

Infra

Imahn Shekhzadeh
`imahn.shekhzadeh@unige.ch`

December 13, 2023

Contents

1	Baobab/Yggdrasil	3
2	Bash	4
3	Linux	6
3.1	Opening Programs from CLI	6
4	Anaconda	8
4.1	Installation of Environments	8
4.2	Export	8
4.3	Installation & Removal of Packages	8
4.4	Usage in VSCode	8
4.5	PyTorch	9
5	CUDA	10
6	Docker	11
6.1	Installation	11
6.2	Basics	11
6.3	Dockerfile	11
6.4	Docker images	12
6.5	Docker containers	13
6.5.1	Basics	13
6.5.2	Passing Arguments	14
6.5.3	Listing & Stopping	15
6.6	Pushing to DockerHub & HPC	15
7	Python	16
7.1	Config File & JSON Files	16
7.2	Jupyter Notebooks	16
7.3	Map (Built-In Function)	16
7.4	PyTorch	17
7.4.1	Leaf Tensors	17
7.4.2	Autograd & Backward	17

8	AWS S3	19
8.1	Installation & Configuration	19
8.2	AWS Credentials (Profiles)	19
8.3	Buckets	20
8.3.1	Creation	20
8.3.2	Listings	20
8.3.3	File Copying	20
8.3.4	Directory Copying	20
8.3.5	Folder/File Deletion	21
8.4	Cloudpathlib	21
9	Git	22
9.1	Cloning	22
9.2	Remote Branch	22
9.3	Merging	22
9.4	Merge Conflicts	23
9.5	Checking History	23
9.6	Removing a File/Folder	24
9.7	Renaming a Repository	24
9.8	Restore File	24
9.9	Delete Branches	24
9.10	List Branches	25
10	Remote Development	26
10.1	Connection	26
10.2	Troubleshooting	26
11	Jax	27
A	<i>.bashrc</i>	28
B	Amazing Programs, Extensions, Plugins & Packages	36
C	VSCoDe	37
C.1	Recommended Extensions	37
C.2	Open the <i>settings.json</i> File	37
C.3	Fix Unresolved Python Imports	37
C.4	Opening a Duplicate Workspace	37
C.5	<i>settings.json</i>	37
D	LibreOffice	39
D.1	Dark Theme	39
E	UNIGE-specific	40

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:

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

- Scp into yggdrasil:

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

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

```
1 squeue
2 squeue -p cms-uhh # partition
3 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)

```
1 conda info
```

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

## 2 Bash

- Downloading file from URL and allowing for redirects,

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

- For this directory structure,

```
1 infra_upd.tex
2 infra_upd.log
3 infra_upd.aux
4 infra_upd.out
5 infra_upd.pdf
```

rename *all* of them via

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

What happens is called a [substring replacement](#).

- Appending line to file,

```
1 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):

```
1 test_sth() {
2 local env_name="$1" # bash starts counting indices from 1
3
4 if [-z "$env_name"]; then # spacing after `[` and before `]` needed
5 echo "The string is empty."
6 return 1 # return value of 1 indicates error
7 fi
8 }
```

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

- For retrieving all but the first argument,

```
1 test_sth(){
2 shift
3
4 echo "all provided args (except the first): $@"
5 }
```

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

```
1 test_sth(){
2 shift
3 ...
4 shift
5
6 echo "all provided args (except the first N): $@"
7 }
```

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,

```
1 dpkg --get-architecture | grep linux-image
```

Listing 3: Find kernel versions in Ubuntu

Finding the currently *active* kernel version,

```
1 uname -a
```

Listing 4: Current kernel 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,

```
1 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,

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

- Seeing the resource consumption,

```
1 htop
```

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

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

### 3.1 Opening Programs from CLI

- Opening the settings from CLI,

```
1 gnome-control-center
```

- Opening VSCode from CLI:

```
1 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,

```
1 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!**):

```
1 import os
2
3 import torch
4
5 if __name__ == "__main__":
6 os.path.dirname(torch.__file__)
```

Afterwards, do

```
1 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!**):

```
1 conda list | grep -E "torch|pytorch"
2 # 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 kernel version, at some point, the newer kernel version was picked up and worked 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_shots 1 --k_shots 1 # example
```

### 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:

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

If you have several notebooks in the same directory,

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

### 7.3 Map (Built-In Function)

- Function signature:

```
1 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,

```
1 from typing import List
2
3 def multiply(x: List, y: List):
4 return x * y
5
6 list_one = [i for i in range(1000)]
7 list_two = [j for j in range(1000, 2000)]
8 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,

```
1 a = np.array([1, 2, 3, 4])
2 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`.

### 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,

```
1 x = torch.tensor([1., 2., 3.], requires_grad=True, device=torch.device("
2 cuda:0"))
3
4 # fwd pass
5 y = 2 * x
6
7 # first backward pass
8 y.sum().backward(retain_graph=True)
9 print(f"Gradients of `x` after first backward pass: {x.grad}") # `torch.
10 tensor([2., 2., 2.])`
11
12 # second backward pass
13 y.sum().backward()
14 print(f"Gradients of `x` after second backward pass: {x.grad}") # `torch.
15 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,

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

- Deleting a file,

```
1 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:

```
1 from cloudpathlib import S3Path, S3Client
2
3 # Create an S3 client with a specific AWS profile
4 s3_client = S3Client(
5 aws_access_key_id=aws_access_key_id,
6 aws_secret_access_key=aws_secret_access_key,
7 endpoint_url=endpoint_url,
8 profile_name="profile1", # specify profile here
9)
10
11 # Make `client` default:
12 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>
```

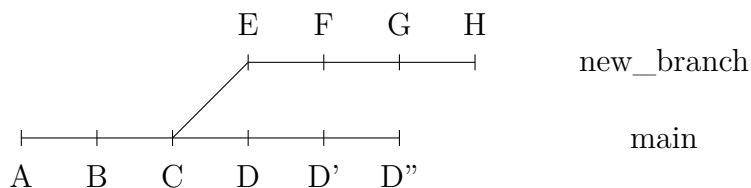
### 9.2 Remote Branch

Creating a local branch and then pushing it remotely,

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

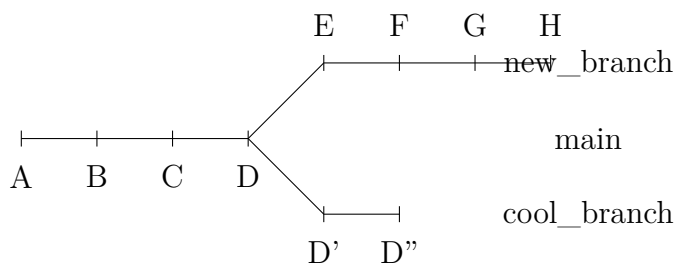
### 9.3 Merging

- 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

```
1 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:

```
1 | LOCAL | BASE | REMOTE |
2 | MERGED
```

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

```
1 | LOCAL | MERGED | REMOTE |
2 LOCAL -- Current branch
3 BASE -- Common ancestor (how did the file look like before both changes?)
4 REMOTE -- File that I am merging into the current branch
5 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,

```
1 :diffg L0
2 :diffg BA
3 :diffg L0
```

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

```
1 :wqa
```

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

```
1 git clean -f
```

Locally restore file

```
1 git restore --source <commit SHA> file
2 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

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

Afterwards, type

```
1 git merge --abort
```

to abort, and no changes will occur.

If you only want to do a fast-forward, do

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

## 9.5 Checking History

- Viewing the history of commits,

```
1 git log
```

- Viewing a specific file,

```
1 git show <commit-hash>:<file-name>
2 # 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:

```
1 git rm --cached <file>
2 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

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

3. and finally

```
1 git remote -v
```

which lists the remote names and their URLs.

## 9.8 Restore File

Resetting specific file to state of previous commit,

```
1 git restore --source=<commit-hash> <file-path>
2 # 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,

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

- Regardless of merge status,

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

- Remote branch,

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



## 9.10 List Branches

Listing all local and remote branches,

```
1 git branch -a
```

## 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:

```
1 import jax
2 from jax import devices, device_put, numpy as jnp
3
4 x = device_put(jnp.arange(10), device=devices("cpu")[0]) # NOTE: put `[0]`
5 # x = device_put(jnp.arange(10), device=devices("gpu")[0]) # NOTE: put
6 `[0]`
7 print(f"Device: {x.device_buffer.device()}")
```

Listing 6: Device specification in Jax

- Specifying the dtype of an array:

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

Listing 7: Jax device retrieval

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

```
1 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):

```
1 from jax import numpy as jnp
2
3 a = jnp.array([1., 2., 3.])
4 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:

```
1 import jax
2 from jax import numpy as jnp
3
4 @jax.jit
5 def selu(x: jnp.array, lamb: float = 1., alpha: float = 0.):
6 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 '{
3 print $1}')"
4
5 # check non-emptiness
6 if [-z "$1"]; then
7 echo "Usage: ca <env_name>"
8 return 1
9 fi
10
11 # check env existence
12 if [! -z "$conda_out"]; then
13 conda activate "$1"
14 else
15 echo "Conda environment '$env_name' does not exist." # single quotes (')
16 only for display
17 return 1
18 fi
19 }
20
21 # ----- CONDA -----
22
23 # activate conda environment
24 # usage: `ca custom-env-name`
25 ca() {
26 conda activate "$@"
27 }
28
29 # deactivate currently activated conda environment
30 cod() {
31 conda deactivate
32 }
33
34 # List all available conda envs:
35 cel() {
36 conda env list
37 }
38
39 # remove conda environment
40 # usage: `crme ant-migrate-dev`
41 crme() {
42
43 # check number of passed arguments via ` $# `
44 if [[$# -ne 1]]; then
45 echo "NOTE: Exactly one argument needs to be provided"
46 else
47 conda deactivate && conda remove -n "$1" --all -y
48 fi
49 }
```

```

49 }
50
51 # alias for `conda__remove_packages`
52 # usage (e.g.): `crm myenv pkg1 pkg2`
53 crm() {
54 conda__remove_packages "$@"
55 }
56
57 # remove conda packages from environment
58 # usage (e.g.): `conda__remove_packages myenv pkg1 pkg2`
59 conda__remove_packages() {
60
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 '{
64 print $1}')"
65
66 # forget first argument (which is saved in `env_name`)
67 shift
68
69 # check non-emptiness
70 if [-z "$env_name"]; then
71 echo "Usage: conda__remove_packages <env_name> [package1] [package2] ... [
72 packageN]"
73 return 1
74 fi
75
76 # check env existence
77 if [! -z "$conda_out"]; then
78 conda remove -n "$env_name" "$@" -y
79 echo "Package(s) '$@" removed from environment '$env_name'"
80 else
81 echo "Conda environment '$env_name' does not exist." # single quotes (')
82 only for display
83 return 1
84 fi
85 }
86
87 # ----- GIT -----
88
89 # list all local and remote branches
90 lb() {
91 git branch -a
92 }
93
94 # delete remote branch
95 lbd() {
96 git push origin --delete "$@"
97 }
98
99 # switch branches and create if non-existent

```

```

98 lsw() {
99 if git rev-parse --verify "$1" >/dev/null 2>&1; then
100 git switch "$1"
101 else
102 git switch -c "$1"
103 fi
104 }
105
106 # example usage: `lsta 2` or `lsta`
107 lsta() {
108 local stash_index=${1:-0} # Default to 0 if no argument provided
109
110 # Check if the provided argument is an integer
111 if ! [[$stash_index =~ ^[0-9]+$]]; then
112 echo "The provided index is not a valid integer."
113 return 1
114 fi
115
116 # Check if the stash index exists
117 if ! git rev-parse --verify stash@{$stash_index} >/dev/null 2>&1; then
118 echo "No stash found at index $stash_index"
119 return 1
120 fi
121
122 # If all checks pass, apply the stash
123 git stash apply "stash@{$stash_index}" --index
124 }
125
126 # Forward commands to `git stash`
127 lst() {
128 git stash "$@"
129 }
130
131 # Stash files, if arguments are provided, they are ignored
132 lstf() {
133 git stash --include-untracked
134 }
135
136 # https://stackoverflow.com/questions/19595067/git-add-commit-and-push-commands-in-one
137 # https://stackoverflow.com/questions/14763608/use-conditional-in-bash-script-to-check-string-argument
138 # if-else statements in bash: https://linuxhandbook.com/if-else-bash/
139 # example usage: lgit "bit" "add ..."
140 lpush() {
141 git add . && git commit -a -m "$1" && git push origin $(bname) && llog
142 }
143
144 # https://stackoverflow.com/questions/3236871/how-to-return-a-string-value-from-a-bash-function
145 bname() {
146 branch=$(git branch --show-current)

```

```

147 echo $branch
148 }
149
150 lupd() {
151 git fetch origin $(bname) && git log HEAD..origin/$(bname) --oneline
152 }
153
154 lpull() {
155 git pull origin $(bname)
156 }
157
158 ldiff() {
159 git status "$@" && git diff --color "$@"
160 }
161
162 lforce() {
163 git push origin $(bname) --force
164 }
165
166 llog() {
167 git log
168 }
169
170 lrm() {
171 git rm -r "$@"
172 }
173
174 lreb() {
175 # Set default value to 5:
176 num1=${1:-5}
177 git rebase -i HEAD~$num1
178 }
179
180 # Reset entire repo to state of `HEAD`, or reset specific file to a specific
181 # commit hash.
182 lres() {
183 if [[$# -eq 0]] || [[$# -eq 1]]; then
184 local commit_hash=${1:-HEAD}
185 git reset --hard "$commit_hash"
186 elif [$# -eq 2]; then
187 local commit_hash="$1"
188 local file_path="$2"
189 git restore --source="$commit_hash" "$file_path"
190 else
191 echo "Usage: lres [commit_hash file_path]"
192 fi
193 }
194
195 lsh(){
196 git show "$@"
197 }

```

```

198 # Usage: lmv /path/to/directory file1 file2 file3 ...
199 lmv() {
200 local target_dir=$1 # The first argument is the directory path
201 if [[! -d "$target_dir"]]; then
202 echo "Target directory does not exist: $target_dir" >&2
203 return 1
204 fi
205
206 # Shift the arguments so that $@ contains only the files to move
207 shift
208
209 # Now, loop through all the remaining arguments
210 for file in "$@"; do
211 if [[-e $file]]; then
212 git mv "$file" "$target_dir"
213 else
214 echo "File does not exist: $file" >&2
215 fi
216 done
217 }
218
219 # ----- PROTONVPN -----
220
221 p() {
222 protonvpn-cli "$@"
223 }
224
225 # ----- MISCELLANEOUS -----
226
227 # pdflatex
228 pd() {
229 /usr/bin/pdflatex "$@"
230 }
231
232 # convert input notebook to PDF
233 jconv() {
234 jupyter nbconvert --to pdf "$1"
235 }
236
237 # `less` with ANSI escape characters
238 less() {
239 /usr/bin/less -R "$@"
240 }
241
242 diff() {
243 /usr/bin/diff --color "$@"
244 }
245
246 # overload `shred` func, allow (recursive) shredding of dirs/files
247 # multiple files/dirs can be provided, mixing allowed
248 # usage (e.g.): `shred 10 <file_name>`
249 # shred <file_name>

```



```

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 }

```

```

301
302 # shortcuts for exiting terminal
303 q() {
304 exit
305 }
306
307 e() {
308 q
309 }
310
311 # tailscale
312 ts() {
313 tailscale status "$@"
314 }
315
316 # xournalpp (https://github.com/xournalpp/xournalpp)
317 xopp() {
318 xournalpp "$@"
319 }
320
321 # strings comparison
322 # usage (e.g.): `str_diff "blub1" "blub1"` or `str_diff blub1 blub1`
323 # or `str_diff $(echo "hey") $(echo "hey")`
324 # NOTE: exactly two arguments need to be provided
325 str_diff() {
326
327 # check number of passed arguments via ` $# `
328 if [[$# -ne 2]]; then
329 echo "NOTE: Exactly two arguments need to be provided"
330 return 1 # return non-zero exit code to indicate error
331 else
332
333 # compare strings
334 if [[$1 == $2]]; then
335 echo -e "Strings '$1' and '$2' \033[92mmatch\033[0m"
336 else
337 echo -e "Strings '$1' and '$2' do \033[91mNOT\033[0m match"
338 fi
339 fi
340 }
341
342
343 # ----- DOCKER -----
344 d() {
345 docker "$@"
346 }
347
348 # ----- CHATGPT -----
349
350 # https://github.com/kardolus/chatgpt-cli/tree/main
351 gpt(){

```

```
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),

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

- Usage:

```
1 chatgpt -i
```

- <https://tailscale.com/download/>

- Once installation is complete, the command

```
1 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

```
1 ssh name@ip_address # find out <name> and <ip_address> via tailscale
 console
2 # 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

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

For directory copying,

```
1 scp -r name@ip_address:/path/to/remote_dir /local/path # find out <
 name> and <ip_address> via tailscale console
2 # 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,

```
1 tailscale ip -4
```

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

```
1 tailscale ip -4 custom-name
2 # 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:

```
1 "python.analysis.diagnosticSeverityOverrides": {
2 "reportMissingImports": "none"
3 }
```

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 in New Window**

### C.5 *settings.json*

```
1 {
2 "workbench.colorTheme": "Default Dark Modern",
3 "telemetry.telemetryLevel": "off",
4 "editor.wordWrap": "wordWrapColumn",
5 "editor.wordWrapColumn": 79,
6 "workbench.editor.enablePreview": false,
7 "gitlens.telemetry.enabled": false,
8 "notebook.lineNumbers": "on",
9 "explorer.confirmDragAndDrop": false,
10 "window.zoomLevel": 1,
11 "python.analysis.diagnosticSeverityOverrides": {
12 "reportMissingImports": "none"
13 },
14 "todo-tree.general.tags": [
15 "BUG",
16 "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->