183.663 Deep Learning for Visual Computing

Table of Contents

Versions

Version	Date	Changes
V1.0	2021-03-30	Initial Release
V1.1	2022-04-15	2022 Update
V1.2	2023-04-24	2023 Update
V1.3	2024-04-02	2024 Update add link to shell command tutorial explicitly mention cifar dataset expand on differences between `srun` and `sbatch` over ssh add common mistakes section

Preface

Welcome to the course Deep Learning for Visual Computing (DLVC). This document is aimed to give you an overview over the infrastructure provided to you as well as how to use it.

This is the fourth time this environment exists in this form and it may still need adaptions. Further questions, constructive feedback or wishes for possible improvement are always welcome, please see *Contact*.

General Remarks

For Windows 10 and later, you can use Windows Subsystem for Linux (WSL) to get a convenient linux command line interface. (Clarification Note: it is not necessary to use WSL for this course!)

If not specified otherwise, command examples are meant to be executed on our server when connected over SSH.

We assume you to be able to use at least some basic level of shell commands. If you are new to this, there are a lot of tutorials online (e.g.: this , sections 3 to 6 should suffice) Everything beyond the above mentioned tutorial in regard to shell commands should either be explained sufficiently within this document or is not strictly necessary.

To avoid a large amount of duplicate data, common resources, such as necessary datasets and examples will be made available under the following path: `/caa/Student/dlvc/public/`

This guide itself in its latest version also should be available as `/caa/Student/dlvc/public/DLVC2024Guide.md` in plaintext.

SSH Connection

You can reach our server `dlvc2.cvl.tuwien.ac.at` via `SSH` using the username (e.g: `dlvc00000000`) and password you should have received per e-mail.

On Linux and macOS, ssh connections can be made using the terminal command `ssh`.

On Windows, you can use Putty or WSL.

Copying Files

You can copy files from and to our server `dlvc2.cvl.tuwien.ac.at` using `SFTP`.

On Linux and macOS, this can be done using the `scp` or `rsync` command. Alternatives using a graphical interface are available as well.

On Windows, you can use FileZilla or WinSCP.

SLURM Cluster

We provide you with SSH access to our job submission node `dlvc2.cvl.tuwien.ac.at`. While you can move around and edit files using terminal commands (`cd`,`mv`,`cp`,`nano`,`vim`,...), the submission node itself provides only very limited computational resources. We ask you kindly not to attempt any serious computational tasks on this host directly. (We have seen students use `vscode` remotely on our login nodes, over the last semesters it was okay but in case it causes issues we'll inform accordingly.) For the above mentioned reason, only very limited programs are installed on the submission node. Instead, use the tools described in the following sections to let our SLURM cluster, consisting of two nodes, `edna` and `skinner`, execute your tasks.

Note: If you try to run the provided example script `dlvc.py` directly on our submission node, it will show an error similar to `ModuleNotFoundError: No module named 'torch'`. Since you are not supposed to work directly on the submission node, it is not installed there. The execution nodes have `python3-torch 1.8.1-4` installed.

Overview

Jobs to be run on either `edna` or `skinner` can be submitted either using:

- `srun`
    - runs interactive, shows output directly
    - blocks/waits until your job is scheduled
    - can be aborted using `ctrl+c` in the terminal
    - will cancel if your ssh connection is closed
- `sbatch`
    - runs non-interactive, output is redirected to a file
    - job is executed in background when scheduled
    - can be aborted by using `squeue` and `scancel`
    - will continue running or beeing scheduled even if you log out of your ssh session

To show the current queue, use: `squeue --partition pDLVC`

If you found a mistake in your code or want to cancel your job, use: `scancel <jobid>` (`jobid` is shown when submitting a job per `srun` and `sbatch` or using `squeue`).

Example

To get you started, we provide a minimal ready-to-run example.

After you logged in per SSH, you can copy it to your homedirectory as follows:

```
cp -r /caa/Student/dlvc/public/example ~/
```

Then change into the directory and list files:

```
cd ~/example && ls -lah ~/example
```

View the contents of the example files, e.g. as follows:

```
cat ~/example/dlvc.py
```

```
cat ~/example/dlvc.slurm
```

Of course you may also use editors (`nano`, `vim`, ...) or similar here to view/edit alternatively.

`dlvc.py` is a minimal example that imports `torch` and prints some information. For your course work, you will likely modify or replace this file with your own. `dlvc.slurm` is a wrapper script that sets some options for our scheduling system and then calls `dlvc.py`. For your course work, you most likely won't need to modify it at all, apart from maybe changing the filename of the python script to be executed.

You can now use `srun` to execute the python script:

```
srun --account=dlvc --partition=pDLVC --mem=8192 --gres=gpu:1 dlvc.py
```

You will notice that quite some parameters are necessary and might find it more convenient to use `sbatch` with a configuration file:

```
sbatch --verbose dlvc.slurm
```

Look at the jobqueue using:

```
squeue
```

After your job has finished, there should be a new file ending with `.log`. Look for it and view its contents:

```
ls -lah ~/example/
```

```
cat ~/example/*.log
```

Limits

We apply some limits to the resources provided. The limits and specific values might be adapted over the course of time as necessary.

Course work should be possible well within these defined limits. If you suspect otherwise, please talk to us.

Storage

To avoid unintentional mistakes, unneccessary disk usage and service disruptions for others, the disk usage of the homedirectories is limited.

Limit	Value	Description
Homedirectory	1GB	Your homedirectory "/caa/Student/dlvc/dlvcYYYYYYY"

You can check your current usage as follows:

```
du -sh ~
```

Note: You do not need to copy datasets to your homedirectory. They are available under:

```
/caa/Student/dlvc/public/datasets
```

At the time of the release of this PDF there is at least:

```
/caa/Student/dlvc/public/datasets/cifar10
```

Additional datasets may be added later as necessary, you may check as follows:

```
ls /caa/Student/dlvc/public/datasets
```

SLURM

To provide an as fair as possible experience sharing computational resources, submitting jobs to our cluster is limited as follows.

Limit	Value	Description
MaxTime	1:00:0	maximum time a submitted job is allowed to run before risking to be killed
MaxMemPerNode	40960	maximum amount of RAM (in MB) that can be allocated per node
MaxSubmit	2	maximum amount of jobs a user is allowed to submit to the queue at the same time
MaxTRES	node=1	maximum amount of nodes a user can occupy at any time
MaxTRESPU	gres/gpu=1	maximum amount of gpus a user can occupy at any time

Common Mistakes

"Permission denied"

If you are getting an error that looks similar to this:

```
slurmstepd: error: execve(): train.py: Permission denied
```

Then probably you are trying to let our cluster execute your python script directly, e.g.: `srun ... train.py` or your batch file contains a call to `./train.py`

Most likely, the execute bit in the File Permissions is missing.
You may check using `ls -lah ./train.py`.
The result for a file where this is the case could look like this: `-rw-r--r--`.

The issue can be corrected using `chmod u+x ./train.py` to set the execution bit for the file owner. Afterwards, a check with `ls -lah ./train.py` should show the x bit as follows: `-rwxr--r--`.

Alternatively, you could invoke the python interpreter with your filename as parameter, e.g.: `srun ... python3 ./train.py` or in your batch file `python3 ./train.py`

"Exec format error"

If you are getting an error that looks similar to this:

```
slurmstepd: error: execve(): train.py: Exec format error
```

Then probably you are trying to let our cluster execute your python script directly, e.g.: `srun ... train.py` or your batch file contains a call to `./train.py`

Most likely, your python script file is missing the Shebang.

The issue can be resolved by ensuring that the first line in your file contains the following:

```
#!/bin/python3
```

Disclaimer

Backups

We do our best to keep everything running and available, but can not make guarantees for the availability of your data stored on our servers. Therefore, please be advised to always keep a local copy (see *Copying Files*) of your work. This should protect you from both, unexpected data loss as well as unexpected unavailability of our infrastructure.

Restarts

To allow for the possibility of maintenance, during the following time windows, restarts or service interruption should be expected. We will do our best to keep already submitted jobs running and if somehow possible not disrupt ongoing work even during listed times.

Maintenance Time Slot
Saturday 18:00 - Sunday 6:00
Wednesday 22:00 - Thursday 6:00

Contact

address	use case
dlvc@cvl.tuwien.ac.at	general questions regarding the lecture
dlvc-trouble@cvl.tuwien.ac.at	technical problems, service unavailability
dlvc-feedback@cvl.tuwien.ac.at	feedback in regard to provided infrastructure (for course feedback, please use TISS)