

## 大数据技术-第三章：HDFS分布式文件系统 HDFS的基础Shell操作（实验4）



# CONTENTS

01. 项目目标

02. 项目描述

03. 知识准备

04. 项目实施

05. 知识拓展

06. 课后实训



# 01

## 项目目标



## 1 项目目标



01

掌握Hadoop平台使用者角色

02

掌握Hadoop普通用户角色的基础Shell操作

03

掌握Hadoop管理员角色的基础Shell操作

# 02

## 项目描述



## 2 项目描述

在Hadoop平台搭建完成后,通过资源分配,租户划分将平台共享给使用者操作,在这过程中,根据不同使用者的角色可以划分为两类:普通平台用户以及平台管理员。普通平台用户主要是根据实际的业务需求对平台进行使用和操作,平台管理员主要是从全局对平台的健康稳定运行提供保证,所有根据两类人员的工作内容差异,对平台的基本shell命令也有不同的侧重点,此案例主要是介绍在实际的生产过程中两内角色会涉及到的一些常用的Shell命令。

实验操作系统为CentOS-7-x86\_64-DVD-1511, Hadoop版本为hadoop-2.6.5, Linux版本JDK为jdk-8u201-linux-x64, 远程连接工具为SecureCRTPortable, 虚拟机工具为VMware-workstation-full-15.5.0-14665864。

本实验所用软件可根据文件名称在资源库内下载。

# 03

## 知识准备







为了完成本实验，学生需要提前掌握以下理论知识内容：

1. Hadoop普通用户基本命令。
2. Hadoop管理员基本命令。



# 04

## 项目实施



## 4.1 实施思路

基于项目描述与知识准备的内容，我们已经对Hadoop伪分布式集群环境搭建知识点有了一定的理解，现在我们学习如何使用Hadoop伪分布式集群的案例，通过对Hadoop集群使用角色的分类理解，来学习普通用户和管理员角色需要掌握的基本命令

我们将按照下述两个步骤来完成训练。

1. Hadoop普通用户的基本命令：创建文件夹命令、上传文件命令、查看文件内容命令、文件复制命令、删除文件命令、删除文件夹命令、显示目录下文件命令等。
2. Hadoop管理员角色的基本命令：查看正在运行的Job、关闭正在运行的Job、检查HDFS是否损坏、检查HDFS状态,平衡集群中的文件。

## 4.2 实施步骤

### 步骤一：Hadoop普通用户的基本命令：

创建文件夹命令mkdir

##参数表示需要创建的目录

```
[hadoop@djt hadoop]$ hadoop dfs -mkdir /test  
DEPRECATED: Use of this script to execute hdfs command is deprecated.  
Instead use the hdfs command for it.
```

通过ls命令查看文件夹是否创建成功

```
[hadoop@djt hadoop]$ hadoop dfs -ls /  
DEPRECATED: Use of this script to execute hdfs command is deprecated.  
Instead use the hdfs command for it.
```

Found 3 items

drwxr-xr-x	- tony	supergroup	0	2020-11-06	10:42	/middle
drwxr-xr-x	- hadoop	supergroup	0	2021-02-21	18:38	/test
drwx-----	- hadoop	supergroup	0	2020-09-16	09:47	/tmp

## 4.2 实施步骤

### 步骤一：Hadoop普通用户的基本命令：

上传文件命令put或copyFromLocal，文件上传是指将Linux文件系统中的文件上传至HDFS文件系统中,有以下两种方式

```
##第一个参数是指linux文件系统中的文件,第二个参数是上传的HDFS文件系统目的地址  
[hadoop@djt ~]$ hadoop dfs -put test.txt /test  
DEPRECATED: Use of this script to execute hdfs command is deprecated.  
Instead use the hdfs command for it.
```

```
##第一个参数是指linux文件系统中的文件,第二个参数是上传的HDFS文件系统目的地址  
[hadoop@djt ~]$ hadoop dfs -copyFromLocal test.txt /test  
DEPRECATED: Use of this script to execute hdfs command is deprecated.  
Instead use the hdfs command for it.
```

### 检查是否上传成功

```
[hadoop@djt ~]$ hadoop dfs -ls /test  
DEPRECATED: Use of this script to execute hdfs command is deprecated.  
Instead use the hdfs command for it.  
  
Found 1 items  
-rw-r--r--  1 hadoop supergroup      36 2021-02-21 18:45 /test/test.txt
```



## 4.2 实施步骤

### 步骤一：Hadoop普通用户的基本命令：

查看文件内容命令cat或tail或text，查看文件内容是指查看HDFS文件系统中文件具体内容,有以下三种方式

```
##参数表示需要显示的文件
[hadoop@djt ~]$ hadoop dfs -cat /test/test.txt
DEPRECATED: Use of this script to execute hdfs command is deprecated.
Instead use the hdfs command for it.

hadoop  test
hadoop  test
hadoop  test

##参数表示需要显示的目录
[hadoop@djt ~]$ hadoop dfs -tail /test/test.txt
DEPRECATED: Use of this script to execute hdfs command is deprecated.
Instead use the hdfs command for it.

hadoop  test
hadoop  test
hadoop  test

##参数表示需要显示的目录
[hadoop@djt ~]$ hadoop dfs -text /test/test.txt
DEPRECATED: Use of this script to execute hdfs command is deprecated.
Instead use the hdfs command for it.

hadoop  test
hadoop  test
hadoop  test
```

## 4.2 实施步骤

### 步骤一：Hadoop普通用户的基本命令：

文件复制命令get或copyToLocal，文件复制是指将文件从HDFS文件系统中复制至Linux文件系统,有两种方式

##第一个参数是指HDFS文件系统中需要复制的文件,第二个参数是Linux文件系统目的地址

```
[hadoop@djt ~]$ hadoop dfs -get /test/test.txt /home/hadoop/test  
DEPRECATED: Use of this script to execute hdfs command is deprecated.  
Instead use the hdfs command for it.
```

##第一个参数是指HDFS文件系统中需要复制的文件,第二个参数是Linux文件系统目的地址

```
[hadoop@djt ~]$ hadoop dfs -copyToLocal /test/test.txt /home/hadoop/test  
DEPRECATED: Use of this script to execute hdfs command is deprecated.  
Instead use the hdfs command for it.
```

##查看复制是否成功

```
[hadoop@djt ~]$ cd test  
[hadoop@djt test]$ ls  
test.txt
```

## 4.2 实施步骤

### 步骤一：Hadoop普通用户的基本命令：

删除文件命令rm，删除文件命令是指删除HDFS文件系统内文件的操作

##参数表示需要删除的文件

```
[hadoop@djt test]$ hadoop dfs -rm /test/test.txt
```

DEPRECATED: Use of this script to execute hdfs command is deprecated.

Instead use the hdfs command for it.

```
21/02/21 19:08:51 INFO fs.TrashPolicyDefault: Namenode trash configuration: Deletion interval = 0 minutes,
```

```
Empty interval = 0 minutes.
```

```
Deleted /test/test.txt
```



## 4.2 实施步骤

### 步骤一：Hadoop普通用户的基本命令：

删除文件命令rm -r，删除文件夹命令是指删除HDFS文件系统内文件夹的操作

##参数表示需要删除的文件夹

```
[hadoop@djt test]$ hadoop dfs -rm -r /test
```

```
DEPRECATED: Use of this script to execute hdfs command is deprecated.  
Instead use the hdfs command for it.
```

```
21/02/21 19:09:17 INFO fs.TrashPolicyDefault: Namenode trash configuration: Deletion interval = 0 minutes,  
Emptier interval = 0 minutes.  
Deleted /test
```

## 4.2 实施步骤

### 步骤一：Hadoop普通用户的基本命令：

显示目录下文件命令，通过该命令查看HDFS文件系统目录下所有的文件或者目录

##参数表示需要显示的目录

```
[hadoop@djt test]$ hadoop dfs -ls /
```

```
DEPRECATED: Use of this script to execute hdfs command is deprecated.  
Instead use the hdfs command for it.
```

Found 3 items

drwxrwxrwx	- tony	supergroup	0	2020-12-07	04:15	/anagram
drwxr-xr-x	- tony	supergroup	0	2020-11-06	10:42	/middle
drwx-----	- hadoop	supergroup	0	2020-09-16	09:47	/tmp

## 4.2 实施步骤

### 步骤二：Hadoop管理员角色的基本命令：

查看正在运行的Job，此命令用于Hadoop平台管理员对平台上运行的Job进行检查，以便于优化调整Job的执行

```
[hadoop@djt test]$ hadoop job -list
```

```
DEPRECATED: Use of this script to execute mapred command is deprecated.
```

```
Instead use the mapred command for it.
```

```
21/02/21 19:17:15 INFO client.RMPProxy: Connecting to ResourceManager at /0.0.0.0:8032
```

```
Total jobs:0
```

	JobId	State	StartTime	UserName	Queue	Priority
UsedContainers	RsvdContainers	UsedMem	RsvdMem	NeededMem	AM info	

## 4.2 实施步骤

### 步骤二：Hadoop管理员角色的基本命令：

关闭正在运行的Job，此命令用于Hadoop平台管理员对平台上运行的异常任务进行关闭

##参数表示需要关闭的JobID

```
[hadoop@djt test]$ hadoop job -kill job_1432108212572_0001
```

```
DEPRECATED: Use of this script to execute mapred command is deprecated.  
Instead use the mapred command for it.
```



## 4.2 实施步骤

### 步骤二：Hadoop管理员角色的基本命令：

检查HDFS块状态，查看是否损坏。此命令用于检查HDFS文件块是否出现异常状态

```
[hadoop@djt test]$ hadoop fsck /  
DEPRECATED: Use of this script to execute hdfs command is deprecated.  
Instead use the hdfs command for it.  
  
Status: HEALTHY  
Total size:      893887313 B  
Total dirs:      18  
Total files:      31  
Total symlinks:      0  
Total blocks (validated): 33 (avg. block size 27087494 B)  
Minimally replicated blocks: 33 (100.0 %)  
Over-replicated blocks: 0 (0.0 %)  
Under-replicated blocks: 12 (36.363636 %)  
Mis-replicated blocks: 0 (0.0 %)  
Default replication factor: 1  
Average block replication: 1.0  
Corrupt blocks: 0  
Missing replicas: 24 (42.105263 %)  
Number of data-nodes: 1  
Number of racks: 1  
FSCK ended at Sun Feb 21 19:20:17 EST 2021 in 23 milliseconds  
  
The filesystem under path '/' is HEALTHY
```

## 4.2 实施步骤

### 步骤二：Hadoop管理员角色的基本命令：

检查HDFS状态，包括DataNode信息。此命令用于查看HDFS文件系统使用情况以及数据节点状态

```
[hadoop@djt test]$ hadoop dfsadmin -report  
DEPRECATED: Use of this script to execute hdfs command is deprecated.  
Instead use the hdfs command for it.
```

```
Configured Capacity: 18746441728 (17.46 GB)  
Present Capacity: 15933657088 (14.84 GB)  
DFS Remaining: 15032573952 (14.00 GB)  
DFS Used: 901083136 (859.34 MB)  
DFS Used%: 5.66%  
Under replicated blocks: 12  
Blocks with corrupt replicas: 0  
Missing blocks: 0
```

```
-----  
  
Live datanodes (1):
```

## 4.2 实施步骤

### 步骤二：Hadoop管理员角色的基本命令：

平衡集群中的文件。此脚本对于HDFS上数据节点的数据库进行平衡处理,防止某些节点出现负载过重而导致服务停止状况

```
[hadoop@djt sbin]$ ./start-balancer.sh
starting balancer, logging to /home/hadoop/app/hadoop-2.6.5/logs/hadoop-hadoop-balancer-djt.out
Time Stamp          Iteration#  Bytes Already Moved  Bytes Left To Move  Bytes Being Moved
```



# 05

## 项目拓展

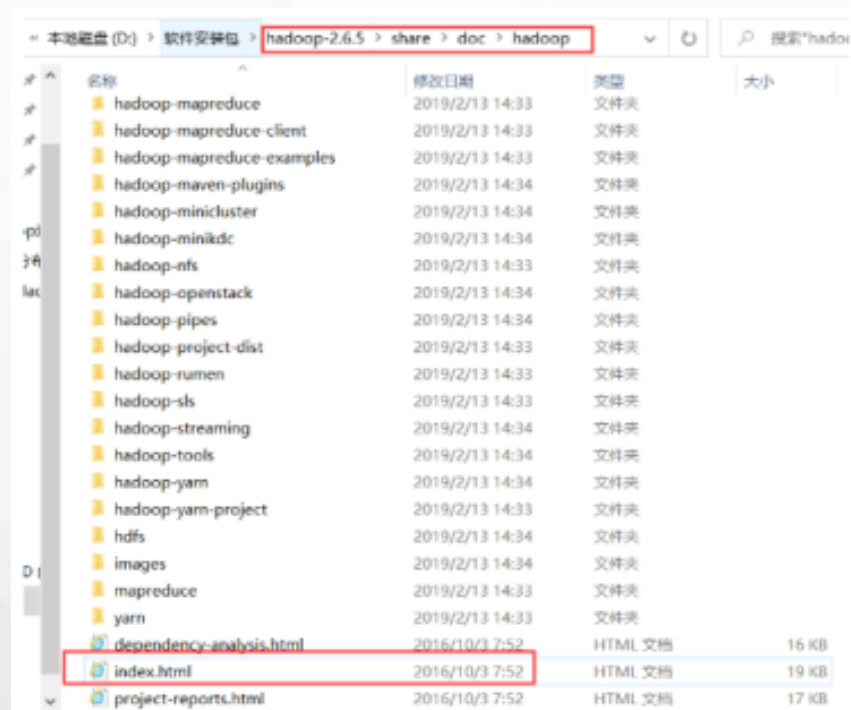
知识点拓展



## 5 知识拓展

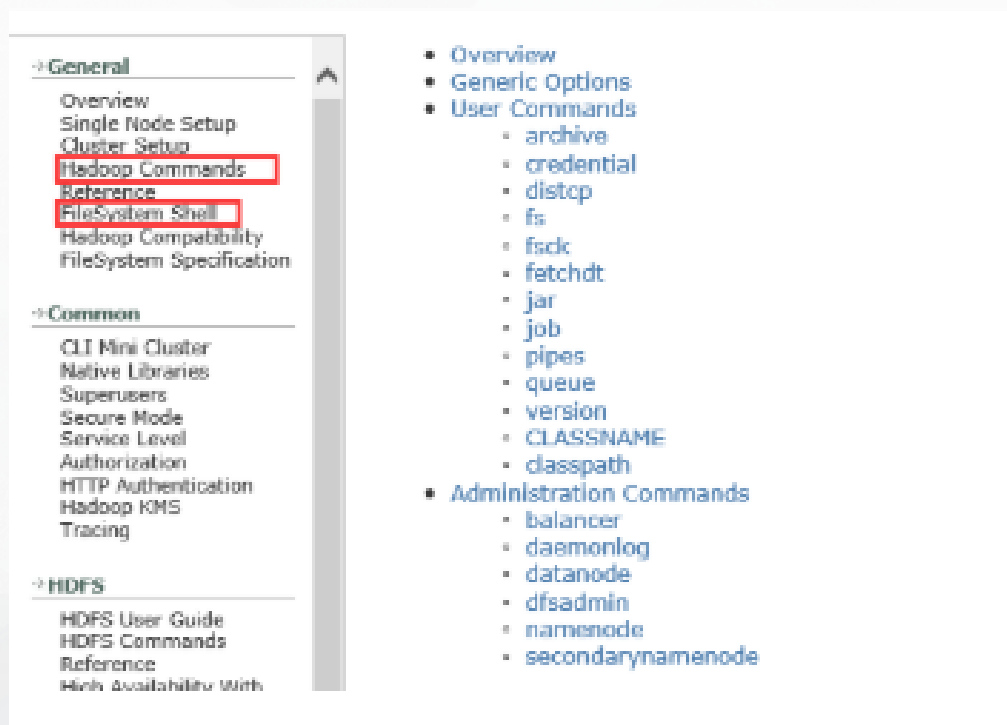
在前面的任务中，我们学习了如何利用Hadoop提供的shell命令来操作平台，而本实验内的只是介绍了常用的一些操作对于在实际工作中会遇到的问题，我们可以采用通过查找Hadoop安装文件内提供的说明文档来解决问题。接下来让我们来了解如何通过查询Hadoop说明文档解决问题（拓展知识点）。

进入Hadoop提供的官方说明文档目录\hadoop-2.6.5\share\doc\hadoop，通过浏览器打开index.html页面



## 5 知识拓展

打开页面后，在左侧菜单栏可以看到Hadoop Commads以及FileSystem Shell两个菜单选项，本实验所介绍的命令在这两个目录内都有详细介绍，除本项目以外的一些Hadoop提供的命令也在文档内都有详细的介绍，后面学习的过程中同学可以参照此方法来查找文档解决问题。



# 06

## 项目实训

课后练习



## 6.1 实训习题

- 简答题：使用HDFS Shell命令如何在HDFS目录内将一个文件拷贝至另一个目录？

Turing AI 万维图灵 | 大数据系列课程

大数据

BIG  
DATA

智 / 能 / 科 / 技      放 / 眼 / 未 / 来

