Hands on: Cloud Computing

ดร. ดำรง ไม้เรียง
ห้องปฏิบัติการเทคโนโลยีชีวภาพทางการแพทย์
ศูนย์พันธุวิศวกรรมและเทคโนโลยีชีวภาพแห่งชาติ (BIOTEC)

After this session...

- Cloud account setup
- Virtual/Cloud server setup
- Accessing and controlling cloud server
- Upload/Download files ←→ cloud server
- Clean up or Pack up

Step 0: Quick walkthrough

- I will show you what we will do in under 10 minutes.
- Please just sit and relax
- We will later repeat this together step-by-step

What do I need to know...

- You will be able to setup and run the pipeline.
  - User with no or little computer skills (and do not plan to develop one)
  - Few exome sequencing data to be analyzed once in a while (i.e. not routine analysis)
  - NOT sensitive data (i.e. data are not from an identifiable patient)

What do I need to know...

- We will explain these slides in details and try to make sure everyone understands.

What do I need to know...

- You want to customize the cloud server and/or the pipeline.
  - User with some computer skills
  - Routine analysis of exome sequencing data
  - Sensitive data or require more security
  - Special exome (non-human exome)

What do I need to know...

- We will briefly explain these slides or provide some hints/links so you may find more information if you want to learn more

STEP 1: Cloud account setup

- Choose your cloud service provider
  - Amazon web services
  - Google cloud platform
  - Microsoft Azure
  - IBM cloud

STEP 1: Cloud account setup

- Our Cloud Partner

- Easy-to-use interface
- Simplified cloud images collection
- Competitive pricing

STEP 1: Cloud account setup

- Let's setup an account for DigitalOcean
- You will need:
  - Email address
  - Credit card, Debit card or PayPal

STEP 1: Cloud account setup

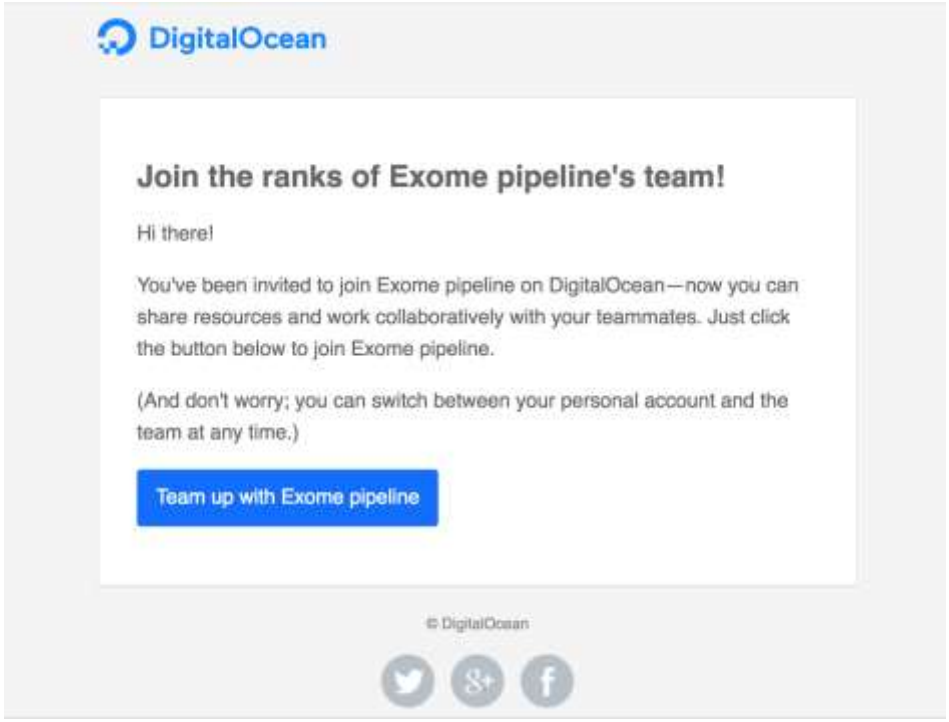
- Before we start
- <u>Do not</u> setup your own account for DigitalOcean yet if you:
  - Do not have credit card, debit card or PayPal account
  - You want to use “free credits” during trial period, but you do not plan to use it now.
- We are providing a temporary access to our team account during the workshop.

STEP 1: Cloud account setup

- Setup with $100 credits for 60 days
- https://try.digitalocean.com/performance/?utm=exome
- https://goo.gl/bUjtzm

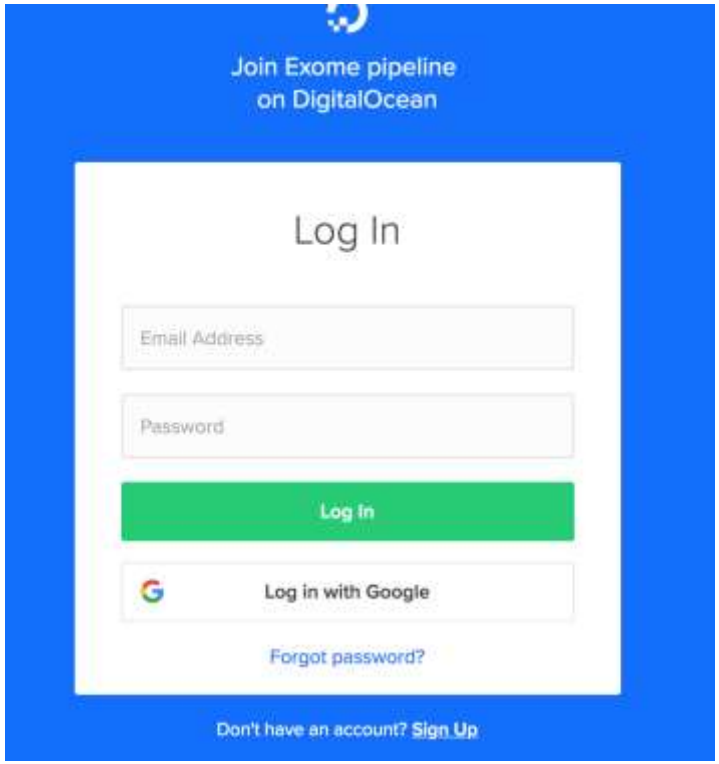
STEP 1: Cloud account setup

- For our temporary team account, please click a link in your email.



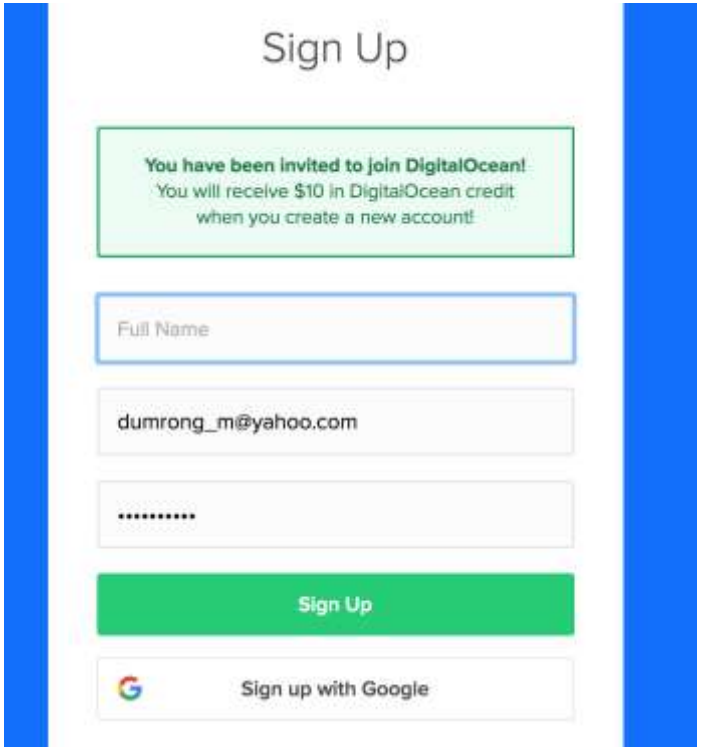
STEP 1: Cloud account setup

- Click “Sign up” or log in if your already have an account



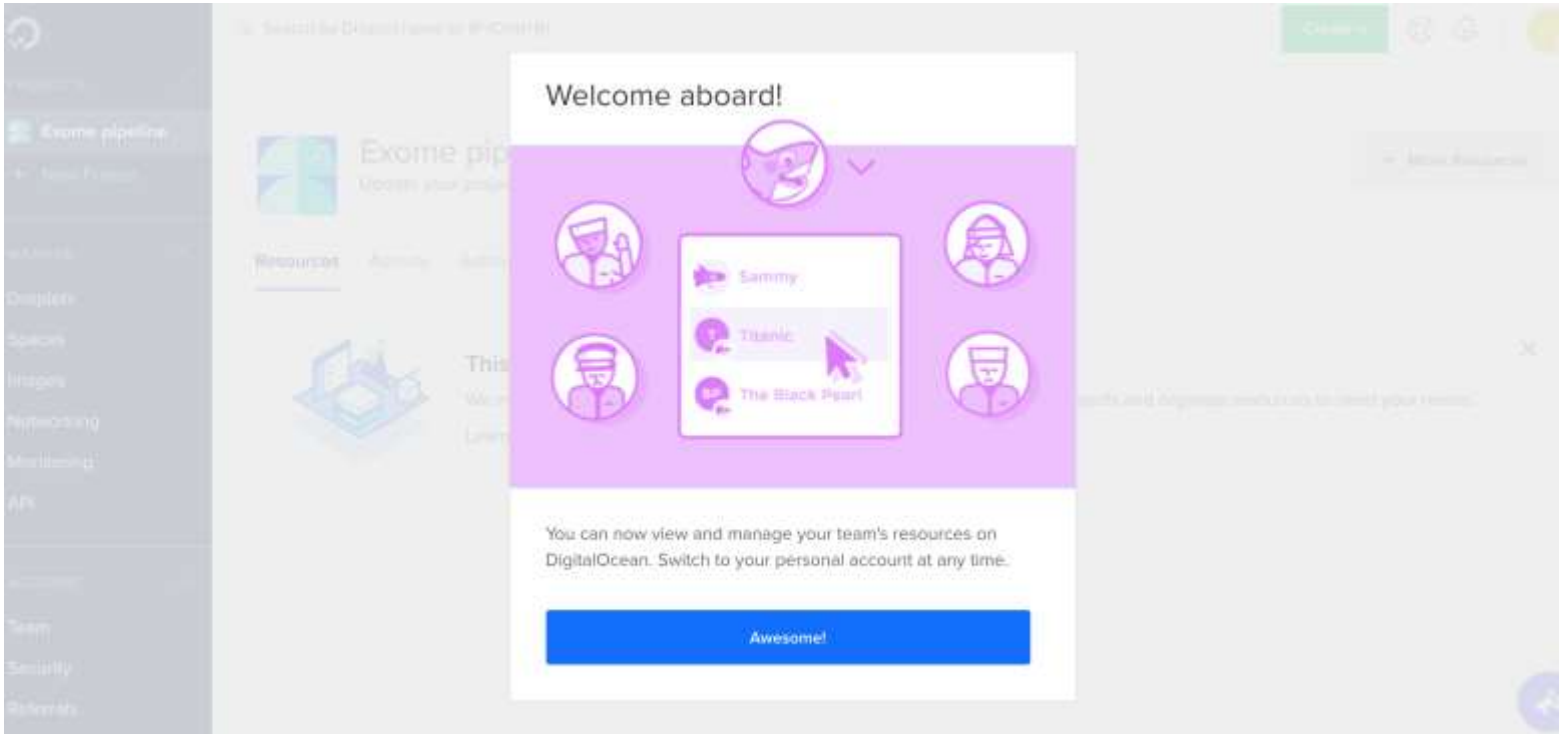
STEP 1: Cloud account setup

- Provide your information and a new password for signing up



STEP 1: Cloud account setup

- You should reach this page



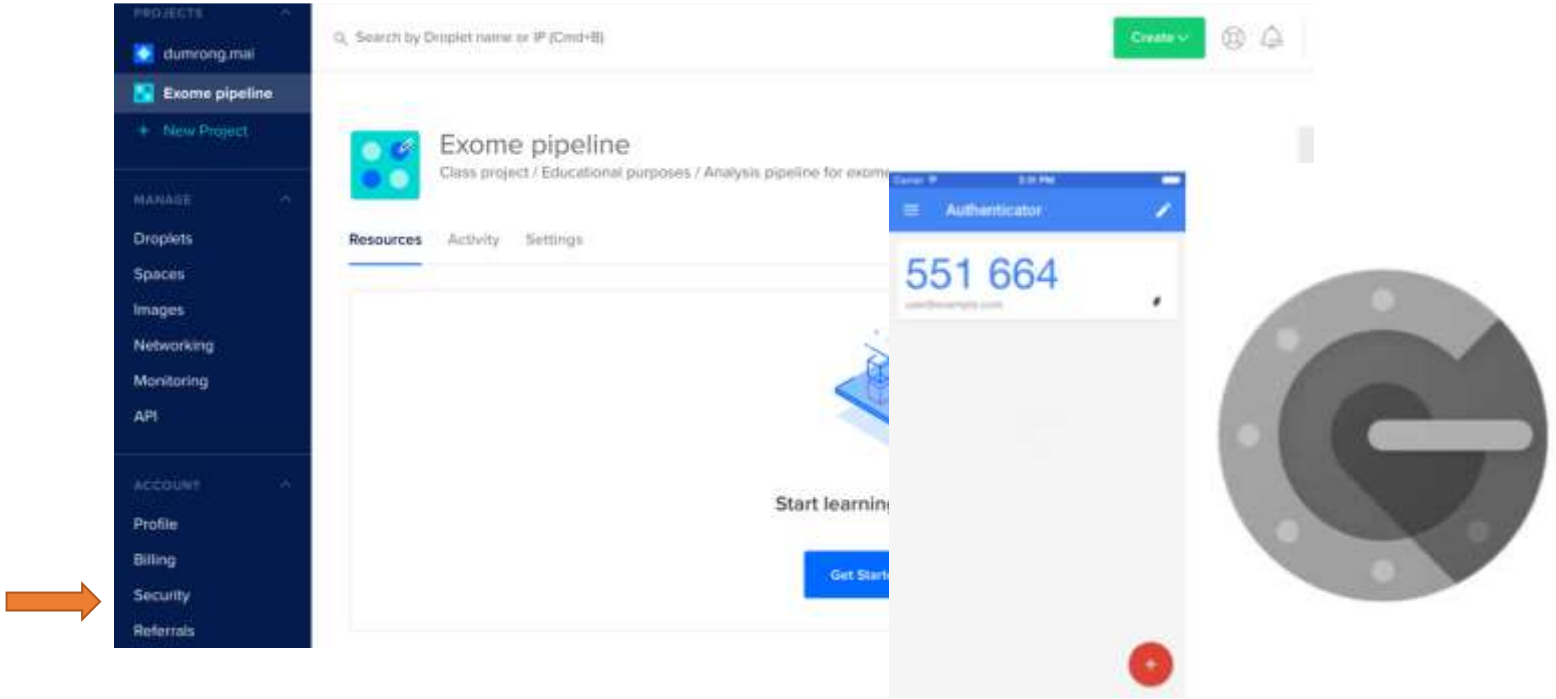
STEP 1: Cloud account setup

Can you access 

?

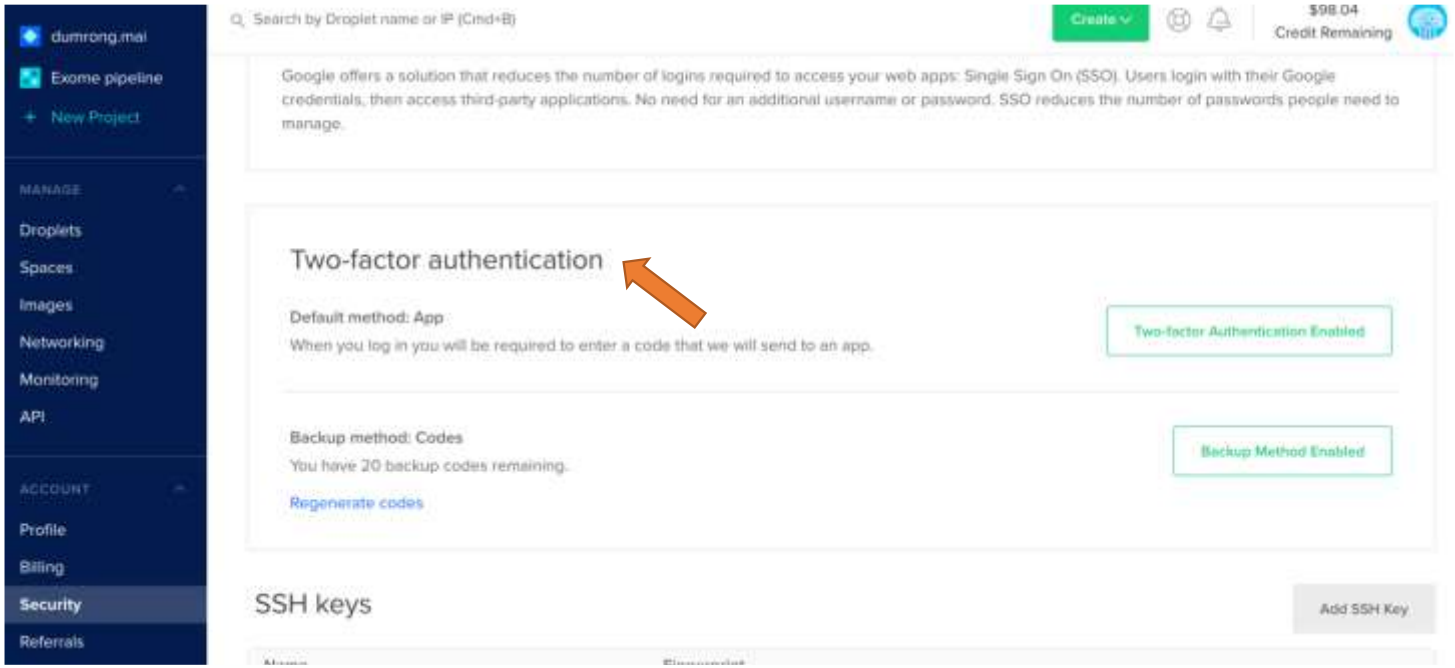
STEP 1: Cloud account setup

- Want a more secured account? (Your own account)
- Setup “two-factor” authenticator with Google app



STEP 1: Cloud account setup

- Want a more secured account? (Your own account)
- Setup “two-factor” authenticator with Google app



STEP 2: Get essential tools

- If you are using Mac or Linux, you can skip this step.
- If you are using Windows, you will need to install these programs:
  - PuTTY 
  - WinSCP 

STEP 2: Get essential tools

- PuTTY is a “Terminal emulator”
- PuTTY connects you to a (cloud) server.
- You can remotely control the server using command lines through PuTTY

STEP 2: Get essential tools

- Download PuTTY from https://www.chiark.greenend.org.uk/~sgtatham/putty/latest.html

Package files

You probably want one of these. They include all the PuTTY utilities.

(Not sure whether you want the 32-bit or the 64-bit version? Read the FAQ entry.)

MSI ('Windows Installer')

32-bit:	putty-0.70-installer.msi	(or by FTP)	(signature)
64-bit:	putty-64bit-0.70-installer.msi	(or by FTP)	(signature)

Unix source archive

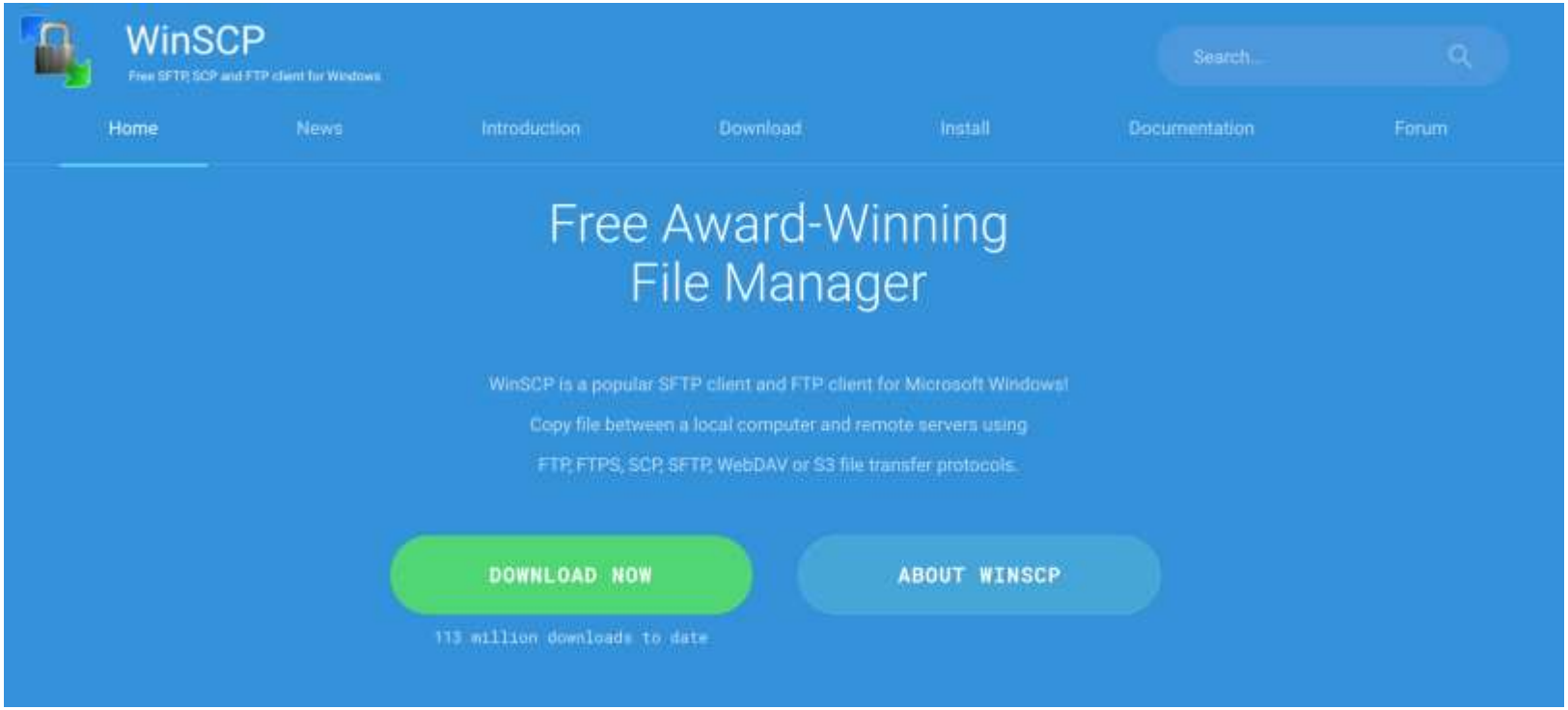
.tar.gz:	putty-0.70.tar.gz	(or by FTP)	(signature)

STEP 2: Get essential tools

- WinSCP = **Win**dows **S**ecure **C**o**p**y
- WinSCP connects you to a (cloud) server.
- You can upload files from your computer to the server via WinSCP
- You can download files from the server to your computer via WinSCP
- WinSCP help you work with files on the server like you are using Windows/Folder explorer

STEP 2: Get essential tools

- Download WinSCP from https://winscp.net/eng/index.php

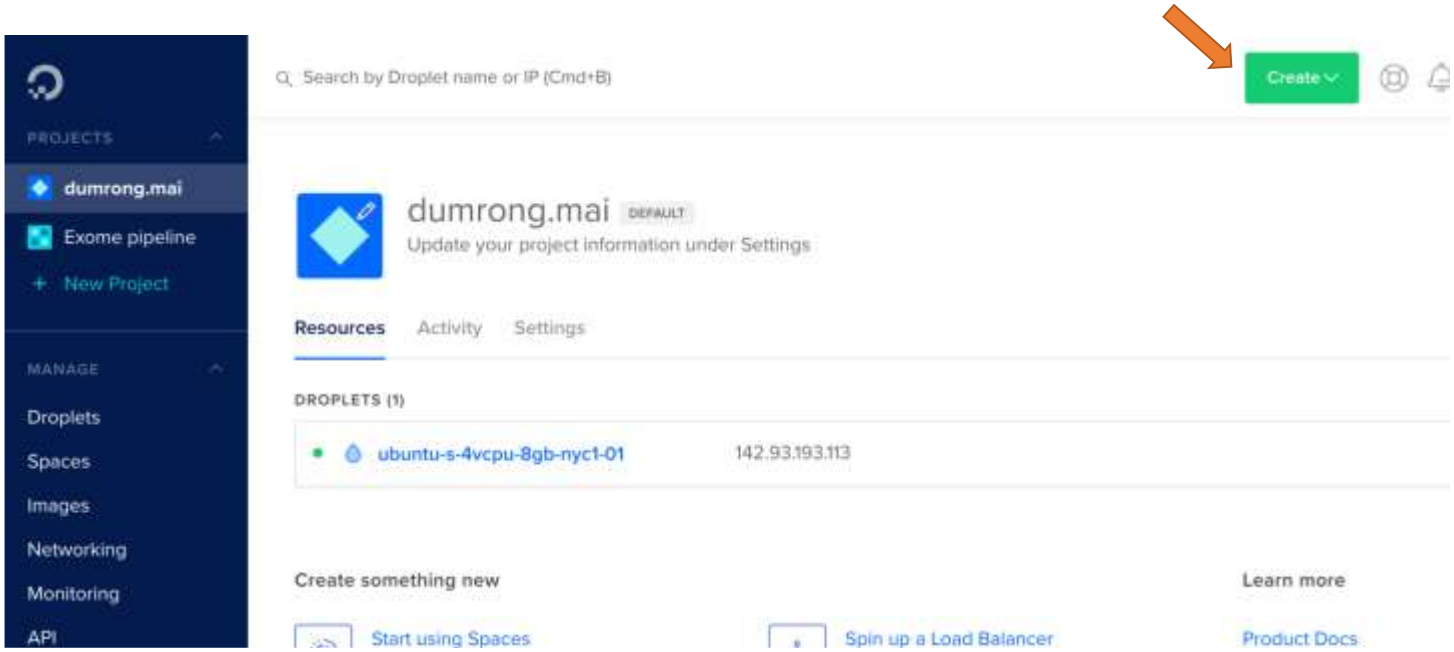


STEP 2: Get essential tools

Can you install  and  ?

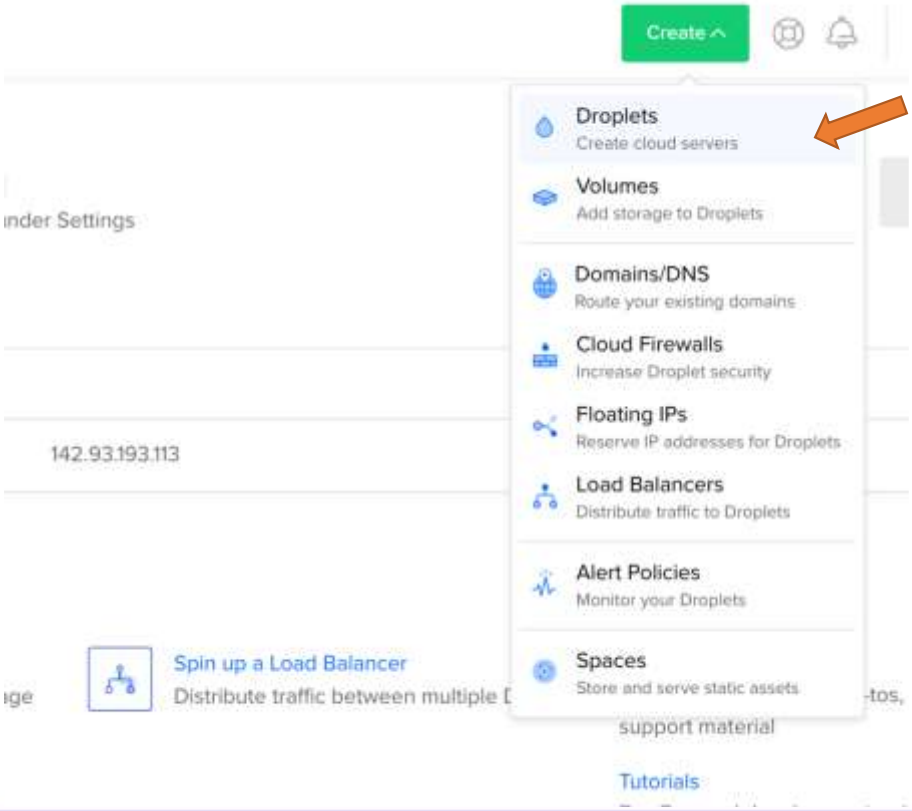
STEP 3: Setup your first droplet/virtual server

- Log in to DigitalOcean
- Click Create



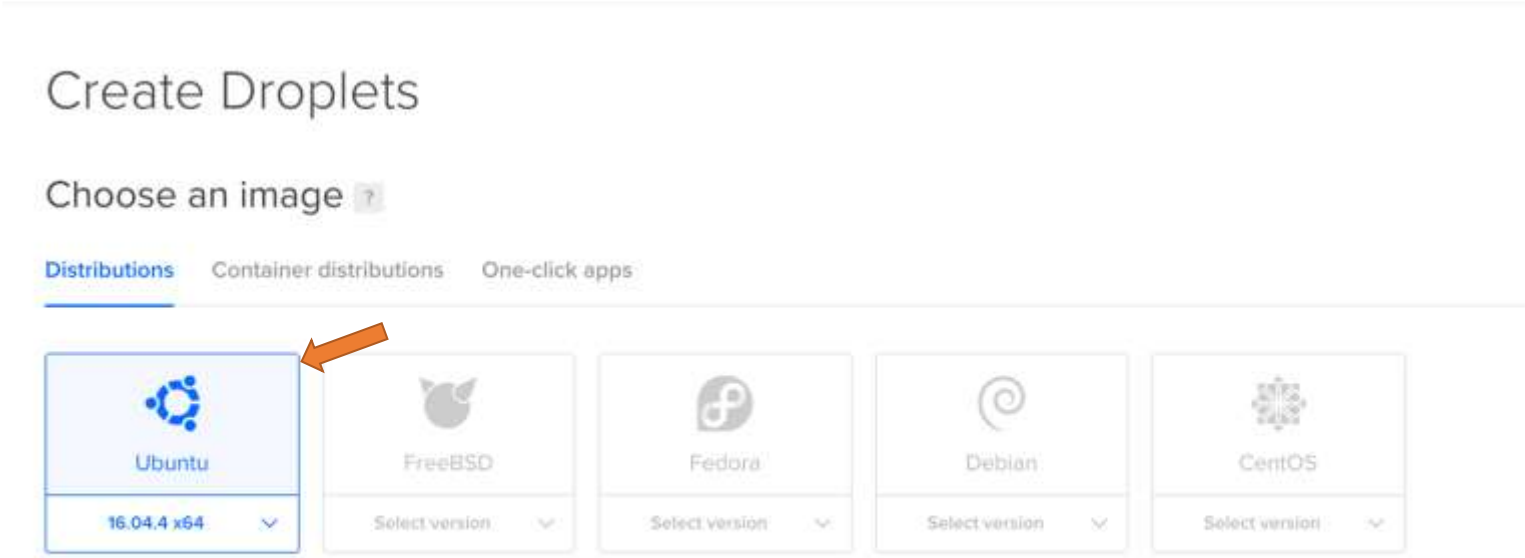
STEP 3: Setup your first droplet/virtual server

- Select “Droplets”



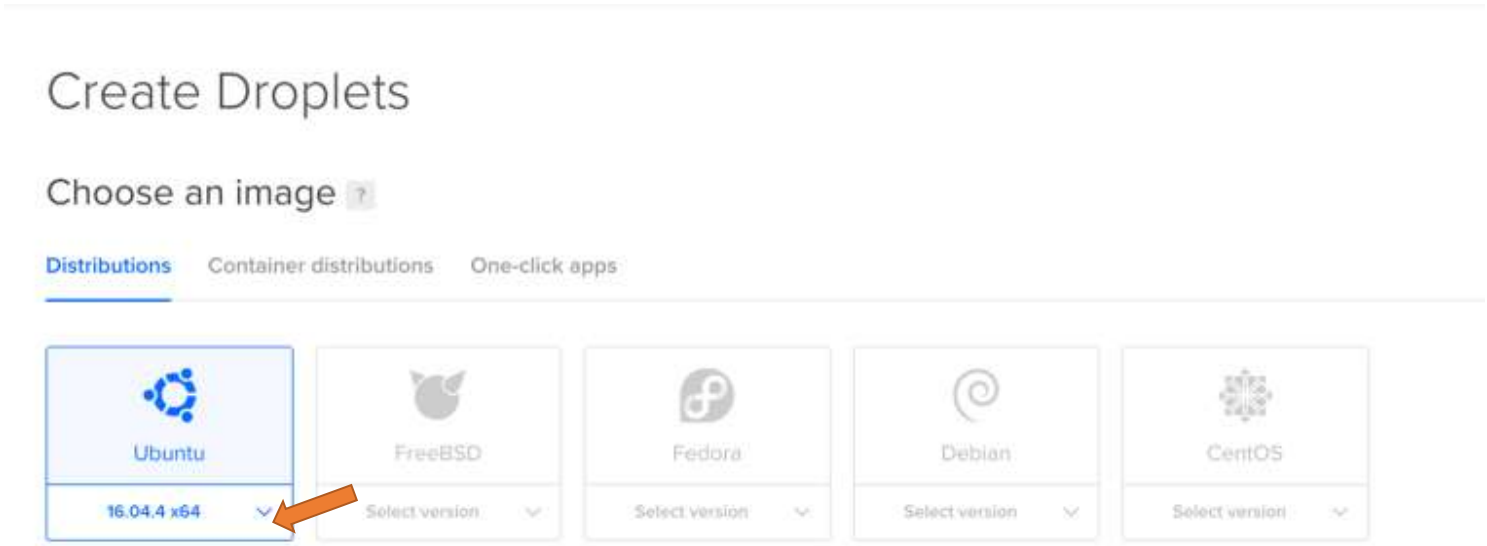
STEP 3: Setup your first droplet/virtual server

- Select “Ubuntu”



STEP 3: Setup your first droplet/virtual server

- Select other version of “Ubuntu” here



STEP 3: Setup your first droplet/virtual server

- Choose an appropriate size of your “droplet”

servers.

MEMORY	vCPUs	SSD DISK	TRANSFER	PRICE
1 GB	1 vCPU	25 GB	1 TB	$5/mo $0.007/hr
2 GB	1 vCPU	50 GB	2 TB	$10/mo $0.015/hr
3 GB	1 vCPU	60 GB	3 TB	$15/mo $0.022/hr
2 GB	2 vCPUs	60 GB	3 TB	$15/mo $0.022/hr
1 GB	3 vCPUs	60 GB	3 TB	$15/mo $0.022/hr
4 GB	2 vCPUs	80 GB	4 TB	$20/mo $0.030/hr
8 GB	4 vCPUs	160 GB	5 TB	$40/mo $0.060/hr

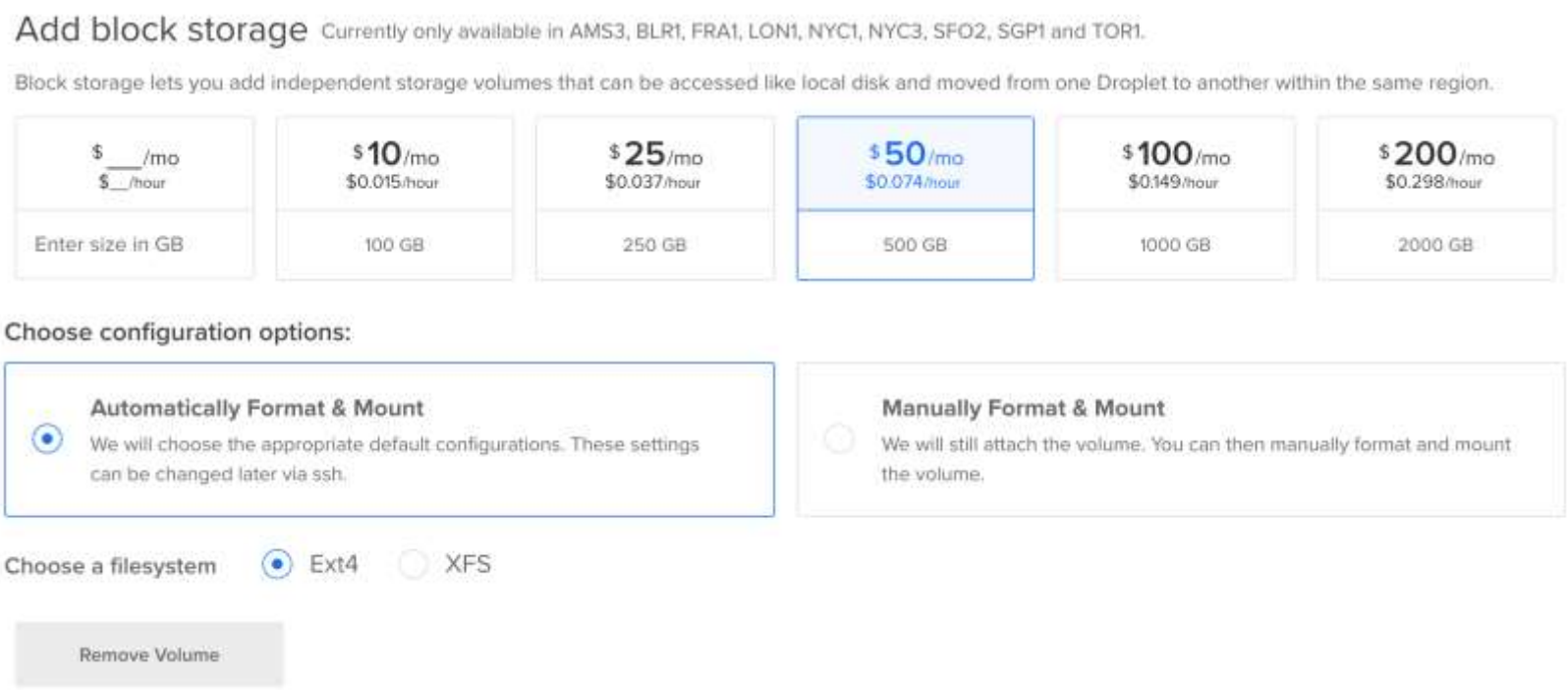
MEMORY	DEDICATED vCPUs	SSD DISK	TRANSFER	PRICE
4 GB	**2 vCPUs**	25 GB	4 TB	$40/mo $0.060/hr
8 GB	**4 vCPUs**	50 GB	5 TB	$80/mo $0.119/hr
16 GB	**8 vCPUs**	100 GB	6 TB	$160/mo $0.238/hr
32 GB	**16 vCPUs**	200 GB	7 TB	$320/mo $0.476/hr
64 GB	**32 vCPUs**	400 GB	9 TB	$640/mo $0.952/hr

STEP 3: Setup your first droplet/virtual server

- A droplet without extra disk space should be able to handle one exome analysis
- Multiple exome analysis should be run with multiple droplets in parallel
- You should NOT need an extra disk space
- But if somehow you need more space, add an extra disk space by:

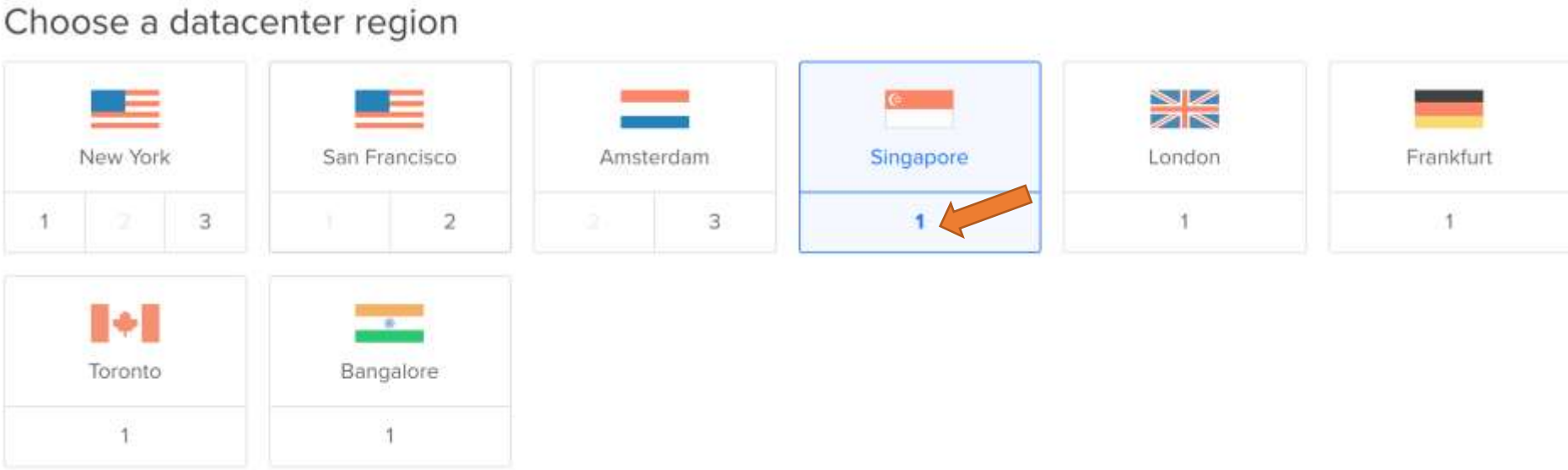
STEP 3: Setup your first droplet/virtual server

- If you need more space:



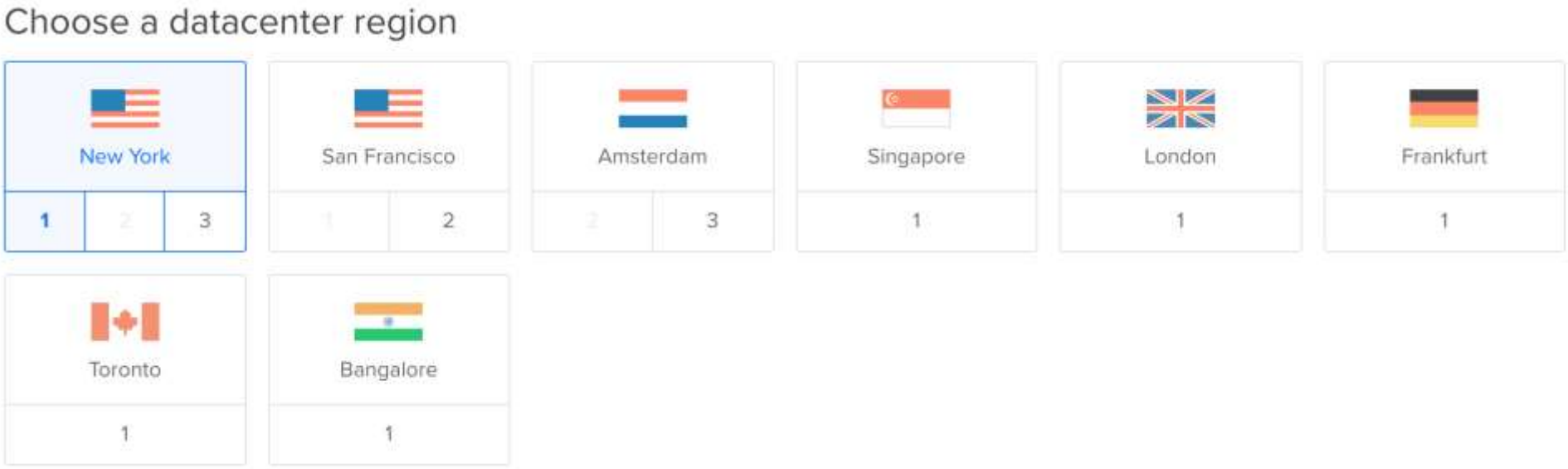
STEP 3: Setup your first droplet/virtual server

- Select an appropriate data center



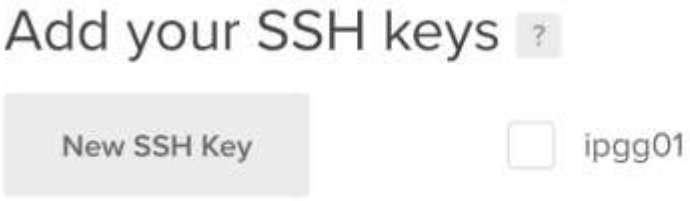
STEP 3: Setup your first droplet/virtual server

- <u>Near you (e.g. Singapore) = upload/download your data faster</u>
- Near databases/applications (e.g. USA) = download databases and applications faster



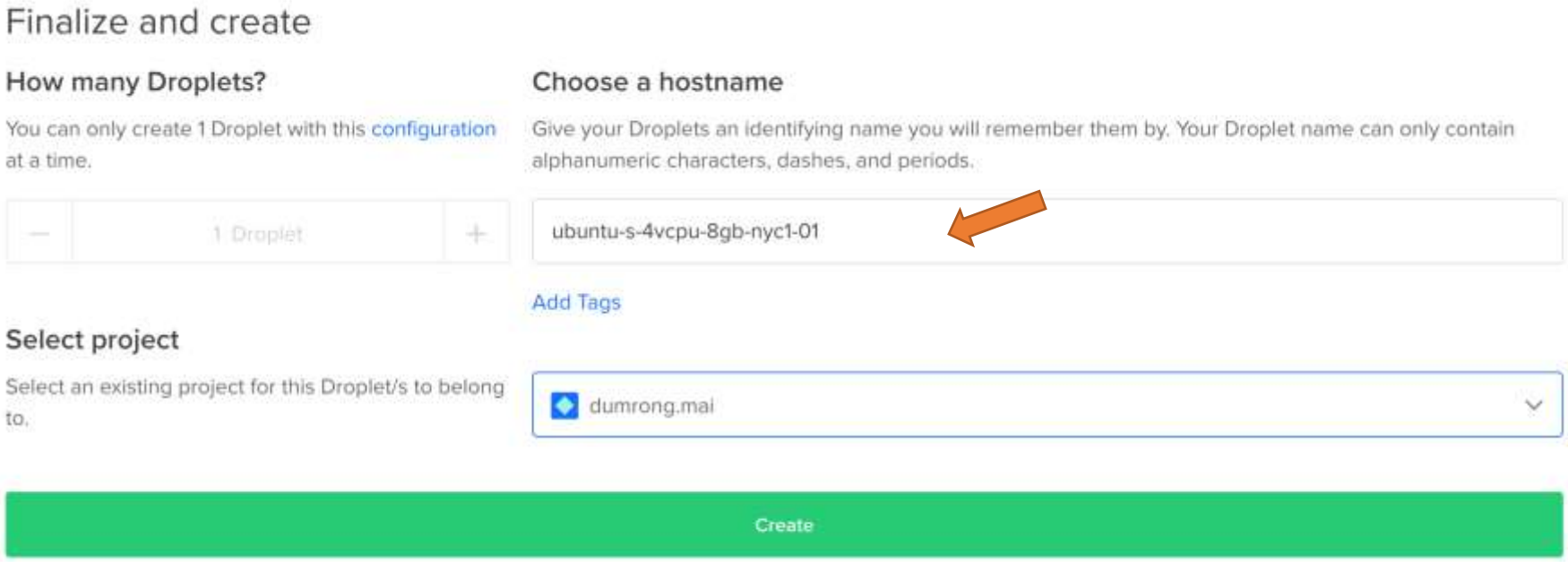
STEP 3: Setup your first droplet/virtual server

- If you want to make sure your droplet is secured, look into “SSH keys” settings
- Otherwise, skip this.



STEP 3: Setup your first droplet/virtual server

- Rename your droplet to your name
- Click ”Create” to start your droplet



STEP 3: Setup your first droplet/virtual server

- You should see this if your droplet is successfully created

STEP 3: Setup your first droplet/virtual server

- Your password should be sent to you email

Your new Droplet is all set to go! You can access it using the following credentials:

Droplet Name: ubuntu-s-4vcpu-8gb-sgp1-01
IP Address: 178.128.82.41
Username: root
Password: 07595c6860eeaddc716ab96f45

For security reasons, you will be required to change this Droplet's root password when you login. You should choose a strong password that will be easy for you to remember, but hard for a computer to guess. You might try creating an alpha-numerical phrase from a memorable sentence (e.g. "I won my first spelling bee at age 7," might become "Iwm#1sbaa7"). Random strings of common words, such as "Mousetrap Sandwich Hospital Anecdote," tend to work well, too.

As an added security measure, we also strongly recommend adding an SSH key to your account. You can do that here:
https://cloud.digitalocean.com/settings/security?i=870202

Once added, you can select your SSH key and use it when creating future Droplets. This eliminates the need for root passwords altogether, and makes your Droplets much less vulnerable to attack.

Happy Coding,
Team DigitalOcean

STEP 3: Setup your first droplet/virtual server

Can you create a “droplet”?

STEP 4: Accessing your droplet

- Before starting your “full-scale” droplet, check your internet connection with the smallest/cheapest droplet.
- Some institutions (e.g. Siriraj) do not allow a connection by SSH to external server
  - VPN or consult your IT

STEP 4: Accessing your droplet

- Mac and Linux users
- Open your terminal
- Use ssh command

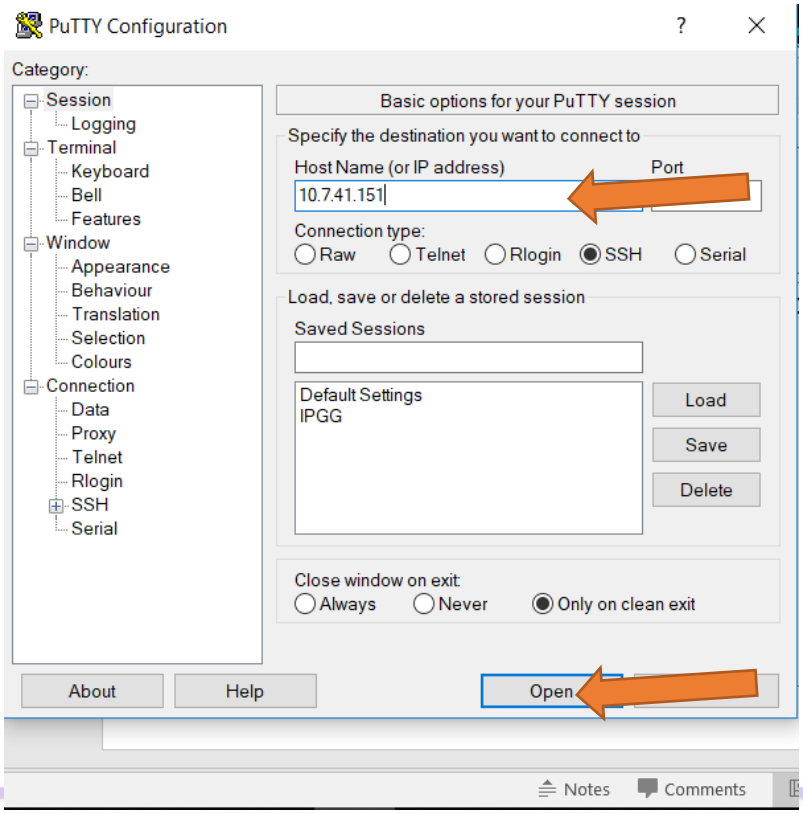
ssh root@<Droplet IP address>

Example: ssh root@123.456.78.90

- Use a password from email or from SSH key setting to login for the first time

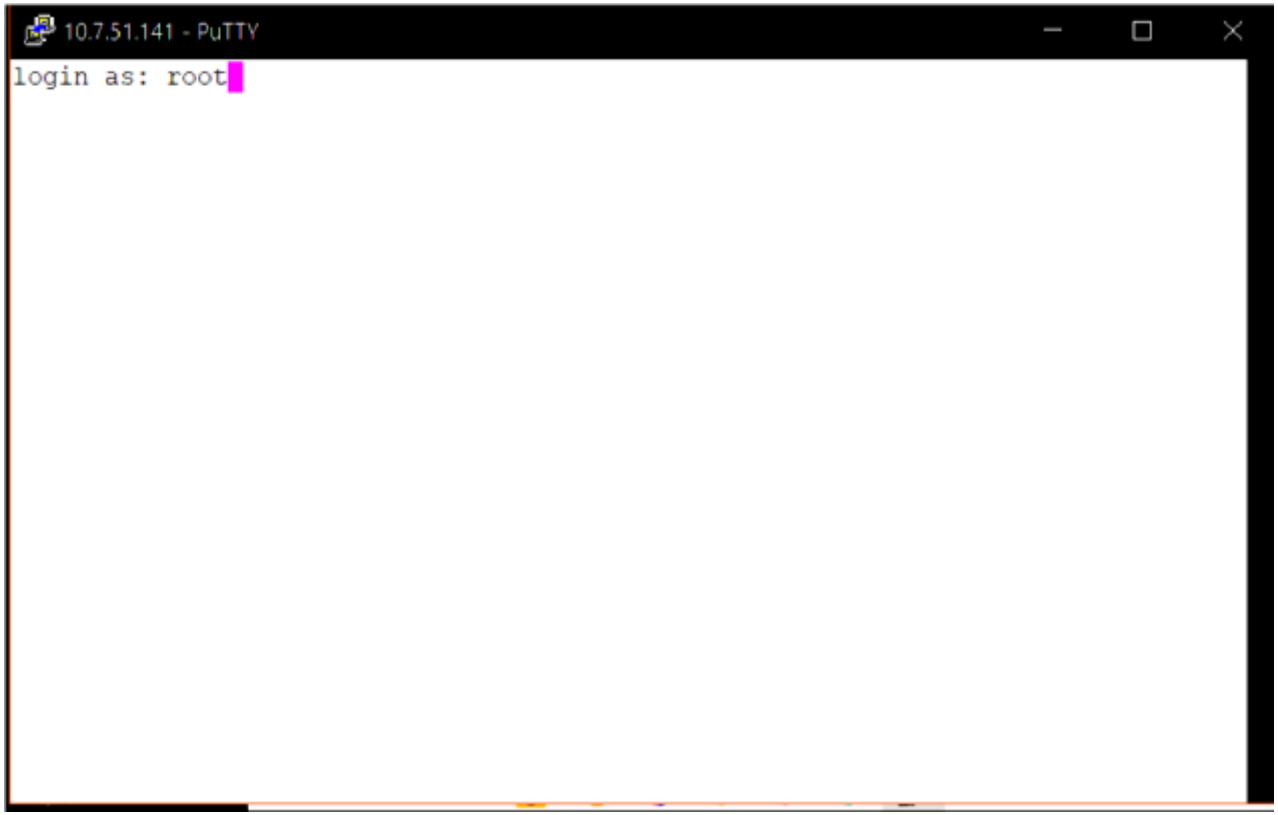
STEP 4: Accessing your droplet

- Windows user: Open PuTTY → enter droplet's IP Address → click "Open"



STEP 4: Accessing your droplet

- Login = root

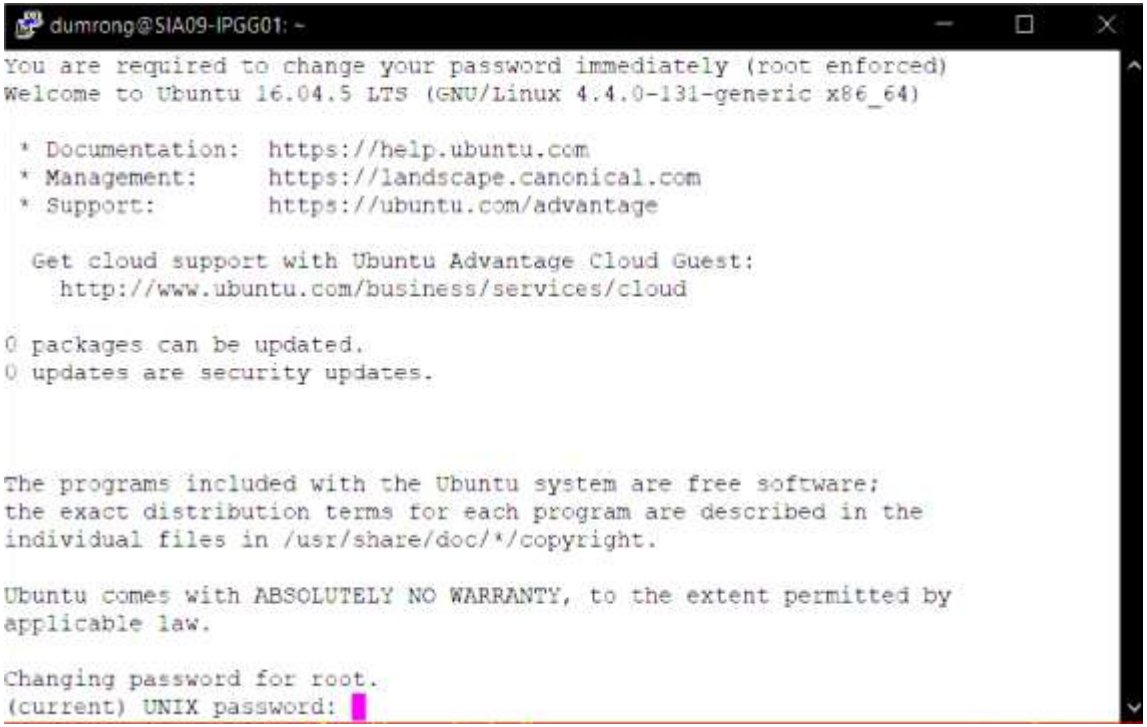


STEP 4: Accessing your droplet

- Password = password from email or SSH key
- Copy → Right click in PuTTY = Paste (You will NOT see any change, but this is OK)

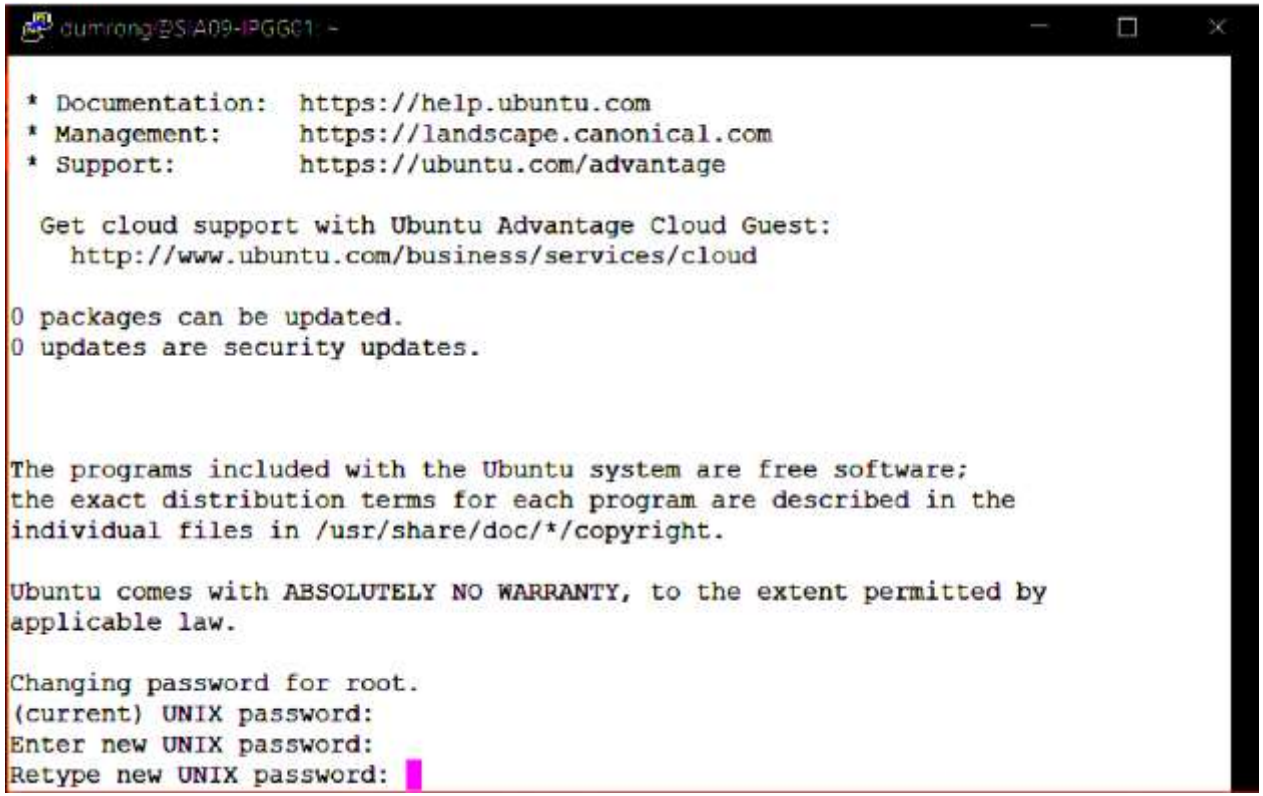
STEP 4: Accessing your droplet

- First log in → Enter password again
- Copy → Right click in PuTTY = Paste (You will NOT see any change, but this is OK)



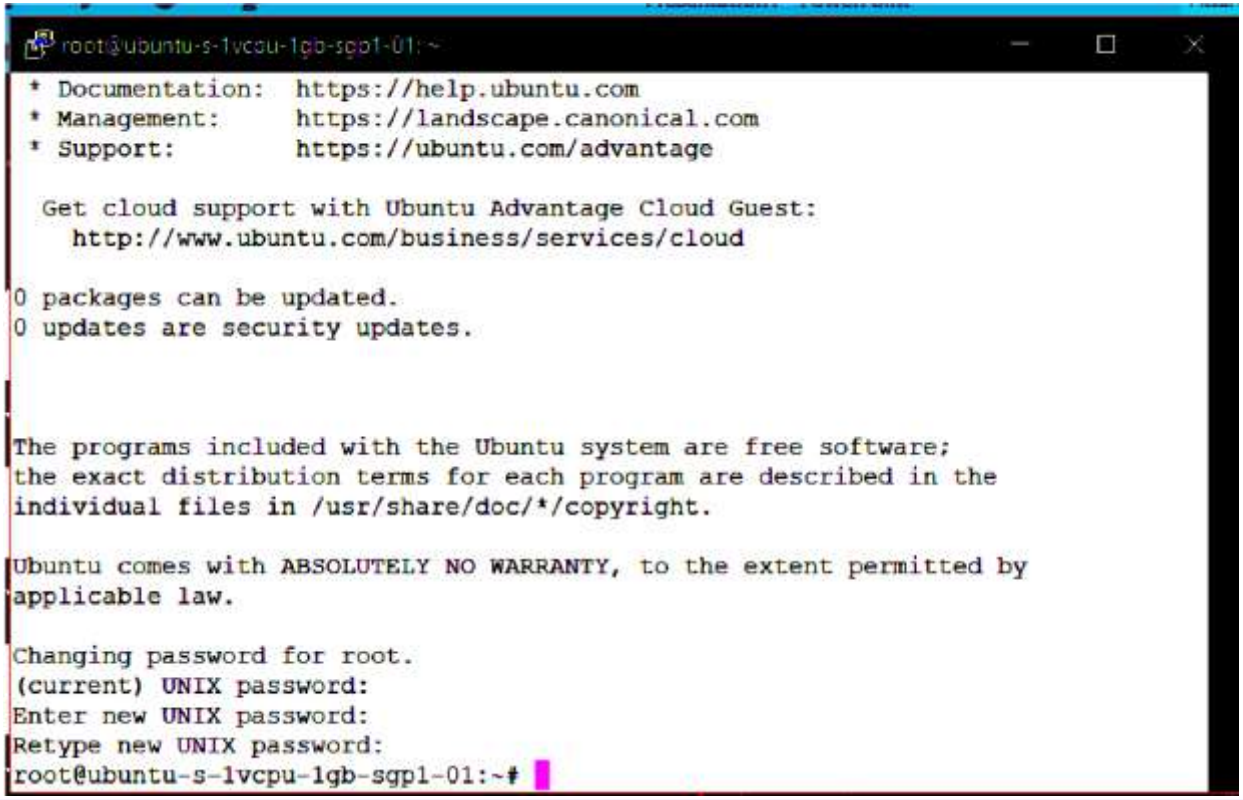
STEP 4: Accessing your droplet

- First log in → Enter password again → Set new password



STEP 4: Accessing your droplet

- You should get this:



STEP 4: Accessing your droplet

- If you set “SSH key,” your password used for setting “SSH key” will be used.
- You do not need to change password at the first log in.

STEP 4: Accessing your droplet

Can you access your “droplet”?

STEP 5: File transfer

- Mac and Linux users
- Open your terminal
- Use scp command

Local → Droplet

scp /path/file root@<Droplet IP address>:~/

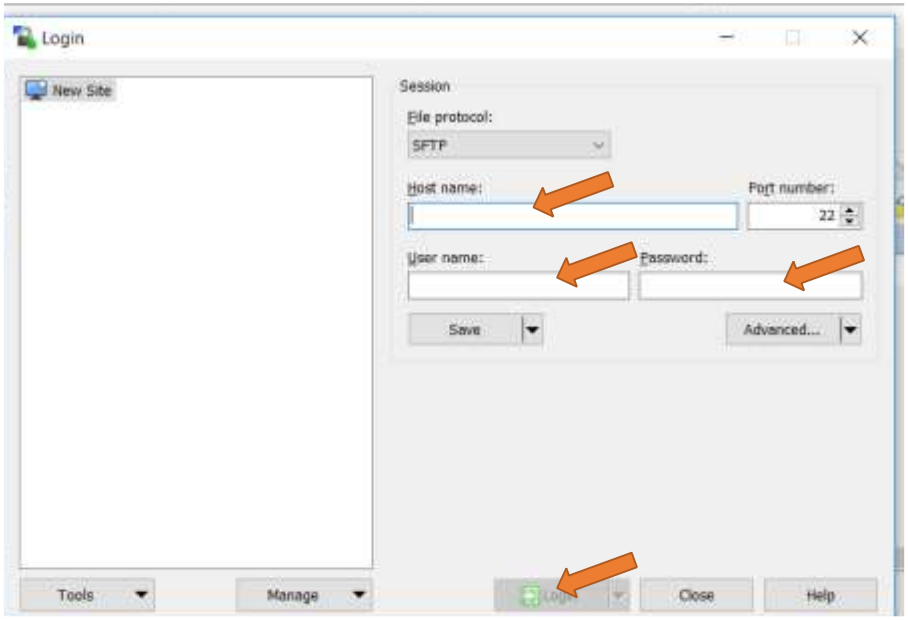
Droplet → Local (current directory)

scp root@<Droplet IP address>:~/path/file .

- Enter a password

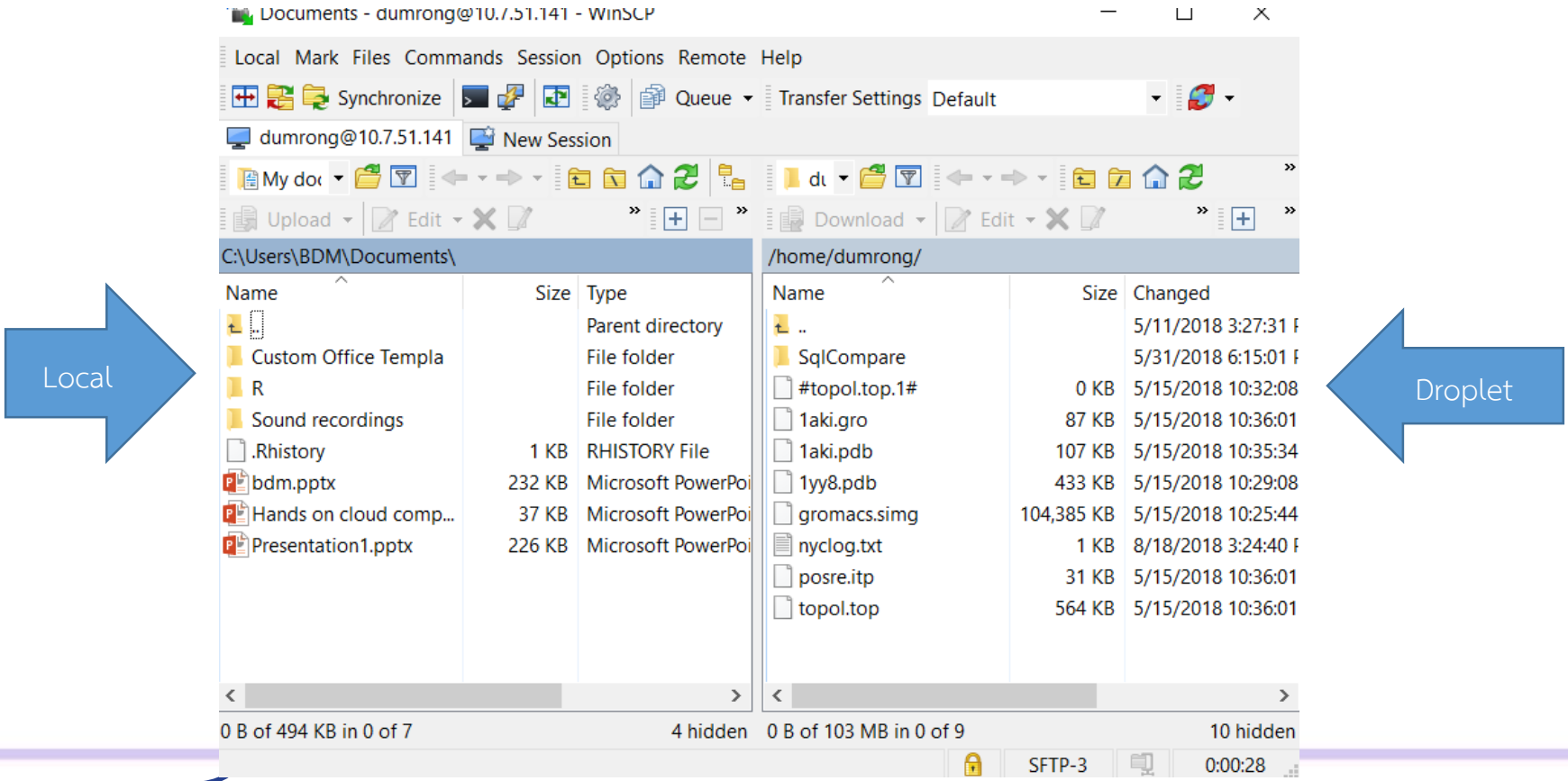
STEP 5: File transfer

- Windows user: Open WinSCP → enter droplet's IP Address → Enter username ('root') and password → click "Login"



STEP 5: File transfer

- You should see this:



STEP 5: File transfer

- Test file upload: Local → Droplet

1. Open notepad/text editor in your computer
2. Write “Thank you.” in a new text file
3. Save the file as `thank.txt`
4. Upload this file to your droplet (WinSCP or scp)
5. Access your droplet (PuTTY or ssh)
6. Type

```
ls
```

see if you can see your file

7. Type

```
nano thank.txt
```

to read the file

STEP 5: File transfer

- Test file download: Droplet → Local

1. Access your droplet (PuTTY or ssh)
2. Type

```
echo "you are welcome" > yaw.txt
```

1. Connect to your droplet with WinSCP (or use scp)
2. Download `yaw.txt` to you computer
3. Use notepad or word to read the file

STEP 5: File transfer

Can you complete file upload/download?

STEP 6: Linux/Unix command

- Update OS and program

```
apt update
```

- Upgrade OS and program

```
apt upgrade
```

- Install a program

```
apt install <program name>
```

Example: `apt install docker.io`

STEP 6: Linux/Unix command

- List files in the current directory

```
ls
```

```
ll
```

- Show the current the directory

```
pwd
```

STEP 6: Linux/Unix command

- Create a new directory

```
mkdir <directory>
```

Example: `mkdir projectA`

- Change to directory

```
cd <directory>
```

Example: `cd projectA`

- Change to “Home” directory

```
cd ~
```

STEP 6: Linux/Unix command

- Copy a file

```
cp /path1/file1 path2/
cp file1 file2
```

- Move a file

```
mv /path1/file1 path2/
```

- Rename a file

```
mv file1 file2
```

STEP 6: Linux/Unix command

- Delete a file

```
rm file1
```

- Delete a directory

```
rm -r path1/
```

- Exit Linux/Unix

```
exit
```

STEP 6: Linux/Unix command

- Screen for running a program in a detached mode
- Create a new screen session

```
screen
```

- Exit the screen session with out terminating the program

PRESS: Ctrl + a + d

- Resume the screen session

```
Screen -r <session ID>
```

STEP 6: Linux/Unix command

- Want to learn more?
- https://maker.pro/linux/tutorial/basic-linux-commands-for-beginners
- http://linuxcommand.org/

STEP 6: Linux/Unix command

Do you have any questions about Linux/Unix commands?

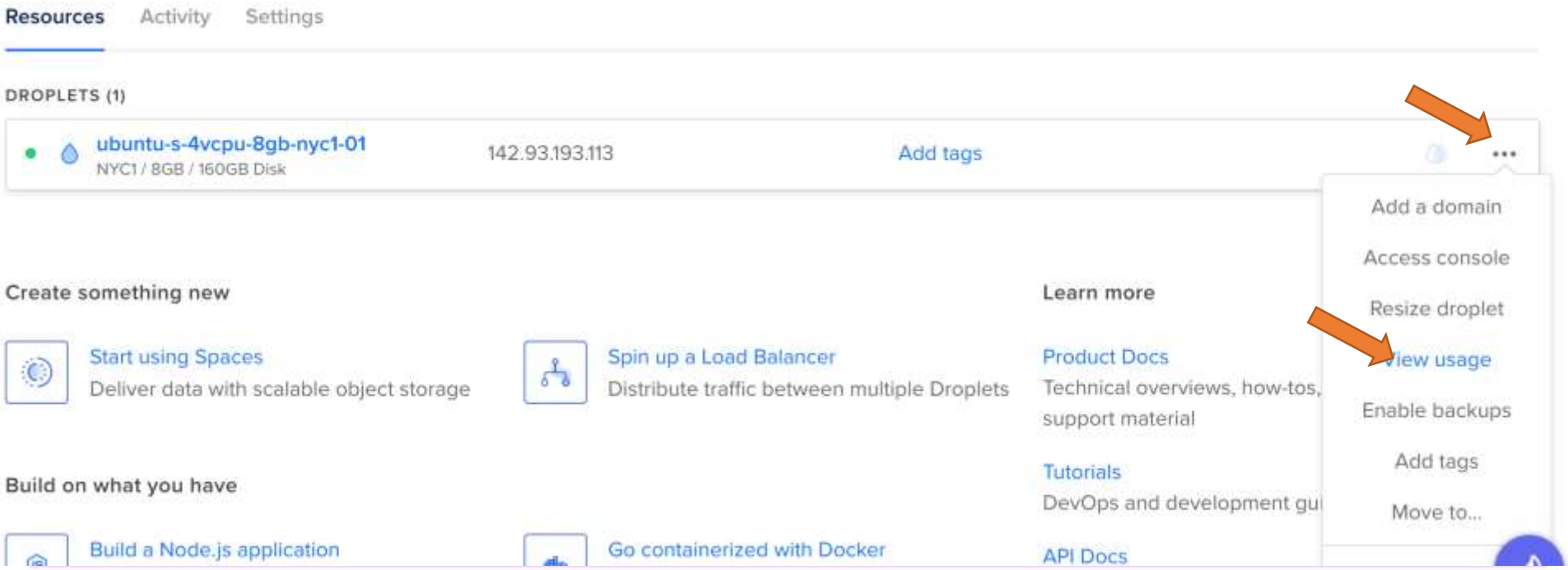
STEP 7: Prepare droplet for the pipeline

- Access your droplet with PuTTY or terminal
- Type the following commands:

```
apt update
apt install docker.io
apt install unzip
apt install tabix
apt install python-all
```

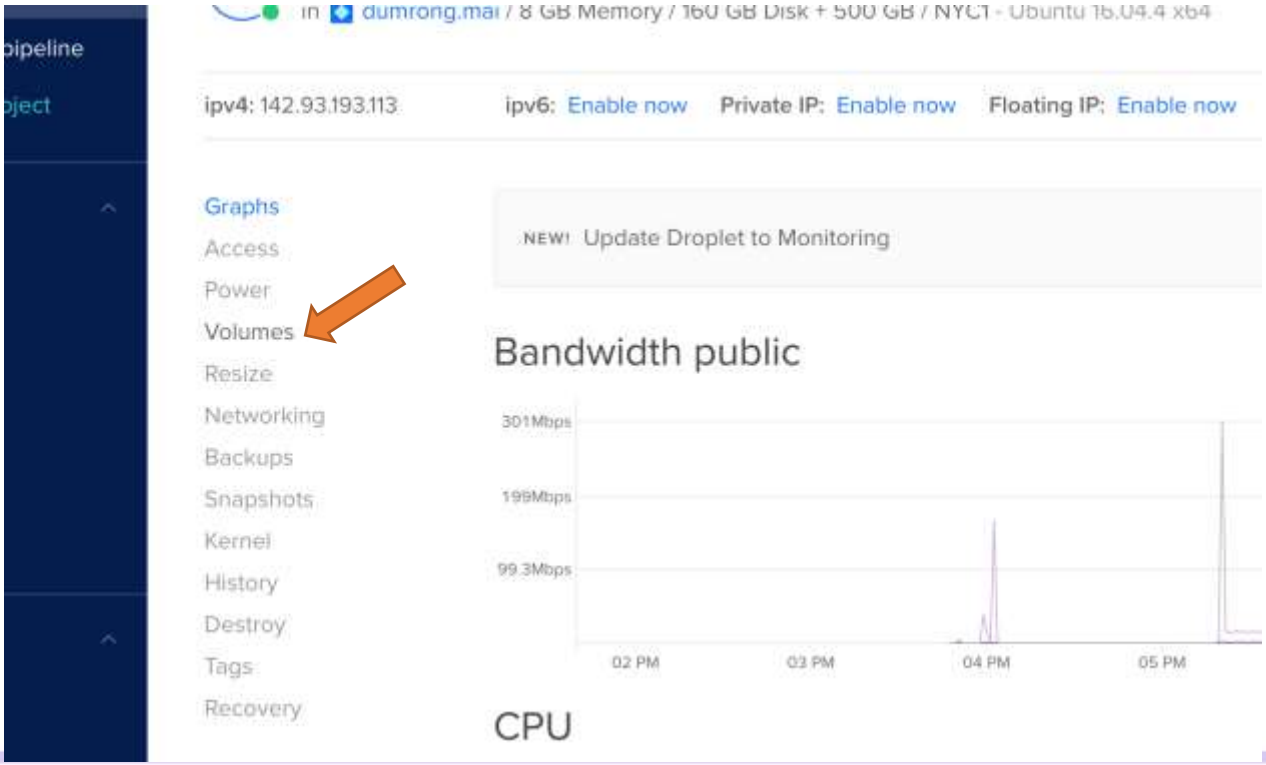
STEP 7: Prepare droplet for the pipeline

- If an extra disk space is required and attached
- Click '...' → Click 'View usage'



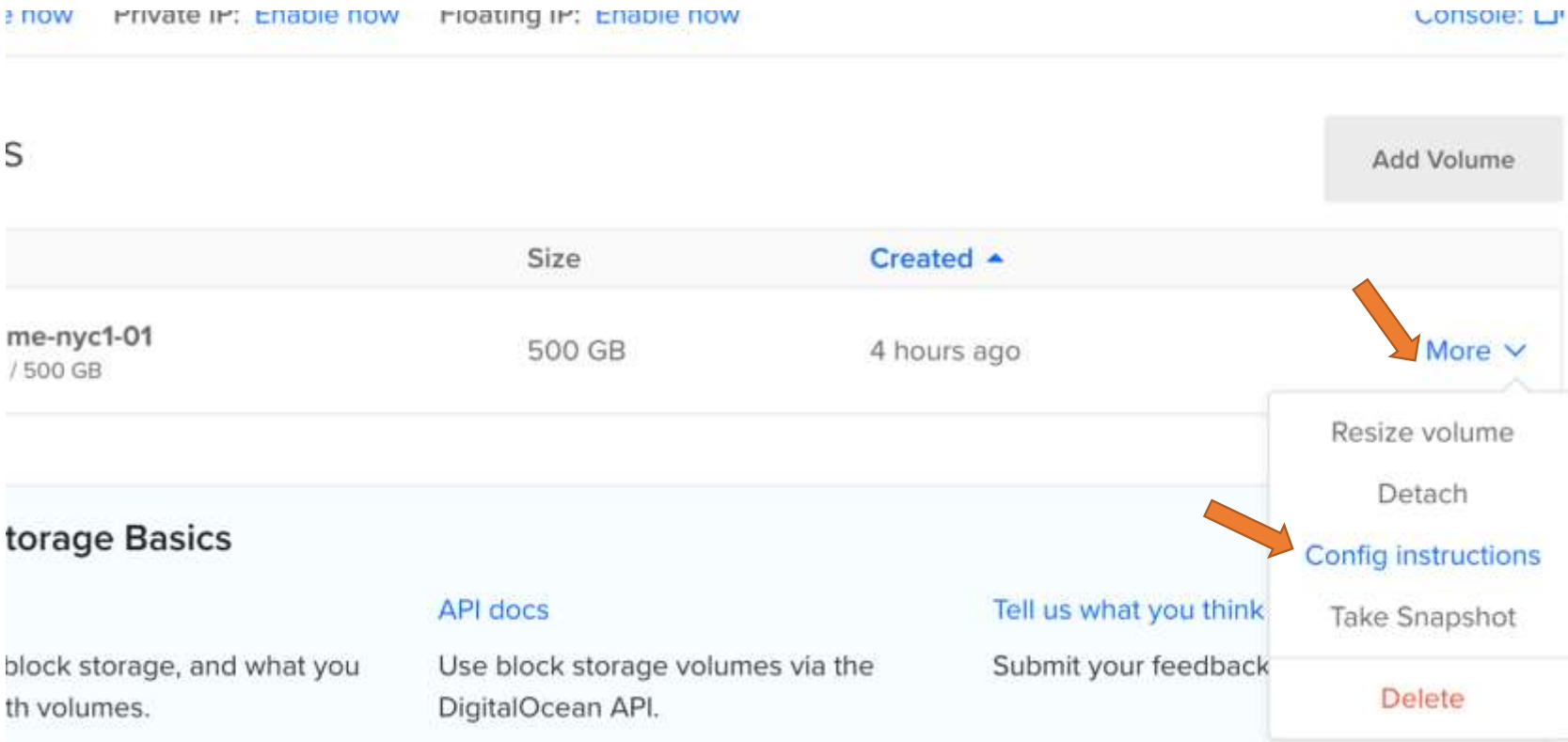
STEP 7: Prepare droplet for the pipeline

- Click ‘Volumes’



STEP 7: Prepare droplet for the pipeline

- Click ‘More’ → Click ‘Config Instructions’



STEP 7: Prepare droplet for the pipeline

- Follow instructions

Log into the Droplet

```
$ ssh root@142.93.193.113
```

Format the volume: First time only

WARNING: This step will destroy all existing data on the volume. Only run this command on a new volume or if you're sure you want to destroy the existing data.

```
$ sudo mkfs.ext4 -F /dev/disk/by-id/scsi-0DO_Volume_volume-nyc1-01
```

Mount the volume

Run the following commands on the Droplet to make your volume accessible:

```
# Create a mount point for your volume:
$ mkdir -p /mnt/volume-nyc1-01

# Mount your volume at the newly-created mount point:
$ mount -o discard,defaults,noatime /dev/disk/by-id/scsi-0DO_Volume_volume-nyc1-01
/mnt/volume-nyc1-01

# Change fstab so the volume will be mounted after a reboot
$ echo '/dev/disk/by-id/scsi-0DO_Volume_volume-nyc1-01 /mnt/volume-nyc1-01 ext4
defaults,nofail,discard 0 0' | sudo tee -a /etc/fstab
```

Copy

Access the volume, write files, and store other data on the volume in the Droplet's **/mnt/volume-nyc1-01** directory. The data will persist if you detach the volume and will be available when the volume is reattached to another Droplet.

STEP 7: Prepare droplet for the pipeline

Can you complete the preparation?

STEP 8: Clean up or Pack up

- After your analysis is complete, you must do one of these two options

1. “Clean up”
2. “Pack up”

STEP 8: Clean up or Pack up

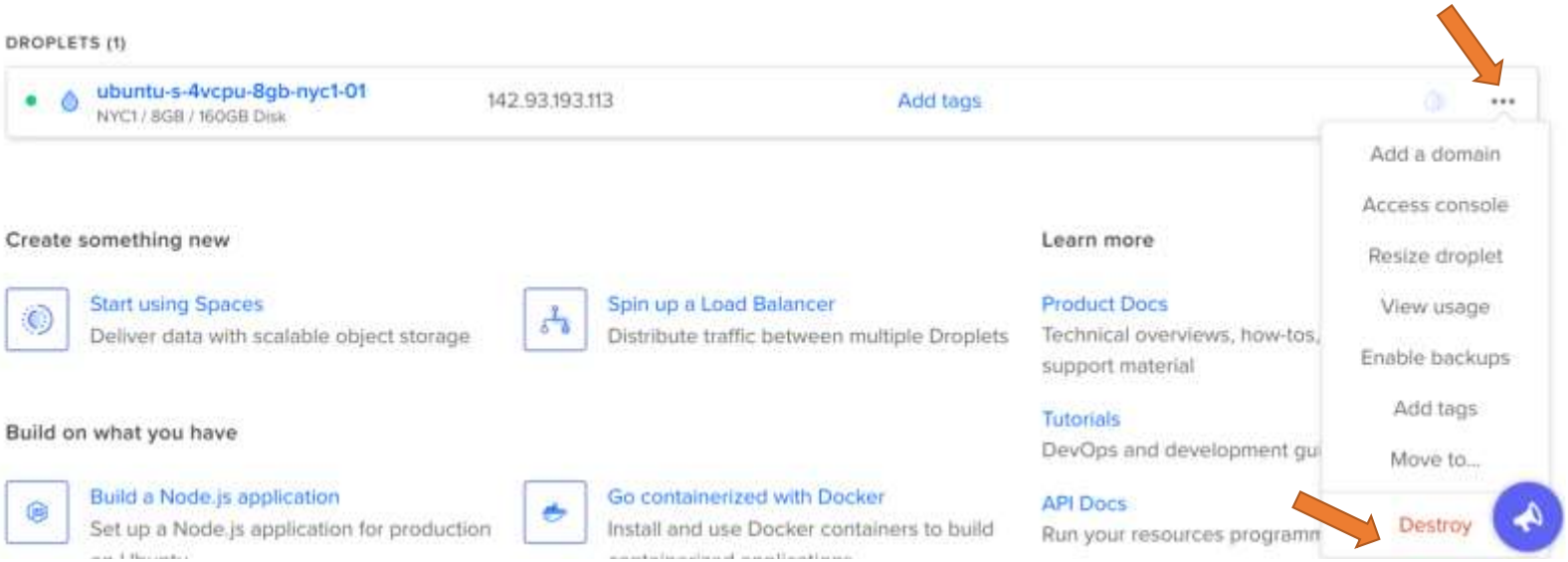
- You should select a “Clean up” option if:
  - You are completely satisfied with the results, and will absolutely not reanalysis these data
  - You do NOT plan to analyze another exome in the next 2-3 months
  - You have no/limited budget to maintain the droplet

STEP 8: Clean up or Pack up

- “Clean up” = Destroy everything!
  - Make sure you have downloaded the results from the droplet.
  - Make sure you have a copy of your data in your local computer
  - Everything in the droplet will be permanently destroyed

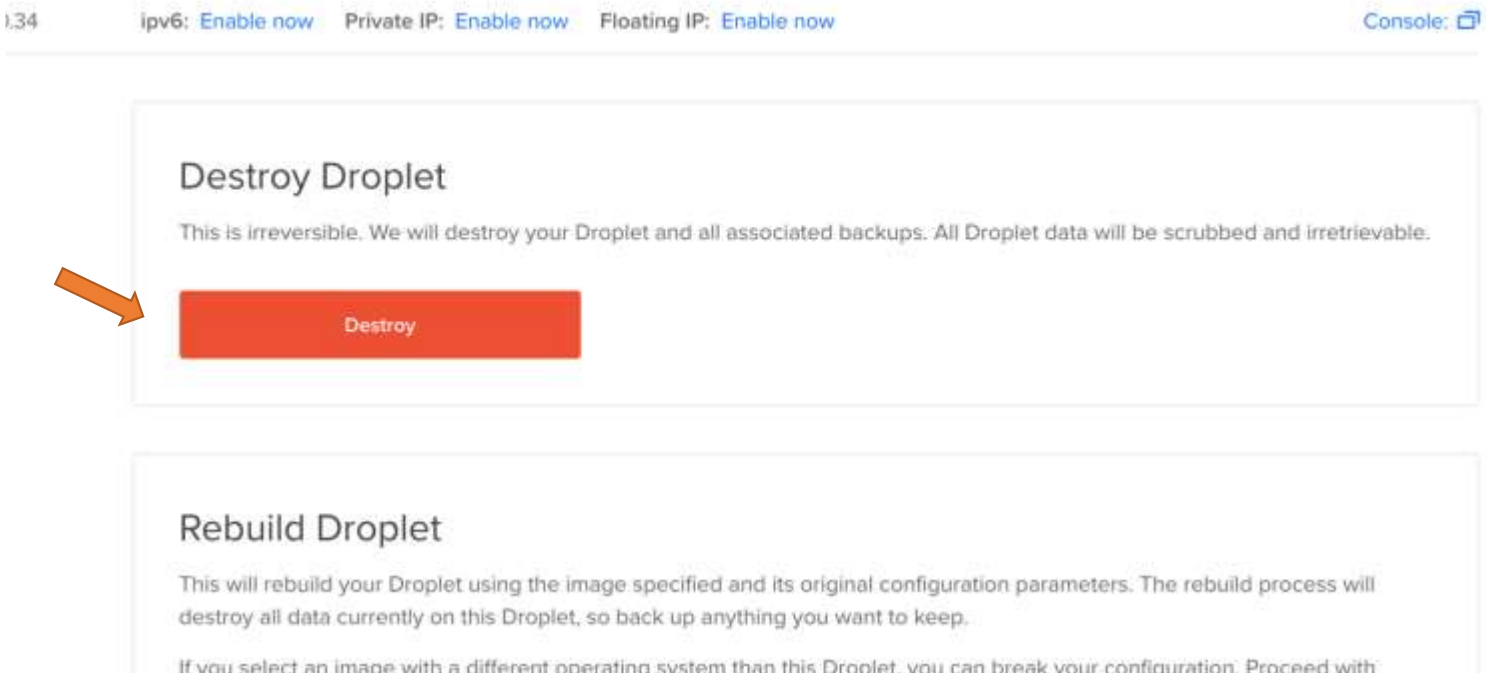
STEP 8: Clean up or Pack up

- “Clean up” = Destroy everything!
- Click “…” → Click “Destroy”



STEP 8: Clean up or Pack up

- “Clean up” = Destroy everything!
- Click “Destroy”



STEP 8: Clean up or Pack up

- You should select a “Pack up” option if:
  - You are not yet satisfied with the results, and might reanalysis these data soon (but not now)
  - You plan to analyze another exome in the next 1-2 months
  - You have some budget to maintain a “snapshot” of the droplet

STEP 8: Clean up or Pack up

- “Pack up” = Taking a snapshot
  - Make sure you have downloaded the results from the droplet.
  - Make sure you have a copy of your data in your local computer
  - Everything in the droplet will be inaccessible until the droplet is reactivated

STEP 8: Clean up or Pack up

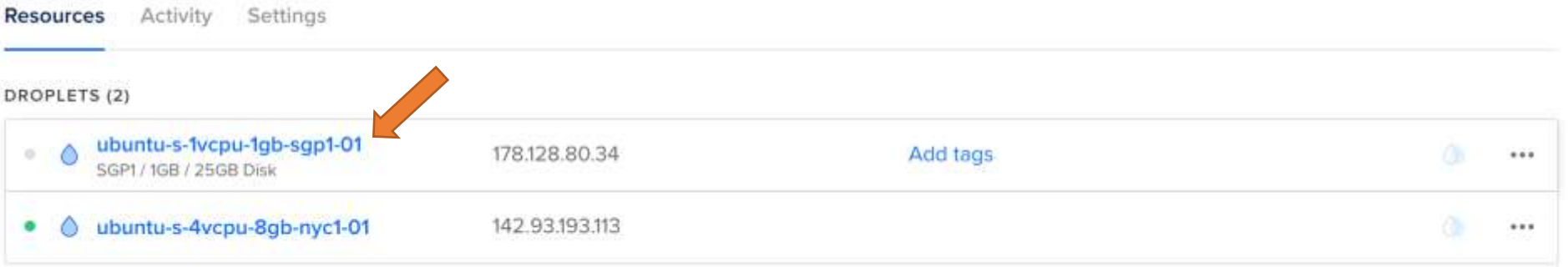
- Login to the droplet and type:

```
poweroff
```

- You can skip the shutdown if you want a live snapshot, but there might be inconsistency in the data

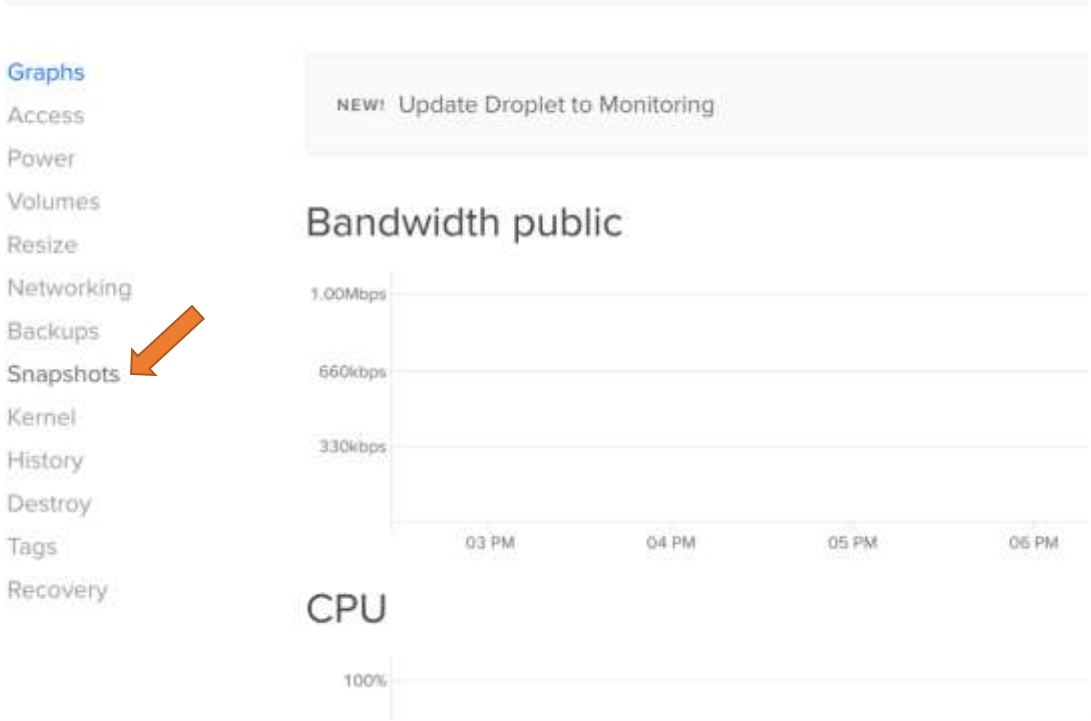
STEP 8: Clean up or Pack up

- Click the droplet



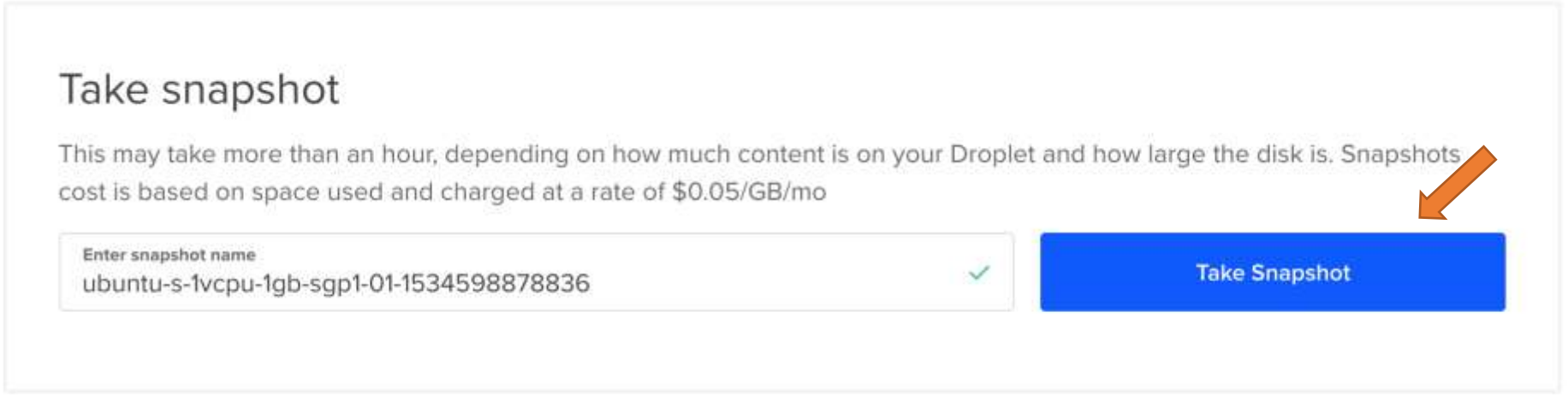
STEP 8: Clean up or Pack up

- Click “Snapshot”



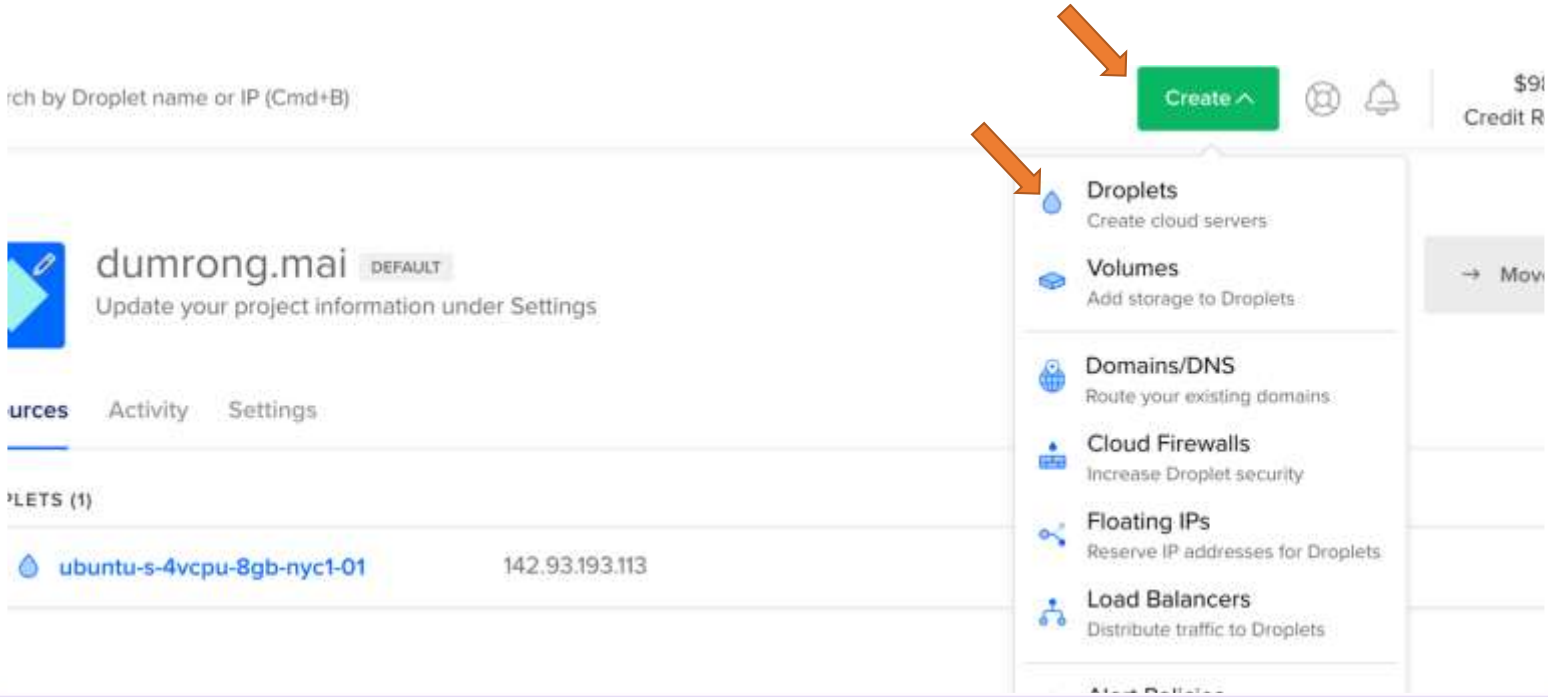
STEP 8: Clean up or Pack up

- Click “Take Snapshot”
- “Destroy” the droplet that you have created a snapshot



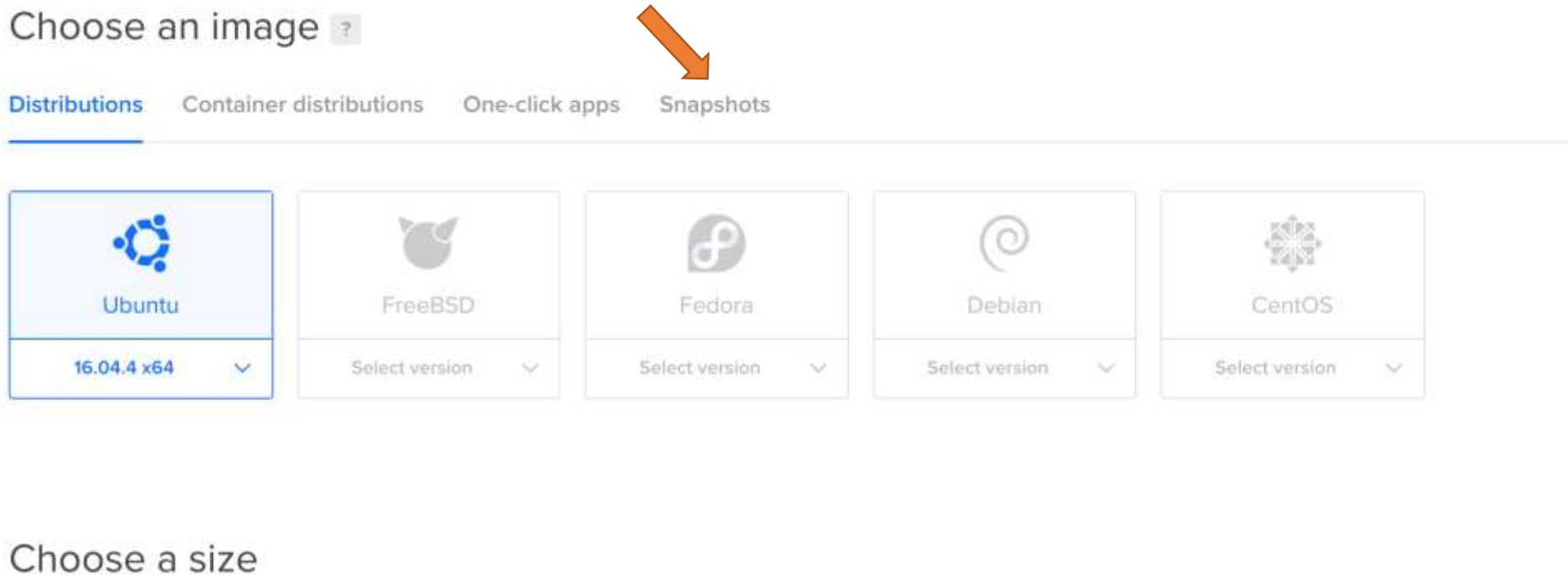
STEP 8: Clean up or Pack up

- To restore the snapshot
- Click “Create” → Click “Droplet”



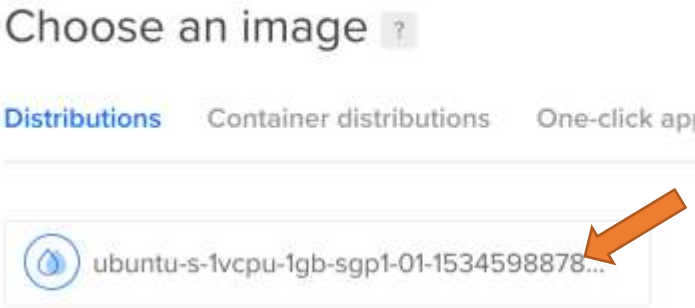
STEP 8: Clean up or Pack up

- Click “Snapshot”



STEP 8: Clean up or Pack up

- Click your “Snapshot”

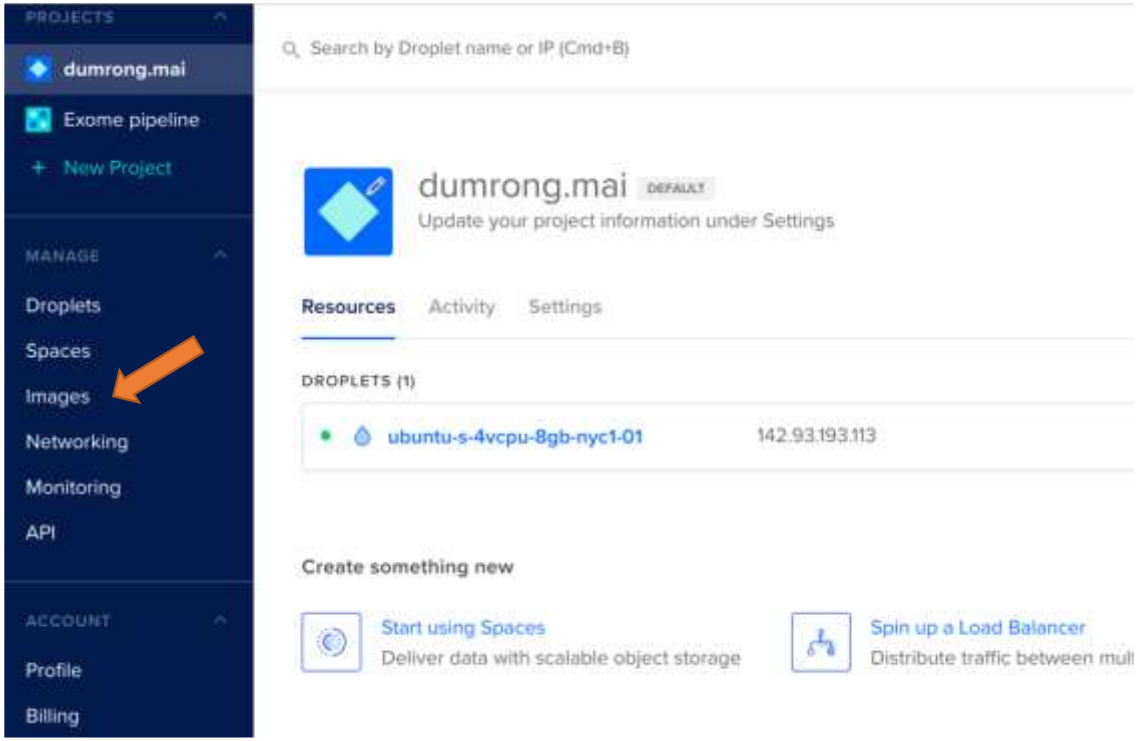


STEP 8: Clean up or Pack up

- Choose droplet size
- Choose optional configuration (e.g. SSH key, backup, block storage)
- Click “Create” to create a new droplet from a snapshot

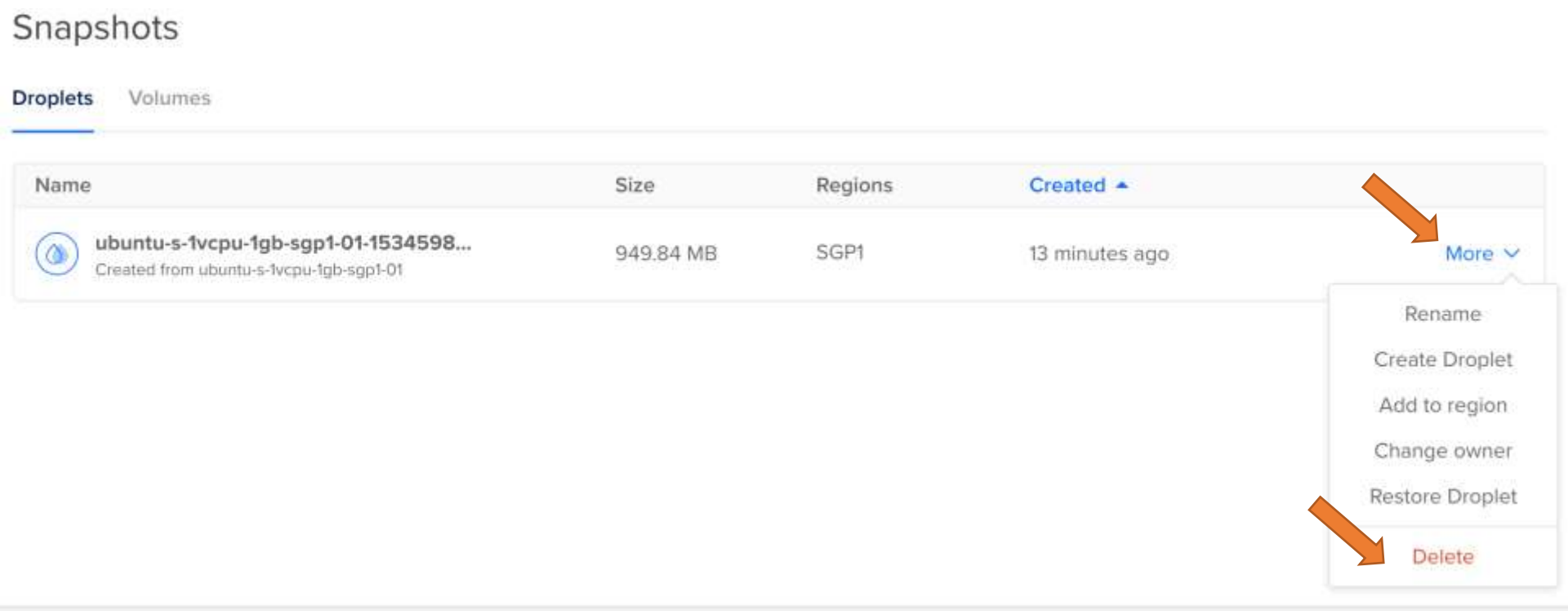
STEP 8: Clean up or Pack up

- To delete a snapshot
- Click “Images”



STEP 8: Clean up or Pack up

- Click “More” → Click “Delete”



STEP 8: Clean up or Pack up

- If you failed to do “Clean up” or “Pack up”, DigitalOcean will keep charging you for all existing droplets
- Power-off droplet = existing droplet = charged the same rate
- Snapshot is not free, but it is cheaper than maintaining a droplet
- Do not waste your money!!!

STEP 8: Clean up or Pack up

Please take a snapshot of your droplet and destroy both the snapshot and the droplet.

STEP 8: Clean up or Pack up

Do you have any question?

Next....

- Droplet specification for 1 exome analysis with the pipeline
  - 4 CPU
  - 8 GB RAM
  - 160 GB HDD

Next....

- For practicing, we will use cloud server sponsored by

  - You may use the server for free until **one month** after the workshop
  - After one month, your data and results in the cloud server may be lost
  - After one month, please use DigitalOcean for exome pipeline (you will have $100 credits for 60 days)