

GPU nodes in Sigma

HARDWARE:

- Two new nodes
- **Note:** 36 CPU cores
- 384 GiB of primary memory (RAM)
- Four NVIDIA Tesla V100 SXM2 GPUs (Volta)
- Two 7680GB NVMe SSD scratch disks (~14TiB total)

Further reading: www.nsc.liu.se/systems/sigma/

TESLA V100 SXM2 SPECIFICATIONS:

compute capability	7.0
tensor cores	640
CUDA cores	5120
double precision performance	7.8 TFLOPS
single precision performance	15.7 TFLOPS
memory	32 GB
power	300 W

- Access requires a dedicated project allocation
- Apply using the [LiU Local GPU 2024](#) round in SUPR
- Can be used for both single and double precision FP

ALLOCATING ONE GPU

Using `interactive`:

```
[torbenr@sigma]$ interactive -n 1 -c 9 --gpus-per-task=1 -t 60 -A liu-gpu-2023-4 --reservation=gpu
salloc: Granted job allocation 3330796
salloc: Waiting for resource configuration
salloc: Nodes n2017 are ready for job
[torbenr@n2017]$
```

`-n 1`

Allocate 1 task

`-c 9`

Generally allocate 9 CPU-cores per GPU

`--gpus-per-task=1`

Allocates the GPU

`-A "slurm account"`

Needed unless you are only included in a `liu-gpu-202X-Y` project!

`--reservation=gpu`

Required flag!

ALLOCATING ONE GPU

Multiple ranks per GPU:

```
[torbenr@sigma]$ interactive -n 2 -c 4 --gpus=1 -t 60 -A liu-gpu-2023-4 --reservation=gpu
salloc: Granted job allocation 3330801
salloc: Waiting for resource configuration
salloc: Nodes n2017 are ready for job
[torbenr@n2017]$
```

ALLOCATING ONE GPU

Batch script header:

```
#!/bin/bash
#SBATCH --ntasks=1
#SBATCH --cpus-per-task=9
#SBATCH --gpus-per-task=1
#SBATCH --time=24:00:00
#SBATCH --account=liu-gpu-2023-4
#SBATCH --reservation=gpu
:
```

Here I've used long options (e.g. `--ntasks=1`), but short options (e.g. `-n 1`) also work!

Further reading: www.nsc.liu.se/support/systems/sigma-GPU-user-guide/

```
[torbenr@sigma]$ interactive -n 1 -c 9 --gpus-per-task=1 -t 60 -A liu-gpu-2023-4 --reservation=gpu
salloc: Granted job allocation 3776330
salloc: Waiting for resource configuration
salloc: Nodes n2017 are ready for job
[torbenr@n2017 ~]$ nvidia-smi
Tue Apr 23 09:06:05 2024
```

NVIDIA-SMI 550.54.14			Driver Version: 550.54.14			CUDA Version: 12.4		
GPU	Name	Persistence-M	Bus-Id	Disp.A	Volatile	Uncorr.	ECC	
Fan	Temp	Perf	Pwr:Usage/Cap	Memory-Usage	GPU-Util	Compute M.	MIG M.	
0	Tesla V100-SXM2-32GB	On	00000000:61:00.0	Off				0
N/A	40C	P0	41W / 300W	0MiB / 32768MiB	0%	Default	N/A	

Processes:							
GPU	GI	CI	PID	Type	Process name	GPU Memory	Usage
	ID	ID					
No running processes found							

```
[torbenr@n2017]$
```


ALLOCATING TWO GPUS

Using `interactive`:

```
[torbenr@sigma]$ interactive -n 1 -c 18 --gpus-per-task=2 -t 60 -A liu-gpu-2023-4 --reservation=gpu
Tue Apr 23 09:11:53 2024
```

```

+-----+-----+-----+-----+-----+-----+-----+
| NVIDIA-SMI 550.54.14                | Driver Version: 550.54.14          | CUDA Version: 12.4          |
+-----+-----+-----+-----+-----+-----+-----+
| GPU  Name                Persistence-M | Bus-Id        Disp.A | Volatile Uncorr. ECC |
| Fan  Temp   Perf         Pwr:Usage/Cap |      Memory-Usage | GPU-Util  Compute M. |
|====+=====+====+=====+=====+=====+=====+
|   0   Tesla V100-SXM2-32GB           On          | 00000000:61:00.0 Off |             0         | |
| N/A   39C    P0              41W / 300W |  0MiB / 32768MiB |      0%    Default  |
|                               |                      |                      | N/A                  |
+-----+-----+-----+-----+-----+-----+-----+
|   1   Tesla V100-SXM2-32GB           On          | 00000000:62:00.0 Off |             0         | |
| N/A   39C    P0              41W / 300W |  0MiB / 32768MiB |      0%    Default  |
|                               |                      |                      | N/A                  |
+-----+-----+-----+-----+-----+-----+-----+
| Processes:                            |
| GPU   GI    CI          PID    Type    Process name          | GPU Memory |
|      ID    ID              |           |         |                      | Usage      |
|====+=====+====+=====+=====+=====+=====+
| No running processes found          |
+-----+-----+-----+-----+-----+-----+-----+

```

```
[torbenr@n2017]$
```

USING SINGULARITY AND NGC

[NGC Catalog](#): Software Hub with containers with a range of GPU-accelerated software for NVIDIA GPUs

```
[torbenr@sigma]$ interactive -n 1 -c 18 --gpus-per-task=v100:2 -t 60 -A liu-gpu-2023-4 --reservation=gpu
salloc: Granted job allocation 1512491
srun: Step created for job 1512491
[torbenr@n2017]$ . ./sourceme.txt
[torbenr@n2017]$ singularity run --nv tf20.09_py3.v3.sif

=====
== TensorFlow ==
=====

NVIDIA Release 20.09-tf2 (build 16003717)
TensorFlow Version 2.3.0

Container image Copyright (c) 2020, NVIDIA CORPORATION. All rights reserved.
Copyright 2017-2020 The TensorFlow Authors. All rights reserved.

Various files include modifications (c) NVIDIA CORPORATION. All rights reserved.
NVIDIA modifications are covered by the license terms that apply to the underlying project or file.

Detected MOFED 5.0-0.

NOTE: MOFED driver was detected, but nv_peer_mem driver was not detected.
Multi-node communication performance may be reduced.

Singularity>
```