Infra

Imahn Shekhzadeh
`imahn.shekhzadeh@unige.ch`

December 14, 2023

Contents

Note

The commands in this document might only run through if you use the *.bashrc* file provided in App. A

1 Baobab/Yggdrasil

- To connect to Baobab from your local machine, just type into a terminal:

```
eval $(ssh-agent)
ssh-add /home/imahn/.ssh/id_ed25519_unige_hpc
ssh shekhza2@login2.baobab.hpc.unige.ch
# ssh shekhza2@login1.yggdrasil.hpc.unige.ch
```

- Scp into yggdrasil:

```
scp file shekhza2@login1.yggdrasil.hpc.unige.ch:/home/users/s/shekhza2/
# scp -r folder_name shekhza2@login1.yggdrasil.hpc.unige.ch:/home/users/s/
    shekhza2/
```

- To see all the machines that are occupied, just type

```
squeue
squeue -p cms-uhh # partition
squeue -u shekhza2 # user
```

Listing 1: Squeue commands

- To find out about your conda environment, just type (e.g. whether you use Anaconda2 or Anaconda3)

```
conda info
```

-
```
pip install ipykernel
python -m ipykernel install --user --name <environment_name> --display-name
    "customStuff"
```

2 Bash

- Downloading file from URL and allowing for redirects,

```
curl -Lo output.out https://url.com
```

- For this directory structure,

```
infra_upd.tex
infra_upd.log
infra_upd.aux
infra_upd.out
infra_upd.pdf
```

  rename *all* of them via

```
for file in infra_upd.*; do mv "$file" "${file/infra_upd/infra}"; done
```

  What happens is called a substring replacement.

- Appending line to file,

```
echo "this is a line" | tee -a output.out # -a: appending, important
```

- Checking whether provided string (e.g. via an argument) is empty or not (typically used within conditional statements):

```
test_sth() {
 local env_name="$1" # bash starts counting indices from 1

 if [ -z "$env_name" ]; then # spacing after `[` and before `]` needed
 echo "The string is empty."
 return 1 # return value of 1 indicates error
 fi
}
```

Listing 2: Check (e.g. in if-clause) whether string is empty or not

- For retrieving all but the first argument,

```
test_sth(){
 shift

 echo "all provided args (except the first): $@"
}
```

- And of course there is nothing stopping us from doing this $N \geq 1$-times . . . Pseudocode:

```
test_sth(){
 shift
 ...
 shift

 echo "all provided args (except the first N): $@"
}
```

4

If *N* arguments are not provided, this is **not** a problem, the code will still run through.

- Example for an alias:

```
1   # forward output
2   ts(){
3    test_sth "$@"
4   }
```

- Finding out the size of a file or directory,

```
1   du -hs <path_to_file_or_dir> # du -hs file.ext
2
3   # for shorter summary (single quotation strings required)
4   du -hs <path_to_file_or_dir> | awk '{print $1}'
```

- When you want to create a new directory and you want all parent directories to be created as well (assuming they don't already exist), do

```
1   mkdir -p <dir>
```

The *-p* option is safe, since if the directory is already existent, no error will be outputted

- Searching for all files with a specific extension, e.g. *.ext*:

```
1   find . -name "*.ext"
2   # find . -name "*.png"
```

Note that this can be nicely combined with *grep*.

- In Bash, using [[ ]] instead of [ ] is preferred, since [[ ]] is safer and more capable within Bash scripts. Also, within [ ] (where word splitting and filename expansion do occur), it's good practice to double-quote your variables. But it's safe to omit the double-quotes for e.g. $# within [[ ]].

- Unzipping a file via the CLI,

```
1   unzip /path/to/file.zip -d /path/to/destination
```

- Opening a file and automatically scrolling to the bottom,

```
1   less +G /path/to/file.ext
```

- Comparing the contents of two directories,

```
1   diff -r --color directory1 directory2 # `-r` for recursive comparison
2   diff -rq --color directory1 directory2 # `-q` suppresses the output of
        differences and only shows which files differ
```

Ignoring files only existent in one of the directories (which treats absent files as empty),

```
1   diff -rq --color --unidirectional-new-file directory1 directory2
```

3 Linux

- Under Ubuntu, listing *all* available kernels,

```
dpkg --list | grep linux-image
```

Listing 3: Find kernel versions in Ubuntu

Finding the currently *active* kernel version,

```
uname -a
```

Listing 4: Current kerel versions in Ubuntu

The *-a* option stands for *appending*, otherwise *tee* overwrites *output.out* (if existent).

- Better colors in CLI:

  1. Use monokai color scheme, i.e. dark gray background (#272822) with light peach color for the text (#F8F8F2)
  2. File paths are still displayed in blue, which is suboptimal, to change the color to the better readable cyan-blue color, click on the three horizontal lines in the CLI, then on **Preferences**, then choose the currently active color, switch to the **Colors** tab, then go to **Palette**, click on the blue color & instead use the color #66D9EF

where *-h* stands for human readability and *-s* for summarizing.

- Retrieving the number of available CPU resources,

```
echo "$(nproc)"
```

- It is possible to use colored outputs in Bash. Check the bash function *str_diff* in App. A. (Note that the *-e* option is mandatory to enable interpretation of the backslash escapes).

- Print day and time from CLI,

```
echo "$(date +%d_%m_%y-%H_%M_%S)"
# echo "$(date +%dp%mp%y-%Hp%Mp%S)"
```

- Seeing the resource consumption,

```
htop
```

- If you did an *sshfs* and the connection hung up, kill the connection via

```
fusermount -zu /path/to/dir
```

3.1 Opening Programs from CLI

- Opening the settings from CLI,

```
gnome-control-center
```

- Opening VSCode from CLI:

```
code path_to_file/file_name.ext
```

If a VSCode editor is already open, use the *-n* flag to open the file in a new editor:

```
1    code -n path_to_file/file_name.ext
```

A folder can also be opened directly:

```
1    code path_to_dir
```

Listing 5: Opening VSCode dir from CLI

- Opening LibreOffice from CLI:

```
1    libreoffice --writer path_to_dir/filename.odt
```

- Opening an image via the CLI:

```
1    eog /path/to/your/image.jpg
```

4 Anaconda

4.1 Installation of Environments

- Installing conda with specific python version,

```
1  # only `myenv` needs to be specified (quotation marks necessary)
2  env_name="myenv" && conda create -n "$env_name" python=3.11.3 -y && conda
       activate "$env_name"
```

  As of Oct 16, I wouldn't recommend installing python 3.12.0 yet (I got a lot of unmet dependency problems when trying to install torch 2.1 with NVIDIA Cuda version 11.8 afterwards).

- Installation of conda environment from bash file:

```
1  conda deactivate # go into base environment
2  source conda/filename.sh
3  touch .env
```

- Completely remove conda environment,

```
1  conda deactivate && conda remove -n custom-env-name --all -y
```

4.2 Export

- Exporting an yml-file to share with others for reproducibility,

```
1  conda env export > environment.yml
```

  At the end of the file, there will be a line starting with "Prefix:", you can safely delete it, for details see here

4.3 Installation & Removal of Packages

- Installation of packages from *pyproject.toml* file,

```
1  pip install -e .
```

- Installing specific conda package version:

```
1  conda install -c conda-forge custom-pkg-name -y
2  # conda install -c conda-forge cloudpathlib=0.15.1 -y
```

- Removing list of packages from conda environment:

```
1  conda remove -n custom-env-name pkg1 pkg2 ... pkgN -y
2  # conda remove -n google_jax matplotlib -y
```

4.4 Usage in VSCode

- Selecting a conda environment in VSCode, do Ctrl + Shift + P and type *Python: select interpreter*.

- Stepping into external code with Python debugger: `https://stackoverflow.com/questions/53594900/visual-studio-code-python-debugging-step-into-the-code-of-external-function`

- Creating a JSON file, here some instructions: `https://code.visualstudio.com/docs/python/debugging`

- Listing all installed environments,

```
conda env list
```

4.5   PyTorch

- Checking whether gpu version of PyTorch is installed, from python shell (**for this, activate the right conda env first!**):

```
import os

import torch

if __name__ == "__main__":
 os.path.dirname(torch.__file__)
```

Afterwards, do

```
ls -larht <path_from_prev_alg> | grep -E "cuda"
```

- If you had installed PyTorch via conda instead of pip, then this is easier, where the *-E* means we are searching for extended regular expressions (**again activate the right conda env first!**):

```
conda list | grep -E "torch|pytorch"
# or `conda list | grep -E "torch|pytorch"`
```

5 CUDA

- When you need to find out the CUDA version installed, install *nvidia-cuda-toolkit*, but do NOT reboot. After its use, immediately remove this package and any package installed alongside with it!

- In case NVIDIA drivers do not allow for boot into Ubuntu (e.g. because you did not uninstall the *nvidia-cuda-toolkit* package):

  1. Boot into an older kernel version of Linux (in order to get there, do a "hard" reboot, and then go into "Advanced options for Ubuntu", and choose an older kernel version).

  2. Once booted into the older kernel version, I removed 'nvidia-cuda-toolkit' and rebooted.

  3. After a few more hard reboots and booting into the older kenel version, at some point, the newer kernel version was picked up and worker again.

  4. Now to fix the monitors (because dual-monitor setup didn't work), I had to open the program "Additional Drivers" and change the driver from the open-source version to an NVIDIA proprietary one.

  5. Then I had to install CUDA according to `https://docs.nvidia.com/cuda/cuda-installation-index.html` again.

  6. For PyTorch to recognize the GPU, I had to reboot.

6 Docker

6.1 Installation

- Follow this great tutorial by DigitalOcean.

- To use NVIDIA GPUs (both in PyTorch & Jax), install the NVIDIA Container Toolkit

- Once done with the installation of the NVIDIA Container Toolkit, proceed with the configuration. During the configuration, it will be necessary to restart the docker daemon, which you can achieve as follows:

```
1    sudo systemctl restart docker
```

6.2 Basics

- Interactive start of containers:

```
1    d ps -a # find out ID (also docker container name)
2    d start -i ID
```

- Copying files from local system to docker container and vice versa; **run both commands from local CLI**

```
1    d cp file_name container_ID:/target_dir # local -> docker
2    d cp container_ID:/file_name dir_name # docker -> local
```

6.3 Dockerfile

- When you find the command for pulling a docker image on `https://hub.docker.com`, e.g.

```
1    d pull ubuntu:jammy-20231004
```

then in the Dockerfile, just write

```
1    FROM ubuntu:jammy-20231004
```

When no tag is specified, by default the *latest* one will be taken. However, using the *latest* tag can potentially cause issues with reproducibility and consistency, because you might pull a different version of the image at different times without knowing it if the latest tag gets updated. **For more predictable builds, it is advised to use a specific version tag.**

- Note that the structure of the *docker pull* command is

```
1    d pull [OPTIONS] NAME[:TAG|@DIGEST]
```

In general, the *NAME* is in the format *repository/image*. If *repository* is not specified, Docker assumes the image is located in the default DockerHub library repository. However, many images (like PyTorch) are hosted under a specific user or organization's namespace on DockerHub, rather than the top-level library. That's why the command for the docker pull (for the latest tag) reads

```
1    d pull pytorch/pytorch
```

- If using a Docker image like *pytorch/pytorch:latest*, conda is already installed. In this case, the default environment is named *base*, which is a common practice in Docker images with conda – unless otherwise stated.

- Copying local scripts into docker container,

```
1    COPY relative/path/to/script.py .
```

From the documentation:

> Multiple ⟨src⟩ resources may be specified but the paths of files and directories will be interpreted as relative to the source of the context of the build.

It is also important to put the . at the end, since it represents the destination in the Docker image where the file should be copied. The dot . refers to the current working directory inside the Docker image, which is determined by the WORKDIR command in the Dockerfile. If WORKDIR is not set, it defaults to the root directory (/) of the image.

Also, each time the script *relative/path/to/script.py* changes, the Dockerfile needs to be rebuilt – **however, a cached version will be used, which speeds things up**.

- Copying local dirs into docker container,

```
1    COPY relative/path/to/dir/ .
```

- Running a Dockerfile:

```
1    d build -f file_name -t img_name .
2    d build -f file_name -t img_name:tag_name . # tag name optional, but
         recommended, e.g. 1.0 (no quotes required)
3    # d build -f file_name --no-cache -t [...] # forcing to rebuild from
         scratch, no cached version is used (only do if really required)
```

where *Image_name* will be the name of the newly created image, *Tag_name* the tag name and *file_name* the name of the docker file.

- Via

```
1    EXPOSE custom-port-number
2    # EXPOSE 80
```

it is possible to expose a port. Note that port exposure is related to network access. Note that even though network access might not be needed, there is still no harm in exposing a port (since an exposure of the port does not make the docker container more vulnerable).

6.4 Docker images

- A Dockerfile does not necessarily need to have the name *Dockerfile*. To pass another name when building the img, do

```
1    d build -f custom_docker_file .
```

The . specifies the context of the build, which is the current directory in this case. **I would recommend running this command from the same dir in which *custom_docker_file* is located**.

- Check all available Docker images via

```
1    d images
```

- Cleaning up dangling docker images (these are the entries with ⟨*none*⟩ in the repository or tag name in the output of the previous algo):

```
1    d image prune -f
```

- Removing a Docker image – **only do this when finished with using the image**

```
1    d image rm Image_name:Tag
2    # d container rm <container_id> # in case some containers are using the
         image
```

6.5 Docker containers

6.5.1 Basics

- Running Docker images – without being able to utilize NVIDIA GPUs:

```
1    d run -it img_name # if `tag_name` was not provided
2    d run -it img_name:tag_name # if `tag_name` was provided during build (
         recommended)
```

- Running Docker images & utilizing GPUs:

```
1    d run --gpus all -it img_name
2    d run --gpus all -it img_name:tag_name # recommended
```

- To mount a local file to the container at runtime, do

```
1    d run -v /absolute/path/to/script.py:/path/to/workdir/script.py --gpus all
         -it img_name
2    d run -v /absolute/path/to/script.py:/path/to/workdir/script.py --gpus all
          -it img_name:tag_name # recommended, provide `img_name` & `tag_name`
```

The mounting expects **absolute** file paths on the side of the host machine.

- Note that you can include the bash command **pwd** to avoid having to manually pass absolute paths for the mounting

```
1    d run -v $(pwd)/script.py:/path/to/workdir/script.py --gpus all -it
         img_name:tag_name # recommended, provide `img_name` & `tag_name`
```

If you need the container to reflect changes made to the scripts on the host without rebuilding the image every time, you would use the *-v* flag to mount the directory. If the scripts won't change, or you don't need to reflect changes in real-time, you don't need to mount the directory, as the necessary scripts have already been copied into the image during the build process.

- It is also possible to directly mount directories:

```
1    d run -v $(pwd)/dir_path:/path/to/workdir --gpus all -it img_name:tag_name
```

Note that the specified directory from the host is mounted into the container at the specified mount point. If there are any existing files or directories in the container at the mount point, they become obscured by the mount.

- In several cases it can be useful to remove the docker container right after execution: When you. . .

  - . . . are running many short-lived containers, like during development or testing,
  - . . . want to avoid manual cleanup of stopped containers later on,
  - . . . are running containers for one-off tasks that do not need to persist any state after they are finished.

  In this case,

```
1   d run --rm -v $(pwd)/dir_path:/path/to/workdir --gpus all -it img_name:
        tag_name
```

- It is also possible to mount two separate host directories to two separate directories within the container,

```
1   d run --rm -v $(pwd)/dir_path1:/path/to/workdir1 -v $(pwd)/dir_path2:/path
        /to/workdir2 --gpus all -it img_name:tag_name
```

  This will not cause any overwriting as each *-v* flag creates a unique mount point inside the container.

- Finding out the python version of the Docker image

```
1   d run -it --rm img_name:tag_name python3 --version
```

  This command will immediately remove the container after execution.

6.5.2 Passing Arguments

It is possible to pass arguments when running a docker container.

1. Assuming you have a bash script *run_scripts.sh*, in which a Python script, e.g.

```
1   #!/bin/sh
2   isort /app/scripts/*.py
3   black /app/scripts/*.py
4
5   python3 -B /app/scripts/test_script.py
6   python3 -B /app/scripts/test_anil.py
```

   Modify this bash script s.t. any arguments passed to the CLI when running the docker container are picked up,

```
1   python3 -B /app/scripts/test_anil.py "$@"
2   # python3 -B /app/scripts/test_script.py "$@" # alternative
```

2. Rebuild (!) the docker image.

3. Now run the docker container as follows:

```
1   d run --rm -v $(pwd)/dir_path:/path/to/workdir --gpus all -it img_name:
        tag_name arg1 arg2
2   # d run --rm -v $(pwd)/dir_path:/path/to/workdir --gpus all -it img_name:
        tag_name --n_ways 1 --k_shots 1 # example
```

14

6.5.3 Listing & Stopping

- Listing all running containers,

```
1    d ps
```

Listing only the container ID (of all running containers),

```
1    d ps -q
```

- Stopping a running container,

```
1    d stop container-ID
```

- Stopping a running container and removing it,

```
1    d stop container-ID && d rm container-ID
```

6.6 Pushing to DockerHub & HPC

1. First login to Docker via

```
1    d login -u user_name -p password
```

2. Then follow the instructions from this tutorial (from minute 17:05 on)

3. And then follow the HPC tutorial from UNIGE

7 Python

7.1 Config File & JSON Files

- When using argparse in combination with a JSON configuration file, the JSON keys need to match the long option names specified in *parser.add_argument* method calls. The argparse module itself does not automatically recognize abbreviated forms from a JSON file.

7.2 Jupyter Notebooks

- Converting jupyter notebooks into PDFs:

```
for nb in /path/one/Notebook1.ipynb /path/two/Notebook2.ipynb [...]
do
 jupyter nbconvert --to pdf "$nb"
done
```

If you have several notebooks in the same directory,

```
for nb in *.ipynb
do
 jupyter nbconvert --to pdf "$nb"
done
```

7.3 Map (Built-In Function)

- Function signature:

```
map(function, iterable, *iterables)
```

Description provided in the documentation:

> Return an iterator that applies function to every item of iterable, yielding the results. If additional iterables arguments are passed, function must take that many arguments and is applied to the items from all iterables in parallel. With multiple iterables, the iterator stops when the shortest iterable is exhausted. For cases where the function inputs are already arranged into argument tuples, see *itertools.starmap()*.

- Example usage: Natively multiplying Python lists elementwise,

```python
from typing import List

def multiply(x: List, y: List):
    return x * y

list_one = [i for i in range(1000)]
list_two = [j for j in range(1000, 2000)]
result = list(map(multiply, list_one, list_two)) # `map` is a built-in
    function, do not use `(list_one, list_two)` in this case
```

- Example usage: Converting NumPy arrays into PyTorch tensors,

```python
a = np.array([1, 2, 3, 4])
tensor_list = list(map(torch.from_numpy, (a,))) # list containing tensor,
    use of additional brackets necessary
```

- Example usage: Converting NumPy arrays into PyTorch tensors,

```
1   a = np.array([1, 2, 3, 4])
2   b = np.array([5, 6, 7, 8])
3   a, b = map(torch.from_numpy, (a, b)) # tuple unpacking
```

7.4 PyTorch

7.4.1 Leaf Tensors

- If `requires_grad=False`, then the tensor will be leaf by convention. If `requires_grad=True`, then the tensor will be leaf if it was created directly by the user and is **not** the result of an operation, e.g. `.to(device)` when the tensor is on `cpu` and `device="cuda:0"`.

- However, by definition, leaf tensors themselves do not have a **gradient function** `.grad_fn` because they are not the result of a differentiable operation applied to other tensors, i.e. `grad_fn` on such tensors will return `None`. The gradient function in neural network libraries like PyTorch or TensorFlow is associated with tensors that are outputs of differentiable operations.

- The `.grad` attribute on leaf tensors that require gradients, i.e. those for which `requires_grad=True`, stores the gradient computed during backpropagation. (For leaf tensors that have `requires_grad=False`, calling the `.grad` attribute outputs `None`.) Note that for non-leaf tensors, calling `.grad` results in a `UserWarning`, since non-leaf tensors are generally intermediate results in the computation graph, and their gradients are usually not needed once the gradients of the leaf tensors have been obtained. However, there are cases where those gradients are needed, which can be enforced by setting `retain_grad=True` on those tensors,

```
1    x = torch.tensor([1., 2., 3.], requires_grad=True, device=torch.device("
         cuda:0"))
2
3    # fwd pass
4    y = x**2
5
6    # retain gradients
7    y.retain_grad()
8
9    # backward pas
10   y.sum().backward()
11
12   # access gradients
13   y.grad # `torch.tensor([1., 1., 1.], device="cuda:0")`
```

Note that in the example of this code snippet, doing `y.grad` means that we access the gradient of the scalar loss function `y.sum()` — on which we performed `.backward()`. Correspondingly, doing `x.grad` implies the gradient of the scalar loss function `y.sum()` with respect to `x`.

- In general, it is **not** possible to perform **in-place** operations on leaf tensors for which `requires_grad=True`, since PyTorch dynamically builds a computational graph during the forward pass, which is used during backpropagation to calculate the gradients. If leaf tensors that have `requires_grad=True` are changed in-place, then the values used during the forward pass are changed, which will affect the gradient calculations in the backward pass. However, note that when no gradients are required for the operations, e.g. when performing parameter updates manually, one can use the context manager `with torch.no_grad()`, in which case in-place operations on leaf tensors **can** be performed, since inside the context manager, `requires_grad=False`.

17

7.4.2 Autograd & Backward

- The function `torch.autograd.grad()` computes the gradient. If the gradient of a scalar (loss function) wrt a (weight) matrix is taken, then the output will also be a matrix, where each element corresponds to the partial derivative of the scalar (loss function) wrt to the (weight) matrix element.

- `torch.autograd.grad()` is particularly useful if more direct control over the gradient computation is desired, in particular compared to `.backward()`.

- Note that the default behavior of `.backward()` accumulates gradients in the `.grad` attribute of tensors,

```python
x = torch.tensor([1., 2., 3.], requires_grad=True, device=torch.device("
    cuda:0"))

# fwd pass
y = 2 * x

# first backward pass
y.sum().backward(retain_graph=True)
print(f"Gradients of `x` after first backward pass: {x.grad}") # `torch.
    tensor([2., 2., 2.])`

# second backward pass
y.sum().backward()
print(f"Gradients of `x` after second backward pass: {x.grad}") # `torch.
    tensor([4., 4., 4.])`, notice how gradients have accumulated
```

However, this behavior can be suppressed by simply zeroing the gradients, i.e. `x.grad.zero_()` — note that `x.grad` returns a tensor, and `<tensor>.zero_()` is a general PyTorch function that sets all elements in-place to 0.

8 AWS S3

8.1 Installation & Configuration

1. Installation instructions: `https://docs.aws.amazon.com/cli/latest/userguide/getting-started.html#getting-started-install-instructions`

2. The CLI will display the path under which the *aws* package was installed, but it might be sufficient to simply run

```
1    aws
```

Double check by running

```
1    which aws
```

3. After installation, configuration is necessary. For this run

```
1    aws configure
```

You can leave these fields empty:

```
1    Default region name [None]:
2    Default output format [None]:
```

A configuration file will be saved under

```
1    ~/.aws/credentials
```

4. In the case you are a member of UNIGE, you can obtain the AWS access key ID and the secret access key as follows:

```
1    echo -n "$user_name" | base64 # the `-n` is important in this context
2    echo -n "$passwd" | md5sum
```

where $user_name and $passwd need to be provided

Otherwise, you need login to the AWS Management Console.

5. **To test the configuration was successful, do this:**

```
1    aws s3 ls --endpoint-url https://your-custom-s3-endpoint.com
```

where you replace the endpoint-url `https://your-custom-s3-endpoint.com` with yours.

8.2 AWS Credentials (Profiles)

- It is possible to use several profiles in the file *~/.aws/credentials*.

- For example:

```
1    [default]
2    aws_access_key_id = YOUR_DEFAULT_ACCESS_KEY
3    aws_secret_access_key = YOUR_DEFAULT_SECRET_KEY
4
5    [profile1]
6    aws_access_key_id = ANOTHER_ACCESS_KEY_ID
7    aws_secret_access_key = ANOTHER_SECRET_ACCESS_KEY
```

```
8
9    [profile2]
10   aws_access_key_id = YET_ANOTHER_ACCESS_KEY_ID
11   aws_secret_access_key = YET_ANOTHER_SECRET_ACCESS_KEY
```

To use a specific profile when running *aws cli* commands, you can use the *–profile* option in the command:

```
1    aws s3 --profile profile1 [...]
2    # aws s3 --profile default [...]
```

8.3   Buckets

One can have several buckets.

8.3.1   Creation

- To create a new bucket:

```
1    aws s3api create-bucket --bucket custom-bucket-name --endpoint-url https://
         custom-s3-endpoint.com --profile default
```

8.3.2   Listings

- Directly showing the file contents of an s3 bucket,

```
1    aws s3 ls s3://custom-bucket-name --recursive --endpoint-url https://custom
         -s3-endpoint.com --profile default # `--recursive` optional
```

8.3.3   File Copying

- Local machine ⟶ S3:

```
1    aws s3 cp path/to/custom_file.ext s3://custom-bucket-name/path/to/
         custom_file.ext --endpoint-url https://custom-s3-endpoint.com --profile
          default
```

- S3 ⟶ local machine:

```
1    aws s3 cp s3://custom-bucket-name/path/to/s3_file.ext custom/destination --
         endpoint-url https://custom-s3-endpoint.com --profile default
```

8.3.4   Directory Copying

- Local machine ⟶ S3:

```
1    aws s3 sync path/to/dir s3://custom-bucket-name/path/to --endpoint-url
2    https://custom-s3-endpoint.com --profile default
```

8.3.5   Folder/File Deletion

- Deleting a folder (which is essentially a prefix in S3) and its contents in an S3 bucket,

```
aws s3 rm s3://your-bucket-name/path-to-your-folder --recursive --endpoint-
    url https://custom-s3-endpoint.com --profile default
```

- Deleting a file,

```
aws s3 rm s3://your-bucket-name/path-to-your-file.out --recursive --
    endpoint-url https://custom-s3-endpoint.com --profile default
```

8.4   Cloudpathlib

- When you use the cloudpathlib module, and you want to specify a profile, do this:

```python
from cloudpathlib import S3Path, S3Client

# Create an S3 client with a specific AWS profile
s3_client = S3Client(
 aws_access_key_id=aws_access_key_id,
 aws_secret_access_key=aws_secret_access_key,
 endpoint_url=endpoint_url,
 profile_name="profile1", # specify profile here
)

# Make `client` default:
client.set_as_default_client()
```

9 Git

9.1 Cloning

- Cloning a repository via *sshfs* and adjusting the folder name,

```
1   git clone <ssh_url> <folder_name>
2   # git clone <https_path> <folder_name>
```

- After moving into the <folder_name> via cd, do

```
1   git switch <desired_branch>
```
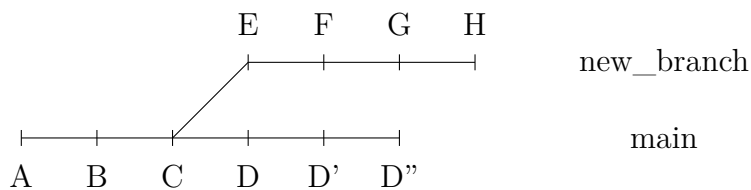
9.2 Remote Branch

Creating a local branch and then pushing it remotely,

```
1   git push origin <local-branch>
2   # git push origin code-dev
```

9.3 Merging

- Assuming the following commit-history is given,

when merging *main* and *new_branch*, the two branches will be either auto-merged or there will be a merge conflict.

- Assuming the following commit history is given,
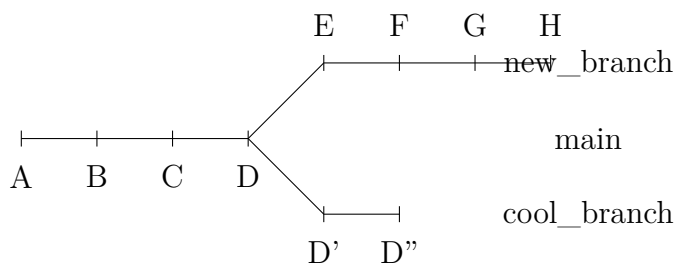
- After merging *main* with *new_feature*, which happens via a fast-forward, you can delete it. New commit structure:

- A – B – C – D – E – F – G – H branch <main>

- Merging *new_feature* into *master*, do

```
1   git switch main
2   git merge <new_feature>
```

9.4 Merge Conflicts

To resolve a merge conflict, type

```
git mergetool
```

Afterwards, confirm with Enter that you want to use vimdiff as default editing tool. The vimdiff display will now resemble the following structure:

```
| LOCAL | BASE | REMOTE |
  MERGED
```

If the file has not already existed in BASE, then we need this view:

```
| LOCAL | MERGED | REMOTE |
LOCAL -- Current branch
BASE -- Common ancestor (how did the file look like before both changes?)
REMOTE -- File that I am merging into the current branch
MERGED -- Merge result
```

It is probably easiest to take the merged view and edit it directly. In the vim editor, an entire line can be deleted by pressing D (no control before!). If I instead wanted the changes from either LOCAL, BASE or REMOTE, you have to do one of these,

```
:diffg LO
:diffg BA
:diffg LO
```

Of course, the merged view can also be edited directly. Regardless of the chosen method, type

```
:wqa
```

into vim. Afterwards, do not forget to commit and push. And if you want, do

```
git clean -f
```

Locally restore file

```
git restore --source <commit SHA> file
git restore --source HEAD file
```

git restore does not overwrite *HEAD*, though. For that, a push would be necessary.

It can sometimes also be useful to run dry merges to proactively check for conflicts during a merge; for this, run

```
git merge --no-commit --no-ff <branch_name>
```

Afterwards, type

```
git merge --abort
```

to abort, and no changes will occur.
If you only want to do a fast-forward, do

```
git merge <branch_to_be_merged> --ff-only
```

9.5 Checking History

- Viewing the history of commits,

```
git log
```

- Viewing a specific file,

```
git show <commit-hash>:<file-name>
# git show 123abc:example.txt
```

9.6 Removing a File/Folder

- To remove a file/folder that is already tracked, adding it to .gitignore won't remove it (though this also needs to happen). For this, do:

```
git rm --cached <file>
git rm -r --cached <folder>
```

- Adding the file/folder to *.gitignore* is still a good idea, though, since the file/dir won't be removed locally with the commands.

9.7 Renaming a Repository

1. Rename the repository remotely first (by going to your repository's URL),

2. then go to the locally cloned version of the repository and do

```
git remote set-url origin <new-url>
# git remote set-url origin https://github.com/username/new-repo-name.git
```

3. and finally

```
git remote -v
```

which lists the remote names and their URLs.

9.8 Restore File

Resetting specific file to state of previous commit,

```
git restore --source=<commit-hash> <file-path>
# git restore --source=HEAD <file_path>
```

Replace ⟨commit-hash⟩ with the commit SHA and ⟨file-path⟩ with the path to the file. This change is local and you would need to commit it if you want it to be reflected in the repository history.

9.9 Delete Branches

- Local git branch,

```
git branch -d <branch_name>
# git branch -d testing
```

- Regardless of merge status,

```
git branch -D <branch_name>
# git branch -D testing
```

- Remote branch,

```
git push origin --delete <branch_name>
# git push origin --delete testing
```

24

9.10 List Branches

Listing all local and remote branches,

```
git branch -a
```

9.11 Remote URL

Obtaining the remote URL,

```
git remote get-url origin
git remote get-url origin | sed 's/\.git$//' # optional: trim output
```

10 Remote Development

10.1 Connection

1. When connecting two machines remotely, install this extension on local machine (also directly in VSCode possible),

2. open VSCode on local machine,

3. press F1-button, choose "Remote-SSH: Connect to Host..." and type for the SSH host (optionally save it in the SSH config file) the same as in Algo. (B),

4. enter the passwd for the remote SSH host.

10.2 Troubleshooting

- If you find you are getting a permission error for saving a file on the remote machine (in VSCode when doing the local coding), try

```
1    sudo chown custom-username path/to/custom/script.ext
```

The *custom-username* here refers to the username on the remote machine.

11 Jax

Try to install via pip first. Only if this doesn't work use conda!

- To put a Jax array onto a specific device, use this:

```
import jax
from jax import devices, device_put, numpy as jnp

x = device_put(jnp.arange(10), device=devices("cpu")[0]) # NOTE: put `[0]`
# x = device_put(jnp.arange(10), device=devices("gpu")[0]) # NOTE: put
    `[0]`
print(f"Device: {x.device_buffer.device()}")
```

Listing 6: Device specification in Jax

- Specifying the dtype of an array:

```
x = jnp.array([1, 2, 3], dtype=jnp.float32)
print(f"Dtype: {x.dtype}")
```

Listing 7: Jax device retrieval

- To find out the device of a Jax array, use this:

```
x.device_buffer.device() # x: Jay array
```

Listing 8: Jax device retrieval

- To make a Jax array out of a Python list or a NumPy array (do not use for tensors):

```
from jax import numpy as jnp

a = jnp.array([1., 2., 3.])
b = jnp.array(np.array([1., 2., 3.]))
```

Listing 9: Jax array creation

- jit (just-in-time compilation): sets up a function with XLA (extended linear algebra): check out the NB *test__jit-compil.ipynb.* To use jit, do this:

```
import jax
from jax import numpy as jnp

@jax.jit
def selu(x: jnp.array, lamb: float = 1., alpha: float = 0.):
 return lamb * jnp.where(x > 0, x, alpha * (jnp.exp(x) - 1.0))
```

Listing 10: Jax array

A .bashrc

```
1   ca() {
2    local conda_out="$(conda env list | grep -E "$env_name" | head -n 1 | awk '{
        print $1}')"
3
4    # check non-emptiness
5    if [ -z "$1" ]; then
6     echo "Usage: ca <env_name>"
7     return 1
8    fi
9
10   # check env existence
11   if [ ! -z "$conda_out" ]; then
12    conda activate "$1"
13   else
14    echo "Conda environment '$env_name' does not exist." # single quotes (')
        only for display
15    return 1
16   fi
17
18  }
19
20  # ------------------------- CONDA -------------------------
21
22  # activate conda environment
23  # usage: `ca custom-env-name`
24  ca() {
25   conda activate "$@"
26  }
27
28  # deactivate currently activated conda environment
29  cod() {
30   conda deactivate
31  }
32
33  # List all available conda envs:
34  cel() {
35   conda env list
36  }
37
38  # remove conda environment
39  # usage: `crme ant-migrate-dev`
40  crme() {
41
42   # check number of passed arguments via `$#`
43   if [[ $# -ne 1 ]]; then
44    echo "NOTE: Exactly one argument needs to be provided"
45   else
46    conda deactivate && conda remove -n "$1" --all -y
47   fi
48
```

```
49    }

51    # alias for `conda__remove_packages`
52    # usage (e.g.): `crm myenv pkg1 pkg2`
53    crm() {
54     conda__remove_packages "$@"
55    }

57    # remove conda packages from environment
58    # usage (e.g.): `conda__remove_packages myenv pkg1 pkg2`
59    conda__remove_packages() {

61     # define local variables first
62     local env_name="$1"
63     local conda_out="$(conda env list | grep -E "$env_name" | head -n 1 | awk '{
          print $1}')"

65     # forget first argument (which is saved in `env_name`)
66     shift

68     # check non-emptiness
69     if [ -z "$env_name" ]; then
70      echo "Usage: conda__remove_packages <env_name> [package1] [package2] ... [
          packageN]"
71      return 1
72     fi

74     # check env existence
75     if [ ! -z "$conda_out" ]; then
76      conda remove -n "$env_name" "$@" -y
77      echo "Package(s) '$@' removed from environment '$env_name'"
78     else
79      echo "Conda environment '$env_name' does not exist." # single quotes (')
          only for display
80      return 1
81     fi

83    }

85    # --------------------------- GIT ---------------------------

87    # list all local and remote branches
88    lb() {
89     git branch -a
90    }

92    # delete remote branch
93    lbd() {
94     git push origin --delete "$@"
95    }

97    # switch branches and create if non-existent
```

```bash
lsw() {
 if git rev-parse --verify "$1" >/dev/null 2>&1; then
   git switch "$1"
 else
   git switch -c "$1"
 fi
}

# example usage: `lsta 2` or `lsta`
lsta() {
 local stash_index=${1:-0} # Default to 0 if no argument provided

 # Check if the provided argument is an integer
 if ! [[ $stash_index =~ ^[0-9]+$ ]]; then
  echo "The provided index is not a valid integer."
  return 1
 fi

 # Check if the stash index exists
 if ! git rev-parse --verify stash@{$stash_index} >/dev/null 2>&1; then
  echo "No stash found at index $stash_index"
  return 1
 fi

 # If all checks pass, apply the stash
 git stash apply "stash@{$stash_index}" --index
}

# Forward commands to `git stash`
lst() {
 git stash "$@"
}

# Stash files, if arguments are provided, they are ignored
lstf() {
 git stash --include-untracked
}

# https://stackoverflow.com/questions/19595067/git-add-commit-and-push-
    commands-in-one
# https://stackoverflow.com/questions/14763608/use-conditional-in-bash-script
    -to-check-string-argument
# if-else statements in bash: https://linuxhandbook.com/if-else-bash/
# example usage: lgit "bit" "add ..."
lpush() {
 git add . && git commit -a -m "$1" && git push origin $(bname) && llog
}

# https://stackoverflow.com/questions/3236871/how-to-return-a-string-value-
    from-a-bash-function
bname() {
 branch=$(git branch --show-current)
```

```
    echo $branch
 }

lupd() {
 git fetch origin $(bname) && git log HEAD..origin/$(bname) --oneline
}

lpull() {
 git pull origin $(bname)
}

ldiff() {
 git status "$@" && git diff --color "$@"
}

lforce() {
 git push origin $(bname) --force
}

llog() {
 git log
}

lrm() {
 git rm -r "$@"
}

lreb() {
 # Set default value to 5:
 num1=${1:-5}
 git rebase -i HEAD~$num1
}

# Reset entire repo to state of `HEAD`, or reset specific file to a specific
    commit hash.
lres() {
 if [[ $# -eq 0 ]] || [[ $# -eq 1 ]]; then
  local commit_hash=${1:-HEAD}
  git reset --hard "$commit_hash"
 elif [ $# -eq 2 ]; then
  local commit_hash="$1"
  local file_path="$2"
  git restore --source="$commit_hash" "$file_path"
 else
  echo "Usage: lres [commit_hash file_path]"
 fi
}

lsh(){
 git show "$@"
}
```

```bash
# Usage: lmv /path/to/directory file1 file2 file3 ...
lmv() {
 local target_dir=$1 # The first argument is the directory path
 if [[ ! -d "$target_dir" ]]; then
  echo "Target directory does not exist: $target_dir" >&2
  return 1
 fi

 # Shift the arguments so that $@ contains only the files to move
 shift

 # Now, loop through all the remaining arguments
 for file in "$@"; do
  if [[ -e $file ]]; then
   git mv "$file" "$target_dir"
  else
   echo "File does not exist: $file" >&2
  fi
 done
}

# ------------------------- PROTONVPN ---------------------

p() {
 protonvpn-cli "$@"
}

# ------------------------- MISCELLANEOUS ------------------

# pdflatex
pd() {
 /usr/bin/pdflatex "$@"
}

# convert input notebook to PDF
jconv() {
 jupyter nbconvert --to pdf "$1"
}

# `less` with ANSI escape characters
less() {
 /usr/bin/less -R "$@"
}

diff() {
 /usr/bin/diff --color "$@"
}

# overload `shred` func, allow (recursive) shredding of dirs/files
# multiple files/dirs can be provided, mixing allowed
# usage (e.g.): `shred 10 <file_name>`
# shred <file_name>
```

```bash
250    # shred <dir_path>
251    # shred <file_name> <dir_path>
252    shred() {
253
254     # check whether first argument is a number
255     if [[ "$1" =~ ^[0-9]+$ ]]; then
256      local iterations="$1"
257      shift
258     else
259      iterations=5 # default
260     fi
261
262     # check file/dir existences
263     for path in "$@"; do
264      if check_existence "$path"; then
265      # check whether passed input is directory or not
266       if [[ -d "$path" ]]; then
267        echo "Files to be shredded in $path:"
268        find "$path" -type f -print0 | xargs -0 ls -ld
269       fi
270      else
271       echo "Error occurred in check_existence for file/dir: $path"
272       return 1
273      fi
274     done
275
276     # prompt user for confirmation
277     read -rp "Do you wish to proceed with shredding all files in $@ for
            $iterations iterations? (yes/no): " confirmation
278
279     if [[ $confirmation = [yY] || $confirmation = [yY][eE][sS] ]]; then
280      for path in "$@"; do
281       if [[ -d "$path" ]]; then
282        # shred all files within the directory
283        find "$path" -type f -exec /usr/bin/shred -uz -n "$iterations" {} \;
284        rm -rf "$path"
285        echo "All files in '$path' have been shredded for $iterations iterations."
286       elif [[ -f "$path" ]]; then
287        # shred the individual file
288        /usr/bin/shred -uz -n "$iterations" "$path"
289        echo "File '$path' has been shredded for $iterations iterations."
290       fi
291      done
292     else
293      echo "Shredding aborted."
294     fi
295    }
296
297    # shortcut for clearing terminal output
298    c() {
299     clear
300    }
```

```bash
# shortcuts for exiting terminal
q() {
 exit
}

e() {
 q
}

# tailscale
ts() {
 tailscale status "$@"
}

# xournalpp (https://github.com/xournalpp/xournalpp)
xopp() {
 xournalpp "$@"
}

# strings comparison
# usage (e.g.): `str_diff "blub1" "blub1"` or `str_diff blub1 blub1`
# or `str_diff $(echo "hey") $(echo "hey")`
# NOTE: exactly two arguments need to be provided
str_diff() {

 # check number of passed arguments via `$#`
 if [[ $# -ne 2 ]]; then
  echo "NOTE: Exactly two arguments need to be provided"
  return 1 # return non-zero exit code to indicate error
 else

  # compare strings
  if [[ $1 == $2 ]]; then
   echo -e "Strings '$1' and '$2' \033[92mmatch\033[0m"
  else
   echo -e "Strings '$1' and '$2' do \033[91mNOT\033[0m match"
  fi
 fi

}



# ------------------------- DOCKER ---------------------
d() {
 docker "$@"
}

# ------------------------- CHATGPT ---------------------

# https://github.com/kardolus/chatgpt-cli/tree/main
gpt(){
```

34

```
353    chatgpt -i
354  }
355
356  export OPENAI_KEY=[...]
357
358  # -------------------------- ALWAYS EXECUTE ----------------
359
360  add_bit
```

Listing 11: Contents of .bashrc file

B Amazing Programs, Extensions, Plugins & Packages

- **https://github.com/charmbracelet/glow**

- **https://github.com/0xacx/chatGPT-shell-cli**

- **https://github.com/kardolus/chatgpt-cli/tree/main**

    - For setting the right model (cf. here for all available models),

    ```
    chatgpt --set-model gpt-4-1106-preview --set-max-tokens 128000
    ```

    - Usage:

    ```
    chatgpt -i
    ```

- **https://tailscale.com/download/**

    - Once installation is complete, the command

    ```
    sudo tailscale up
    ```

    should be run to login, though this command will also display after installation in the CLI. The signing in should happen via GitHub. To be able to use Tailscale from a new device, it must be added as a device under `https://login.tailscale.com/admin/machines`. Once this is done, open a CLI and type

    ```
    ssh name@ip_address # find out <name> and <ip_address> via tailscale
        console
    # ssh ellie@100.xx.xxx.xx
    ```

    NOTE that if the file already exists locally, it will be overwritten.

    - For file copying (e.g. from the host machine to the currently used machine), do this

    ```
    scp name@ip_address:/path/to/remote_file.ext /local/path # find out <
        name> and <ip_address> via tailscale console
    # ssh ellie@100.xx.xxx.xx
    ```

    For directory copying,

    ```
    scp -r name@ip_address:/path/to/remote_dir /local/path # find out <
        name> and <ip_address> via tailscale console
    # ssh ellie@100.xx.xxx.xx
    ```

- `https://tailscale.com/kb/1080/cli/` (no separate installation necessary, only tailscale needs to be installed)

    - Finding out the IPv4 address of the currently active machine,

    ```
    tailscale ip -4
    ```

    - Finding out the IPv4 address of another machine connected via the Tailscale network,

    ```
    tailscale ip -4 custom-name
    # tailscale ip -4 ellie
    ```

- `https://github.com/aws/aws-cli`

- `https://github.com/termcolor/termcolor`

C VSCode

C.1 Recommended Extensions

- `https://marketplace.visualstudio.com/items?itemName=ms-vscode-remote.vscode-remote-ex`

- `https://marketplace.visualstudio.com/items?itemName=Gruntfuggly.todo-tree`

C.2 Open the *settings.json* File

1. press Ctrl + Shift + P to the Command Palette,

2. type **Open User Settings (JSON)** and select it to open the **settings.json** file.

C.3 Fix Unresolved Python Imports

- If you run a docker container where a conda environment is installed (with packages that you do not have locally), then VSCode will show those imports as unresolved. To fix this, open the **settings.json**, cf. App. C.2, and add the following setting:

```
"python.analysis.diagnosticSeverityOverrides": {
    "reportMissingImports": "none"
}
```

  How to incorporate this into the **settings.json** file is shown in App. C.5.

- Note that if you have an SSH connection to another machine going on (e.g. in the Remote Development extension), then putting the above lines into the *settings.json* file will not have an immediate effect, for this the SSH connection needs to be restarted.

C.4 Opening a Duplicate Workspace

1. press Ctrl + Shift + P to the Command Palette,

2. then type **Workspaces: Duplicate As Workspace is New Window**

C.5 *settings.json*

```
{
  "workbench.colorTheme": "Default Dark Modern",
  "telemetry.telemetryLevel": "off",
  "editor.wordWrap": "wordWrapColumn",
  "editor.wordWrapColumn": 79,
  "workbench.editor.enablePreview": false,
  "gitlens.telemetry.enabled": false,
  "notebook.lineNumbers": "on",
  "explorer.confirmDragAndDrop": false,
  "window.zoomLevel": 1,
  "python.analysis.diagnosticSeverityOverrides": {
      "reportMissingImports": "none"
  },
  "todo-tree.general.tags": [
    "BUG",
    "HACK",
```

```
17        "FIXME",
18        "TODO",
19        "NOTE",
20        "XXX",
21        "[ ]",
22        "[x]"
23      ],
24
25    "files.associations": {"*.log": "plaintext"},
26    "[plaintext]": {"editor.wordWrap": "off"},
27    "workbench.editor.tabSizing": "shrink"
28    }
```

Listing 12: Contents of settings.json file

D   LibreOffice

D.1   Dark Theme

Go to *Tools → Options → LibreOffice → Application Colors → Custom Colors → General →
Document Background* and choose a dark color.

E UNIGE-specific

- Setting up UNIGE e-mail in TB:
  `https://plone.unige.ch/distic/pub/messagerie/configuration/comment-configurer-compte-`