

数据科学导论实验 指南

实验 4

熟悉常用的 HDFS 操作

(版本号: 2021 年 02 月 21 日版本)

目录

1. 实验目的	1
2. 实验平台	1
3. 实验内容和要求	1
4. 实验报告	2

实验 4

熟悉常用的 HDFS 操作

1 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。

2 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

3 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：
 - (1) 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；
 - (2) 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；
 - (3) 将 HDFS 中指定文件的内容输出到终端中；
 - (4) 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；
 - (5) 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；
 - (6) 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；
 - (7) 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；
 - (8) 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；
 - (9) 删除 HDFS 中指定的文件；
 - (10) 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；
 - (11) 在 HDFS 中，将文件从源路径移动到目的路径。
2. 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。
3. 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。

4 实验报告

《数据科学导论》课程实验报告				
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 and 解决办法，列出没有解决的问题）：				