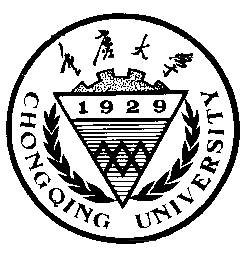
数据科学导论实验报告

实验一 python数据基础实践



学 生：严奕凡\_20220815

王国祥\_20221704

郑俊博\_20220652

张天议\_20220847

年 级：2022级

专 业：数据科学与大数据技术

重庆大学大数据与软件学院

2024年 5月 24日

## 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

## 实验要求

操作系统：Linux

Hadoop 版本：2.6.0 或以上版本 实测3.2.2

JDK 版本：1.6 或以上版本

Java IDE：Eclipse

## 开发环境：

Windows下Vmware虚拟机搭载Linux Ubunto 64位

## 实验内容：

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HD FS 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.hado

op.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.h

adoop.fs.FsURLStreamHandlerFactory”编程完成输出HDFS 中指定文件

的文本到终端中。

## 实验结果分析(重点)

## 问题和解决

我的代码仓库地址：[mozhongzhou/CQU\_Python\_experiment (github.com)](https://github.com/mozhongzhou/CQU_Python_experiment)