Hadoop HDFS 실습 가이드
실험 2-2: Hadoop 의사분산 모드 실행 및 HDFS 조작

📋 실습 개요

전제 조건: Hadoop 의사분산 모드 설치가 완료된 상태
(조동사 (Auxiliary Verb): 将 (~할 것이다)
이번 실습에서는 다음 내용을 학습합니다:
- Hadoop 클러스터 시작 및 종료
- HDFS 기본 명령어 사용법
- WordCount 예제 실행
- 파일 시스템 조작 실습

🚀 1. Hadoop 클러스터 시작

1.1 Hadoop 서비스 시작

```
# hadoop 사용자로 로그인 후 실행

# HDFS 서비스 시작 (NameNode, DataNode, SecondaryNameNode)
start-dfs.sh (시간 걸림)

# YARN 서비스 시작 (ResourceManager, NodeManager)
start-yarn.sh

# 또는 모든 서비스를 한 번에 시작
# start-all.sh
```

Hadoop HDFS 实验指南
实验 2-2：Hadoop 伪分布式模式运行及 HDFS 操作

📋 实验概要

前提条件：已完成 Hadoop 伪分布式模式安装

本次实验将学习以下内容：
- Hadoop 集群启动与关闭
- HDFS 基本命令使用方法
- WordCount 示例运行
- 文件系统操作实验

🚀 1. Hadoop 集群启动

1.1 启动 Hadoop 服务

```
# 以 hadoop 用户登录后执行

# 启动 HDFS 服务（NameNode、DataNode、SecondaryNameNode）
$start-dfs.sh

# 启动 YARN 服务（ResourceManager、NodeManager）
$start-yarn.sh

# 或者一次启动所有服务
# start-all.sh
```

명령어	설명	사용 예시
-ls [경로]	디렉토리 내용 나열	hdfs dfs -ls /data
-mkdir [경로]	디렉토리 생성	hdfs dfs -mkdir /user/hadoop
-put [로컬파일] [HDFS경로]	로컬 → HDFS 파일 업로드	hdfs dfs -put file.txt /data/
-get [HDFS파일] [로컬경로]	HDFS → 로컬 파일 다운로드	hdfs dfs -get /data/file.txt ./
-cp [원본] [대상]	HDFS 내 파일 복사	hdfs dfs -cp /data/a.txt /backup/
-mv [원본] [대상]	HDFS 내 파일 이동	hdfs dfs -mv /data/a.txt /archive/
-rm [-r] [경로]	파일/디렉토리 삭제 (-r: 재귀 삭제)	hdfs dfs -rm /data/old.log
-cat [파일경로]	파일 내용 출력	hdfs dfs -cat /data/log.txt
-tail [-f] [파일경로]	파일 끝 부분 실시간 조회 (-f: 실시간)	hdfs dfs -tail -f /data/log.txt

1. 文件/目录操作

命令	说明	示例
-ls [路径]	列出目录内容	hdfs dfs -ls /data
-mkdir [路径]	创建目录	hdfs dfs -mkdir /user/hadoop
-put [本地文件] [HDFS路径]	上传本地文件到HDFS	hdfs dfs -put file.txt /data/
-get [HDFS文件] [本地路径]	下载HDFS文件到本地	hdfs dfs -get /data/file.txt ./
-cp [源路径] [目标路径]	HDFS内复制文件	hdfs dfs -cp /data/a.txt /backup/
-mv [源路径] [目标路径]	HDFS内移动文件	hdfs dfs -mv /data/a.txt /archive/
-rm [-r] [路径]	删除文件/目录 (-r递归删除)	hdfs dfs -rm /data/old.log
-cat [文件路径]	查看文件内容	hdfs dfs -cat /data/log.txt
-tail [-f] [文件路径]	实时查看文件尾部 (-f持续跟踪)	hdfs dfs -tail -f /data/log.txt

2. 文件状态检查

命令	说明	示例
-du [-h] [路径]	查看文件/目录大小 (-h人性化显示)	hdfs dfs -du -h /data
-df [-h]	查看HDFS总磁盘使用量	hdfs dfs -df -h
-count [路径]	统计文件/目录数量和大小	hdfs dfs -count /user
-stat [格式] [路径]	显示文件状态 (如修改时间)	hdfs dfs -stat "%y" /data/file.txt

1.2 Hadoop 프로세스 확인

```
# Java 프로세스 확인
jps
```

정상 실행 시 예상 출력:

12345 NameNode
12456 DataNode
12567 SecondaryNameNode
12678 ResourceManager
12789 NodeManager
12890 Jps

1.3 웹 인터페이스 확인

- HDFS 웹 UI: http://localhost:9870
- YARN 웹 UI: http://localhost:8088

2. WordCount 테스트 실행

2.1 테스트 디렉토리 및 파일 준비

```
# HDFS에 입력 디렉토리 생성
hdfs dfs -mkdir /input

# Hadoop 설정 파일들을 입력 데이터로 사용
hdfs dfs -put $HADOOP_HOME/etc/hadoop/*.xml /user/hadoop/input

# 입력 파일 확인
hdfs dfs -ls input
```

1.2 确认 Hadoop 进程

```
# 检查 Java 进程
jps
```

正常运行时预期输出：

12345 NameNode
12456 DataNode
12567 SecondaryNameNode
12678 ResourceManager
12789 NodeManager
12890 Jps

1.3 确认 Web 界面

- HDFS Web UI：http://localhost:9870
- YARN Web UI：http://localhost:8088

● RPC포트는 9000번임

2. WordCount 测试运行

2.1 准备测试目录和文件

```
# 在 HDFS 中创建输入目录
hdfs dfs -mkdir /input

# 使用 Hadoop 配置文件作为输入数据
hdfs dfs -put $HADOOP_HOME/etc/hadoop/*.xml /user/hadoop/input

# 确认输入文件
hdfs dfs -ls input
```

2.2 WordCount 실행

```
# WordCount MapReduce 작업 실행
hadoop jar $HADOOP_HOME/share/hadoop/mapreduce/hadoop-mapreduce-examples-*.jar wordcount input output

# 실행 결과 확인
hdfs dfs -ls output
hdfs dfs -cat output/part-r-00000 | head -20
hdfs dfs -cat output/part-r-00000 | wc -l
hdfs dfs -cat /output/part-r-00000 | grep "hadoop"
```

3. HDFS 기본 명령어

3.1 디렉토리 관련 명령어

`hdfs dfs -mkdir` - 디렉토리 생성

```
# 기본 사용법
hdfs dfs -mkdir /user/hadoop/test

# 중간 디렉토리도 함께 생성 (-p 옵션)
hdfs dfs -mkdir -p /user/hadoop/data/input
```

`hdfs dfs -ls` - 디렉토리 내용 확인

```
# 기본 사용법
hdfs dfs -ls /

# 특정 디렉토리 확인
hdfs dfs -ls /user

# 재귀적으로 모든 하위 디렉토리 확인
hdfs dfs -ls -R /user
```

2.2 运行 WordCount

```
# 执行 WordCount MapReduce 作业
hadoop jar $HADOOP_HOME/share/hadoop/mapreduce/hadoop-mapreduce-examples-*.jar wordcount input output

# 确认执行结果
hdfs dfs -ls output
hdfs dfs -cat output/part-r-00000 | head -20
hdfs dfs -cat output/part-r-00000 | wc -l
hdfs dfs -cat /output/part-r-00000 | grep "hadoop"
```

3. HDFS 基本命令

3.1 目录相关命令

`hdfs dfs -mkdir` - 创建目录

```
# 基本用法
hdfs dfs -mkdir /user/hadoop/test

# 同时创建中间目录（-p 选项）
hdfs dfs -mkdir -p /user/hadoop/data/input
```

`hdfs dfs -ls` - 查看目录内容

```
# 基本用法
hdfs dfs -ls /

# 查看特定目录
hdfs dfs -ls /user/hadoop

# 递归查看所有子目录
hdfs dfs -ls -R /user
```

3.2 파일 전송 명령어

`hdfs dfs -put` - 로컬에서 HDFS로 업로드

```
# 기본 사용법
hdfs dfs -put /home/hadoop/localfile.txt /user/hadoop/

# 여러 파일 업로드
hdfs dfs -put /home/hadoop/*.txt /user/hadoop/

# 디렉토리 전체 업로드
hdfs dfs -put /home/hadoop/localdir /user/hadoop/
```

`hdfs dfs -get` - HDFS에서 로컬로 다운로드

```
# 기본 사용법
hdfs dfs -get /user/hadoop/test.txt /home/hadoop/

# 특정 디렉토리로 다운로드
hdfs dfs -get /user/hadoop/test.txt /home/hadoop/다운로드/
```

3.3 파일 조작 명령어

`hdfs dfs -cat` - 파일 내용 출력

```
# 파일 내용 출력
hdfs dfs -cat /user/hadoop/test.txt

# 파일 첫 부분만 출력
hdfs dfs -cat /user/hadoop/test.txt | head -10
```

3.2 文件传输命令

`hdfs dfs -put` - 从本地上传到 HDFS

```
# 基本用法
hdfs dfs -put /home/hadoop/localfile.txt /user/hadoop/

# 上传多个文件
hdfs dfs -put /home/hadoop/*.txt /user/hadoop/

# 上传整个目录
hdfs dfs -put /home/hadoop/localdir /user/hadoop/
```

`hdfs dfs -get` - 从 HDFS 下载到本地

```
# 基本用法
hdfs dfs -get /user/hadoop/test.txt /home/hadoop/

# 下载到指定目录
hdfs dfs -get /user/hadoop/test.txt /home/hadoop/下载/
```

3.3 文件操作命令

`hdfs dfs -cat` - 输出文件内容

```
# 输出文件内容
hdfs dfs -cat /user/hadoop/test.txt

# 只输出文件开头部分
hdfs dfs -cat /user/hadoop/test.txt | head -10
```

`hdfs dfs -cp` - HDFS 내에서 파일 복사

```
# 파일 복사
hdfs dfs -cp /user/hadoop/test.txt /user/hadoop/backup/

# 디렉토리 복사
hdfs dfs -cp /user/hadoop/data /user/hadoop/backup/
```

`hdfs dfs -cp` - 在 HDFS 内复制文件

```
# 复制文件
hdfs dfs -cp /user/hadoop/test.txt /user/hadoop/backup/

# 复制目录
hdfs dfs -cp /user/hadoop/data /user/hadoop/backup/
```

`hdfs dfs -rm` - 파일/디렉토리 삭제

```
# 파일 삭제
hdfs dfs -rm /user/hadoop/test.txt

# 디렉토리 삭제 (재귀적)
hdfs dfs -rm -r /user/hadoop/input

# 강제 삭제 (확인 없이)
hdfs dfs -rm -r -f /user/hadoop/temp
```

`hdfs dfs -rm` - 删除文件/目录

```
# 删除文件
hdfs dfs -rm /user/hadoop/test.txt

# 删除目录（递归）
hdfs dfs -rm -r /user/hadoop/input

# 强制删除（无需确认）
hdfs dfs -rm -r -f /user/hadoop/temp
```

🛠 4. 단계별 실습 과제

 # 단계 (1): HDFS 사용자 디렉토리 생성

```
# Hadoop 서비스가 실행 중인지 확인
jps

# 사용자 디렉토리 생성
hdfs dfs -mkdir -p /user/hadoop/temp

# 생성 확인
hdfs dfs -ls
```

예상 출력:

```
Found 1 items
drwxr-xr-x   - hadoop supergroup          0 2024-xx-xx xx:xx /user/hadoop
```

 # 단계 (2): 로컬 파일 생성 및 HDFS 업로드

```
# 로컬 홈 디렉토리로 이동
cd /home/hadoop

# test.txt 파일 생성
cat > test.txt << EOF
안녕하세요, Hadoop 실습입니다!
HDFS 파일 시스템을 배우고 있습니다.
빅데이터 처리를 위한 첫 걸음입니다.
Hadoop Distributed File System
MapReduce 프로그래밍 모델
YARN 리소스 관리자
EOF

# 파일 내용 확인
cat test.txt
# HDFS로 업로드
hdfs dfs -put test.txt /user/hadoop/

# 업로드 확인
hdfs dfs -ls /user/hadoop/
hdfs dfs -ls -h /user/hadoop/test.txt
```

🛠 4. 分步实验任务

 # 步骤 (1)：创建 HDFS 用户目录

```
# 确认 Hadoop 服务正在运行
jps

# 创建用户目录
hdfs dfs -mkdir -p /user/hadoop/temp

# 确认创建结果
hdfs dfs -ls
```

预期输出：

```
Found 1 items
drwxr-xr-x   - hadoop supergroup          0 2024-xx-xx xx:xx /user/hadoop
```

 # 步骤 (2)：创建本地文件并上传到 HDFS

```
# 移动到本地主目录
cd /home/hadoop

# 创建 test.txt 文件
cat > test.txt << EOF
> 您好，这是 Hadoop 实验！
> 我们正在学习HDFS 文件系统。
> 这是大数据处理的第一步。
> Hadoop Distributed File System
> MapReduce 编程模型
> YARM 资源管理器
> EOF

# 确认文件内容
cat test.txt
# 上传到 HDFS
hdfs dfs -put test.txt /user/hadoop/

# 确认上传结果
hdfs dfs -ls /user/hadoop/
hdfs dfs -ls -h /user/hadoop/test.txt
```

```
# 단계 (3): HDFS에서 로컬로 파일 다운로드
# 다운로드 디렉토리 생성
mkdir -p /home/hadoop/다운로드

# HDFS에서 파일 다운로드
hdfs dfs -get /user/hadoop/test.txt /home/hadoop/다운로드/

# 다운로드 확인
ls -la /home/hadoop/다운로드/
echo "=== 다운로드된 파일 내용 ==="
cat /home/hadoop/다운로드/test.txt
```

```
# 步骤 (3)：从 HDFS 下载文件到本地
# 创建下载目录
mkdir -p /home/hadoop/下载

# 从 HDFS 下载文件
hdfs dfs -get /user/hadoop/test.txt /home/hadoop/下载/

# 确认下载结果
ls -la /home/hadoop/下载/
echo "=== 下载文件内容 ==="
cat /home/hadoop/下载/test.txt
```

```
# 단계 (4): HDFS 파일 내용 터미널 출력
# 파일 내용을 터미널에 직접 출력
echo "=== HDFS 파일 내용 출력 ==="
hdfs dfs -cat /user/hadoop/test.txt

# 줄 번호와 함께 출력
echo "=== 줄 번호와 함께 출력 ==="
hdfs dfs -cat /user/hadoop/test.txt | nl
```

```
# 步骤 (4)：在终端中输出 HDFS 文件内容
# 直接在终端输出文件内容
echo "=== HDFS 文件内容输出 ==="
hdfs dfs -cat /user/hadoop/test.txt

# 带行号输出
echo "=== 带行号输出 ==="
hdfs dfs -cat /user/hadoop/test.txt | nl
```

```
# 단계 (5): 하위 디렉토리 생성 및 파일 복사
# input 하위 디렉토리 생성
hdfs dfs -mkdir /user/hadoop/input

# 디렉토리 생성 확인
hdfs dfs -ls /user/hadoop/

# test.txt 파일을 input 디렉토리로 복사
hdfs dfs -cp /user/hadoop/test.txt /user/hadoop/input/

# 복사 결과 확인
echo "=== input 디렉토리 내용 ==="
hdfs dfs -ls /user/hadoop/input/

echo "=== 복사된 파일 내용 확인 ==="
hdfs dfs -cat /user/hadoop/input/test.txt
```

```
# 步骤 (5)：创建子目录并复制文件
# 创建 input 子目录
hdfs dfs -mkdir /user/hadoop/input

# 确认目录创建
hdfs dfs -ls /user/hadoop/

# 将 test.txt 文件复制到 input 目录
hdfs dfs -cp /user/hadoop/test.txt /user/hadoop/input/
# 确认复制结果
echo "=== input 目录内容 ==="
hdfs dfs -ls /user/hadoop/input/

echo "=== 确认复制文件内容 ==="
hdfs dfs -cat /user/hadoop/input/test.txt
```

 # 단계 (6): 파일 및 디렉토리 삭제

```
# 복사 전 현재 상태 확인
echo "=== 삭제 전 상태 ==="
hdfs dfs -ls -R /user/hadoop/

# /user/hadoop 디렉토리 하위의 test.txt 파일 삭제
hdfs dfs -rm /user/hadoop/test.txt

# input 디렉토리와 그 안의 모든 내용 삭제
hdfs dfs -rm -r /user/hadoop/input

# 삭제 후 상태 확인
echo "=== 삭제 후 상태 ==="
hdfs dfs -ls /user/hadoop/
```

예상 출력:

ls: `/user/hadoop': No such file or directory
또는 빈 디렉토리 표시

 # 步骤 (6)：删除文件和目录

```
# 删除前确认当前状态
echo "=== 删除前状态 ==="
hdfs dfs -ls -R /user/hadoop/

# 删除 /user/hadoop 目录下的 test.txt 文件
hdfs dfs -rm /user/hadoop/test.txt

# 删除 input 目录及其所有内容
hdfs dfs -rm -r /user/hadoop/input

# 确认删除后状态
echo "=== 删除后状态 ==="
hdfs dfs -ls /user/hadoop/
```

预期输出：

ls: `/user/hadoop': No such file or directory
或显示空目录

☑ 실습 완료 체크리스트

실습을 진행하면서 각 단계를 완료했는지 확인하세요:

- [ ] Hadoop 서비스 시작 (start-dfs.sh, start-yarn.sh)
- [ ] jps 명령으로 모든 프로세스 실행 확인
- [ ] WordCount 예제 실행 성공
- [ ] `/user/hadoop` 디렉토리 생성
- [ ] 로컬에서 test.txt 파일 생성
- [ ] test.txt 파일을 HDFS에 업로드
- [ ] HDFS에서 로컬 다운로드 디렉토리로 파일 다운로드
- [ ] `hdfs dfs -cat` 명령으로 파일 내용 터미널 출력
- [ ] input 하위 디렉토리 생성
- [ ] test.txt 파일을 input 디렉토리로 복사
- [ ] 파일 및 디렉토리 삭제 완료

☑ 实验完成检查清单

在进行实验时，请检查每个步骤是否完成：

- [ ] Hadoop 服务启动（start-dfs.sh、start-yarn.sh）
- [ ] 用 jps 命令确认所有进程运行
- [ ] WordCount 示例运行成功
- [ ] 创建 `/user/hadoop` 目录
- [ ] 在本地创建 test.txt 文件
- [ ] 将 test.txt 文件上传到 HDFS
- [ ] 从 HDFS 下载文件到本地下载目录
- [ ] 用 `hdfs dfs -cat` 命令在终端输出文件内容
- [ ] 创建 input 子目录
- [ ] 将 test.txt 文件复制到 input 目录
- [ ] 完成文件和目录删除

🔧 5. Hadoop 클러스터 종료

실습 완료 후 Hadoop 서비스를 정상적으로 종료합니다:

```
# YARN 서비스 종료
stop-yarn.sh

# HDFS 서비스 종료
stop-dfs.sh

# 또는 모든 서비스를 한 번에 종료
# stop-all.sh

# 프로세스 종료 확인
jps
```

🔧 5. Hadoop 集群关闭

实验完成后，正常关闭 Hadoop 服务：

```
# 关闭 YARN 服务
stop-yarn.sh

# 关闭 HDFS 服务
stop-dfs.sh

# 或一次关闭所有服务
# stop-all.sh

# 确认进程关闭
jps
```

🔍 6. 추가 유용한 HDFS 명령어

파일 시스템 상태 확인

```
# HDFS 전체 상태 보고
hdfs dfsadmin -report

# 파일 시스템 검사
hdfs fsck /

# Safe Mode 확인 및 해제
hdfs dfsadmin -safemode get
hdfs dfsadmin -safemode leave
```

파일 정보 확인

```
# 파일 상세 정보
hdfs dfs -stat "%n %o %r %u %g %b %y" /user/hadoop/test.txt

# 디스크 사용량 확인
hdfs dfs -du -h /user/hadoop/

# 파일 체크섬 확인
hdfs dfs -checksum /user/hadoop/test.txt
```

파일 권한 관리

```
# 파일 권한 변경
hdfs dfs -chmod 755 /user/hadoop/test.txt

# 파일 소유자 변경
hdfs dfs -chown hadoop:hadoop /user/hadoop/test.txt

# 그룹 변경
hdfs dfs -chgrp supergroup /user/hadoop/test.txt
```

🔍 6. 其他有用的 HDFS 命令

文件系统状态检查

```
# HDFS 整体状态报告
hdfs dfsadmin -report

# 文件系统检查
hdfs fsck /

# Safe Mode 确认与解除
hdfs dfsadmin -safemode get
hdfs dfsadmin -safemode leave
```

文件信息查看

```
# 文件详细信息
hdfs dfs -stat "%n %o %r %u %g %b %y" /user/hadoop/test.txt

# 磁盘使用量查看
hdfs dfs -du -h /user/hadoop/

# 文件校验和查看
hdfs dfs -checksum /user/hadoop/test.txt
```

文件权限管理

```
# 修改文件权限
hdfs dfs -chmod 755 /user/hadoop/test.txt

# 修改文件所有者
hdfs dfs -chown hadoop:hadoop /user/hadoop/test.txt

# 修改组
hdfs dfs -chgrp supergroup /user/hadoop/test.txt
```

⚠ 문제 해결 가이드

자주 발생하는 문제들

1. Safe Mode 문제

```
# Safe Mode 상태 확인
hdfs dfsadmin -safemode get

# Safe Mode 강제 해제
hdfs dfsadmin -safemode leave
```

2. 권한 문제

```
# Hadoop 디렉토리 소유권 변경
sudo chown -R hadoop:hadoop $HADOOP_HOME
sudo chown -R hadoop:hadoop /tmp/hadoop-*
```

3. 포트 충돌

```
# 포트 사용 상태 확인
netstat -tulpn | grep :9000
netstat -tulpn | grep :9870
```

4. 프로세스가 시작되지 않는 경우

```
# 로그 파일 확인
tail -f $HADOOP_HOME/logs/hadoop-hadoop-namenode-*.log
tail -f $HADOOP_HOME/logs/hadoop-hadoop-datanode-*.log
```

⚠ 问题解决指南

常见问题

1. Safe Mode 问题

```
# 检查 Safe Mode 状态
hdfs dfsadmin -safemode get

# 强制解除 Safe Mode
hdfs dfsadmin -safemode leave
```

2. 权限问题

```
# 修改 Hadoop 目录所有权
sudo chown -R hadoop:hadoop $HADOOP_HOME
sudo chown -R hadoop:hadoop /tmp/hadoop-*
```

3. 端口冲突

```
# 检查端口使用状态
netstat -tulpn | grep :9000
netstat -tulpn | grep :9870
```

4. 进程无法启动

```
# 检查日志文件
tail -f $HADOOP_HOME/logs/hadoop-hadoop-namenode-*.log
tail -f $HADOOP_HOME/logs/hadoop-hadoop-datanode-*.log
```

실습 팁

- 명령어 실행 전: 항상 `jps` 명령으로 필요한 프로세스가 실행 중인지 확인
- 경로 확인: HDFS 경로는 `/`로 시작하고, 로컬 경로는 절대경로나 상대경로 모두 가능
- 안전한 실습: 중요한 데이터는 백업 후 실습 진행
- 웹 UI 활용: 명령줄과 함께 웹 인터페이스도 활용하여 상태 확인

실습 중 문제가 발생하면 로그 파일을 확인하고, 각 단계를 차근차근 따라해보세요!

实验技巧

- 执行命令前：始终用 `jps` 命令确认必要的进程正在运行
- 路径确认：HDFS 路径以 `/` 开头，本地路径可以是绝对路径或相对路径
- 安全实验：重要数据请备份后再进行实验
- 利用 Web UI：结合命令行和 Web 界面

实验中遇到问题时，请检查日志文件，并按步骤仔细操作！

!