**Implementation Notes**

**Emulated HDFS Web Application**

Group 7: Yaohao Wang, Xinyuan Zhang, Chenyu Li

### 

### **1. Project overview**

#### **1.1 Required Package**

aiohttp == 3.8.4

json == 0.9.6

requests == 2.28.1

lxml == 4.9.1

glob == 0.7

quart == 0.18.4

urllib == 5.4.0

#### **1.2 Project Files**

DSCI-551-FinalProject-main

-bin

edfs

start-dfs

-conf

Config.py

configuration.xml # From here, you can change the replication ,EC2 IP and saving path

-fsimage

fsimage\_test.xml # Here is the file system, you can get the detailed information about files and dictionaries

-logs

namenode.log

datanode.log

-src

-fstree

fstree.py

-server

datanode.py

namenode.py

-static

background1.jpg

background2.jpg

-templates

index\_new.html

edfs.log

edfs.py

start\_dfs.py

web\_ui.py

readme.md

requirement.txt

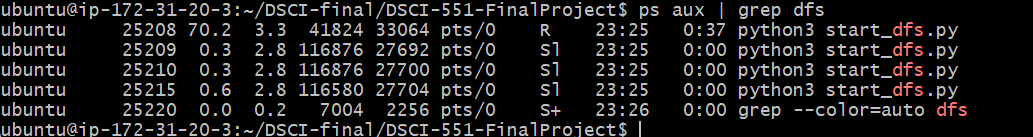
### **2. Start server on EC2 and DFS client**

#### **2.1. Start server**

1. git clone https://github.com/francescazhang92/DSCI-551-FinalProject
2. cd DSCI-551-FinalProject
3. lnohup python3 start\_dfs.py > start\_dfs.log &

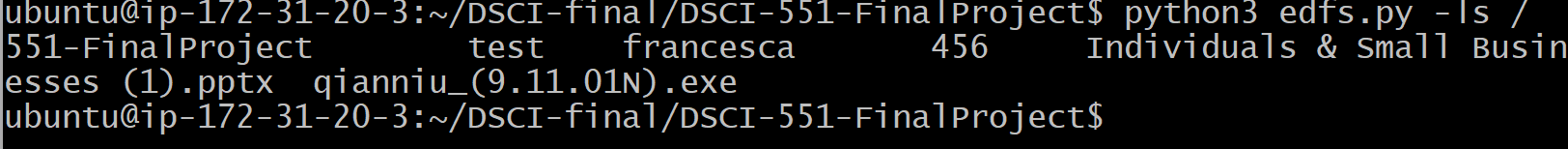


1. ps aux | grep start\_dfs.py

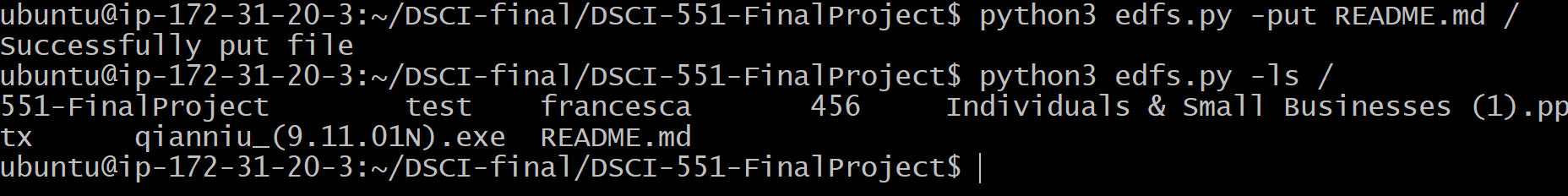
We can see there are 4 processes, including start\_dfs.py itself and 1 namenode, 2 datanode

#### **2.2. Command Shell**

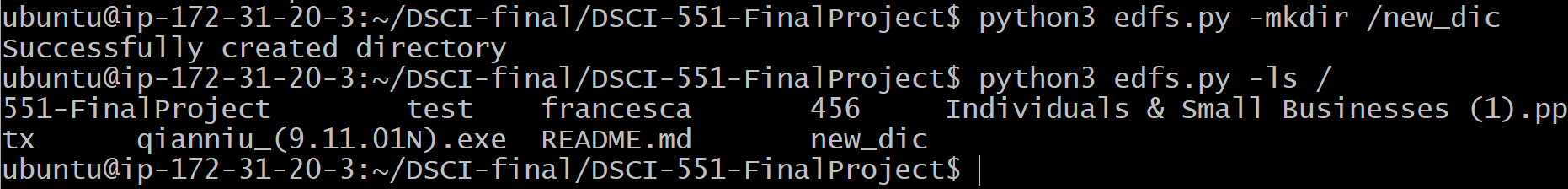
1. - ls：show directory



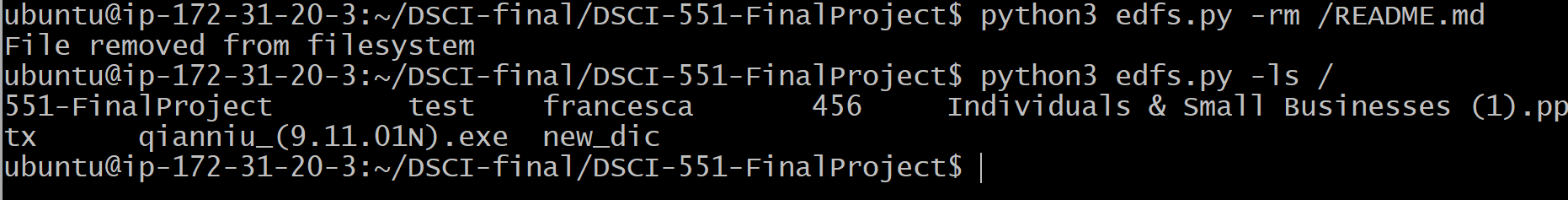
1. - put：create new files



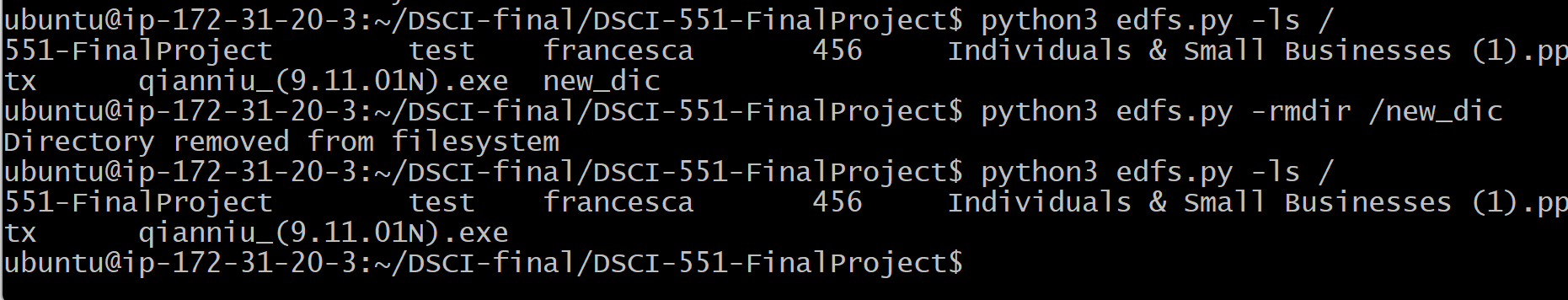
1. - mkdir：create a new directory



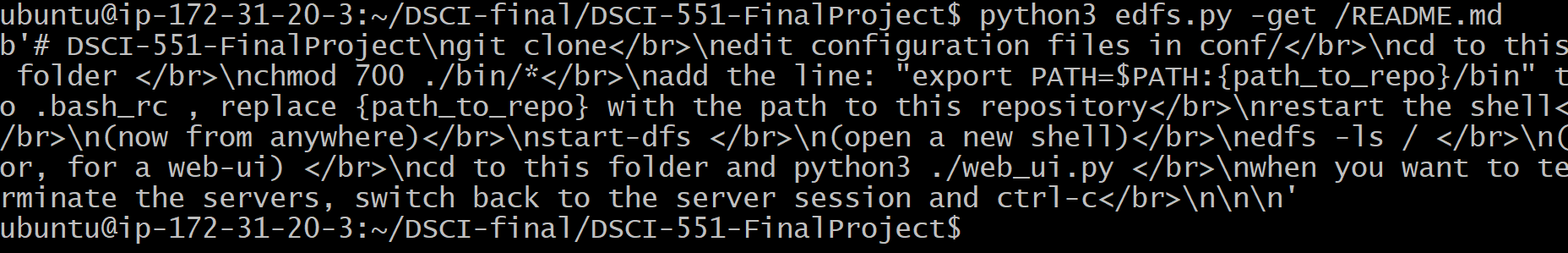
1. - rm：remove a file



1. - rmdir：remove a directory

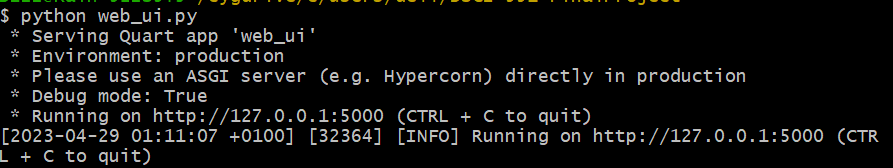


1. - cat：reads data from the file and gives their content as output

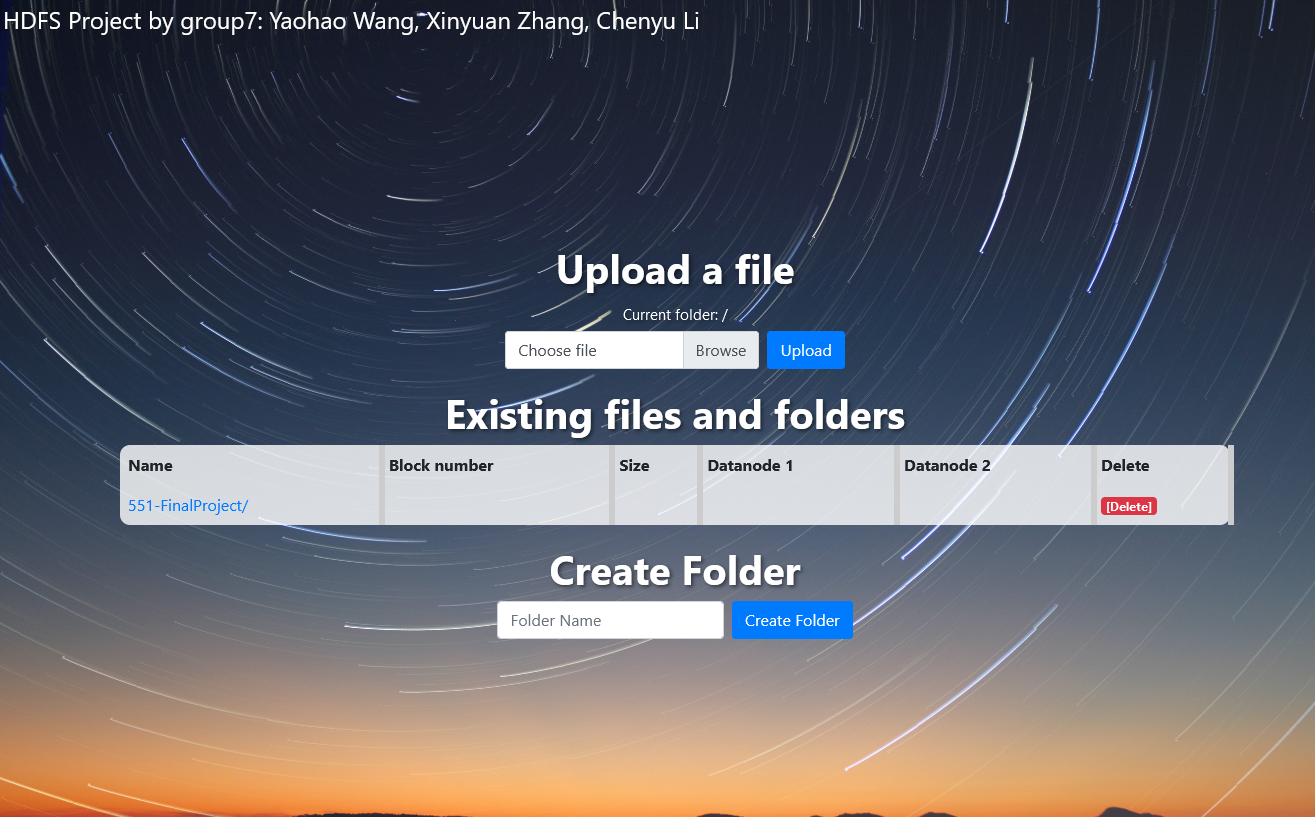


#### **2.3. Web ui implementation**

1. python web\_ui.py

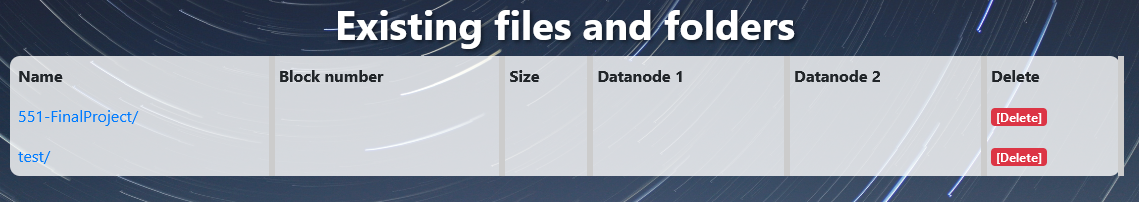


1. Enter into page 127.0.0.1:5000

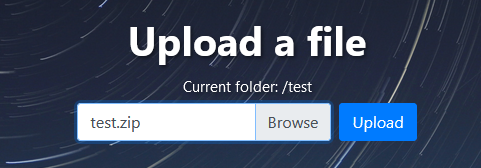


1. Using the **“Create Folder”** button to create a new folder in the current directory. Click on the name of the folder to enter it.

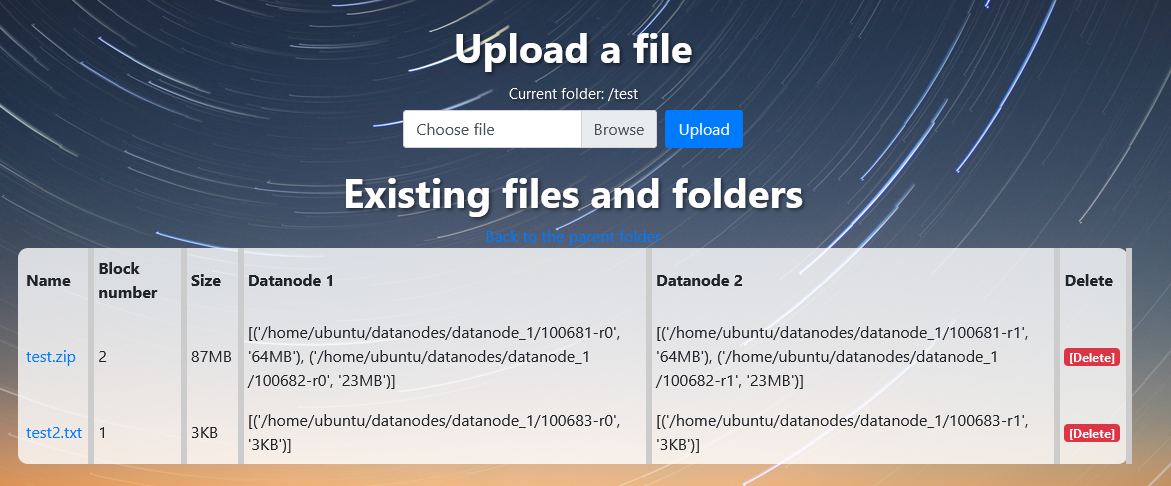




1. Using the **“Browse file”** and **“Upload”** button to upload a file. After choosing the file, a script updates the label of the file input with the selected file's name.



1. The **“Existing files and folders”** section displays a table of files and folders. For file, the section will display file name, block number, size, datanodes that stores the files.



1. Using the **“Delete”** button in red to delete files and empty folders.

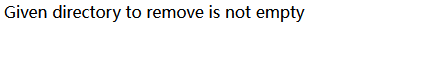


1. Using the **“Back to the parent folder”** link is displayed if the current folder is not the root folder.



1. Error message:

* If you delete a folder that is not empty, an error message will be displayed.



* If the filename of the file the user wants to upload is invalid (e.g. no files selected) , an error message will be displayed.

