

# 实验二 HDFS 基础入门实验报告

江昱峰 21009200038

2023 年 11 月 13 日

## 1 背景介绍

之前通过 Hadoop 伪分布式集群搭建，已经将我们的基础环境配置完成。我们在此基础上，将 HDFS 分布式文件系统作为 Hadoop 的文件存储系统，结合自身高容错、高吞吐以及可扩展的特点，为我们大数据的大数据集存储提供保障。

## 2 实验目的

实践并掌握 HDFS 基础，具体包括以下两部分内容：

- HDFS 的 shell 命令——增删改查；
- HDFS 的 shell 管理命令。

## 3 实验知识

HDFS 概论：Hadoop 分布式文件系统 (HDFS) 是指被设计成适合运行在通用硬件 (commodity hardware) 上的分布式文件系统 (Distributed File System)。它和现有的分布式文件系统有很多共同点。但同时，它和其他的分布式文件系统的区别也是很明显的。HDFS 是一个高度容错性的系统，适合部署在廉价的机器上。HDFS 能提供高吞吐量的数据访问，非常适合大规模数据集上的应用。HDFS 放宽了一部分 POSIX 约束，来实现流式读取文件系统数据的目的。HDFS 在最开始是作为 Apache Nutch 搜索引擎项目的基础架构而开发的。HDFS 是 Apache Hadoop Core 项目的一部分。

## 4 实验要求

完成 HDFS 基础，具体包括以下两部分头文：

- HDFS 的 shell 命令——增删改查；
- HDFS 的 shell 管理命令。

## 5 实验环境

本次实验实验环境为青椒课堂平台的 Linux（Ubuntu 20.04）操作系统。

## 6 实验步骤

### 6.1 HDFS 的 shell 命令——增删改查

#### 6.1.1 任务 1：操作 HDFS 文件或目录

shell 命令实操：

1. 查看 HDFS 根目录下结构。

```
→ ~ hadoop fs -ls /  
→ ~
```

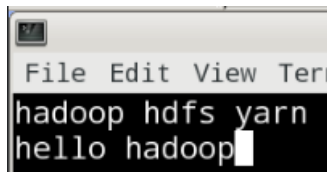
因为我们根目录下没有问价或文件夹，因此不显示内容。

2. 在 HDFS 根目录下创建 root 文件夹。

```
→ ~ hadoop fs -mkdir /root  
→ ~
```

3. 在本地/root 目录下创建 hadoop.txt 文件, 添加如下内容：

```
hadoop hdfs yarn  
hello hadoop
```



4. 将本地 `hadoop.txt` 上传到 HDFS 目录 `/root/` 下。

```
→ root hadoop fs -put /root/hadoop.txt /root/  
→ root
```

5. 将 HDFS 目录文件 `/root/hadoop.txt` 复制到根目录下并查看内容。

6. 删除 HDFS 目录文件 `/root/hadoop.txt`。

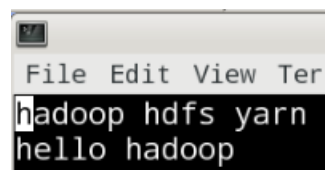
```
→ root hadoop fs -rm /root/hadoop.txt  
23/11/11 14:14:20 INFO fs.TrashPolicyDefault: Namenode trash configuration: Dele  
tion interval = 0 minutes, Empty interval = 0 minutes.  
Deleted /root/hadoop.txt  
→ root
```

7. 将 HDFS 目录文件 `/hadoop.txt` 迁移到 HDFS 目录 `/root/` 下并查看迁移是否成功。

```
→ root hadoop fs -mv /hadoop.txt /root/  
→ root hadoop fs -ls /root/  
Found 1 items  
-rw-r--r-- 1 root supergroup 30 2023-11-11 14:20 /root/hadoop.txt  
→ root
```

8. 将 HDFS 目录文件 `/root/hadoop.txt` 复制到本地根目录下并查看。

```
→ root hadoop fs -get /root/hadoop.txt /  
→ root vim /hadoop.txt
```



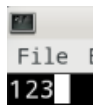
### 6.1.2 任务 2：查看、追加、合并文本

appendToFile 命令实操：

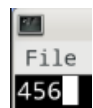
1. 在本地当前目录（/headless）下创建 a.txt, b.txt, c.txt 文件。

```
→ ~ touch a.txt  
→ ~ touch b.txt  
→ ~ touch c.txt  
→ ~
```

2. 分别添加内容 123, 456, 789。



File  
123



File  
456



File  
789

3. 在 HDFS 根目录下创建 abc.txt 文件并查看。

```
→ ~ hadoop fs -touchz /abc.txt  
→ ~
```

4. 将本地 a.txt, b.txt, c.txt 追加到 abc.txt 文件。

```
→ ~ hadoop fs -appendToFile a.txt b.txt c.txt /abc.txt  
→ ~
```

5. 查看 abc.txt 文件。

```
→ ~ hadoop fs -cat /abc.txt  
123  
456  
789  
→ ~
```

getmerge 命令实操:

1. 将刚才创建的 a.txt, b.txt, c.txt 文件上传到 HDFS 根目录。

```
→ ~ hadoop fs -put a.txt b.txt c.txt /  
→ ~
```

2. 将 HDFS 根目录下 \*.txt 文件下载到本地/root/sum.txt。

```
→ ~ hadoop fs -getmerge /*.txt /root/sum.txt  
→ ~
```

## 6.2 HDFS 的 shell 管理命令

### 6.2.1 任务 1: 任务一: 修改权限命令

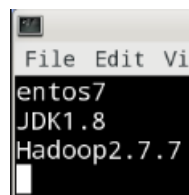
修改权限命令实操:

1. 在本地当前目录创建测试文件 a.txt 并添加内容如下:

Centos7

JDK1.8

Hadoop2.7.7



2. 将文件上传到 HDFS 根目录。

```
→ ~ hadoop fs -put a.txt /  
→ ~
```

3. 修改 a.txt 文件权限为所有用户可读不可写可执行并进行查看。

```
→ ~ hadoop fs -chmod -R 555 /a.txt  
→ ~
```

4. 修改 a.txt 文件所属用户和用户组为 root 并查看。

```
→ ~ hadoop fs -chmod -R 555 /a.txt
→ ~ hadoop fs -ls -R /
-r-xr-xr-x  1 root supergroup          27 2023-11-12 10:33 /a.txt
→ ~
```

```
→ ~ hadoop fs -chown root:root /a.txt
→ ~ hadoop fs -ls -R /
-r-xr-xr-x  1 root root                27 2023-11-12 10:33 /a.txt
→ ~
```

### 6.2.2 任务 2：任务二：统计、设置副本命令

统计命令实操：

1. 统计根目录下目录数，文件数和字节数。

- (1) 创建 HDFS 目录/tmp。

```
→ ~ hadoop fs -mkdir /tmp
→ ~
```

- (2) 将上节任务 a.txt 移动到 HDFS 目录/tmp 下。

```
→ ~ hadoop fs -mv /a.txt /tmp
→ ~
```

2. 统计文件系统的容量、可用空间和已用空间信息。

```
→ ~ hadoop fs -df -h /tmp
Filesystem              Size  Used  Available  Use%
hdfs://localhost:9000  68.8 G  44 K    59.5 G     0%
→ ~
```

3. 统计/tmp 目录下所有文件和文件夹的大小。

```
→ ~ hadoop fs -du -h /tmp
27  /tmp/a.txt
→ ~
```

设置副本命令实操：

设 a.txt 副本数为 2。

```
→ ~ hadoop fs -setrep 2 /tmp/a.txt  
Replication 2 set: /tmp/a.txt  
→ ~
```

## 7 实验结果截图

## 8 困难解决

本次实验较为简单，没有遇到困难。

## 9 心得体会

做完本次实验，除了掌握了实验目的部分中所有内容的收获之外，我还有以下几点心得体会：

- 实践并掌握了 HDFS 查看、追加、合并文本，修改权限命令，统计、设置副本命令等内容；
- 对比分析了 HDFS 与一般 Linux 系统中 Shell 命令的异同点。