

# MiniDFS

---

## 1 项目介绍

---

项目设计了一个迷你分布式文件系统(Mini-DFS)，系统包含一个Client，一个Name server，四个Data servers（可以通过系统参数调整Data servers的数量）。其中Client通过进程运行，Name server和Data servers是该进程的不同线程。

本迷你分布式文件系统有两种模式，本地模式和分布式模式。本地模式下，Data servers是本地的若干个文件夹，分布式模式下Data servers是若干台服务器。

## 2 功能

---

系统功能：

- 读/写/删除文件。上传文件成功并返回文件 ID;根据文件 ID 和偏移量读取文件内容。
- 文件条带化。将文件切分成几个 chunks，每个 chunk 的大小为 2MB，并在四个 data servers 之间均匀分布。
- 副本。每个 chunk 包含三个副本，并且分布在不同的 Data server 中。保证不同的Data Server中 chunk 使用MD5 checksum 的结果相同。

系统结构：

- **Client**: 提供文件的读/写/删除/下载接口。
- **NameServer**: 列出文件和chunks之间的关系，列出副本和data servers的关系；管理Data server。
- **DataServer**: 读/写/删除本地chunk。

# 3 代码结构

## 3.1 客户端 (Client, Name server)

### main.py

main.py是系统的启动代码，通过调用core/init\_server.py启动name节点和data节点，然后通过core/deal\_cmd.py对用户输入的指令进行处理，得到用户的指令类型。

方法	说明
run(events, params_data)	循环运行，根据用户输入的命令进行处理

### core/config.py

系统的配置信息，主要存储系统变量。

变量 (部分)	值	说明
CHUNK_SIZE	2 * 1024 * 1024	chunk大小 (字节)
NUM_DATA_SERVER	4	data server数量
SYSTEM_Mode	'DISTRIBUTED'/'LOCAL'	系统运行模式(分布式/本地)
NAME_NODE_META_PATH	'./dfs/namenode/meta.pkl'	元数据文件存储位置
operation_names	('put', 'read', 'fetch', 'quit', 'ls', 'delete', 'll')	系统支持的指令集

### core/init\_server.py

方法	说明
init_name_node(name, events, params_data)	初始化Name Node
init_data_node(name, server_id, events, params_data)	初始化Data node，赋予相应的server_id

### core/deal\_cmd.py

方法	说明
class ParamsData()	存放全局变量，用于线程间传递信息
def deal_cmd(cmd_str, params_data)	处理用户输入的命令，将处理结果存储到ParamsData对象中

## core/namenode.py

方法	指令格式	说明
def run(self)	无	接受进程的通知并调用相应方法处理任务
def load_meta(self)	无	加载元数据
def update_meta(self)	无	更新元数据
def ls(self)	ls	打印系统文件清单
def ll(self)	ll	打印系统chunk文件详细信息
def put(self)	put source_file_path	上传文件，切分成chunk并通知datanode线程将其上传至服务器
def delete(self)	delete file_id	删除文件，通知相应的线程删除Data servers上的chunk文件
def read(self)	read file_id offset count	根据文件id读取给定偏移位置后的给定字节大小的数据
def fetch(self)	fetch file_id save_path	下载文件，通知相应的线程删除Data servers上的chunk文件
def create_socket(self, server_id)	无	创建连接Data servers的socket

## core/datanode.py

方法	说明
def save(self)	接收Name server的通知并存储接收到的chunk内容
def delete(self)	接收Name server的通知并删除指定的chunk
def read(self)	接收Name server的通知并传输的chunk中指定内容
def fetch(self)	接收Name server的通知并传输指定的chunk
def create_socket(self)	创建和Data servers的socket连接

## 3.2 服务器端 (Data Server)

### DataServer.py

方法	说明
def run()	等待data node线程的通知，并在Data servers上执行相应任务
def create_socket()	创建和Data servers的socket连接
def receive_socket( <i>conn</i> )	

# 4 功能实现及运行实例

## 4.1 分布式模式

- 启动

```
+ MiniDFS git:(master) python3 main.py
NameServer 启动
DataServer0 启动
DataServer1 启动
DataServer2 启动
DataServer3 启动
MiniDFS >>> █
```

- 上传文件---put

Client:

```
MiniDFS >>> put test.txt
DataServer-0 save file-0-0 0 2097152
DataServer-1 save file-0-0 0 2097152
DataServer-2 save file-0-0 0 2097152
DataServer-3 save file-0-0 0 2097152
DataServer-0 save file-0-1 2097152 2097152
DataServer-2 save file-0-1 2097152 2097152
DataServer-1 save file-0-1 2097152 2097152
DataServer-3 save file-0-1 2097152 2097152
DataServer-2 save file-0-2 4194304 2097152
DataServer-0 save file-0-2 4194304 2097152
DataServer-3 save file-0-2 4194304 2097152
DataServer-1 save file-0-2 4194304 2097152
DataServer-0 save file-0-3 6291456 2097152
DataServer-1 save file-0-3 6291456 2097152
DataServer-2 save file-0-3 6291456 2097152
DataServer-3 save file-0-3 6291456 2097152
DataServer-0 save file-0-4 8388608 1898764
DataServer-2 save file-0-4 8388608 1898764
DataServer-1 save file-0-4 8388608 1898764
DataServer-2 save file-0-4 8388608 1898764
Put succeed! File_ID is 0
MiniDFS >>> put test-2.txt
DataServer-2 save file-1-0 0 2097152
DataServer-3 save file-1-0 0 2097152
DataServer-1 save file-1-0 0 2097152
DataServer-0 save file-1-0 0 2097152
DataServer-2 save file-1-1 2097152 2097152
DataServer-0 save file-1-1 2097152 2097152
DataServer-3 save file-1-1 2097152 2097152
DataServer-1 save file-1-1 2097152 2097152
DataServer-2 save file-1-2 4194304 2097152
DataServer-0 save file-1-2 4194304 2097152
DataServer-2 save file-1-2 4194304 2097152
DataServer-3 save file-1-2 4194304 2097152
DataServer-2 save file-1-3 6291456 2097152
DataServer-0 save file-1-3 6291456 2097152
DataServer-1 save file-1-3 6291456 2097152
DataServer-3 save file-1-3 6291456 2097152
DataServer-0 save file-1-4 8388608 1898764
DataServer-2 save file-1-4 8388608 1898764
DataServer-1 save file-1-4 8388608 1898764
DataServer-3 save file-1-4 8388608 1898764
Put succeed! File_ID is 1
```

## Data server:

The image shows four terminal windows side-by-side, each displaying the output of a HDFS save operation. The logs show multiple connections from the same IP address (27.115.122.232) attempting to save different parts of a file ('file-0-0' through 'file-0-4'). Each part is successfully saved, and the process ends with an error message indicating a list index out of range.

```
[root@ecs-a310-5a76-0003: /home/hadoop (ssh)] [root@ecs-a310-5a76-0004: /home/hadoop (ssh)] [root@ecs-a310-5a76-0001: /home/hadoop (ssh)] [root@ecs-a310-5a76-0000: /home/hadoop (ssh)]
```

```
[info]: wait for connection...  
[info]: Connected by 27.115.122.232:40447  
-----file-0-0-----save successful-----  
[info]: wait for connection...  
[info]: Connected by 27.115.122.232:40451  
-----file-0-1-----save successful-----  
[info]: wait for connection...  
[info]: Connected by 27.115.122.232:40458  
-----file-0-2-----save successful-----  
[info]: wait for connection...  
[info]: Connected by 27.115.122.232:40462  
-----file-0-3-----save successful-----  
[info]: wait for connection...  
[info]: Connected by 27.115.122.232:40464  
-----file-0-4-----save successful-----  
[info]: wait for connection...  
[info]: Connected by 27.115.122.232:40471  
[Error]: list index out of range  
[info]: wait for connection...  
  
[info]: wait for connection...  
[info]: Connected by 27.115.122.232:40444  
-----file-0-0-----save successful-----  
[info]: wait for connection...  
[info]: Connected by 27.115.122.232:40452  
-----file-0-1-----save successful-----  
[info]: wait for connection...  
[info]: Connected by 27.115.122.232:40456  
-----file-0-2-----save successful-----  
[info]: wait for connection...  
[info]: Connected by 27.115.122.232:40460  
-----file-0-3-----save successful-----  
[info]: wait for connection...  
[info]: Connected by 27.115.122.232:40463  
-----file-0-4-----save successful-----  
[info]: wait for connection...  
[info]: Connected by 27.115.122.232:40476  
[Error]: list index out of range  
[info]: wait for connection...  
[info]: Connected by 27.115.122.232:40478  
-----file-1-0-----save successful-----  
  
[info]: wait for connection...  
[info]: Connected by 27.115.122.232:40449  
-----file-0-0-----save successful-----  
[info]: wait for connection...  
[info]: Connected by 27.115.122.232:40453  
-----file-0-1-----save successful-----  
[info]: wait for connection...  
[info]: Connected by 27.115.122.232:40455  
-----file-0-2-----save successful-----  
[info]: wait for connection...  
[info]: Connected by 27.115.122.232:40459  
-----file-0-3-----save successful-----  
[info]: wait for connection...  
[info]: Connected by 27.115.122.232:40466  
-----file-0-4-----save successful-----  
[info]: wait for connection...  
[info]: Connected by 27.115.122.232:40473  
[Error]: list index out of range  
[info]: wait for connection...  
  
[info]: wait for connection...  
[info]: Connected by 27.115.122.232:40450  
-----file-0-0-----save successful-----  
[info]: wait for connection...  
[info]: Connected by 27.115.122.232:40454  
-----file-0-1-----save successful-----  
[info]: wait for connection...  
[info]: Connected by 27.115.122.232:40457  
-----file-0-2-----save successful-----  
[info]: wait for connection...  
[info]: Connected by 27.115.122.232:40461  
-----file-0-3-----save successful-----  
[info]: wait for connection...  
[info]: Connected by 27.115.122.232:40465  
-----file-0-4-----save successful-----  
[info]: wait for connection...  
[info]: Connected by 27.115.122.232:40474  
[Error]: list index out of range  
[info]: wait for connection...  
[info]: Connected by 27.115.122.232:40477  
-----file-1-0-----save successful-----
```

- 查看文件信息---ls, ---查看chunk分布情况---ll

```
MiniDFS >>> ls  
File total number 3  
0 test.txt 10287372  
1 test-2.txt 10287372  
2 test.txt 10287372  
MiniDFS >>> ll  
id name size chunk chunk_server  
0 test.txt 2097152 file-0-0 [0, 1, 2, 3]  
0 test.txt 2097152 file-0-1 [1, 2, 3, 0]  
0 test.txt 2097152 file-0-2 [2, 3, 0, 1]  
0 test.txt 2097152 file-0-3 [3, 0, 1, 2]  
0 test.txt 1898764 file-0-4 [0, 1, 2, 3]  
1 test-2.txt 2097152 file-1-0 [1, 2, 3, 0]  
1 test-2.txt 2097152 file-1-1 [2, 3, 0, 1]  
1 test-2.txt 2097152 file-1-2 [3, 0, 1, 2]  
1 test-2.txt 2097152 file-1-3 [0, 1, 2, 3]  
1 test-2.txt 1898764 file-1-4 [1, 2, 3, 0]  
2 test.txt 2097152 file-2-0 [2, 3, 0, 1]  
2 test.txt 2097152 file-2-1 [3, 0, 1, 2]  
2 test.txt 2097152 file-2-2 [0, 1, 2, 3]  
2 test.txt 2097152 file-2-3 [1, 2, 3, 0]  
2 test.txt 1898764 file-2-4 [2, 3, 0, 1]
```

- 删除文件---delete

Client:

```

MiniDFS >>> delete 1
Usage: put|read|fetch|quit|ls|delete
MiniDFS >>> delete 1
DataServer-0 file-1-0 delete successful
DataServer-3 file-1-0 delete successful
DataServer-2 file-1-0 delete successful
DataServer-1 file-1-0 delete successful
DataServer-3 file-1-1 delete successful
DataServer-1 file-1-1 delete successful
DataServer-0 file-1-1 delete successful
DataServer-2 file-1-1 delete successful
DataServer-0 file-1-2 delete successful
DataServer-2 file-1-2 delete successful
DataServer-3 file-1-2 delete successful
DataServer-1 file-1-2 delete successful
DataServer-1 file-1-3 delete successful
DataServer-0 file-1-3 delete successful
DataServer-3 file-1-3 delete successful
DataServer-2 file-1-3 delete successful
DataServer-3 file-1-4 delete successful
DataServer-2 file-1-4 delete successful
DataServer-0 file-1-4 delete successful
DataServer-1 file-1-4 delete successful
Delete succeed! File_ID is 1
MiniDFS >>> ls
File total number 2
0 test.txt 10287372
2 test.txt 10287372
MiniDFS >>> ll
id name size chunk chunk_server
0 test.txt 2097152 file-0-0 [0, 1, 2, 3]
0 test.txt 2097152 file-0-1 [1, 2, 3, 0]
0 test.txt 2097152 file-0-2 [2, 3, 0, 1]
0 test.txt 2097152 file-0-3 [3, 0, 1, 2]
0 test.txt 1898764 file-0-4 [0, 1, 2, 3]
2 test.txt 2097152 file-2-0 [2, 3, 0, 1]
2 test.txt 2097152 file-2-1 [3, 0, 1, 2]
2 test.txt 2097152 file-2-2 [0, 1, 2, 3]
2 test.txt 2097152 file-2-3 [1, 2, 3, 0]
2 test.txt 1898764 file-2-4 [2, 3, 0, 1]

```

Data server:

```

root@ecs-a310-5a76-0003: /home/hadoop (ssh)
root@ecs-a310-5a76-0004: /home/hadoop (ssh)
root@ecs-a310-5a76-0001: /home/hadoop (ssh)
root@ecs-a310-5a76-0002: /home/hadoop (ssh)

[info]: wait for connection...
[info]: Connected by 27.115.122.232:40628
[Error]: list index out of range
[info]: wait for connection...
[info]: Connected by 27.115.122.232:40631
-----/home/hadoop/file-1-1-----delete successful-----
[info]: wait for connection...
[info]: Connected by 27.115.122.232:40635
[Error]: list index out of range
[info]: wait for connection...
[info]: Connected by 27.115.122.232:40639
-----/home/hadoop/file-1-2-----delete successful-----
[info]: wait for connection...
[info]: Connected by 27.115.122.232:40514
[Error]: list index out of range
[info]: wait for connection...
[info]: Connected by 27.115.122.232:40519
-----/home/hadoop/file-1-3-----delete successful-----
[info]: wait for connection...
[info]: Connected by 27.115.122.232:40514
[Error]: list index out of range
[info]: wait for connection...
[info]: Connected by 27.115.122.232:40519
-----/home/hadoop/file-1-4-----delete successful-----
-----/home/hadoop/file-1-1-----delete successful-----
[info]: wait for connection...
[info]: Connected by 27.115.122.232:40629
[Error]: list index out of range
[info]: wait for connection...
[info]: Connected by 27.115.122.232:40633
-----/home/hadoop/file-1-2-----delete successful-----
[info]: wait for connection...
[info]: Connected by 27.115.122.232:40636
[Error]: list index out of range
[info]: wait for connection...
[info]: Connected by 27.115.122.232:40638
-----/home/hadoop/file-1-3-----delete successful-----
[info]: wait for connection...
[info]: Connected by 27.115.122.232:40515
[Error]: list index out of range
[info]: wait for connection...
[info]: Connected by 27.115.122.232:40518
-----/home/hadoop/file-1-4-----delete successful-----
-----/home/hadoop/file-1-1-----delete successful-----
[info]: wait for connection...
[info]: Connected by 27.115.122.232:40622
-----/home/hadoop/file-1-2-----delete successful-----
[info]: wait for connection...
[info]: Connected by 27.115.122.232:40626
[Error]: list index out of range
[info]: wait for connection...
[info]: Connected by 27.115.122.232:40630
-----/home/hadoop/file-1-2-----delete successful-----
[info]: wait for connection...
[info]: Connected by 27.115.122.232:40637
[Error]: list index out of range
[info]: wait for connection...
[info]: Connected by 27.115.122.232:40512
-----/home/hadoop/file-1-3-----delete successful-----
[info]: wait for connection...
[info]: Connected by 27.115.122.232:40516
[Error]: list index out of range
[info]: wait for connection...
[info]: Connected by 27.115.122.232:40521
-----/home/hadoop/file-1-4-----delete successful-----

```

- 读取文件内容---read

```

MiniDFS >>> read 0 3000000 20
DataServer-1:
but publishers i un

root@ecs-a310-5a76-0003: /home/hadoop (ssh)      root@ecs-a310-5a76-0004: /home/hadoop (ssh)
[info]: Connected by 27.115.122.232:40519          [info]: Connected by 27.115.122.232:40518
-----/home/hadoop/file-1-4-----delete successful-----
[info]: wait for connection...
[info]: Connected by 27.115.122.232:40528          [info]: wait for connection...
[Error]: list index out of range                   [Error]: list index out of range
[info]: wait for connection...                      [info]: wait for connection...
[info]: Connected by 27.115.122.232:40665          [info]: Connected by 27.115.122.232:40658
[Error]: list index out of range                   [Error]: list index out of range
[info]: wait for connection...                      [info]: wait for connection...
[info]: Connected by 27.115.122.232:40666          [info]: Connected by 27.115.122.232:40659
-----file-2-3-----fetch successful-----        -----file-2-0-----fetch successful-----
[info]: wait for connection...                      [info]: wait for connection...
[info]: Connected by 27.115.122.232:40824          [info]: Connected by 27.115.122.232:40671
[Error]: list index out of range                   [Error]: list index out of range
[info]: wait for connection...                      [info]: wait for connection...
[info]: Connected by 27.115.122.232:40825          [info]: Connected by 27.115.122.232:40672
-----file-0-1-----read successful-----         -----file-2-4-----fetch successful-----
[info]: wait for connection...                      [info]: wait for connection...
[info]: list index out of range                   [info]: list index out of range
[info]: wait for connection...                      [info]: wait for connection...
[info]: Connected by 27.115.122.232:40513          [info]: Connected by 27.115.122.232:40512
-----/home/hadoop/file-1-3-----delete successful-----
[info]: wait for connection...
[info]: Connected by 27.115.122.232:40520          [info]: wait for connection...
[Error]: list index out of range                   [Error]: list index out of range
[info]: wait for connection...                      [info]: wait for connection...
[info]: Connected by 27.115.122.232:40520          [info]: Connected by 27.115.122.232:40521
-----/home/hadoop/file-1-4-----delete successful-----
[info]: wait for connection...
[info]: Connected by 27.115.122.232:40527          [info]: wait for connection...
[Error]: list index out of range                   [Error]: list index out of range
[info]: wait for connection...                      [info]: wait for connection...
[info]: Connected by 27.115.122.232:40662          [info]: Connected by 27.115.122.232:40658
[Error]: list index out of range                   [Error]: list index out of range
[info]: wait for connection...                      [info]: wait for connection...
[info]: Connected by 27.115.122.232:40663          [info]: Connected by 27.115.122.232:40659
-----file-2-2-----fetch successful-----        -----file-2-1-----fetch successful-----
[info]: wait for connection...                      [info]: wait for connection...
[info]: list index out of range                   [info]: list index out of range
[info]: wait for connection...                      [info]: wait for connection...
[info]: list index out of range                   [info]: list index out of range
[info]: wait for connection...                      [info]: wait for connection...
[info]: Connected by 27.115.122.232:40666          [info]: Connected by 27.115.122.232:40663
-----file-2-3-----fetch successful-----        Finished download!

```

- 下载文件---fetch

```

MiniDFS >>> fetch 2 a.txt
DataServer-2  fetch file-2-0
DataServer-3  fetch file-2-1
DataServer-0  fetch file-2-2
DataServer-1  fetch file-2-3
DataServer-2  fetch file-2-4
Finished download!

```

```

root@ecs-a310-5a76-0003:/home/hadoop (ssh) [root@ecs-a310-5a76-0003:/home/hadoop (ssh)]
[info]: Connected by 27.115.122.232:40639
-----/home/hadoop/file-1-3-----delete successful-----
[info]: wait for connection...
[info]: Connected by 27.115.122.232:40514
[Error]: list index out of range
[info]: wait for connection...
[info]: Connected by 27.115.122.232:40519
-----/home/hadoop/file-1-4-----delete successful-----
[info]: wait for connection...
[info]: Connected by 27.115.122.232:40528
[Error]: list index out of range
[info]: wait for connection...
[info]: Connected by 27.115.122.232:40665
-----/home/hadoop/file-1-5-----delete successful-----
[info]: wait for connection...
[info]: Connected by 27.115.122.232:40666
[Error]: list index out of range
[info]: wait for connection...
[info]: Connected by 27.115.122.232:40666
-----file-2-3-----fetch successful----- ←
[info]: wait for connection...
[ ] ←

root@ecs-a310-5a76-0003:/home/hadoop (ssh) [root@ecs-a310-5a76-0004:/home/hadoop (ssh)]
[info]: Connected by 27.115.122.232:40518
-----/home/hadoop/file-1-4-----delete successful-----
[info]: wait for connection...
[info]: Connected by 27.115.122.232:40529
-----/home/hadoop/file-1-5-----delete successful-----
[info]: wait for connection...
[info]: Connected by 27.115.122.232:40655
-----/home/hadoop/file-1-6-----delete successful-----
[info]: wait for connection...
[info]: Connected by 27.115.122.232:40656
-----file-2-0-----fetch successful----- ←
[info]: wait for connection...
[info]: Connected by 27.115.122.232:40671
-----/home/hadoop/file-1-7-----delete successful-----
[info]: wait for connection...
[info]: Connected by 27.115.122.232:40672
-----file-2-1-----fetch successful----- ←
[info]: wait for connection...
[ ] ←

root@ecs-a310-5a76-0001:/home/hadoop (ssh) [root@ecs-a310-5a76-0001:/home/hadoop (ssh)]
[Error]: list index out of range
[info]: wait for connection...
[info]: Connected by 27.115.122.232:40513
-----/home/hadoop/file-1-3-----delete successful-----
[info]: wait for connection...
[info]: Connected by 27.115.122.232:40517
[Error]: list index out of range
[info]: wait for connection...
[info]: Connected by 27.115.122.232:40520
-----/home/hadoop/file-1-4-----delete successful-----
[info]: wait for connection...
[info]: Connected by 27.115.122.232:40527
[Error]: list index out of range
[info]: wait for connection...
[info]: Connected by 27.115.122.232:40662
[Error]: list index out of range
[info]: wait for connection...
[info]: Connected by 27.115.122.232:40663
-----file-2-2-----fetch successful----- ←
[info]: wait for connection...
[ ] ←

root@ecs-a310-5a76-0001:/home/hadoop (ssh) [root@ecs-a310-5a76-0001:/home/hadoop (ssh)]
[Error]: list index out of range
[info]: wait for connection...
[info]: Connected by 27.115.122.232:40512
-----/home/hadoop/file-1-3-----delete successful-----
[info]: wait for connection...
[info]: Connected by 27.115.122.232:40516
[Error]: list index out of range
[info]: wait for connection...
[info]: Connected by 27.115.122.232:40521
-----/home/hadoop/file-1-4-----delete successful-----
[info]: wait for connection...
[info]: Connected by 27.115.122.232:40530
-----/home/hadoop/file-1-5-----delete successful-----
[info]: wait for connection...
[info]: Connected by 27.115.122.232:40658
[Error]: list index out of range
[info]: wait for connection...
[info]: Connected by 27.115.122.232:40659
-----file-2-1-----fetch successful----- ←
[info]: wait for connection...
[ ] ←

```

- 退出系统--quit

```

MiniDFS >>> quit
Bye: Exiting miniDFS...
→ MiniDFS git:(master) x

```

- md5sum

```

root@ecs-a310-5a76-0003:/home/hadoop (ssh) [root@ecs-a310-5a76-0003:/home/hadoop (ssh)]
[info]: Connected by 27.115.122.232:41439
-----/home/hadoop/file-3-2-----delete successful-----
[info]: wait for connection...
[info]: Connected by 27.115.122.232:41461
-----/home/hadoop/file-3-3-----delete successful-----
[info]: wait for connection...
[info]: Connected by 27.115.122.232:41468
-----/home/hadoop/file-3-4-----delete successful-----
[info]: wait for connection...
[info]: Connected by 27.115.122.232:41469
-----/home/hadoop/file-3-5-----delete successful-----
[info]: wait for connection...
[info]: Connected by 27.115.122.232:41470
-----/home/hadoop/file-3-6-----delete successful-----
[info]: wait for connection...
^CTraceback (most recent call last):
  File "DataServer.py", line 89, in <module>
    run()
  File "DataServer.py", line 17, in run
    conn, addr = sock.accept()
  File "/usr/lib/python2.7/socket.py", line 206, in accept
    sock, addr = self._sock.accept()
KeyboardInterrupt
root@ecs-a310-5a76-0003:/home/hadoop# md5sum file-0-0
367cc4d32cba0980e758fae79c2dd5d file-0-0
root@ecs-a310-5a76-0003:/home/hadoop# [ ] ←

root@ecs-a310-5a76-0004:/home/hadoop (ssh) [root@ecs-a310-5a76-0004:/home/hadoop (ssh)]
[info]: Connected by 27.115.122.232:41460
-----/home/hadoop/file-3-3-----delete successful-----
[info]: wait for connection...
[info]: Connected by 27.115.122.232:41465
-----/home/hadoop/file-3-4-----delete successful-----
[info]: wait for connection...
[info]: Connected by 27.115.122.232:41467
-----/home/hadoop/file-3-5-----delete successful-----
[info]: wait for connection...
^CTraceback (most recent call last):
  File "DataServer.py", line 89, in <module>
    run()
  File "DataServer.py", line 17, in run
    conn, addr = sock.accept()
  File "/usr/lib/python2.7/socket.py", line 206, in accept
    sock, addr = self._sock.accept()
KeyboardInterrupt
root@ecs-a310-5a76-0004:/home/hadoop# md5sum file-0-0
367cc4d632cba0980e758fae79c2dd5d file-0-0
root@ecs-a310-5a76-0004:/home/hadoop# [ ] ←

root@ecs-a310-5a76-0001:/home/hadoop (ssh) [root@ecs-a310-5a76-0001:/home/hadoop (ssh)]
[Error]: list index out of range
[info]: wait for connection...
[info]: Connected by 27.115.122.232:41462
-----/home/hadoop/file-3-3-----delete successful-----
[info]: wait for connection...
[info]: Connected by 27.115.122.232:41463
[Error]: list index out of range
[info]: wait for connection...
[info]: Connected by 27.115.122.232:41466
-----/home/hadoop/file-3-4-----delete successful-----
[info]: wait for connection...
^CTraceback (most recent call last):
  File "DataServer.py", line 89, in <module>
    run()
  File "DataServer.py", line 17, in run
    conn, addr = sock.accept()
  File "/usr/lib/python2.7/socket.py", line 206, in accept
    sock, addr = self._sock.accept()
KeyboardInterrupt
root@ecs-a310-5a76-0001:/home/hadoop# md5sum file-0-0
367cc4d632cba0980e758fae79c2dd5d file-0-0
root@ecs-a310-5a76-0001:/home/hadoop# [ ] ←

```

## 4.2 分布式模式 (data server 0 掉线后)

- 查看文件---断连的服务器会用-1代替

```
MiniDFS >>> ls
File total number 2
0           test.txt      10287372
2           test.txt      10287372

MiniDFS >>> ll
DataNode 0 [Error]: [Errno 61] Connection refused
id      name       size    chunk    chunk_server
0       test.txt   2097152 file-0-0   [-1, 1, 2, 3]
0       test.txt   2097152 file-0-1   [1, 2, 3, -1]
0       test.txt   2097152 file-0-2   [2, 3, -1, 1]
0       test.txt   2097152 file-0-3   [3, -1, 1, 2]
0       test.txt   1898764 file-0-4   [-1, 1, 2, 3]
2       test.txt   2097152 file-2-0   [2, 3, -1, 1]
2       test.txt   2097152 file-2-1   [3, -1, 1, 2]
2       test.txt   2097152 file-2-2   [-1, 1, 2, 3]
2       test.txt   2097152 file-2-3   [1, 2, 3, -1]
2       test.txt   1898764 file-2-4   [2, 3, -1, 1]
```

- 上传文件---会提示断连的服务器

```
MiniDFS >>> put test.txt
DataNode 0 [Error]: timed out
DataServer-2 save file-4-0 0 2097152
DataServer-1 save file-4-0 0 2097152
DataServer-3 save file-4-0 0 2097152
DataServer-2 save file-4-1 2097152 2097152
DataServer-1 save file-4-1 2097152 2097152
DataServer-3 save file-4-1 2097152 2097152
DataServer-2 save file-4-2 4194304 2097152
DataServer-3 save file-4-2 4194304 2097152
DataServer-1 save file-4-2 4194304 2097152
DataServer-1 save file-4-3 6291456 2097152
DataServer-2 save file-4-3 6291456 2097152
DataServer-3 save file-4-3 6291456 2097152
DataServer-2 save file-4-4 8388608 1898764
DataServer-1 save file-4-4 8388608 1898764
DataServer-3 save file-4-4 8388608 1898764
Put succeed! File_ID is 4
```

- 读取文件---会先从存储此chunk的服务器读取，若断连则会从备份服务器读取

```
MiniDFS >>> read 0 200000 20
DataNode 0 [Error]: timed out
DataServer-1:
    aer banknote berlit
```

```

root@ecs-a310-5a76-0003: /home/hadoop (ssh)
[info]: wait for connection...
[info]: Connected by 27.115.122.232:41225
-----file-4-4-----save successful-----
[info]: wait for connection...
[info]: Connected by 27.115.122.232:41246
[Error]: list index out of range
[info]: wait for connection...
[info]: Connected by 27.115.122.232:41256
[Error]: list index out of range
[info]: wait for connection...
[info]: Connected by 27.115.122.232:41257
-----file-0-1-----read successful-----
[info]: wait for connection...
[info]: Connected by 27.115.122.232:41337
[Error]: list index out of range
[info]: wait for connection...
[info]: Connected by 27.115.122.232:41338
-----file-0-0-----read successful-----
[info]: wait for connection...

root@ecs-a310-5a76-0004: /home/hadoop (ssh)
[info]: wait for connection...
[info]: Connected by 27.115.122.232:41212
-----file-4-0-----save successful-----
[info]: wait for connection...
[info]: Connected by 27.115.122.232:41215
-----file-4-1-----save successful-----
[info]: wait for connection...
[info]: Connected by 27.115.122.232:41218
-----file-4-2-----save successful-----
[info]: wait for connection...
[info]: Connected by 27.115.122.232:41220
-----file-4-3-----save successful-----
[info]: wait for connection...
[info]: Connected by 27.115.122.232:41226
-----file-4-4-----save successful-----
[info]: wait for connection...
[info]: Connected by 27.115.122.232:41247
[Error]: list index out of range
[info]: wait for connection...

root@ecs-a310-5a76-0001: /home/hadoop (ssh)
[info]: Connected by 27.115.122.232:41211
[Error]: list index out of range
[info]: wait for connection...
[info]: Connected by 27.115.122.232:41213
-----file-4-0-----save successful-----
[info]: wait for connection...
[info]: Connected by 27.115.122.232:41216
-----file-4-1-----save successful-----
[info]: wait for connection...
[info]: Connected by 27.115.122.232:41219
-----file-4-2-----save successful-----
[info]: wait for connection...
[info]: Connected by 27.115.122.232:41222
-----file-4-3-----save successful-----
[info]: wait for connection...
[info]: Connected by 27.115.122.232:41227
-----file-4-4-----save successful-----
[info]: wait for connection...
[info]: Connected by 27.115.122.232:41228
[Error]: list index out of range
[info]: wait for connection...
[info]: Connected by 27.115.122.232:40517
-----/home/hadoop/file-1-3-----delete successful-----
[info]: wait for connection...
[info]: Connected by 27.115.122.232:40517
[Error]: list index out of range
[info]: wait for connection...
[info]: Connected by 27.115.122.232:40520
-----/home/hadoop/file-1-4-----delete successful-----
[info]: wait for connection...
[info]: Connected by 27.115.122.232:40527
[Error]: list index out of range
[info]: wait for connection...
[info]: Connected by 27.115.122.232:40662
[Error]: list index out of range
[info]: wait for connection...
[info]: Connected by 27.115.122.232:40663
-----file-2-2-----fetch successful-----
[info]: wait for connection...
Connection to 121.37.138.6 closed by remote host.
Connection to 121.37.138.6 closed.

```

- 删除文件

Client:

```

MiniDFS >>> delete 3
DataNode 0 [Error]: timed out
DataServer-3 file-3-0 delete successful
DataServer-2 file-3-0 delete successful
DataServer-1 file-3-0 delete successful
DataNode 0 [Error]: timed out
DataServer-3 file-3-1 delete successful
DataServer-1 file-3-1 delete successful
DataServer-2 file-3-1 delete successful
DataNode 0 [Error]: timed out
DataServer-2 file-3-2 delete successful
DataServer-3 file-3-2 delete successful
DataServer-1 file-3-2 delete successful
DataNode 0 [Error]: timed out
DataServer-2 file-3-3 delete successful
DataServer-3 file-3-3 delete successful
DataServer-1 file-3-3 delete successful
DataNode 0 [Error]: timed out
DataServer-2 file-3-4 delete successful
DataServer-1 file-3-4 delete successful
DataServer-2 file-3-4 delete successful
DataServer-3 file-3-4 delete successful
Delete succeed! File_ID is 3

MiniDFS >>> ls
File total number 3
0           test.txt      10287372
2           test.txt      10287372
4           test.txt      10287372

```

- 下载文件---会先从存储此chunk的服务器下载，若断连则会从备份服务器读取

Client:

```

MiniDFS >>> fetch 0 q.txt
DataNode 0 [Error]: [Errno 61] Connection refused
DataServer-1 fetch file-0-0
DataServer-1 fetch file-0-1
DataServer-2 fetch file-0-2
DataServer-3 fetch file-0-3
DataNode 0 [Error]: [Errno 61] Connection refused
DataServer-1 fetch file-0-4
Finished download!

```

Data servers:

```

root@ecs-a310-5a76-0003:/home/hadoop (ssh) [root@ecs-a310-5a76-0003 ~]# 
[info]: wait for connection...
[info]: Connected by 27.115.122.232:57724
[Error]: list index out of range
[info]: wait for connection...
[info]: Connected by 27.115.122.232:57725
-----file-0-0-----fetch successful-----
[info]: wait for connection...
[info]: Connected by 27.115.122.232:57745
[Error]: list index out of range
[info]: wait for connection...
[info]: Connected by 27.115.122.232:57746
-----file-0-1-----fetch successful-----
[info]: wait for connection...
[info]: Connected by 27.115.122.232:57770
[Error]: list index out of range
[info]: wait for connection...
[info]: Connected by 27.115.122.232:57771
-----file-0-4-----fetch successful-----
[info]: wait for connection...
[info]: wait for connection...
[info]: wait for connection...
[info]: Connected by 27.115.122.232:57774
-----file-0-2-----fetch successful-----
[info]: wait for connection...
[info]: Connected by 27.115.122.232:57775
-----file-0-3-----fetch successful-----
[info]: wait for connection...
[info]: Connected by 27.115.122.232:57776
-----file-0-4-----delete successful-----
[info]: wait for connection...
[info]: Connected by 27.115.122.232:56308
-----/home/hadoop/file-1-3-----delete successful-----
[info]: wait for connection...
[info]: Connected by 27.115.122.232:56313
[Error]: list index out of range
[info]: wait for connection...
[info]: Connected by 27.115.122.232:56316
-----/home/hadoop/file-1-4-----delete successful-----
[info]: wait for connection...
[info]: Connected by 27.115.122.232:56322
[Error]: list index out of range
[info]: wait for connection...
^CTraceback (most recent call last):
  File "DataServer.py", line 89, in <module>
    run()
  File "DataServer.py", line 17, in run
    conn, addr = sock.accept()
  File "/usr/lib/python2.7/socket.py", line 206, in accept
    sock, addr = self._sock.accept()
KeyboardInterrupt
root@ecs-a310:/home/hadoop# 

```

本地文件:

```

→ MiniDFS git:(master) ✘ ll
total 60304
drwxr-xr-x@ 3 jiang staff 96B 6 20 09:46 DataServer
-rw xr-xr-x@ 1 jiang staff 0B 6 20 15:39 LICENSE.txt
-rw r--r--@ 1 jiang staff 32B 6 20 09:49 README.md
drwxr-xr-x@ 8 jiang staff 256B 6 17 08:54 core
drwxr-xr-x@ 3 jiang staff 96B 6 29 10:38 dfs
-rwxrwxrwx@ 1 jiang staff 2.5K 6 29 09:35 main.py
-rw r--r--@ 1 jiang staff 9.8M 6 29 11:12 q.txt
-rwxrwxrwx@ 1 jiang staff 9.8M 6 20 14:34 test-2.txt
-rwxrwxrwx@ 1 jiang staff 9.8M 6 20 14:30 test.txt

```