

**大数据处理实验报告**

实验一：HDFS的基本操作

专业班级： CS 2002

学 号： U202015324

姓 名： 屈绍博

指导教师： 石宣化

报告日期： 2022.3.18

**计算机科学与技术学院**

**《大数据处理》课程实验报告**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 实验地点 | | 南一楼804 | 课程名称 | | 大数据处理 | | |
| 实验题目 | | HDFS的基本操作 | 成绩 |  | | 指导教师 |  |
| 教师评价 | □ 实验过程正确； □ 源程序/实验内容提交； □ 程序结构/实验步骤合理；  □ 实验结果正确； □ 语法、语义/命令正确； □ 报告规范；  其他： | | | | | | |
| **一、实验目的**   1. 掌握云平台的配置和使用 2. 熟悉常用Linux命令 3. 掌握HDFS的基本操作 4. 了解HDFS存储数据的原理   **二、实验内容**   1. 实验环境配置 2. 文件准备 (20’) 3. 元数据及副本查看 (30’) 4. DataNode故障模拟 (40’) 5. 实验总结 (10’)   **三、实验环境**  **Master节点** 通用计算增强型 32 vCPUs | 128 GB | ac7.8xlarge.4 系统盘 通用型SSD 480 GB x 1 数据盘 通用型SSD 600 GB x 1  **分析Core节点** 通用计算增强型 8 vCPUs | 32 GB | ac7.2xlarge.4 系统盘 通用型SSD 480 GB x 1 数据盘 通用型SSD 600 GB x 1  **四、实验过程或步骤（源程序）**   1. **文件准备** 2. **创建文件，创建一个名称为U202015324的文件**     **图1.1 文件的创建**  利用dd if=/dev/zero of=U202015324 bs=200M count=1指令创建文件  在 HDFS 中创建文件夹，将文件移动到 hdfs 并显示    **图1.2 文件夹的创建**  利用下列指令在 HDFS 中创建文件夹，将文件移动到 hdfs  hdfs dfs -mkdir /test  hdfs dfs -put filename /test  hdfs dfs -ls /test   1. **查看 hdfs 文件信息**     **图1.3 查看hdfs文件信息**  利用指令hdfs fsck /test/U202015324 -files -blocks -replicaDetails 查看文件信息  可以看到文件大小（Total size），文件数量(Total files)，文件副本，线程等信息。   1. **通过 ssh 进入 0 号块第一个副本所在的数据节点。**     **图1.4 进入0号块数据节点**  通过指令ssh @root+0号块对应的ip地址可以访问该节点，输入密码后即可登入。   1. **查找文件名为块 ID，后缀为.meta的块。**     **图1.5 块的查找**   1. **进入该文件的上层目录，查看该目录下的的块文件。**     **图1.6 查看上层目录下的块文件**   1. **DataNode 故障模拟** 2. **在华为云MRS服务中进入集群**     **图2.1 集群的查看**  并在core 节点中找到上述文件 0 号块第一个副本所在的数据节点。    **图2.2 找到0号块对应节点的ip**  在控制台中查看0号块对应的ip地址为192.168.0.103，并在core节点中找到该节点为第三个    **图2.3 在结点中找到0号块对应的节点**   1. **强制关闭该节点模拟数据离线。**   在节点界面选择关机，即可对节点进行强制关闭。    **图2.4 强制关闭节点**  **图2.5 强制关闭节点**   1. **等待一段时间后，连接 master 节点查看文件详细信息**     **图2.6 等待一段时间后的文件信息**    **图2.7 原始文件信息**  通过比对文件信息中的节点ip，我们可以发现在强制关闭0号节点后文件的ip发生了改变，这说明服务器在另外的节点上生成了新的副本，以保证副本数量不变。   1. **实验结束后删除集群并释放IP**   **图3.1 集群的删除**  **五、出现的问题与解决方案**  在通过ssh进入0号块第一个数据节点的那一步，我在输入密码时在电脑屏幕上并没有给我相应的回馈，我本以为是网络出现了延迟或是电脑发生了卡顿，但在我等待了很长时间并且换了网络后再次尝试还是出现了相同的问题，我才想起来在Linux系统下密码是不会被显示的，最终我正常输入密码后顺利完成了验证登入。  **六、实验总结**  这次是大数据的第一次实验课，我们第一次从头至尾，从配置及购买集群，开辟ip，到进入cloudshell中利用Linux指令对云服务器进行简单的操作，让我对云服务器的工作流有了粗浅的了解，也熟悉了一些Linux的常用指令及系统特点，掌握了一些HDFS的基本操作，对于HDFS存储数据的原理有了一些了解。虽然内容比较基础，但这毕竟是我们第一次上手操作这类云服务器，也需要一个适应的过程，有了这次实验的基础，相信我们在之后的实验中可以对云服务器的使用更加得心应手。 | | | | | | | |