



Instruksi

Bentuk kelompok dengan anggota minimal 4 orang dan lakukan setup *cluster* Hadoop di 4 komputer yang berbeda. Apabila di kelas Anda sudah ada kelompok yang dibentuk di mata kuliah PBL, maka Anda dapat menggunakan kelompok yang sama dengan kelompok PBL tersebut, tentunya dengan seizin dan sepengetahuan pengampu MK Big Data di kelas Anda.

Pada 4 komputer milik masing-masing anggota kelompok Anda tersebut pilih 1 diantaranya untuk dijadikan sebagai namenode, dan 3 yang lainnya sebagai datanode.

- Untuk menjadi namenode pilihlah komputer yang setidaknya memiliki 8 GB RAM dan 3 inti (*core*) CPU.
- Sedangkan untuk menjadi datanode, setidaknya suatu komputer harus memiliki 4 GB RAM dan 2 inti (*core*) CPU.

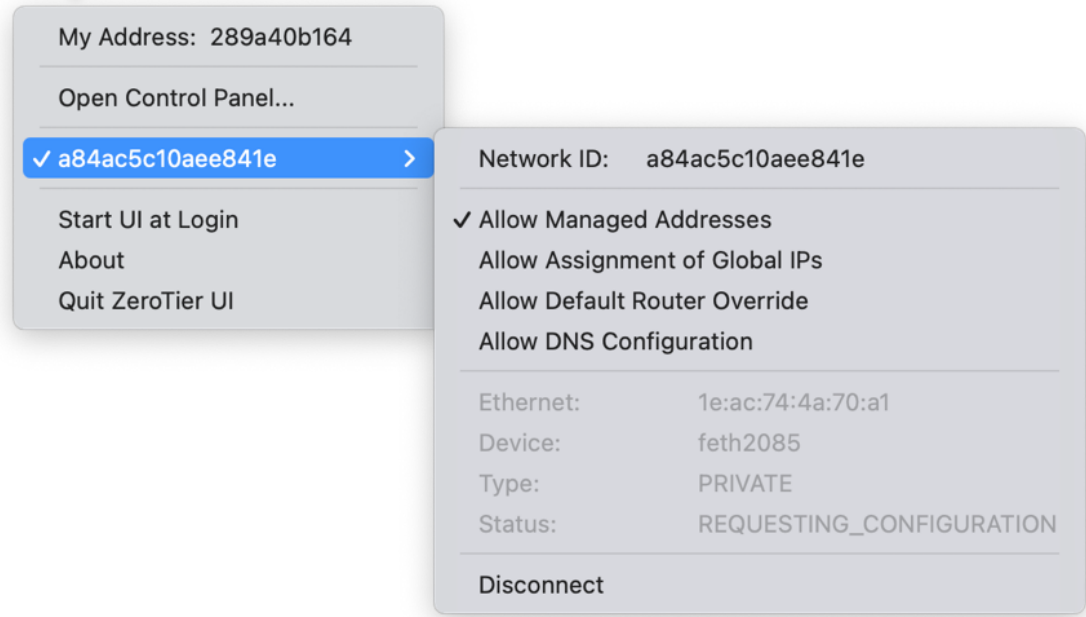
Kerjakan instruksi di bagian “**Soal**”, dan demokan ke dosen pengampu jika sudah selesai.

Catatan:

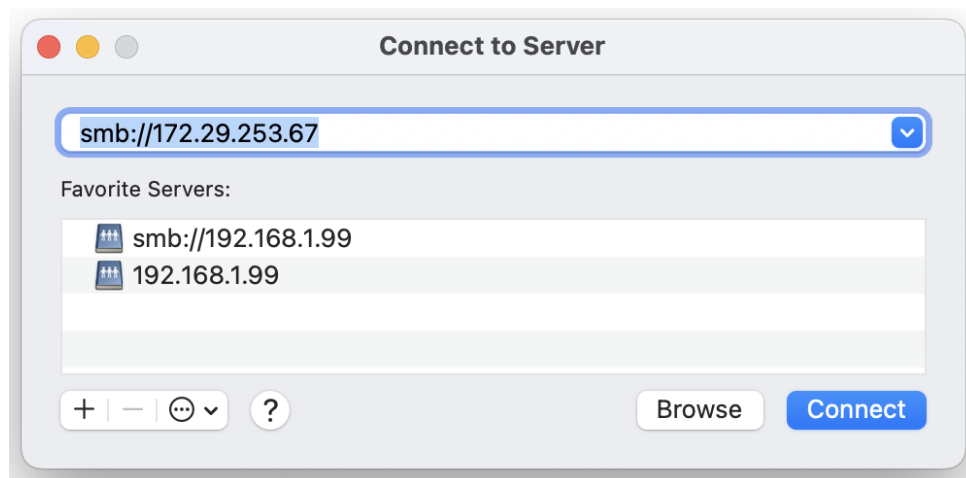
- Langkah-langkah pada soal kuis-1 ini dibuat berdasarkan spesifikasi komputer dengan sistem operasi Ubuntu Server yang ada di ruang lab Teknologi Data.
- Apabila terdapat permasalahan maupun error pada saat Anda mencoba langkah-langkah ini, maka itu adalah hal yang wajar mengingat perbedaan spesifikasi komputer yang Anda miliki.
- Apabila Anda menemui permasalahan/error tersebut, **gunakanlah segala kreativitas Anda untuk mencoba mengatasinya sendiri.**

Bagian-1: Mengunduh VM Hadoop Polinema

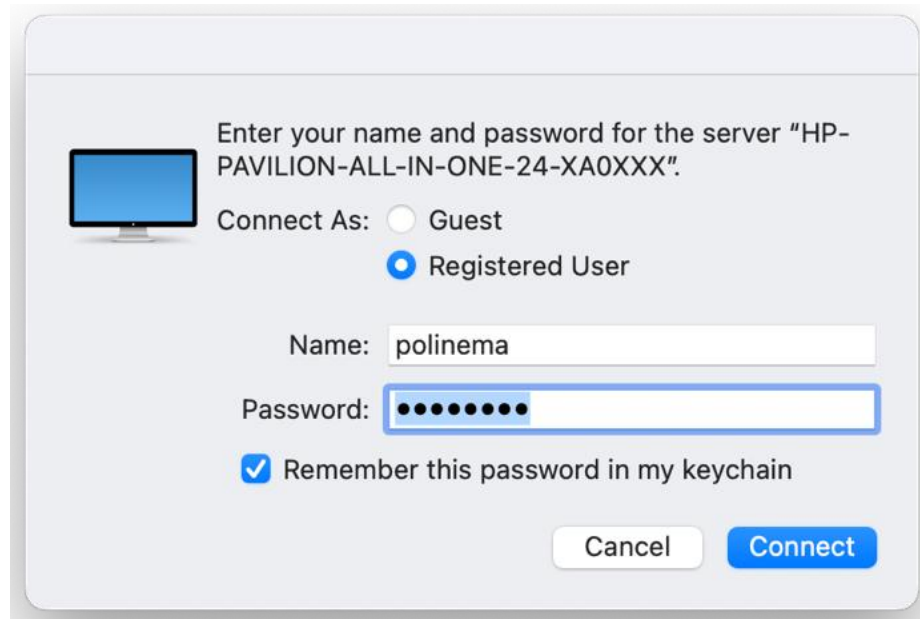
1. Terhubunglah ke jaringan VPN Zerotier **polinema_hadoop_cluster** dengan menggunakan Network ID berikut: **a84ac5c10aee841e**.



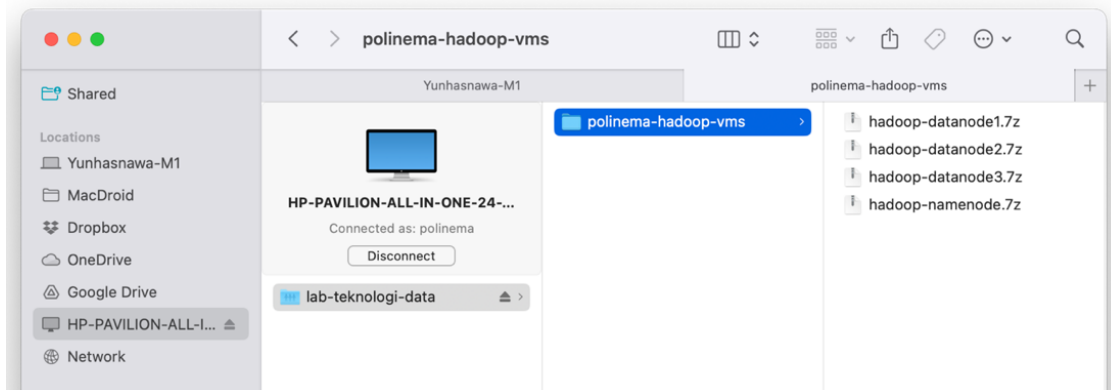
2. Dengan menggunakan aplikasi file *explorer* di komputer Anda masing-masing, *browse* ke server Linux ubuntu lab Teknologi Data di alamat IP Berikut: **172.29.253.67**. Jika Anda menggunakan sistem operasi Microsoft Windows 10 ataupun 11, ikuti langkah-langkah seperti pada laman berikut (mulai dari langkah-3): <https://www.simplehelp.net/2021/01/24/how-access-the-files-on-your-mac-from-windows/>.



3. Gunakan kredensial berikut untuk bisa terhubung ke server tersebut:
- Username : **polinema**
 - Password : **polinema**



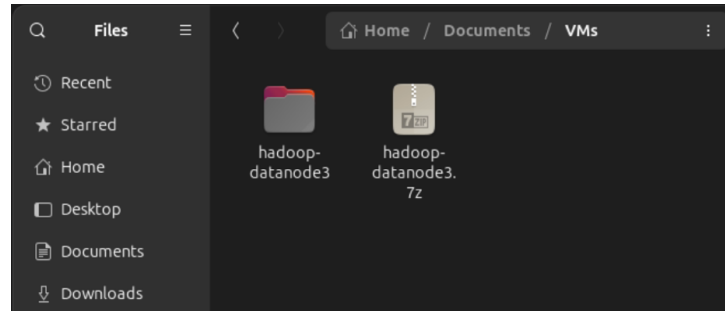
4. Jika Anda berhasil terhubung, masuklah ke folder **lab-teknologi-data** → **polinema-hadoop-vms**. Pada folder tersebut terdapat 4 buah file yang terkompresi dengan format 7zip. Masing-masing file berukuran cukup besar yaitu lebih dari 4GB.



5. Silahkan salin/unduh file **yang sesuai** ke masing-masing komputer milik anggota kelompok Anda yang akan dijadikan sebagai **namenode**, **datanode1**, **datanode2**, dan **datanode3**. Ekstraksi file 7zip tersebut ke sembarang direktori yang mudah Anda ingat.

Pada contoh berikut ini, file **hadoop-datanode-3.zip** diunduh dan diekstraksi di PC/komputer yang akan dijadikan sebagai datanode nomor-3.

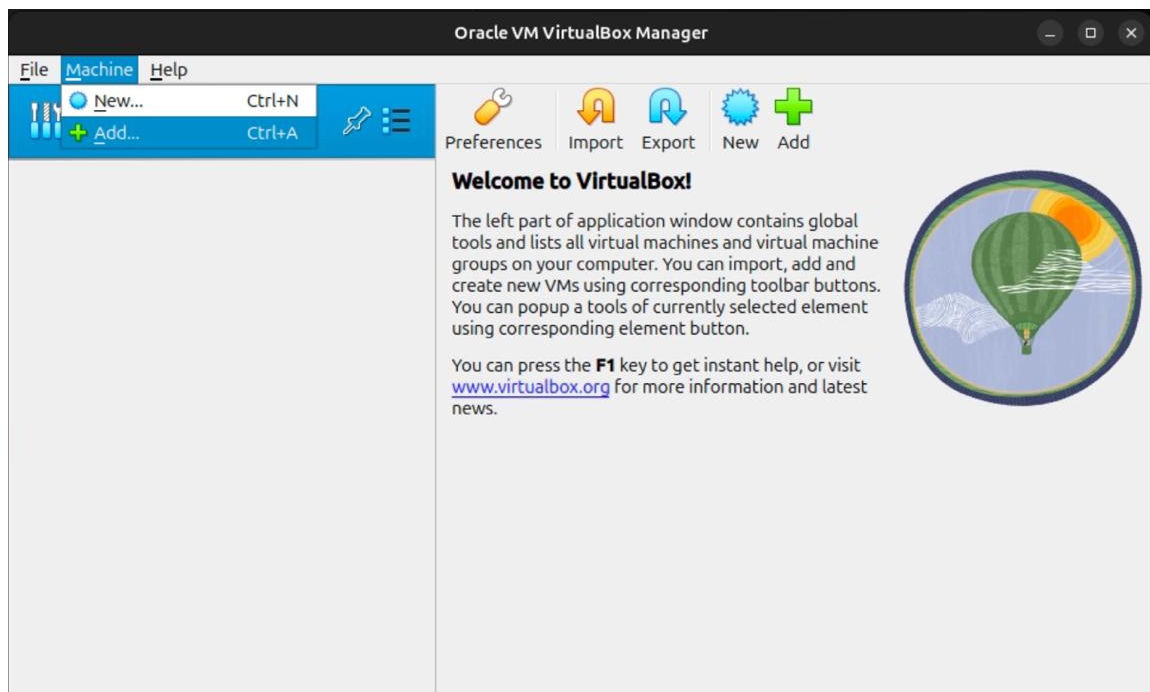
Catatan: Ketika sudah selesai diekstraksi, file 7zip-nya boleh dihapus agar tidak memakan penyimpanan, karena cukup besar.



6. Sampai di sini, matikan koneksi VPN ZeroTier Anda. Koneksi VPN tidak diperlukan lagi karena file VM sudah selesai Anda unduh.

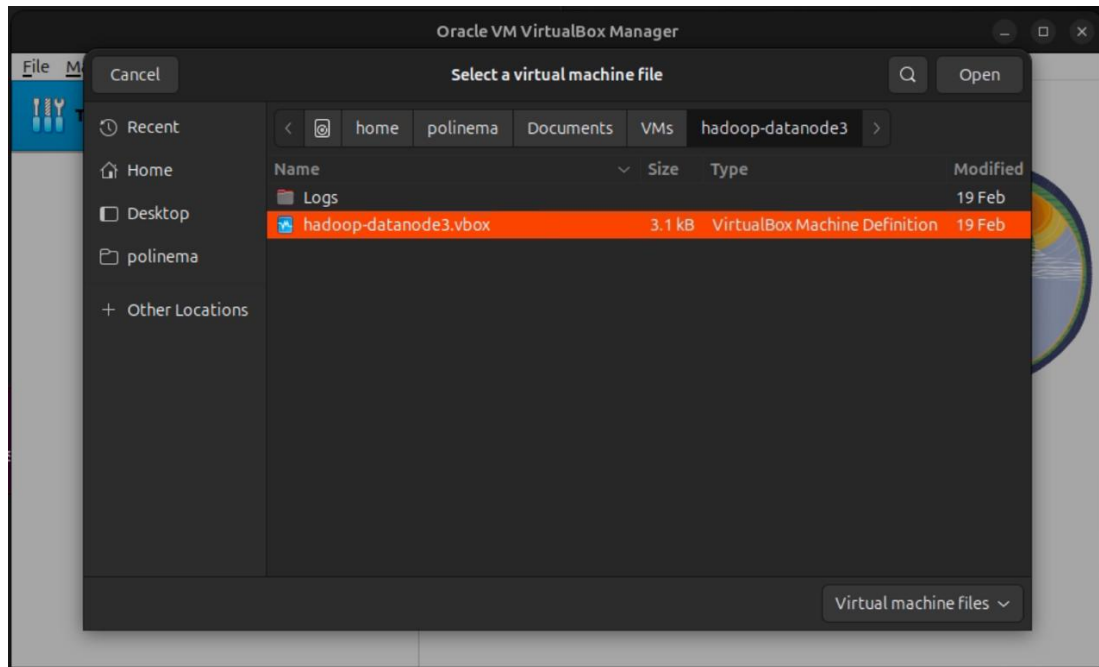
Bagian-2: Menjalankan VM Hadoop

1. Pasang dan jalankan VirtualBox di **semua** komputer yang akan dijadikan cluster Hadoop. Anda dapat mengunduh aplikasi VirtualBox di website resminya pada URL berikut ini: <https://www.virtualbox.org/wiki/Downloads>
2. Pada VirtualBox di masing-masing komputer, impor VM yang sudah diekstraksi pada bagian sebelumnya. Lakukan dengan cara klik menu **Machine** → **Add...**

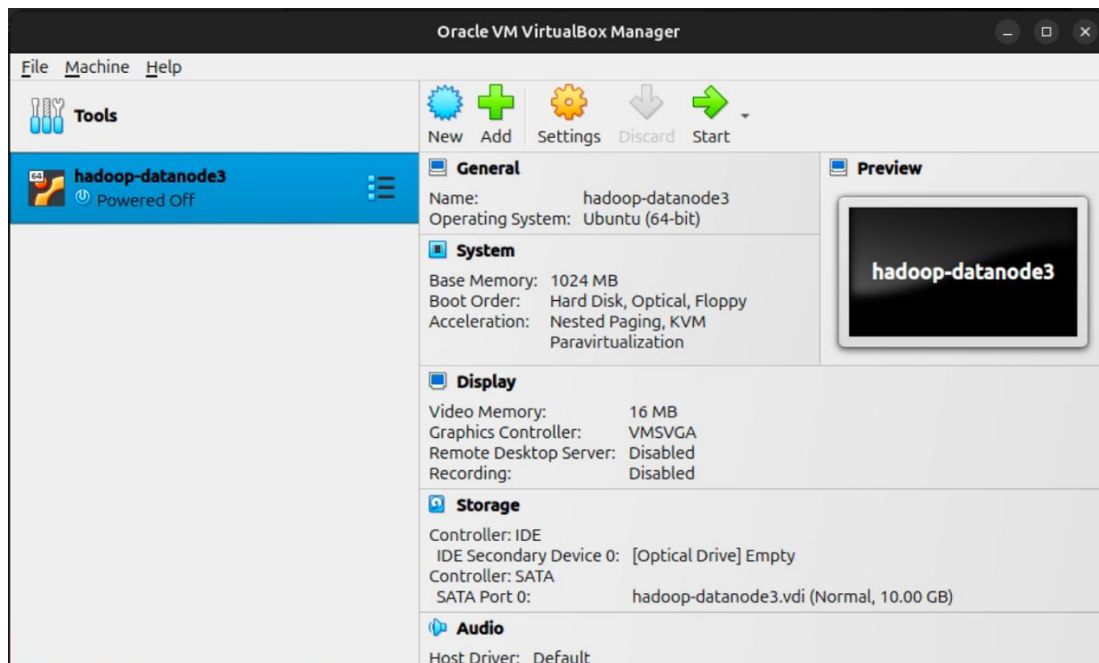




3. Arahkan ke folder yang telah diekstraksi di langkah sebelumnya tadi. Pilih file dengan ekstensi ***.vbox**.



4. Apabila benar, maka VM yang baru ditambahkan tersebut akan muncul di jendela VirtualBox. Jika sudah muncul, klik 2x VM tersebut untuk membuatnya *booting*.





5. Apabila tidak ada masalah yang terjadi, maka sistem operasi Ubuntu Server yang ada pada VM tersebut akan dimulai dan tampilannya akan menjadi seperti berikut.



6. **Lakukan langkah yang sama untuk setiap komputer** yang akan dijadikan **namenode**, **datanode1**, **datanode2**, dan **datanode3**. Pastikan setiap komputer yang akan dijadikan node memiliki spesifikasi setidaknya minimal:
- RAM 4 GB atau lebih. Khusus Namenode sebaiknya 8 GB atau lebih.
 - Dua inti (core) prosesor. Khusus Namenode sebaiknya 4 core atau lebih.
 - Penyimpanan 20 GB atau lebih.**

Bagian-3: Konfigurasi *Cluster*

- Pastikan keempat komputer fisik teman-teman dalam satu kelompok **berada dalam satu jaringan yang sama**. Lalu **nyalakan semua VM** yang ada di semua komputer fisik tersebut.
- Untuk dapat menjadi sebuah cluster, **semua VM yang menjadi node** harus mengenali **alamat IP lokal** masing-masing. Untuk memastikan hal tersebut, terlebih dahulu **cek semua IP address dari masing-masing VM dan catat!** Untuk melakukan hal tersebut, login-lah dahulu ke masing-masing VM dengan menggunakan kredensial berikut:
 - Username : **hadoopuser**
 - Password : **hadoop**



3. Setelah berhasil login, pada masing-masing VM ketikkan perintah: **ip addr**. Pada contoh ini, IP lokal VM-nya adalah yang nomor 2 (enp0s3), yaitu: **192.168.2.150**.

```
hadoop-datanode3 [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
hadoopuser@hadoop-datanode3:~$ ip addr
1: lo: <LOOPBACK,UP,LOWER_UP> mtu 65536 qdisc noqueue state UNKNOWN group default qlen 1000
    link/loopback 00:00:00:00:00:00 brd 00:00:00:00:00:00
    inet 127.0.0.1/8 scope host lo
        valid_lft forever preferred_lft forever
    inet6 ::1/128 scope host noprefixroute
        valid_lft forever preferred_lft forever
2: enp0s3: <BROADCAST,MULTICAST,UP,LOWER_UP> mtu 1500 qdisc fq_codel state UP group default qlen 1000
    link/ether 08:00:27:a6:58:94 brd ff:ff:ff:ff:ff:ff
    inet 192.168.2.150/25 metric 100 brd 192.168.2.255 scope global dynamic enp0s3
        valid_lft 449sec preferred_lft 449sec
    inet6 fe80::a00:27ff:fea6:5894/64 scope link dadfailed tentative
        valid_lft forever preferred_lft forever
hadoopuser@hadoop-datanode3:~$
```

4. Catat **IP address lokal semua VM**. Lalu di semua VM, edit file **/etc/hosts** dengan menggunakan perintah **nano**. Berikut perintahnya:

- **sudo nano /etc/hosts**

Apabila benar, maka tampilannya akan menjadi seperti berikut ini.

```
hadoop-datanode3 [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
GNU nano 7.2 /etc/hosts
127.0.0.1 localhost

# The following lines are desirable for IPv6 capable hosts
::1 ip6-localhost ip6-loopback
fe00::0 ip6-localnet
ff00::0 ip6-mcastprefix
ff02::1 ip6-allnodes
ff02::2 ip6-allrouters

# Hadoop cluster nodes
192.168.2.204 hadoop-namenode
192.168.2.216 hadoop-datanode1
192.168.2.217 hadoop-datanode2
192.168.2.218 hadoop-datanode3
```

5. Pada file tersebut, ubahlah daftar alamat IP yang ada di bawah komentar **# Hadoop cluster nodes**. Ubah hanya alamat IP-nya saja, jangan ubah yang lainnya. Sesuaikan alamat IP lokal dari masing-masing node, **samakan dengan daftar alamat IP yang telah Anda catat** di langkah sebelumnya! Pastikan isi file tersebut **sama persis** di keempat VM Anda.
6. Jika semua file **/etc/hosts** di semua VM sudah sama dan sudah sesuai daftar alamat IP-nya, **di VM namenode**, ketikkan perintah berikut:

- **start-dfs.sh**
- **start-yarn.sh**



- **jps**

Bila semua langkah yang dilakukan benar, maka tampilan di VM **namenode** akan menjadi seperti berikut:

```
hadoop-namenode (Namenode-OK) [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
hadoopuser@hadoop-namenode:~$ start-dfs.sh
Starting namenodes on [hadoop-namenode]
Starting datanodes
Starting secondary namenodes [hadoop-namenode]
hadoopuser@hadoop-namenode:~$ start-yarn.sh
Starting resourcemanager
Starting nodemanagers
hadoopuser@hadoop-namenode:~$ jps
44194 NameNode
44436 SecondaryNameNode
44615 ResourceManager
44905 Jps
hadoopuser@hadoop-namenode:~$
```

7. Pada semua **VM yang menjadi datanode**, jalankan juga perintah **jps**. Pastikan hasilnya seperti berikut ini:

```
hadoop-datanode3
File Machine View Input Devices Hel
hadoopuser@hadoop-datanode3:~$ jps
3124 Jps
2277 DataNode
2425 NodeManager
hadoopuser@hadoop-datanode3:~$
```

Soal

1. Terhubunglah ke cluster kelompok Anda melalui terminal dari komputer Anda masing-masing.
2. Buatlah sebuah direktori di cluster Hadoop kelompok Anda masing-masing, dan beri nama sesuai nomor absen dan nama masing-masing dengan format **NoAbs_NamaLengkap**, sesuai contoh berikut: 13_DanaAnagataNusantara.
3. Unggahlah file TXT biasa yang berisi nama dan nomor absen Anda serta informasi lain yang tidak privat ke folder Anda masing-masing.

~ Selamat Bekerja ~