About NTNU’s HPC group

We play with the largest, fastest, newest (“state-of-the-art”), and most expensive machines in Norway.
About NTNU’s HPC group

Responsible for NTNU’s and Norway’s HPC resources
About NTNU’s HPC group

Responsible for NTNU’s and Norway’s HPC resources

Main activities:

- Procurement and maintenance of NTNU’s and Norway’s HPC infrastructure
- Scientific computing & researcher support
- HPC research
- ...

We play with the largest, fastest, newest (“state-of-the-art”), and most expensive machines in Norway.
About NTNU’s HPC group

Responsible for NTNU’s and Norway’s HPC resources

Main activities:

- Procurement and maintainance of NTNU’s and Norway’s HPC infrastructure
- Scientific computing & researcher support
- HPC research
- ...

We play with the largest, fastest, newest ("state-of-the-art"), and most expensive machines in Norway.
What is High Performance Computing?

1. Connect computers such that we can use them as one big machine.

1https://insidehpc.com/hpc-basic-training/what-is-hpc/
What is High Performance Computing?

High Performance Computing most generally refers to the practice of aggregating computing power in a way that delivers much higher performance than one could get out of a typical desktop computer or workstation in order to solve large problems in science, engineering, or business. ¹

¹https://insidehpc.com/hpc-basic-training/what-is-hpc/
What is High Performance Computing?

High Performance Computing most generally refers to the practice of aggregating computing power in a way that delivers much higher performance than one could get out of a typical desktop computer or workstation in order to solve large problems in science, engineering, or business. ¹

=> Connect computers such that we can use them as one big machine.

¹https://insidehpc.com/hpc-basic-training/what-is-hpc/
European HPC Pyramid

Tier 0: Continental infrastructure
Tier 1: National facilities
Tier 2: Regional facilities
Tier 3: Private computers

European HPC
Fram & Betzy
Idun
Workstations & Laptops
Idun cluster

1. 70 nodes with a total of ca. 2000 cores
2. High-throughput and low latency infiniband interconnect
3. Lustre parallel file system
1. 70 nodes with a total of ca. 2000 cores
2. High-throughput and low latency infiniband interconnect
3. Lustre parallel file system
Idun cluster
What is IDI’s share in Idun?

<table>
<thead>
<tr>
<th>#Nodes</th>
<th>#Cores</th>
<th>Memory [Gb]</th>
<th>GPUs</th>
<th>Extras</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>36</td>
<td>128</td>
<td>2 P100 16Gb</td>
<td>Power meters</td>
</tr>
<tr>
<td>19</td>
<td>24</td>
<td>128</td>
<td>2 P100 16Gb</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>28</td>
<td>768</td>
<td>2 V100 16Gb</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>20</td>
<td>768</td>
<td>8 V100 32Gb</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>20</td>
<td>768</td>
<td>10 V100 32Gb</td>
<td></td>
</tr>
<tr>
<td>35</td>
<td>944</td>
<td></td>
<td></td>
<td>90</td>
</tr>
</tbody>
</table>
What are the next steps?

1. Get access to Idun
   • Ask supervisor/professor for access

2. Transfer (input) data to the cluster storage
   • 1 Tb of working storage
   • 2 million files

3. Start submitting jobs
   • Cluster is a shared resource
   • Researchers/students from all over NTNU use it
   • Need to be able to deal with a cluster’s resource manager (slurm)
What are the next steps?

1. Get access to Idun
   • Ask supervisor/professor for access
What are the next steps?

1. Get access to Idun
   - Ask supervisor/professor for access
2. Transfer (input) data to the cluster storage
   - 1 Tb of working storage
   - 2 million files
What are the next steps?

1. Get access to Idun
   - Ask supervisor/professor for access
2. Transfer (input) data to the cluster storage
   - 1 Tb of working storage
   - 2 million files
3. Start submitting jobs
   - Cluster is a shared resource
   - Researchers/students from all over NTNU use it
   - Need to be able to deal with a cluster’s resource manager (slurm)
Slurm - Workflow

- Resource Description
  - .slurm

- Application

- Submit Job

- Scheduler

- Node1
- Node2
- Node3
Slurm - Simple job script

```bash
>$ cat test.slurm
#!/bin/bash
#SBATCH --partition=TEST
#SBATCH --account=support
#SBATCH --nodes=1
#SBATCH --cores=1
#SBATCH --time=0-00:05:00
#SBATCH --mem=1GB
#SBATCH --output=slurm-test.out
#SBATCH --error=slurm-test.err

module load foss/2019b
hostname
sleep 5

>$ sbatch test.slurm
```

Norwegian University of Science and Technology
## Slurm - Nodes and Partitions

**Partition:** Logical collection of nodes with the same/similar properties

```bash
> $ sinfo
PARTITION  AVAIL  TIMELIMIT  NODES   STATE  NODELIST
WORKQ      up      21-00:00:0  23     mix    compute-2-0-[3,5-8]
WORKQ      up      21-00:00:0 13      alloc  compute-2-0-[1-2,4]
TEST       up      30:00      1       idle   compute-2-0-43
EPIC       up      7-00:00:0  7       mix    compute-3-0-[1-5,7-8]
EPIC2      up      7-00:00:0 19      mix    compute-4-0-[1-19]
EPICALL    up      7-00:00:0 26      mix    compute-3-0-[1-5,7-8]
V100       up      30:00      4       mix    compute-5-0-[1,3-5]
V100       up      30:00      1       idle   compute-5-0-2
V100-IDI   up      7-00:00:0  4       mix    compute-5-0-[1,3-5]
V100-IDI   up      7-00:00:0  1       idle   compute-5-0-2
STORAGE    up      1-00:00:0  2       idle   idun-samba[1-2]
```
Slurm - Job Queue

```
>$ squeue

Tue Nov 26 15:11:51 2019

<table>
<thead>
<tr>
<th>JOBID</th>
<th>PARTITION</th>
<th>NAME</th>
<th>USER</th>
<th>STATE</th>
<th>TIME</th>
<th>NODES</th>
<th>NODELIST(REASON)</th>
</tr>
</thead>
<tbody>
<tr>
<td>492537</td>
<td>WORKQ</td>
<td>...</td>
<td>...</td>
<td>PENDING</td>
<td>0:00</td>
<td>16</td>
<td>(Resources)</td>
</tr>
<tr>
<td>497457</td>
<td>WORKQ,EPI</td>
<td>...</td>
<td>...</td>
<td>PENDING</td>
<td>0:00</td>
<td>1</td>
<td>(Resources)</td>
</tr>
<tr>
<td>496655</td>
<td>WORKQ</td>
<td>...</td>
<td>...</td>
<td>PENDING</td>
<td>0:00</td>
<td>4</td>
<td>(Priority)</td>
</tr>
<tr>
<td>496651</td>
<td>WORKQ</td>
<td>...</td>
<td>...</td>
<td>PENDING</td>
<td>0:00</td>
<td>4</td>
<td>(Priority)</td>
</tr>
<tr>
<td>496653</td>
<td>WORKQ</td>
<td>...</td>
<td>...</td>
<td>PENDING</td>
<td>0:00</td>
<td>4</td>
<td>(Priority)</td>
</tr>
<tr>
<td>496912</td>
<td>WORKQ,EPI</td>
<td>...</td>
<td>...</td>
<td>PENDING</td>
<td>0:00</td>
<td>1</td>
<td>(Priority)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>497550</td>
<td>V100-IDI</td>
<td>...</td>
<td>...</td>
<td>RUNNING</td>
<td>3:26</td>
<td>1</td>
<td>compute-5-0-</td>
</tr>
<tr>
<td>496794</td>
<td>WORKQ</td>
<td>...</td>
<td>...</td>
<td>RUNNING</td>
<td>18:16:20</td>
<td>2</td>
<td>compute-2-0-</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
```
Slurm - Using GPUs

#!/bin/sh
#SBATCH --partition=EPICALL
#SBATCH --time=00:02:00
#SBATCH --nodes=2
#SBATCH --ntasks-per-node=1
#SBATCH --gres=gpu:1
#SBATCH --job-name="mpi-cuda-test"
#SBATCH --output=mpi-cuda-test.out

module purge
module load fosscuda/2018b

mpirun hostname
mpirun ./a.out
Questions!?

Slurm commands:

- **sbatch**: Submit job to scheduler.
- **sinfo**: View information about nodes and partitions.
- **squeue**: View information about scheduling queue.
- **sprio**: View job priority factors.
- **scontrol**: View and modify slurm job specifications.
- ...

Idun Documentation:

https://www.hpc.ntnu.no/display/hpc/Idun+Cluster