
notes Documentation

Yi Hsiao

Jun 27, 2021

Contents:

1	git	1
1.1	Resources	1
1.2	FAQ	1
2	Python	3
2.1	Resources	3
2.2	Articles	3
2.3	Pandas	3
2.4	FAQ	3
3	Shell script	5
3.1	Cheatsheet	5
3.2	FAQ	5
4	SSH	7
4.1	Copy ssh id to a server	7
5	AWS (Linux)	9
5.1	Enable password login	9
5.2	Add a user	9
5.3	Change user name	9
5.4	Add sudo user	9
5.5	Steps to expand EBS disk size	9
6	Docker	11
6.1	FAQ	11
7	Kubernetes-related	13
8	Bioinformatics	15
8.1	Environment setup	15
8.2	Reference Genome Build	15
8.3	Misc	15
8.4	Manuscript-related	16
9	Indices and tables	17

1.1 Resources

- Basic:
 - Git
 - git
 - * Supplementary
 - Understanding git for real by exploring the .git directory
 - Git - (garbage collection)
- Advanced
 - git hook
 - git reflog
 - ORIG_HEAD
- Others
 - learn interactively `learn-git-branching.js`

1.2 FAQ

- Rename master branch for both local and remote Git repositories
- How to remove tag locally and remotely
- How to break a commit into multiple commits

2.1 Resources

- The “Python Machine Learning” book code repository and info resource

2.2 Articles

- Dictionaries in Python (not only default one)

2.3 Pandas

- 10 Minutes to pandas
- Creating Pandas DataFrames from Lists and Dictionaries

2.4 FAQ

- How do I check whether a file exists using Python?
- Does Python have a ternary conditional operator?
- How to check if a directory exists and create it if necessary?
- What does `if __name__ == "__main__":` do?
- Calling an external command in Python
- Python join: why is it `string.join(list)` instead of `list.join(string)`?
- How can I remove (chomp) a newline in Python?

- `rstrip("\n")`
 - Universal line break
- What is the purpose of `self`?
- How can I print all arguments passed to a python script?
- Delete an item from a dictionary
- Difference between `__repr__` and `__str__`

3.1 Cheatsheet

- get current directory of the script that is performed

```
DIR="$( cd "$( dirname "$0" )" && pwd )"
```

- rsync

```
rsync -avzh /mypath/myfile.gz pi@192.168.1.12:/mybackup/
```

3.2 FAQ

- How to redirect output to a file and stdout?
 - unbuffer and tee
- How to count lines in a document?
- How to count differences between two files on linux?
- Checking for the correct number of arguments
- Propagate all arguments in a bash shell script

4.1 Copy ssh id to a server

```
ssh-copy-id -i ~/.ssh/mykey user@host
```


5.1 Enable password login

- Check [this](#) first

5.2 Add a user

- `sudo adduser [user name]`
- `sudo password [user name]`

5.3 Change user name

- Check [this](#)

5.4 Add sudo user

- Check [this](#)

5.5 Steps to expand EBS disk size

1. Modify the size in the Volume page
2. `sudo yum install xfsprogs`

```

3. [ec2-user ~]$ lsblk
NAME        MAJ:MIN RM SIZE RO TYPE MOUNTPOINT
xvda        202:0    0  30G  0 disk
└─xvda1     202:1    0  30G  0 part /
xvdb        202:16    0  30G  0 disk /mnt
xvdf        202:80    0  35G  0 disk
└─xvdf1     202:81    0   8G  0 part

[ec2-user ~]$ df -h
Filesystem      Size  Used Avail Use% Mounted on
/dev/xvda1       8.0G  943M   6.9G  12% /
tmpfs            1.9G     0   1.9G   0% /dev/shm
/dev/xvdf       1014M   33M  982M   4% /mnt

[ec2-user ~]$ sudo growpart /dev/xvdf 1
CHANGED: disk=/dev/xvdf partition=1: start=4096 old: size=16773086,end=16777182_
↪new: size=73396190,end=73400286

[ec2-user ~]$ sudo xfs_growfs -d /mnt
meta-data=/dev/xvdf          isize=256    agcount=4, agsize=65536 blks
=                               sectsz=512   attr=2
data      =                   bsize=4096   blocks=262144, imaxpct=25
=                               sunit=0      swidth=0 blks
naming    =version 2          bsize=4096   ascii-ci=0
log       =internal           bsize=4096   blocks=2560, version=2
=                               sectsz=512   sunit=0 blks, lazy-count=1
realtime  =none               extsz=4096   blocks=0, rtextents=0
data blocks changed from 262144 to 26214400

```

6.1 FAQ

- Enter detached container (for example, a development server)

```
docker run -d -p 8888:8888 python:3.5 python -m http.server 8888 --name python_
↪http
docker exec -it python_http /bin/bash
```

- Mount the data to the container

```
docker run -ti -v "$PWD":/work_dir ubuntu:14.04 /bin/bash
```

- Remove the images

```
docker images -f "dangling=true" -q
```

- Rename the image

```
docker tag <old_name> <new_name>
docker rmi <old_name>
```


CHAPTER 7

Kubernetes-related

- Be careful about the `WORK_DIR` when using `kubect1` [cp https://github.com/kubernetes/kubernetes/issues/58692](https://github.com/kubernetes/kubernetes/issues/58692)

8.1 Environment setup

- R
 - On MacOS, use `brew cask install r-apps` **rather than** `brew install r`

8.2 Reference Genome Build

- Reference genomes and Gene models: Impact on variant interpretation
- Which human reference genome to use?
- hs37d5 information

8.3 Misc

- Frequent File Formats
- Installation of samtools
- Human Reference Discrepancies
- BAF
- VCF files: Change Chromosome Notation
- Generate `vcf.gz` file and its index file `vcf.gz.tbi`
- Tools for coverage analysis: 1, 2
- WES and RNA-seq difference
- BAM Tricks

- [How to check if duplicate reads are marked in bam file?](#)

8.4 Manuscript-related

- [Which Font Looks Best in a Figure?](#)
- [How to write a manuscript by Springer Nature](#)
- [Rainbow Color Map \(Still\) Considered Harmful](#)
 - Related materials:
 - * `viridis`, `magma`, `plasma`, and `inferno` are colormaps recommended to be used in R.
 - * for automatic alignment in R:
 - `ggplot` facet layout: `facet_wrap`
 - `Cowplot::plot_grid`
 - `gridExtra::arrange.grid`

CHAPTER 9

Indices and tables

- `genindex`
- `modindex`
- `search`