第一次实验 熟悉常用的HDFS操作

# 实验目的

1. 理解HDFS在Hadoop体系结构中的角色；

2. 熟练使用HDFS操作常用的Shell命令；

3. 熟悉HDFS操作常用的Java API。

# 实验平台

操作系统：Linux

Hadoop版本：2.6.0或以上版本

JDK版本：1.6或以上版本

Java IDE：Eclipse

# 实验内容和要求

1. 编程实现以下指定功能，并利用Hadoop提供的Shell命令完成相同任务：
2. 向HDFS中上传任意文本文件，如果指定的文件在HDFS中已经存在，由用户指定是追加到原有文件末尾还是覆盖原有的文件；
3. 从HDFS中下载指定文件，如果本地文件与要下载的文件名称相同，则自动对下载的文件重命名；
4. 将HDFS中指定文件的内容输出到终端中；
5. 显示HDFS中指定的文件的读写权限、大小、创建时间、路径等信息；
6. 给定HDFS中某一个目录，输出该目录下的所有文件的读写权限、大小、创建时间、路径等信息，如果该文件是目录，则递归输出该目录下所有文件相关信息；
7. 提供一个HDFS内的文件的路径，对该文件进行创建和删除操作。如果文件所在目录不存在，则自动创建目录；
8. 提供一个HDFS的目录的路径，对该目录进行创建和删除操作。创建目录时，如果目录文件所在目录不存在则自动创建相应目录；删除目录时，由用户指定当该目录不为空时是否还删除该目录；
9. 向HDFS中指定的文件追加内容，由用户指定内容追加到原有文件的开头或结尾；
10. 删除HDFS中指定的文件；
11. 删除HDFS中指定的目录，由用户指定目录中如果存在文件时是否删除目录；
12. 在HDFS中，将文件从源路径移动到目的路径。
13. 编程实现一个类“MyFSDataInputStream”，该类继承“org.apache.hadoop.fs.FSDataInputStream”，要求如下：实现按行读取HDFS中指定文件的方法“readLine()”，如果读到文件末尾，则返回空，否则返回文件一行的文本。
14. 查看Java帮助手册或其它资料，用“java.net.URL”和“org.apache.hadoop.fs.FsURLStreamHandlerFactory”编程完成输出HDFS中指定文件的文本到终端中。

# 实验报告

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 《大数据技术基础》课程机房上机实验报告 | | | | |
| 题目： |  | 姓名 |  | 日期 |
| 实验环境： | | | | |
| 实验内容与完成情况： | | | | |
| 出现的问题： | | | | |
| 解决方案（列出遇到的问题和解决办法，列出没有解决的问题）： | | | | |