# Quick reference for SLURM

NEBIUS

## Cluster status commands

**sinfo -Nel** list all nodes

```
NODELIST   NODES PARTITION       STATE
worker-0       1     main*       idle
worker-1       1     main*       idle
```

## Drain/Undrain node

**scontrol update NodeName=**<node_name> **State=drain Reason=**"<reason>"

**scontrol update NodeName=**<node_name> **State=resume**

## Node information

**scontrol show node** <node_name>

```
NodeName=worker-1 Arch=x86_64 CoresPerSocket=32
   CPUAlloc=2 CPUEfctv=128 CPUTot=128 CPULoad=8.54
   Gres=gpu:nvidia_h100_80gb_hbm3:8(S:0-1)
   NodeAddr=10.0.93.132 NodeHostName=worker-1 Version=24.05.2
```

## Gathering job information

**squeue -a** list all jobs

```
JOBID  PARTITION  NAME      USER    ST   TIME      NODES
670    main       nccl_tes  root    PD   0:00      1
641    main       bash      decart  R    4:39:39   8
640    main       bash      decart  R    4:40:06   8
```

The **squeue** command shows the list of jobs which are currently running (they are in the RUNNING state, noted as '**R**') or waiting for resources (noted as '**PD**', short for PENDING).

**squeue -tR** list running jobs

**scontrol show job** [job_id] detailed job info

## Job scheduling commands

**sbatch** submit a slurm job

```
sbatch [script]
```

**scancel** delete slurm batch job

```
scancel [job_id]
```

## Interactive jobs

If you need interactive Bash session on a compute node, with the same environment set as the batch jobs, run the following command:

**srun --pty bash -l**

**srun -N 2 --gpus-per-node=8 nvidia-smi**

## Useful Slurm aliases for Bash

```
alias si="sinfo -o \"%20P %5D %14F %8z %10m %10d %11l %16f %N\""
alias sq="squeue -o \"%8i %12j %4t %10u %20q %20a %10g %20P %10Q %5D %11l %11L %R\""
```

## Job script basics

A typical job script will look like this:

```
#!/bin/bash
#SBATCH --job-name=my_ml_job
#SBATCH --output=output_%j.txt
#SBATCH --error=error_%j.txt
#SBATCH --ntasks=1
#SBATCH --time=01:00:00
#SBATCH --gres=gpu:1
#SBATCH --error=JobName.%J.err
#SBATCH --output=JobName.%J.out

# Activate environments here
source .venv/bin/activate

# Run your application
srun python train.py
```

## Common Node States

**IDLE** The node is not currently running any jobs and is available for scheduling.
Causes: No jobs assigned yet

**ALLOCATED** The node is actively running one or more jobs.
Causes: A job has been assigned and is executing.

**MIXED** Some CPUs on the node are allocated to jobs, while others remain idle.
Causes: Partial job allocations

**DOWN** The node is unavailable due to issues or administrative.
Causes: Hardware failure, maintenance, or manual marking by an admin.

**DRAINED** The node has been taken offline and is not available
Causes: Completed maintenance or permanent removal from use.

**UNKNOWN** The node state cannot be determined.
Causes: Communication issues between the controller and the node.

**FAIL** The node has failed and cannot execute jobs.
Causes: Critical hardware or software issues.