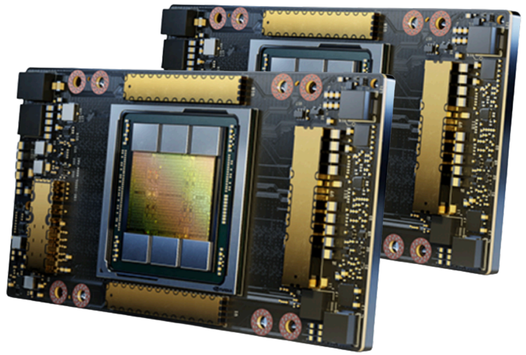
Kisski: Training Platform

Introduction to Deep Learning Workflows on Kisski Infrastructure

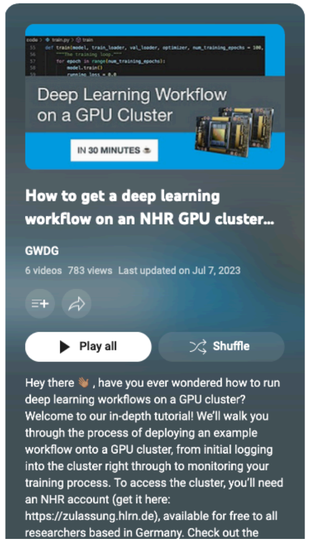
KISSKI
KI-Servicezentrum für sensible
und kritische Infrastrukturen

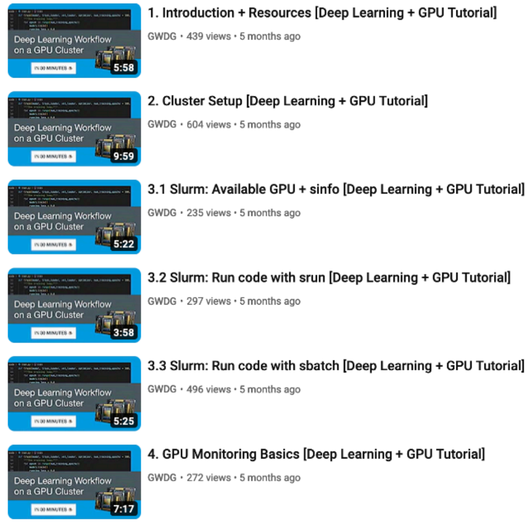


Dorothea Sommer | GWDG | 20.12.2023

Look this up

1. 30 Minute Step-by-step Youtube Tutorial

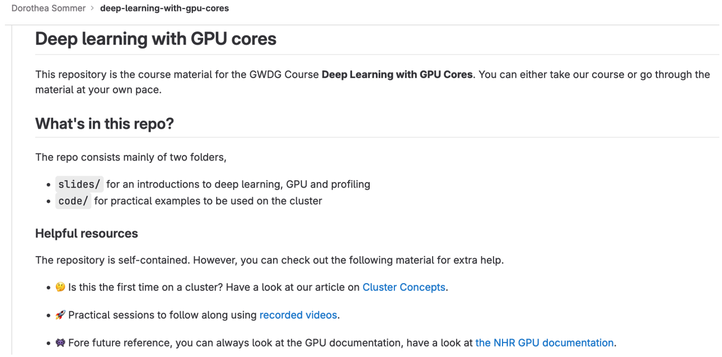




GWDG -> Playlist
-> Deep learning workflow on a GPU cluster

https://www.youtube.com/playlist?list=PLvcoSsXFNRblM4AG5PZwY1AfYEW3EbD9O

2. Example DL Code Repository



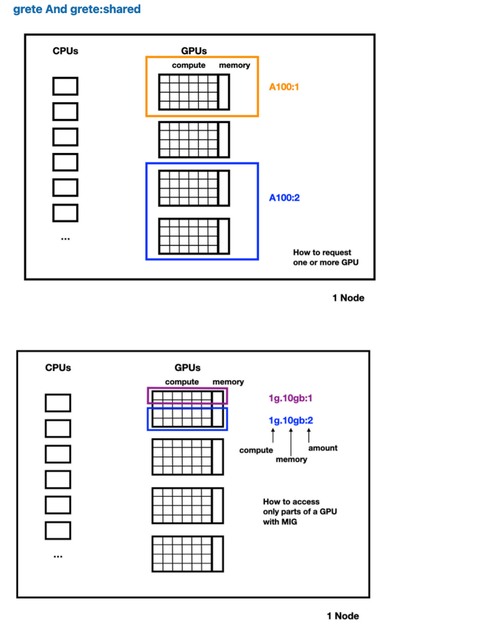
https://gitlab-ce.gwdg.de/dmuelle3/deep-learning-with-gpu-cores

4. Cluster concepts

“What scientist should know to efficiently
use the Scientific Computing Cluster”

https://gitlab-ce.gwdg.de/hpc-team-public/science-domains-blog/-/blob/main/20230417_cluster-practical.md

3. GPU + Slurm Documentation



https://www.hlrn.de/doc/display/PUB/GPU+Usage

Helpful recipes for similar use cases!

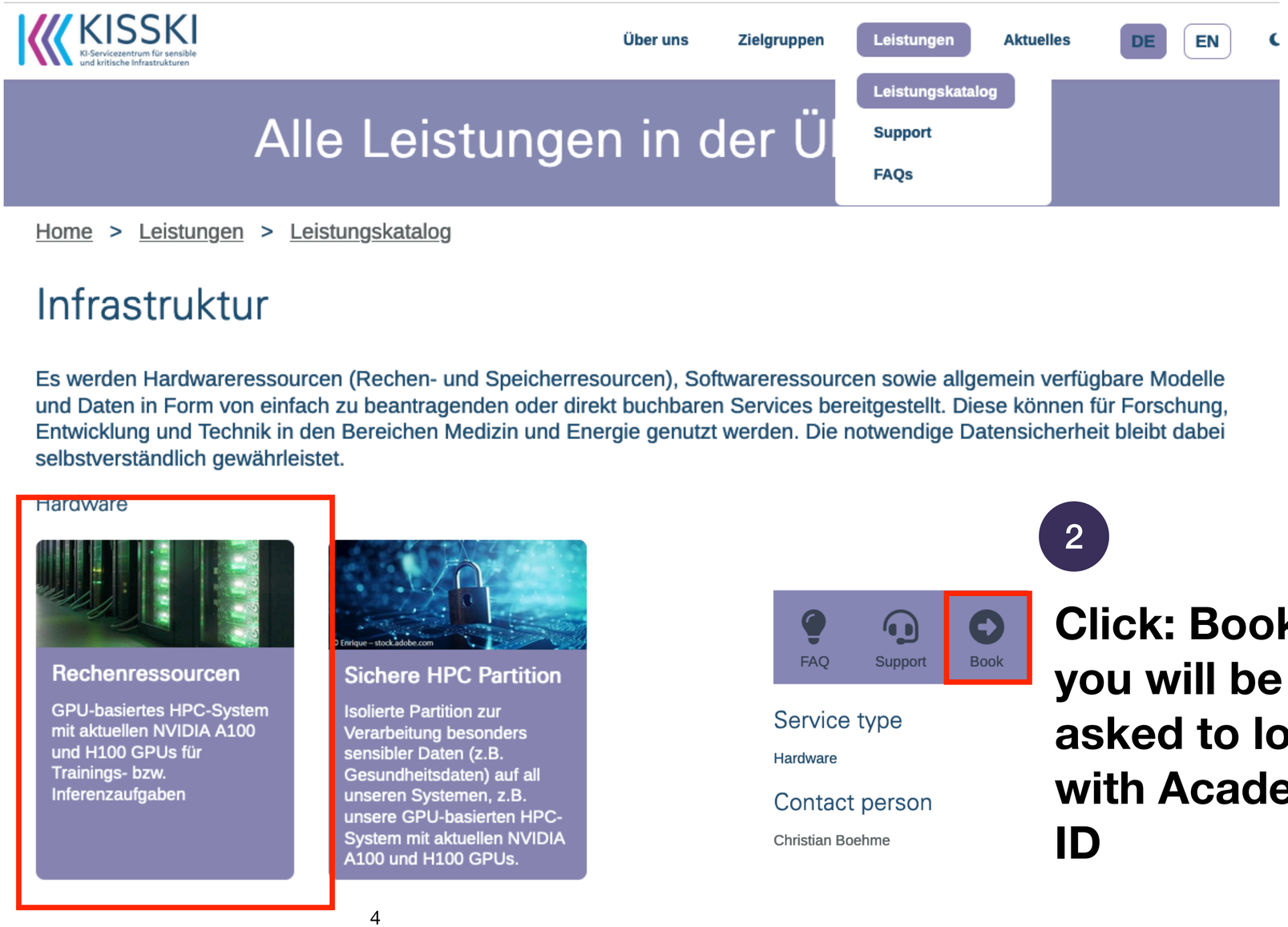
What we'll cover

1. **Setup**
   - Getting a user account
   - Connecting to the cluster
2. **Kisski Training Platform Setup**
   - Cluster Orientation + data perspective
   - Loading DL environments
3. **Running Code with Slurm**
   - Interactive working
   - Scheduling longer jobs
4. **Q&A**

1. How to get access

1 **Kisski website:**

Service catalogue/
Leistungskatalog



Select:
Computing Resources/
Rechenressourcen

2

Click: Book, you will be asked to login with Academic ID

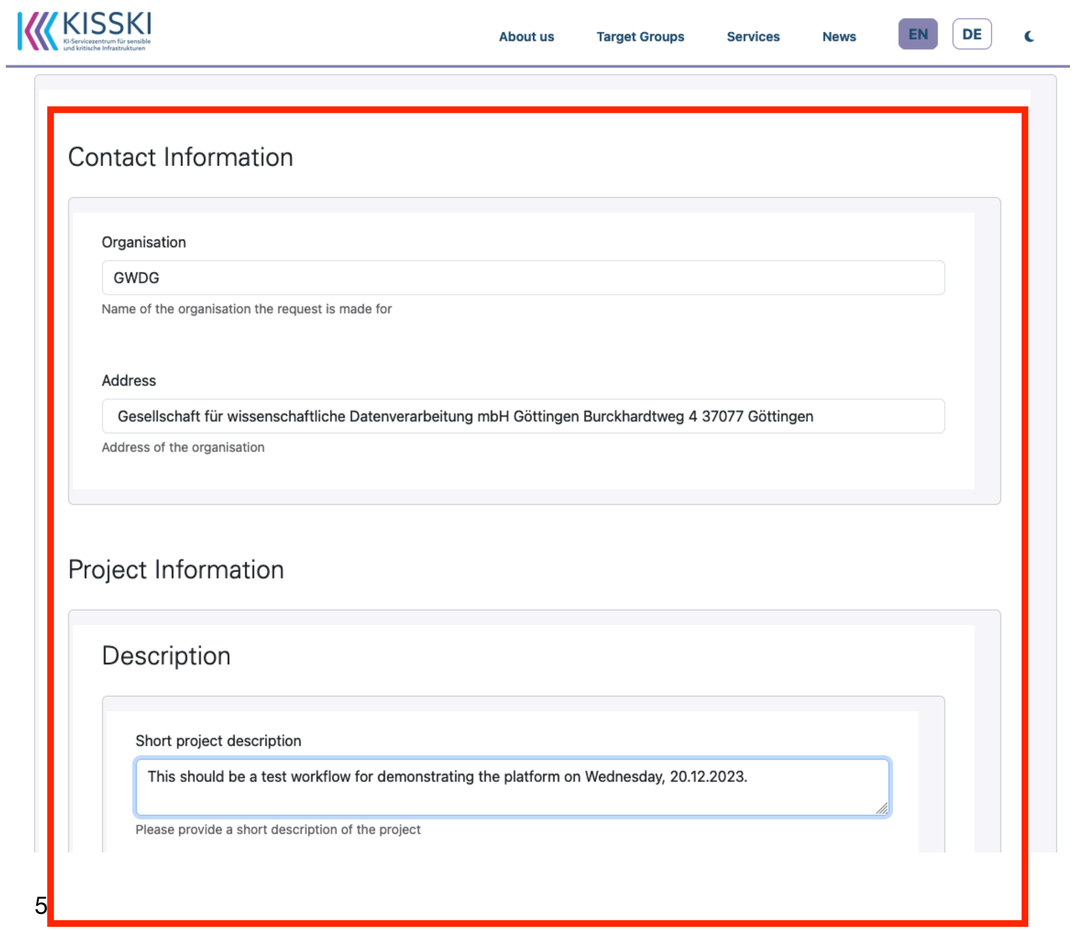
1. How to get access

3 **Make an IDM account if needed (follow the steps) at https://idm.gwdg.de**

4 **Fill out:**

Project Information
+ which resources you need

5 **Wait… ☕**
until you have been granted resources



1. How to get access

hpc-project-service@gwdg.de
Einladung zum HPC Projekt KISSKI Project 4567 / invitation for HPC project KISSKI Project 4567
To: Dorothea Müller

Inbox - Exchange 8. December 2023, 10:52

Please find the English version below.

Hallo dmuelle3,

dies ist eine automatisierte Benachrichtigung vom HPC-Projektmanagementsystem der GWDG.

Sie wurden zum HPC-Projekt "KISSKI Project 4567" hinzugefügt, bitte folgen Sie dem Link, um weitere Informationen über Ihr Projekt zu erhalten:

HPC-Projektmanagementsystem

Sie können sich mit dem Anmeldename

u10792

auf den Loginknoten des HPC Systems anmelden.

Mit freundlichen Grüßen

GWDG

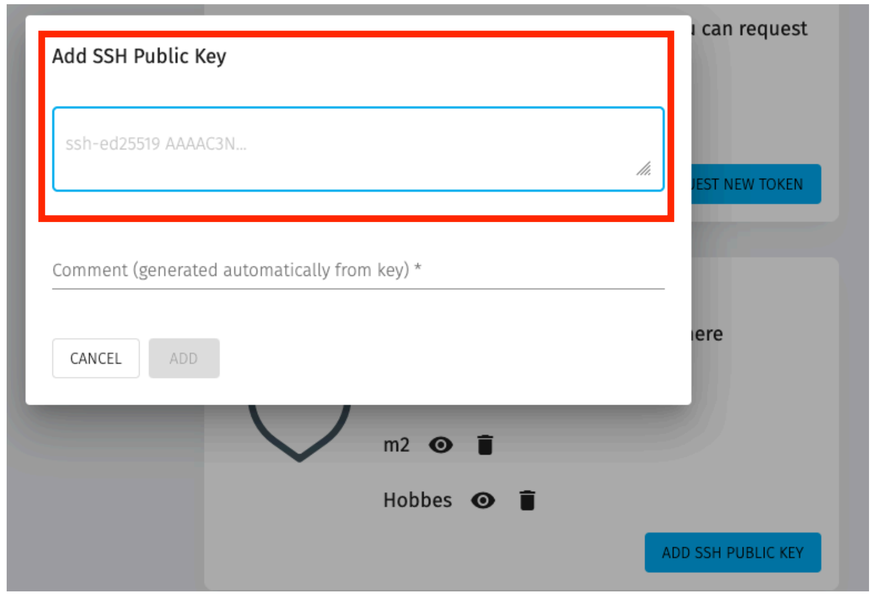
Wir freuen uns auf Ihren Besuch!

6 **Save your credentials from the mail you get**

- the project
- your user name

1. How to get access

7 **Make an ssh key. Upload under https://id.academiccloud.de under “Security” (left bar), scroll down:**



8 **Log in!**

Example: In `.ssh/config`

```
Host kisski
  HostName glogin9.hlrn.de
  User u10792
  IdentityFile ~/.ssh/scc
```

`$ssh kisski`

What we'll cover

1 Setup

- Getting a user account
- Connecting to the cluster

2 Kisski Training Platform Setup

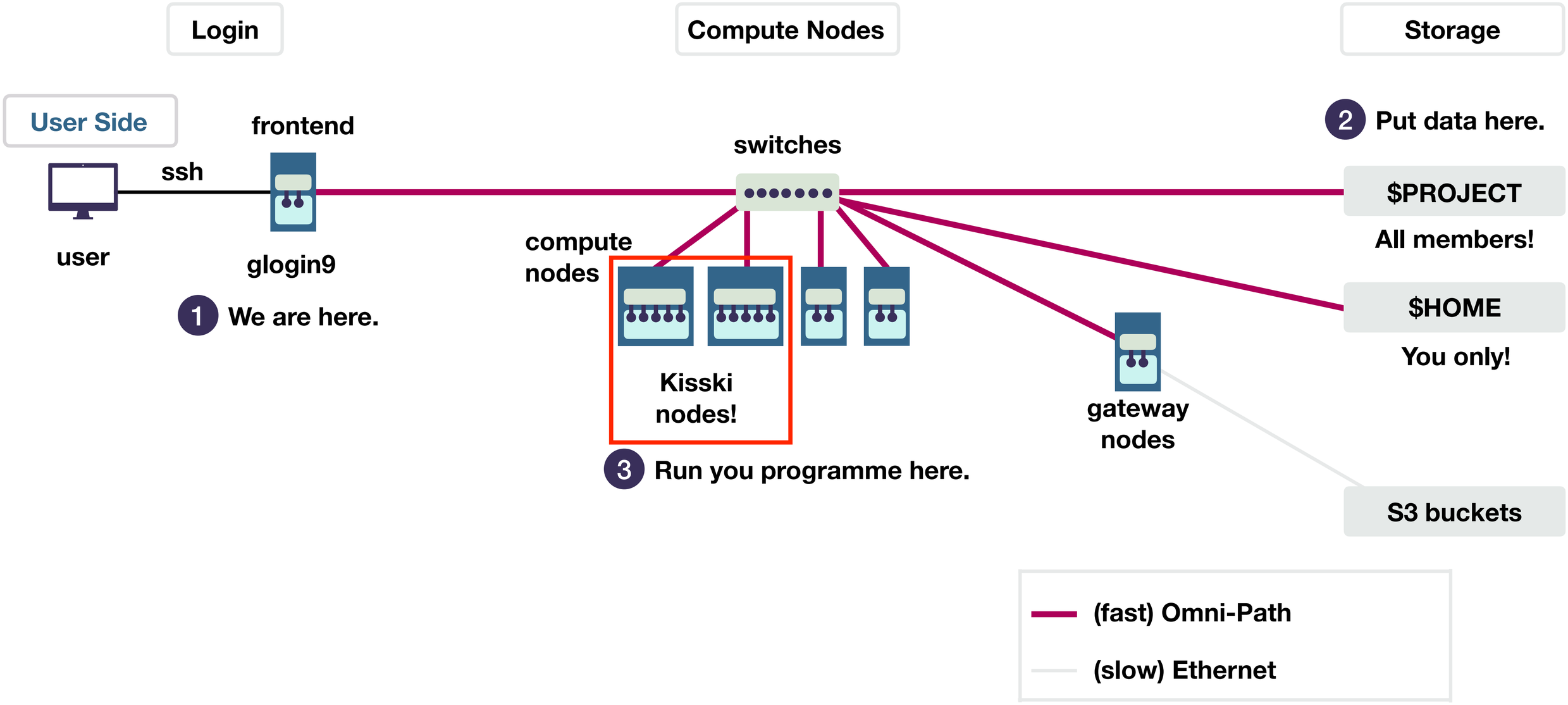
- Cluster Orientation + data perspective
- Loading DL environments

3 Running Code with Slurm

- Interactive working
- Scheduling longer jobs

4 Q&A

2. Setup


Login
Compute Nodes
Storage
User Side
frontend
switches
2 Put data here.
ssh
$PROJECT
All members!
user
glogin9
compute nodes
$HOME
You only!
1 We are here.
Kisski nodes!
gateway nodes
3 Run you programme here.
S3 buckets
(fast) Omni-Path
(slow) Ethernet

2. Setup

Our Use Case

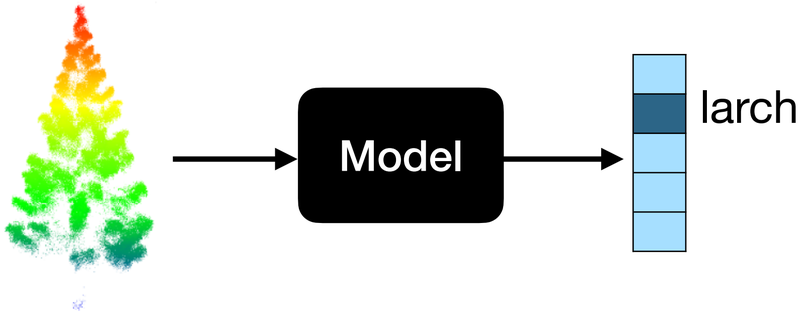
tree species classification

Data

$\underline{x}_i$ tree photo

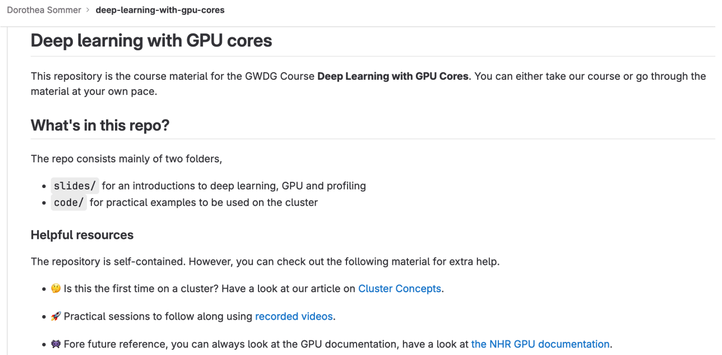
y_i corresponding tree species

Aim



classification
predict tree species given LiDAR scan of the tree

Example DL Code Repository



Dorothea Sommer › **deep-learning-with-gpu-cores**

Deep learning with GPU cores

This repository is the course material for the GWDG Course **Deep Learning with GPU Cores**. You can either take our course or go through the material at your own pace.

What's in this repo?

The repo consists mainly of two folders,

- `slides/` for an introductions to deep learning, GPU and profiling
- `code/` for practical examples to be used on the cluster

Helpful resources

The repository is self-contained. However, you can check out the following material for extra help.

- 🐣 Is this the first time on a cluster? Have a look at our article on Cluster Concepts.
- 🚀 Practical sessions to follow along using recorded videos.
- 🌌 Fore future reference, you can always look at the GPU documentation, have a look at the NHR GPU documentation.

https://gitlab-ce.gwdg.de/dmuelle3/deep-learning-with-gpu-cores

Try this out at home :-)

What we'll cover

1 Setup

- Getting a user account
- Connecting to the cluster

2 Kisski Training Platform Setup

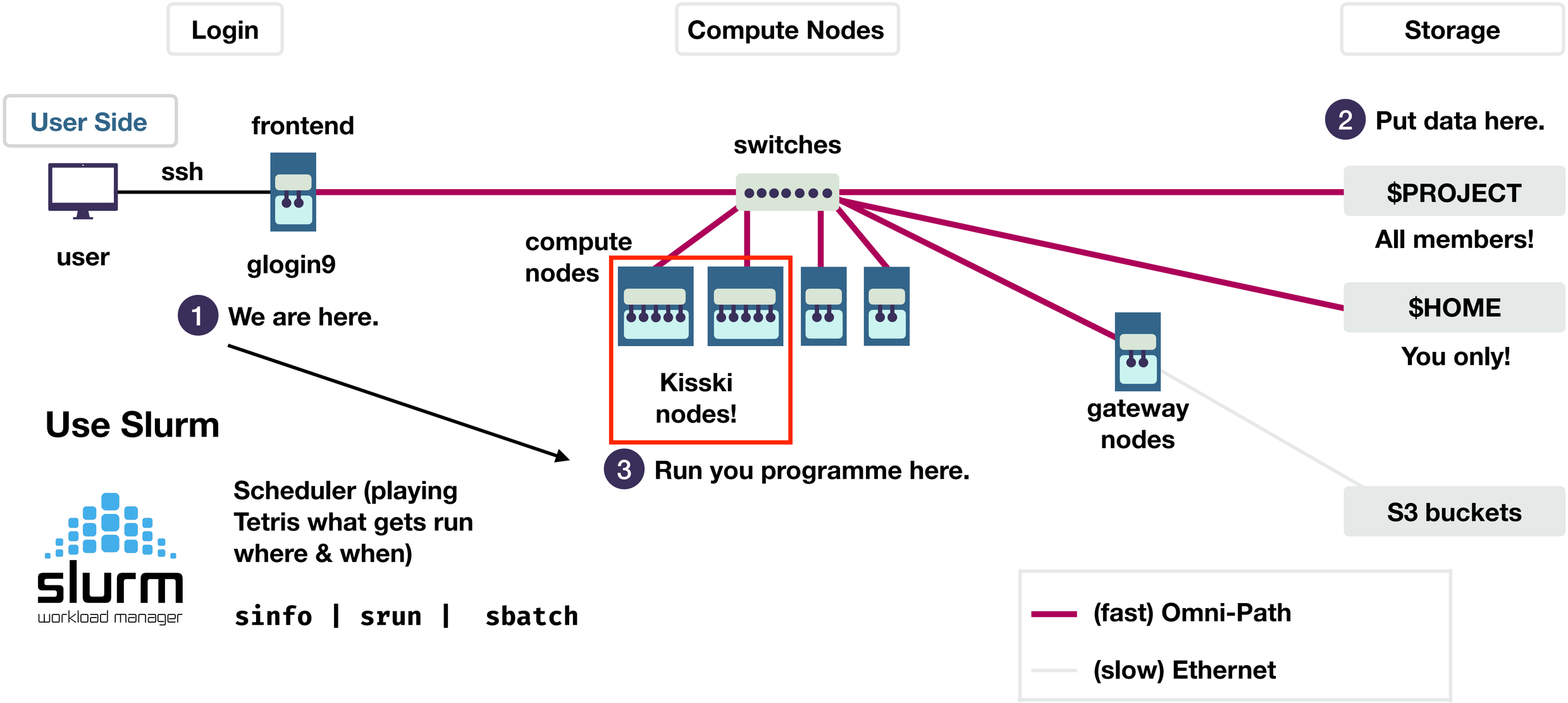
- Cluster Orientation + data perspective
- Loading DL environments

3 Running Code with Slurm

- Interactive working
- Scheduling longer jobs

4 Q&A

3. Running Code



3. Example Script

code > $ submit_train.sh

```
#!/bin/bash
#SBATCH --job-name=train-nn-gpu
#SBATCH -t 05:00:00                    # estimated time # TODO: adapt to your needs
#SBATCH -p kisski                      # the partition you are training on (i.e., which nodes)
#SBATCH -G 1
#SBATCH --nodes=1                      # total number of nodes
#SBATCH --ntasks=1                     # total number of tasks
#SBATCH --mail-type=all                # send mail when job begins and ends
#SBATCH --mail-user=dorothea.sommer@gwdg.de  # TODO: change this to your mailaddress!
#SBATCH --output=./slurm_files/slurm-%x-%j.out     # where to write output, %x give job name
#SBATCH --error=./slurm_files/slurm-%x-%j.err      # where to write slurm error

export TMPDIR=$LOCAL_TMPDIR # Temporarily needed.

module load anaconda3
source activate dl-gpu-home # Or whatever you called your environment.
python -u train.py
```

$sbatch submit_train.sh

What we have covered

KISSKI
KI-Servicezentrum für sensible und kritische Infrastrukturen

1 Setup

- Getting a user account
- Connecting to the cluster

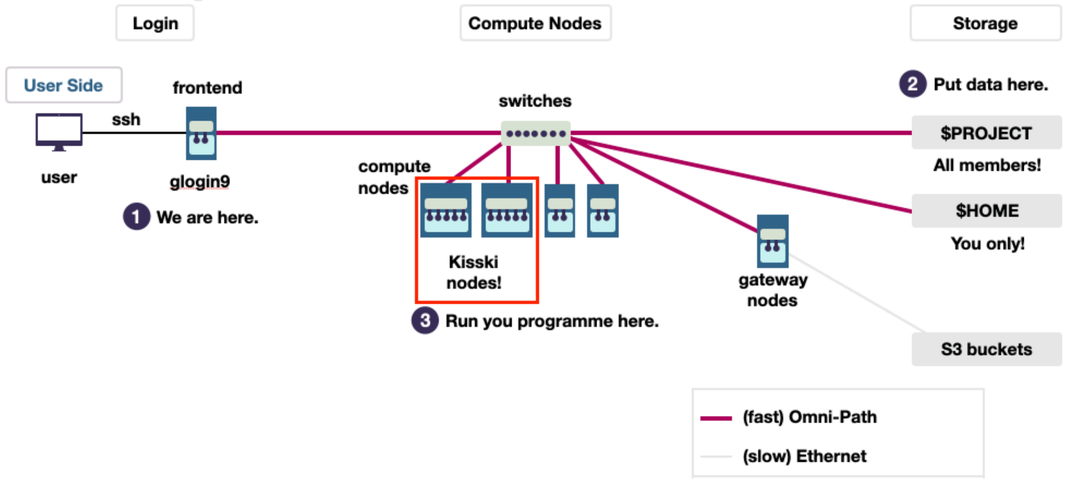
2 Kisski Training Platform Setup

- Cluster Orientation + data perspective
- Loading DL environments

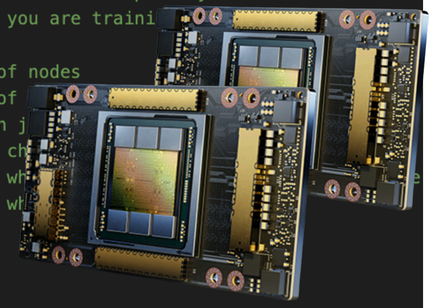
3 Running Code with Slurm

- Interactive working
- Scheduling longer jobs

4 Q&A

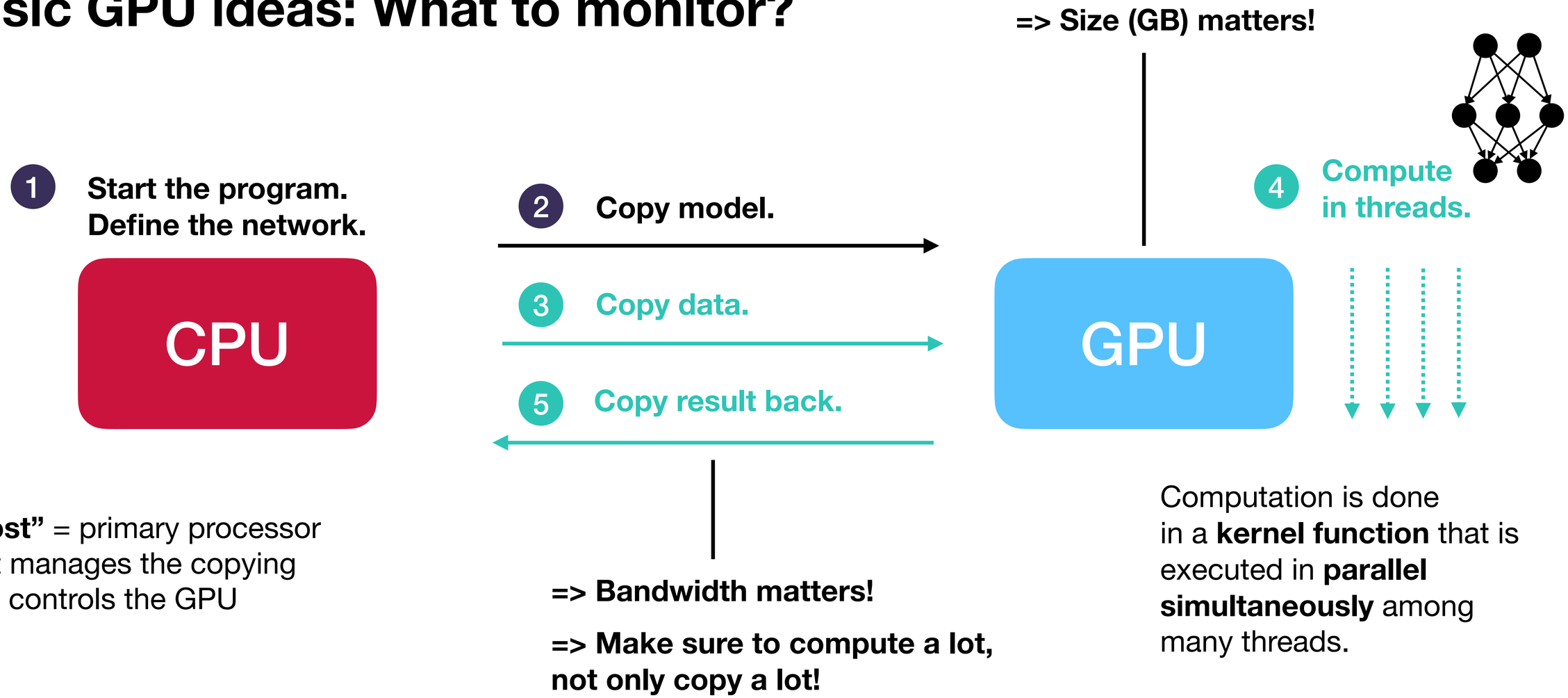



```
#SBATCH --job-name=train-nn-gpu
#SBATCH -t 05:00:00                  # estimated time # TODO: adapt to your needs
#SBATCH -p kisski                    # the partition you are traini
#SBATCH -G 1
#SBATCH --nodes=1                    # total number of nodes
#SBATCH --ntasks=1                   # total number of
#SBATCH --mail-type=all              # send mail when j
#SBATCH --mail-user=dorothea.sommer@gwdg.de  # TODO: ch
#SBATCH --output=./slurm_files/slurm-%x-%j.out     # wh
#SBATCH --error=./slurm_files/slurm-%x-%j.err      # wh

export TMPDIR=$LOCAL_TMPDIR # Temporarily needed.

module load anaconda3
source activate dl-gpu-home # Or whatever you called your environment.
python -u train.py
```



Monitoring

Basic GPU ideas: What to monitor?



“Host” = primary processor that manages the copying and controls the GPU

=> Bandwidth matters!

=> Make sure to compute a lot, not only copy a lot!

Computation is done in a **kernel function** that is executed in **parallel simultaneously** among many threads.