. 在主节点 执行 spark-shell 命令并进行截图（1分）

# spark-shell 命令截图





## 任务三、Hbase集群搭建（5分）

**HBase安装**

**我们在搭建好Hadoop集群之后就可以搭建HBase数据仓库了.在安装HBase之前检查Hadoop集群是否处于启动状态,如图**

**下面我们开始HBase的搭建**

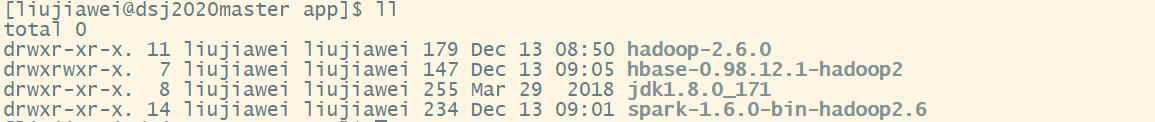
### 1、解压Hase、修改HBase的配置文件（3分）

1）HBase安装包并解压（0.5分）

命令截图 ：



结果截图：

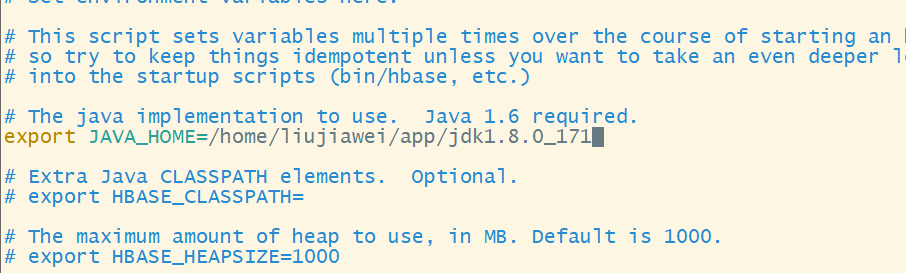


1. 修改hbase-env.sh，（0.5分）

命令截图 ：



结果截图：

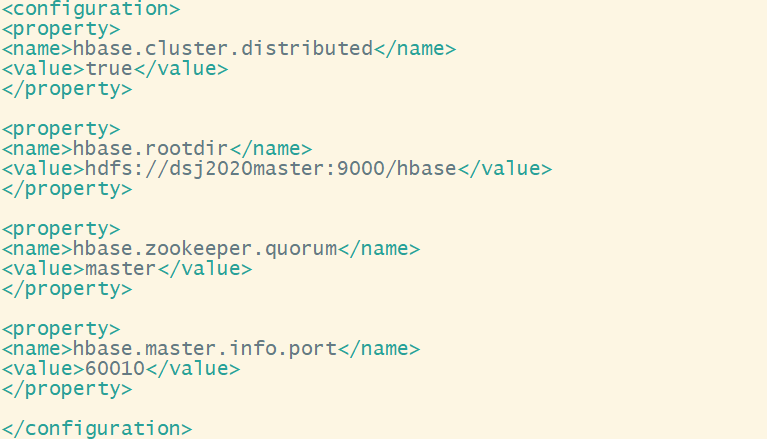


1. 修改hbase-site.xml，（1分）

命令截图 ：



结果截图：



4）配置环境变量(/etc/profile) （0.5分）

命令截图 ：



结果截图：

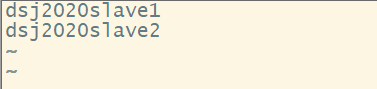


5）添加子节点的主机名，（0.5分）

命令截图 ：



结果截图：



### 2、启动hbase集群（2分）

1）启动命令:（1分）

./start-hbase.sh

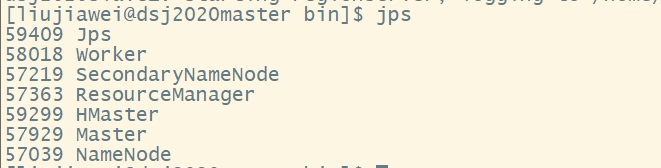
启动命令截图：



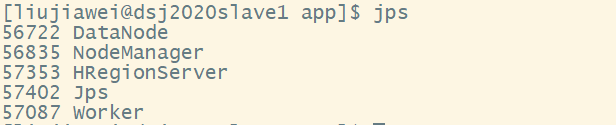
2）分别在主节点 和从节点执行 jps 命令并进行截图（1分）

#jps

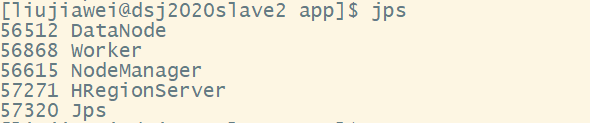
master节点执行jps 截图：



slave1节点执行jps 截图：



slave2节点执行jps 截图：



## 任务四、大数据平台参数优化（5分）

### 1：配置Hadoop时，JAVA\_HOME包含在哪一个配置文件中 请写出 文件路径； （0.5分）

**答：**

**/home/liujiawei/app/hadoop-2.6.0/etc/hadoop/hadoop-env.sh**

**/home/liujiawei/app/hadoop-2.6.0/etc/hadoop/yarn-env.sh**

### 2：Hadoop集群中的HDFS的默认的数据块的大小是多少 请写出大小 包括单位；（0.5分）

**答：**

128MB

### 3：hive中清空表的数据是什么命令；（0.5分）

**答：**

Truncate table 表名

### 4、Spark默认的存储级别 (A )（0.5分）

**答：**

MEMORY\_ONLY

### 5、Spark 中 每个executor进程使用的内存数（0.5分）

**答：**

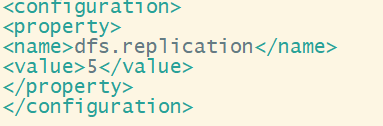
### 6、Hive中删除数据库命令? （0.5分）

**答：**

Drop database 数据库

### 7、设置HDFS数据块的备份数为5（请将关键配置项写出并截图）；（1分）

**答：**



### 8、hive中设置 每个作业的reduce任务数 40（请将关键配置项写出并截图）。（1分）

**答：**