INTRODUCTION TO HIGH-PERFORMANCE COMPUTING AT THE UNIVERSITY OF MARIBOR

Miran Ulbin

University of Maribor Open Science Summer School

Agenda

Why do we need supercomputers?

Supercomputers at the University of Maribor

Software and Usage

Useful contacts and links

Solving system of linear equations

Matrix order N	Required memory (64-bit) $= 8N^2$ bytes	Required floating point operations [FLOPS] $= N^3/6$	Solving time on high end personal computer	Solving time on HPC
10^4	800 MB	$10^{12}/6$	3 min	2 s
10^5	80 GB	$10^{15}/6$	46 hours	30 min
10^6	8 TB	$10^{18}/6$	5 years	21 days

Purpose of HPC at University of Maribor

High-performance computing (HPC)

Parallel engineering simulations

Artificial intelligence and machine learning

Data storage and data mining

Basic requirements:

Fast and efficient simulation

Fast data transfer

First computers

IBM 1130 – 1969 first computer at the University of Maribor https://blog.cobiss.si/2021/05/12/pionirski-cas-mariborskega-racunalnistva-do-ustanovitve-univerze-leta-1975/

CONVEX – 1992? - first vector supercomputer at the University of Maribor - https://www.old-computers.com/Museum/computer.asp?c=983&st=1

Computer clusters

1992 – 2012 - Nblue, Rheat – personal computers connected into computer clusters

Hpc-core – 2012 - first compact computer cluster at the University of Maribor

- 2 TFLOPS
- 14 computer nodes with a total of 224 cores
- Most nodes have 512GB RAM
- 16TB SSD storage
- 10Gb interconnect network

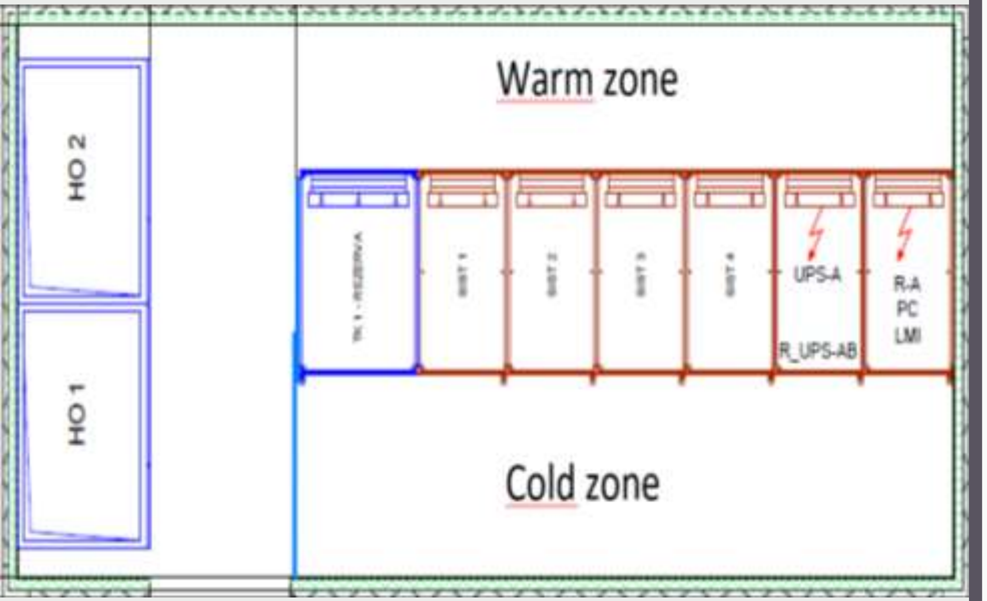
Univerza v Mariboru

IZUM
Institut informacijskih znanosti, Maribor

FIS
Fakulteta za informacijske študije
Faculty of information studies

HPC RIVR 2018

- HPC RIVR project with a budget of 20 million EUR
- Consortium of UM, IZUM, and FIS in cooperation with SLING
- Prototype HPC - UM
- Dislocated nodes and remote access – FIS
- Peta-scale HPC - IZUM



HPC Maister since 2019

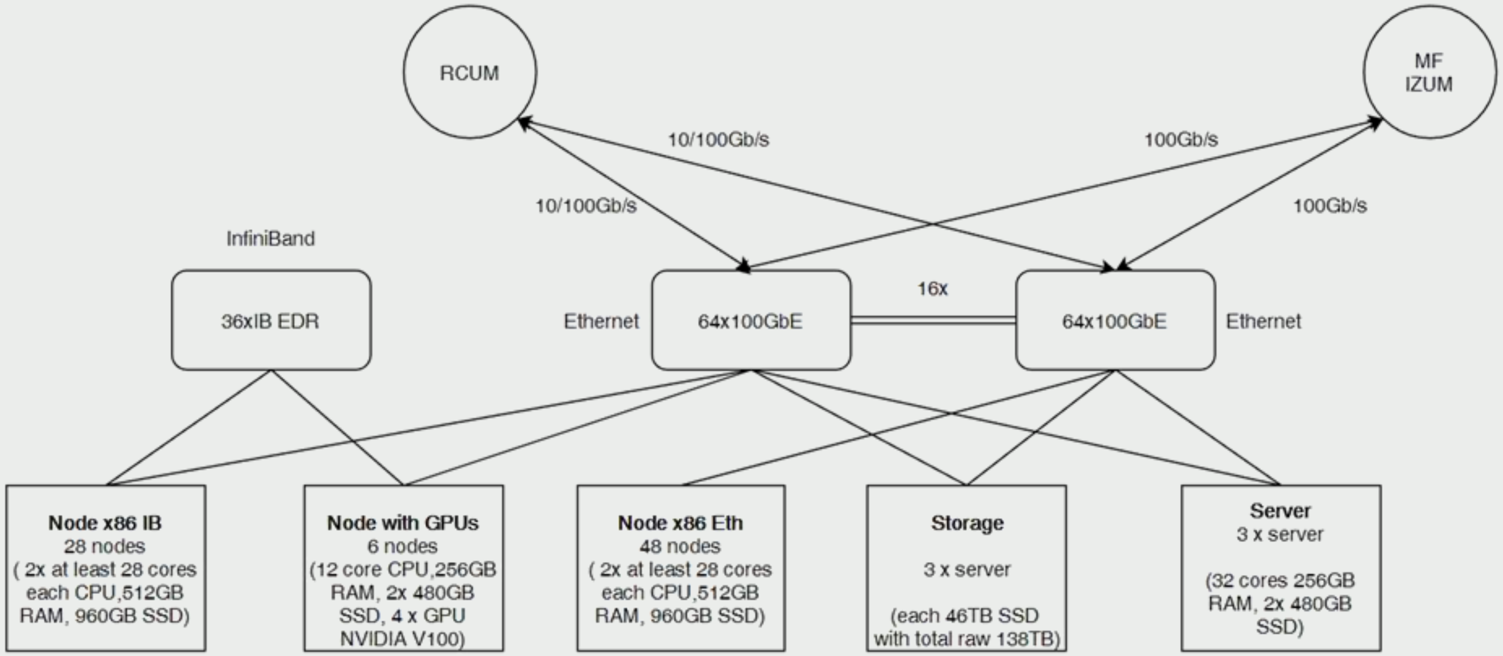
- A large number of nodes at a small place with small energy consumption
- Dimensions - 6,5 x 2,9 x 2,9 m
- Capacity – 4 system racks for max. 168 units
- Cooling – redundant mechanical cooling 2x40kW
- Power supply – up to 80kVA, 20kW UPS (for critical devices only) min 6min

HPC Maister location

Placed at the parking of Technical faculties

An electrical transformer in close vicinity due to peak consumption of almost 100KW

Connected with optical cable with a capacity of 200Gb/s



Node x86 IB
28 nodes
(2x at least 28 cores each CPU,512GB RAM, 960GB SSD)

Node with GPUs
6 nodes
(12 core CPU,256GB RAM, 2x 480GB SSD, 4 x GPU NVIDIA V100)

Node x86 Eth
48 nodes
(2x at least 28 cores each CPU,512GB RAM, 960GB SSD)

Storage
3 x server
(each 46TB SSD with total raw 138TB)

Server
3 x server
(32 cores 256GB RAM, 2x 480GB SSD)

Maister configuration

Network

- Ethernet 100Gb/s
- Infiniband 100Gb/s

Storage

- SSD storage - 138TB SSD
- HDD storage – 2,88PB

GPU nodes

- 6 nodes, each with 4x NVIDIA TESLA V100 GPUs

•CPU nodes

- 76 nodes, each 2x AMD EPYC 32C, 512GB RAM, 2x 960GB SSD, 100Gb/s network

•Servers, login nodes

- 3 servers, containing number of virtual servers and login nodes

Maister HPC

Servers and login nodes

- maister.hpc-rivr.um.si – Centos 8 head node for cluster (not directly accessible)
- rmaister.hpc-rivr.um.si – Centos 8 login node (alias rmaister1,rmaister2, rmaister3, rmaister4)
- spremljevalec1.hpc-rivr.um.si – Centos 8 web server for slurm user interface

Ethernet workers

- cn01-cn48 – Centos 8 for job execution (accessible using Slurm)

Infiniband/Ethernet workers

- dpcn01-dpcn28 – Centos 8 for job execution (accessible using Slurm)

•GPU workers

- gpu01-gpu06 – Centos 8 for job execution (accessible using Slurm)

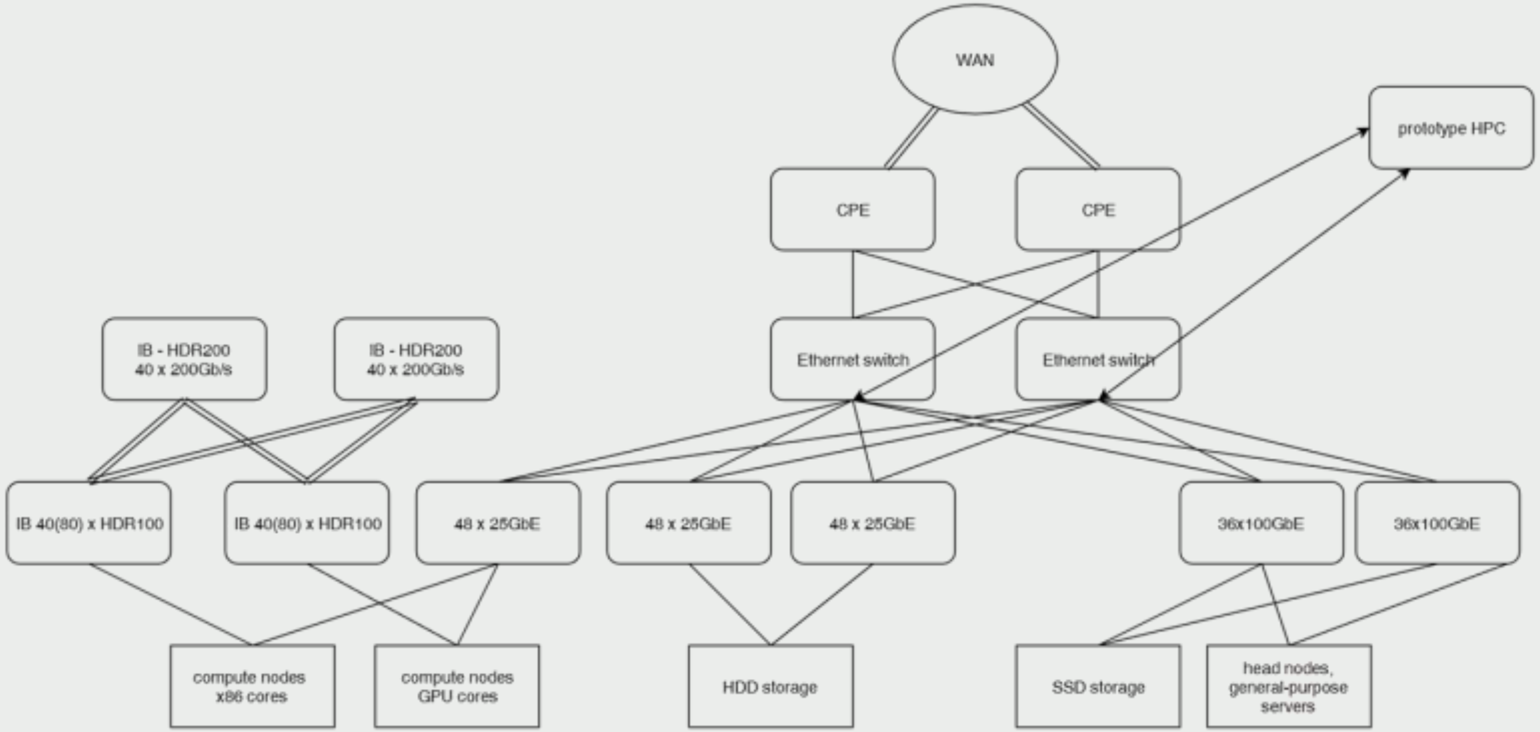
•Storage SSD

- /ceph/grid/home/username – 200GB disk quota (extendible on request)

Storage HDD

- /ceph/grid/home/username/CEPH2 – 200GB disk quota (extendible on request)
- Amazon cloud S3 storage - 200GB disk quota (extendible on request)

https://www.hpc-rivr.si/



Vega initial configuration

An involvement of EURO HPC granting additional 7M€

The design should be augmented, and the projected power consumption was too high

Liquid cooling was required, and the Euro HPC tender required a complete solution

Vega

HPC

2020

Servers and login nodes

- CPU login – 4 servers
- GPU login – 4 servers
- Virtualization and service – 30 servers

CPU partition

- 768 standard nodes (2x AMD Rome 7H12 (64c, 2.6GHz, 280W) 256GB RAM 1x HDR100 single port mezzanine 1x 1.92TB M.2 SSD)
- 192 large memory nodes (2x AMD Rome (64c, 2.6GHz, 280W) 1TB RAM 1x HDR100 single port mezzanine 1x 1.92TB M.2 SSD)

GPU partition

- 60 GPU nodes (4x Nvidia A100, 2x AMD Rome 7H12, 512 GB RAM, 2x HDR dual port mezzanine, 1x 1.92TB M.2 SSD)

•HPST - High-performance storage tier

- 1,5 PB (NVMe, 8 InfiniBand HDR100, 4 embedded Lustre VMs)

•LCST - Large Capacity Storage tier

- 23PB (61 nodes with HDD with parallel file system)

https://doc.vega.izum.si/architecture/

Trdina

HPC

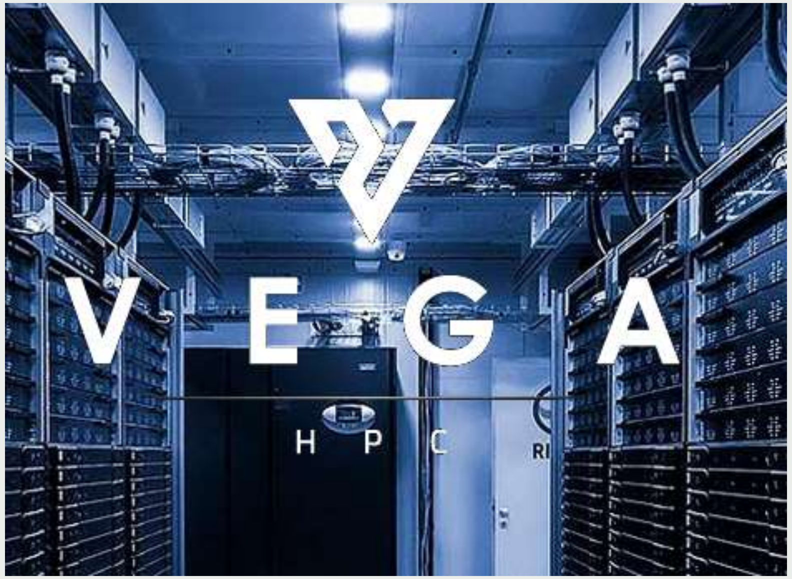
2020

Servers and working nodes

- 2 CPU large memory node
- 4 GPU NVIDIA V100 node
- SSD distributed disk server 138TB

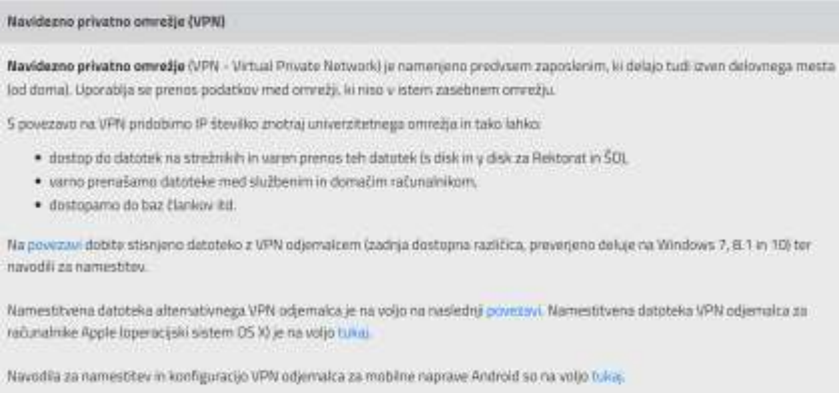
Access to Vega HPC

- Access to Vega should be granted with application on tender
- Usually using Nordugrid ARC middleware (not presented in this lecture)
- More information: https://doc.vega.izum.si/cluster-access/



Access to HPC at University of Maribor

- Due to security reason access to Maister HPC and hpc-core is restricted to university network
- From outside access requires Virtual Private Network (VPN)
- User should install and use VPN client on home computer to get access to university network
- More information: https://it.um.si/zaposleni/Strani/default.aspx



Navidezno privatno omrežje (VPN)

Navidezno privatno omrežje (VPN - Virtual Private Network) je namenjeno predvsem zaposlenim, ki delajo tudi izven delovnega mesta (od doma). Uporablja se prenos podatkov med omrežji, ki niso v istem zasebnem omrežju.

S povezavo na VPN pridobimo IP številko znotraj univerzitetnega omrežja in tako lahko:

- dostop do datotek na strežnikih in varen prenos teh datotek (s disk in y disk za Rektorat in ŠO),
- varno prenašamo datoteke med službenim in domačim računalnikom,
- dostopamo do baz člankov itd.

Na povezavi dobite stisnjeno datoteko z VPN odjemalcem (zadnja dostopna različica, preverjeno deluje na Windows 7, 8.1 in 10) ter navodili za namestitev.

Namestitvena datoteka alternativnega VPN odjemalca je na voljo na naslednji povezavi. Namestitvena datoteka VPN odjemalca za računalnike Apple (operacijski sistem OS X) je na voljo tukaj.

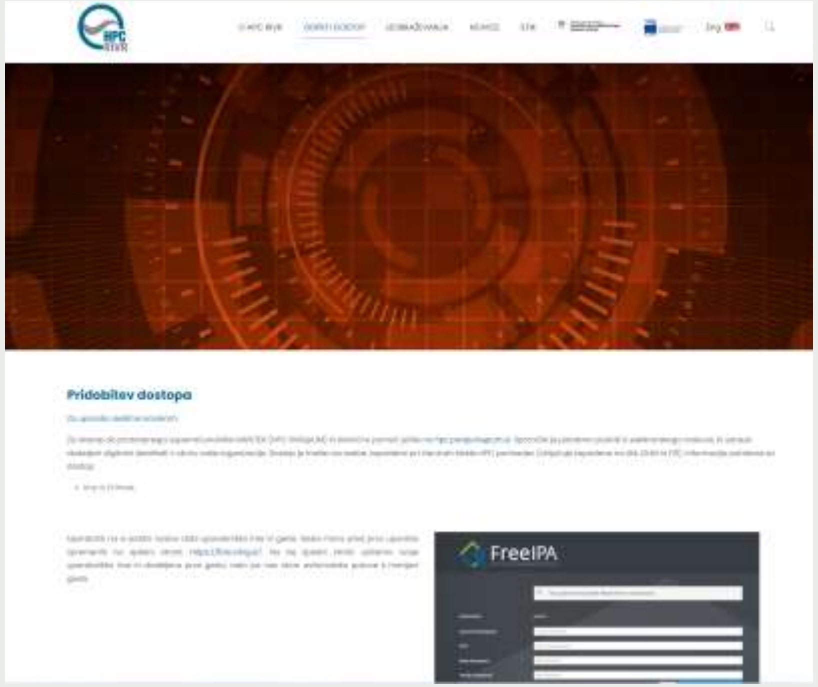
Navodila za namestitev in konfiguracijo VPN odjemalca za mobilne naprave Android so na voljo tukaj.

hpc-core.um.si

- Access is possible for small and specific jobs due to small available number of cores
- Request should be addresed to zoran.ren@um.si or miran.ulbin@um.si

Maister HPC

- Detailed instructions on https://www.hpc-rivr.si/uporaba-hpc/
- Usually the mail to hpc.podpora@um.si is required
- The user will obtain username and password for ssh access
- Profile settings (pasword change etc.) is done on webpage https://fido.sling.si/



Pridobitev dostopa

ssh

Putty

- https://www.putty.org/

Windows terminal

- Microsoft Store
- https://docs.microsoft.com/en-us/windows/terminal/install

Usage

- ssh rmaister.hpc-rivr.um.si
- Username: xyz
- Password: ********

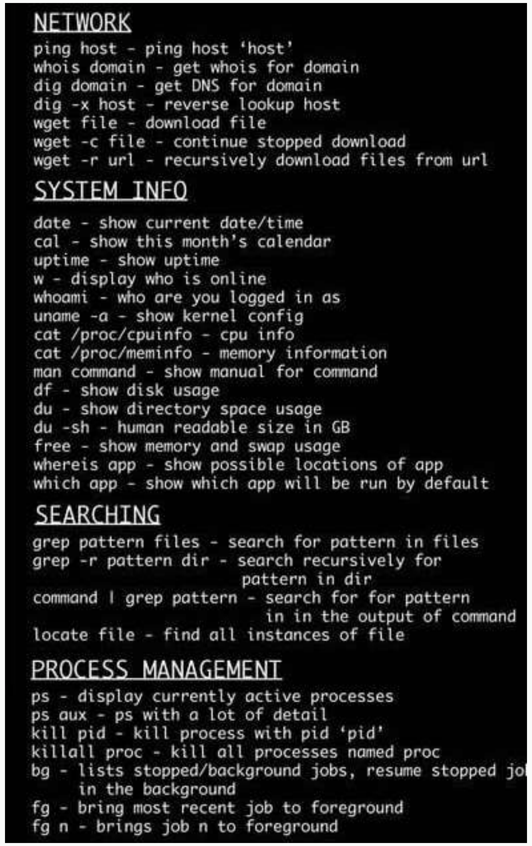
•Using SSH public keys

- Putty - https://tavu.io/en/tutorials/how-to-use-ssh-keys-with-putty-on-windows
- Microsoft - https://docs.microsoft.com/en-us/windows-server/administration/openssh/openssh_keymanagement

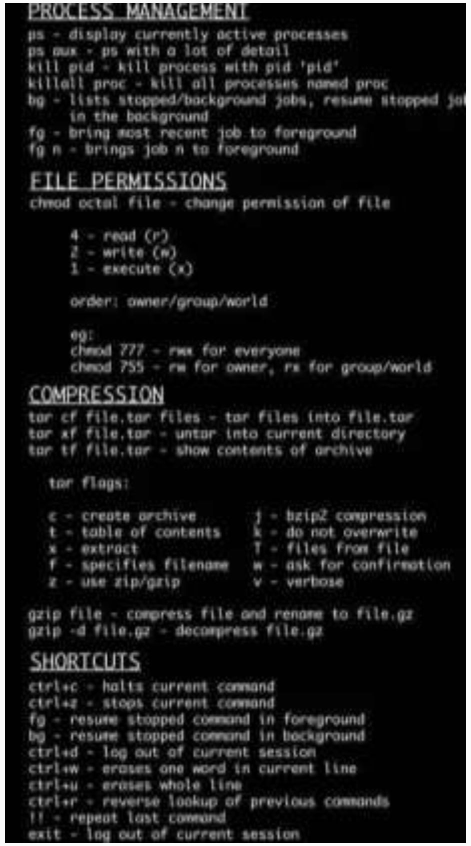
•Password change or setting of SSH public keys

- Maister - https://fido.sling.si/
- Hpc-core – add the key into *~/.ssh/authorized_keys*

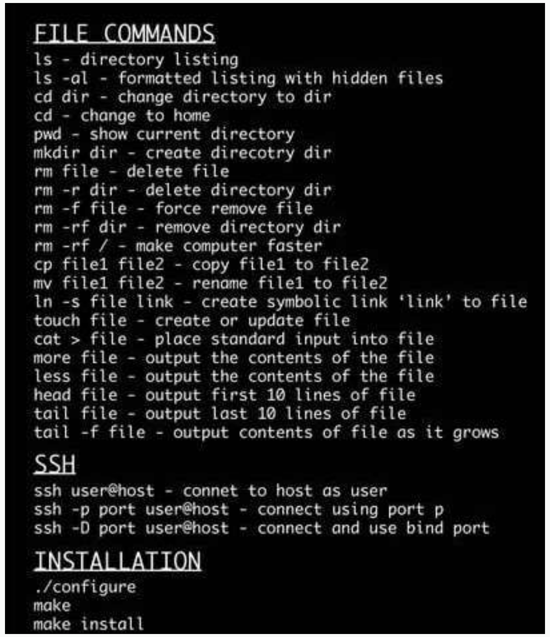
Linux (Centos 8)



```
NETWORK
ping host - ping host 'host'
whois domain - get whois for domain
dig domain - get DNS for domain
dig -x host - reverse lookup host
wget file - download file
wget -c file - continue stopped download
wget -r url - recursively download files from url
SYSTEM INFO
date - show current date/time
cal - show this month's calendar
uptime - show uptime
w - display who is online
whoami - who are you logged in as
uname -a - show kernel config
cat /proc/cpuinfo - cpu info
cat /proc/meminfo - memory information
man command - show manual for command
df - show disk usage
du - show directory space usage
du -sh - human readable size in GB
free - show memory and swap usage
whereis app - show possible locations of app
which app - show which app will be run by default
SEARCHING
grep pattern files - search for pattern in files
grep -r pattern dir - search recursively for
                      pattern in dir
command | grep pattern - search for for pattern
                         in in the output of command
locate file - find all instances of file
PROCESS MANAGEMENT
ps - display currently active processes
ps aux - ps with a lot of detail
kill pid - kill process with pid 'pid'
killall proc - kill all processes named proc
bg - lists stopped/background jobs, resume stopped jo
     in the background
fg - bring most recent job to foreground
fg n - brings job n to foreground
```



```
PROCESS MANAGEMENT
ps - display currently active processes
ps aux - ps with a lot of detail
kill pid - kill process with pid 'pid'
killall proc - kill all processes named proc
bg - lists stopped/background jobs, resume stopped jo
     in the background
fg - bring most recent job to foreground
fg n - brings job n to foreground
FILE PERMISSIONS
chmod octal file - change permission of file

     4 - read (r)
     2 - write (w)
     1 - execute (x)

     order: owner/group/world

     eg:
     chmod 777 - rwx for everyone
     chmod 755 - rw for owner, rx for group/world
COMPRESSION
tar cf file.tar files - tar files into file.tar
tar xf file.tar - untar into current directory
tar tf file.tar - show contents of archive

  tar flags:

  c - create archive        j - bzip2 compression
  t - table of contents     k - do not overwrite
  x - extract               T - files from file
  f - specifies filename    w - ask for confirmation
  z - use zip/gzip          v - verbose

gzip file - compress file and rename to file.gz
gzip -d file.gz - decompress file.gz
SHORTCUTS
ctrl+c - halts current command
ctrl+z - stops current command
fg - resume stopped command in foreground
bg - resume stopped command in background
ctrl+d - log out of current session
ctrl+w - erases one word in current line
ctrl+u - erases whole line
ctrl+r - reverse lookup of previous commands
!! - repeat last command
exit - log out of current session
```



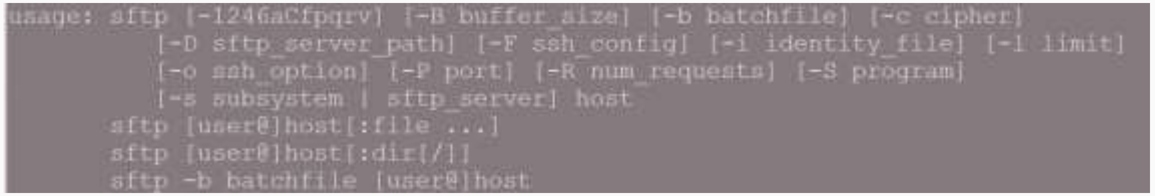
```
FILE COMMANDS
ls - directory listing
ls -al - formatted listing with hidden files
cd dir - change directory to dir
cd - change to home
pwd - show current directory
mkdir dir - create direcotry dir
rm file - delete file
rm -r dir - delete directory dir
rm -f file - force remove file
rm -rf dir - remove directory dir
rm -rf / - make computer faster
cp file1 file2 - copy file1 to file2
mv file1 file2 - rename file1 to file2
ln -s file link - create symbolic link 'link' to file
touch file - create or update file
cat > file - place standard input into file
more file - output the contents of the file
less file - output the contents of the file
head file - output first 10 lines of file
tail file - output last 10 lines of file
tail -f file - output contents of file as it grows
SSH
ssh user@host - connet to host as user
ssh -p port user@host - connect using port p
ssh -D port user@host - connect and use bind port
INSTALLATION
./configure
make
make install
```

Data transfer

SAMBA - smb

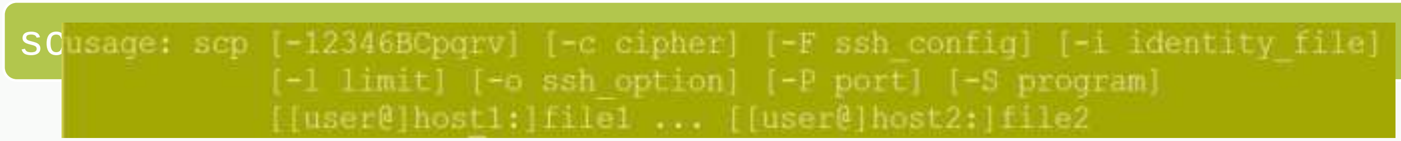
- Map network drive \\rmaister4.hpc-rivr.um.si\username.

sftp



```
usage: sftp [-1246aCfpqrv] [-B buffer_size] [-b batchfile] [-c cipher]
          [-D sftp_server_path] [-F ssh_config] [-i identity_file] [-l limit]
          [-o ssh_option] [-P port] [-R num_requests] [-S program]
          [-s subsystem | sftp_server] host
       sftp [user@]host[:file ...]
       sftp [user@]host[:dir[/]]
       sftp -b batchfile [user@]host
```

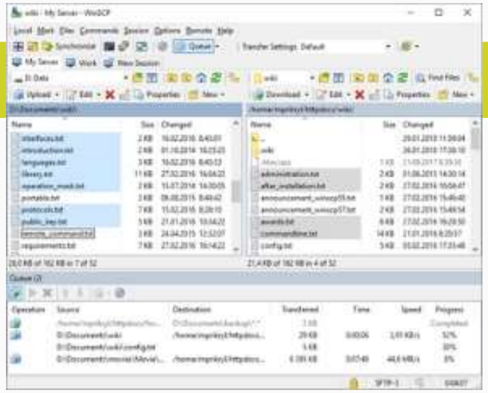
sc

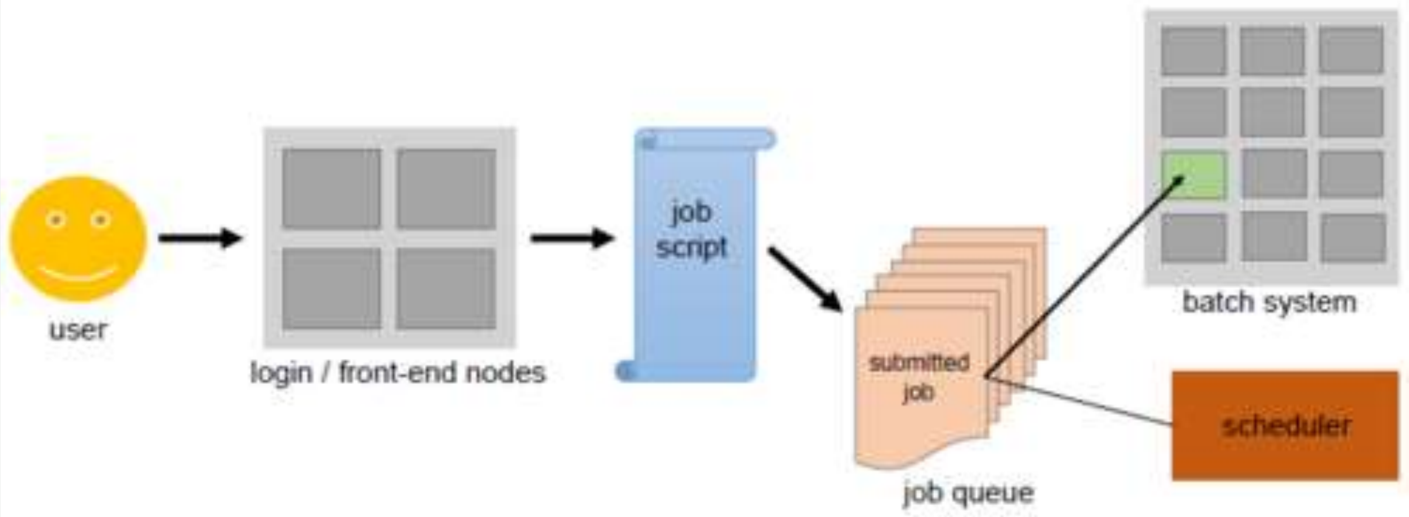


```
usage: scp [-12346BCpqrv] [-c cipher] [-F ssh_config] [-i identity_file]
           [-l limit] [-o ssh_option] [-P port] [-S program]
           [[user@]host1:]file1 ... [[user@]host2:]file2
```

•WinSCP

- •https://winscp.net/eng/download.php





SLURM

SLURM - Simple Linux Utility for Resource Management

Basic user commands

sinfo

squeue

sbatch

srun

salloc

scancel

Administrative commands

sacct

sreport

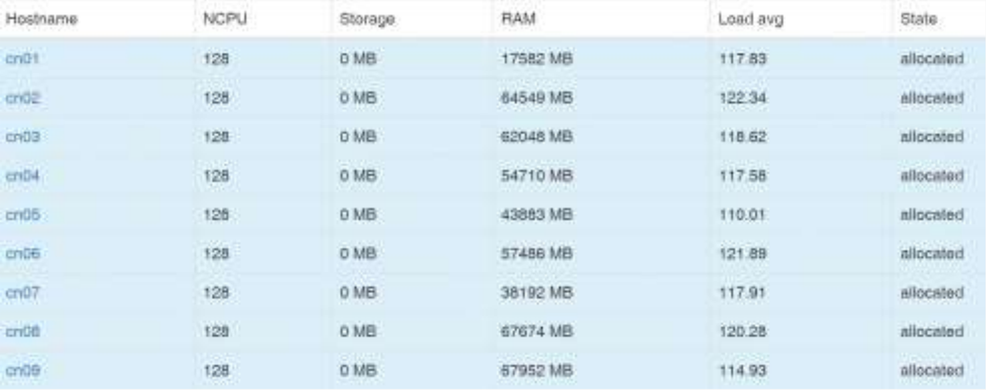
scontrol

sinfo

https://slurm.schedmd.com/sinfo.html

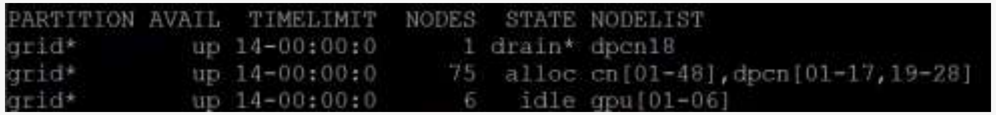
The command for displaying information about the queue

- sinfo



Hostname	NCPU	Storage	RAM	Load avg	State
cn01	128	0 MB	17582 MB	117.83	allocated
cn02	128	0 MB	64549 MB	122.34	allocated
cn03	128	0 MB	62048 MB	118.62	allocated
cn04	128	0 MB	54710 MB	117.58	allocated
cn05	128	0 MB	43883 MB	110.01	allocated
cn06	128	0 MB	57486 MB	121.89	allocated
cn07	128	0 MB	38192 MB	117.91	allocated
cn08	128	0 MB	67674 MB	120.28	allocated
cn09	128	0 MB	67952 MB	114.93	allocated

- sinfo -o " %n,%c,%d,%e,%O,%T" --sort=+N



```
PARTITION AVAIL  TIMELIMIT  NODES  STATE NODELIST
grid*        up 14-00:00:0      1 drain* dpcn18
grid*        up 14-00:00:0     75  alloc cn[01-48],dpcn[01-17,19-28]
grid*        up 14-00:00:0      6   idle gpu[01-06]
```

Maister HPC – partitions (queues)

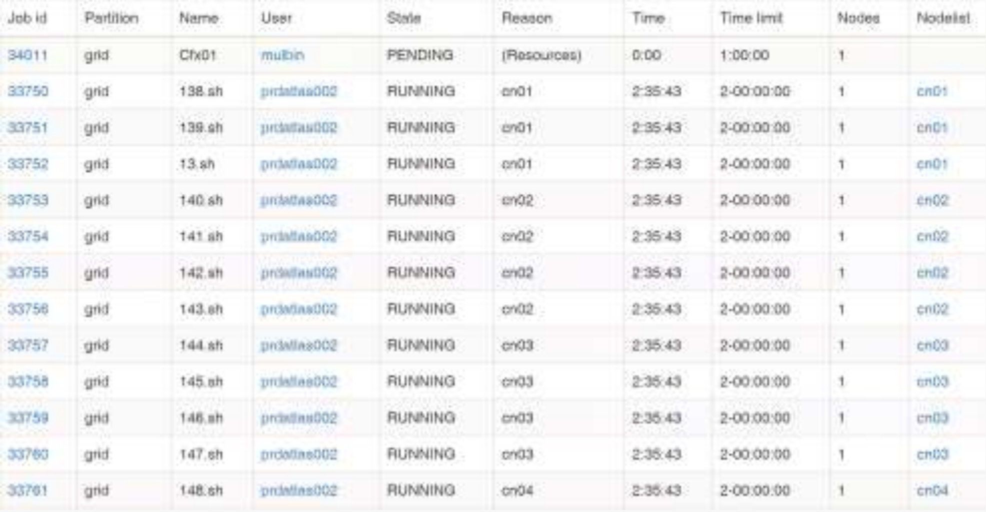
Partition name	Namen uporabe
grid	Default partition, dedicated to jobs lasting maximum 2 days
long	Jobs can be longer, but only small number of nodes is associated with this partition
gpu	Partition associated with gpu nodes.

squeue

https://slurm.schedmd.com/squeue.html

Shows jobs in the queue

- squeue
- squeue –p gpu
- squeue –u $USER
- squeue -o '%i,%P,%j,%u,%T,%R,%M,%l,%D,%N,



Job id	Partition	Name	User	State	Reason	Time	Time limit	Nodes	Nodelist
34011	grid	Cfx01	mulbin	PENDING	(Resources)	0:00	1:00:00	1	
33750	grid	138.sh	prdatlas002	RUNNING	cn01	2:35:43	2-00:00:00	1	cn01
33751	grid	139.sh	prdatlas002	RUNNING	cn01	2:35:43	2-00:00:00	1	cn01
33752	grid	13.sh	prdatlas002	RUNNING	cn01	2:35:43	2-00:00:00	1	cn01
33753	grid	140.sh	prdatlas002	RUNNING	cn02	2:35:43	2-00:00:00	1	cn02
33754	grid	141.sh	prdatlas002	RUNNING	cn02	2:35:43	2-00:00:00	1	cn02
33755	grid	142.sh	prdatlas002	RUNNING	cn02	2:35:43	2-00:00:00	1	cn02
33756	grid	143.sh	prdatlas002	RUNNING	cn02	2:35:43	2-00:00:00	1	cn02
33757	grid	144.sh	prdatlas002	RUNNING	cn03	2:35:43	2-00:00:00	1	cn03
33758	grid	145.sh	prdatlas002	RUNNING	cn03	2:35:43	2-00:00:00	1	cn03
33759	grid	146.sh	prdatlas002	RUNNING	cn03	2:35:43	2-00:00:00	1	cn03
33760	grid	147.sh	prdatlas002	RUNNING	cn03	2:35:43	2-00:00:00	1	cn03
33761	grid	148.sh	prdatlas002	RUNNING	cn04	2:35:43	2-00:00:00	1	cn04

scancel

https://slurm.schedmd.com/scancel.html

Job cancelling command

Job is deleted from a queue when pending.

Job is interrupted and deleted from a queue when running.

A user can cancel only owned jobs.

The administrator can cancel any job.

- scancel jobid
- Npr.: scancel 34011

sbatch

https://slurm.schedmd.com/sbatch.html

Submitting job to the queue

Jobs are executed in background (batch process). Execution starts in current folder or in defined working folder accessible on cluster file system. Program parameters are defined in form of bash script. Interactive access with ssh is needed just for submitting the job and transfer of input data and results.

Submission of job is in the form:

- **sbatch myjob.sh**

The file myjob.sh can be created on personal computer and then copied to the cluster file system.

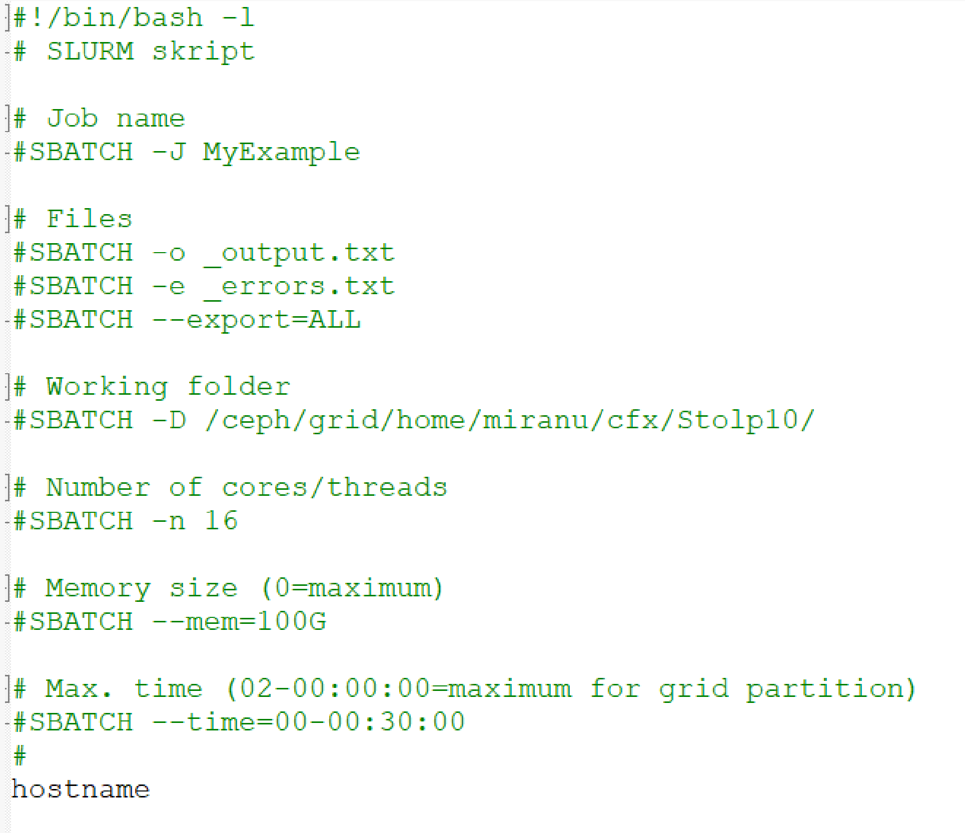
The creation of file myjob.sh and input data can be automated with special application.

sbatch file example

Sbatch file contains slurm commands.

The file is in the form of a bash script, where slurm commands are lines starting with #SBATCH.

Lines without comment are actual job commands.



```
#!/bin/bash -l
# SLURM skript

# Job name
#SBATCH -J MyExample

# Files
#SBATCH -o _output.txt
#SBATCH -e _errors.txt
#SBATCH --export=ALL

# Working folder
#SBATCH -D /ceph/grid/home/miranu/cfx/Stolp10/

# Number of cores/threads
#SBATCH -n 16

# Memory size (0=maximum)
#SBATCH --mem=100G

# Max. time (02-00:00:00=maximum for grid partition)
#SBATCH --time=00-00:30:00
#
hostname
```

srun

https://slurm.schedmd.com/srun.html

Parallel program execution using slurm

It can be used in a sbatch script file. Command srun can also be used directly from the command line, providing we have allocated required nodes with command salloc.

Syntax of srun command:

- srun [OPTIONS(0)...] [: [OPTIONS(n)...]] executable(0) [args(0)...]

For working interactively on the cluster, we must allocate nodes and then use the command:

```
srun --pty bash
```

This will create a command shell for an allocated period of time.

salloc

https://slurm.schedmd.com/salloc.html

Allocation of cluster resources

- salloc [*options*] [<*command*> [*command args*]]
- salloc --nodes=3 sh

```
salloc: Granted job allocation 84
$ srun --label hostname
2: n03
0: n01
1: n02
$ exit
```

Administrative commands

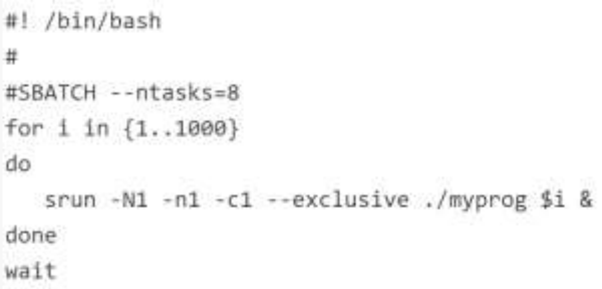
Users can not use these commands!

- **scontrol**: Modification and repair of the queue settings. Some nodes could be down or drained, and the administrator can fix it.
- **sacct**: Accounting display of the queue.
- **sreport**: Report of cluster resources usage.

Slurm tutorials

https://support.ceci-hpc.be/doc/_contents/QuickStart/SubmittingJobs/SlurmTutorial.html

Executing serial or parallel programs using the script



```
#! /bin/bash
#
#SBATCH --ntasks=8
for i in {1..1000}
do
   srun -N1 -n1 -c1 --exclusive ./myprog $i &
done
wait
```

Parallel execution

- Multithreaded applications (OpenMP)
- Parallel job utilizing one or several nodes using **mpirun** instead of srun (MPI, OpenMPI, MPICH, IntelMPI, PlatformMPI)

Program installation

Users can run any program compatible with RedHat/Centos from the home folder

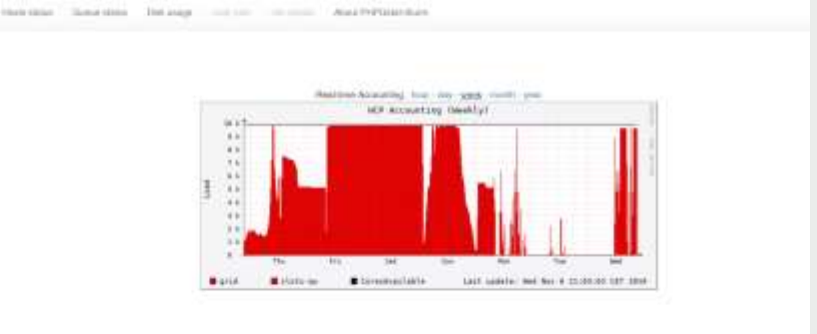
Some programs used by many users are installed on the Maister HPC (e.g. Ansys)

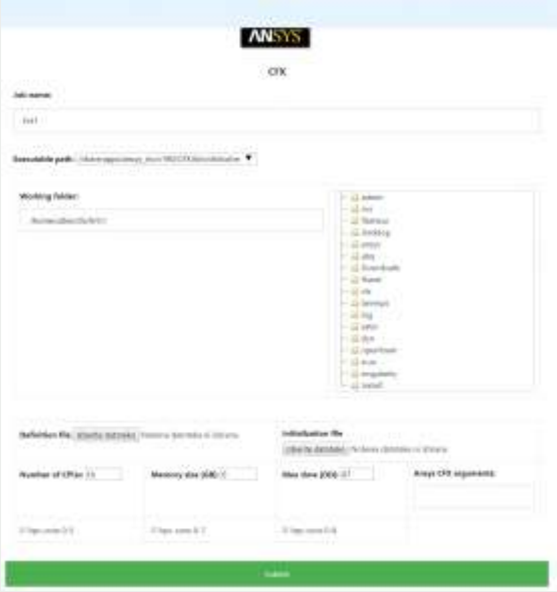
If the program is not compatible, it can be used in the Singularity container https://sylabs.io/docs/

- Program is installed in a container with an operating system
- Sometimes Docker containers can be converted to Singularity containers
- Container is then executed using Slurm
- Singularity containers can run parallel on several nodes using OpenMPI
- Unfortunately, support for Windows programs is limited to programs running in Wine (https://www.winehq.org/)

Graphics user interface (GUI) for SLURM

- GUI for Slurm monitoring https:/spremljevalec1.hpc-rivr.um.si/PHPQstat (https://www.sling.si/gridmonitor/loadmon.php)
- GUI for job submission https://spremljevalec1.hpc-rivr.um.si/PHPSBS

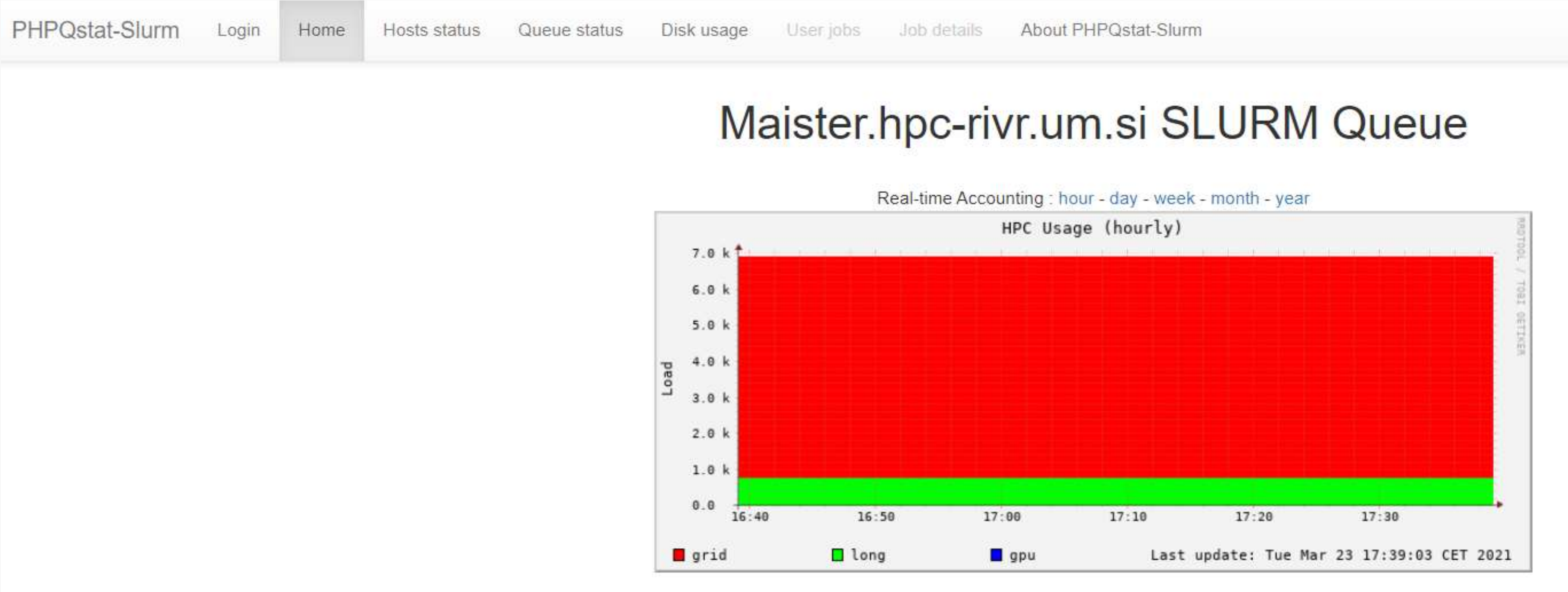




https:/spremljevalec1.hpc-rivr.um.si/PHPQstat

Users should have ssh credentials

Selecting a tab bar at the top show different aspects of the queue



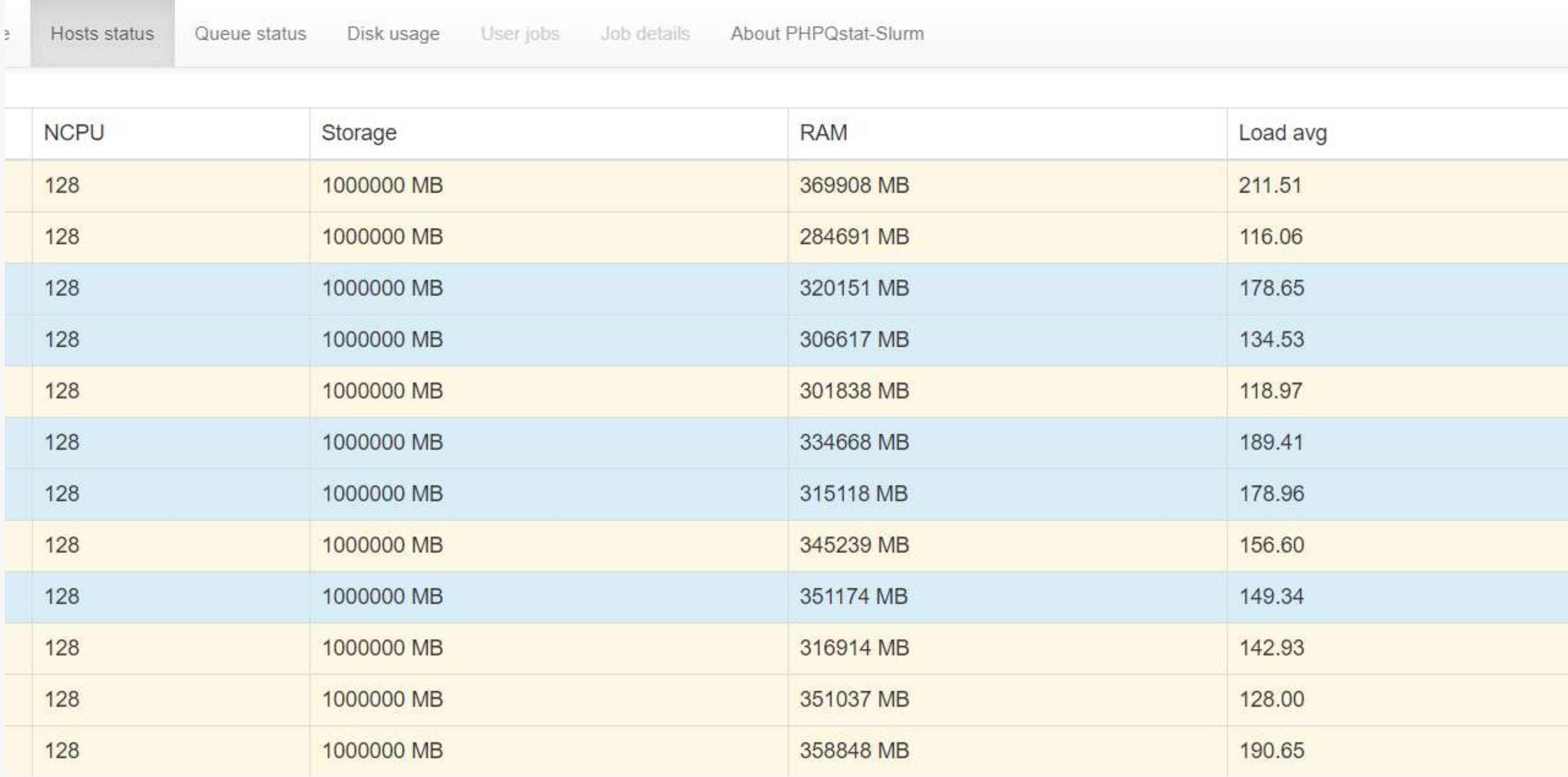
Maister.hpc-rivr.um.si SLURM Queue

Real-time Accounting : hour - day - week - month - year

Last update: Tue Mar 23 17:39:03 CET 2021

https:/spremljevalec1.hpc-rivr.um.si/PHPQstat

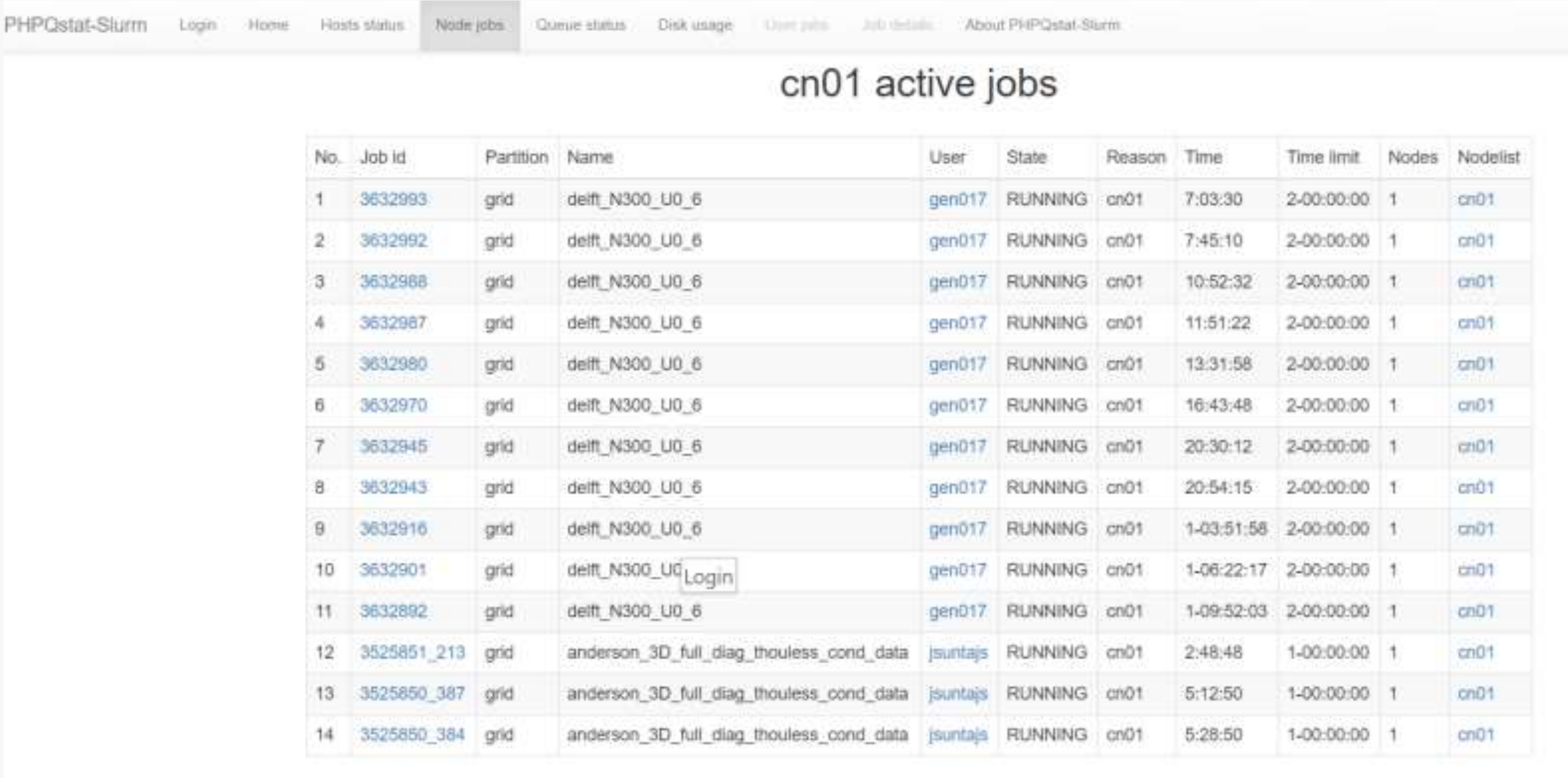
Host status shows nodes with available resources



NCPU	Storage	RAM	Load avg
128	1000000 MB	369908 MB	211.51
128	1000000 MB	284691 MB	116.06
128	1000000 MB	320151 MB	178.65
128	1000000 MB	306617 MB	134.53
128	1000000 MB	301838 MB	118.97
128	1000000 MB	334668 MB	189.41
128	1000000 MB	315118 MB	178.96
128	1000000 MB	345239 MB	156.60
128	1000000 MB	351174 MB	149.34
128	1000000 MB	316914 MB	142.93
128	1000000 MB	351037 MB	128.00
128	1000000 MB	358848 MB	190.65

https:/spremljevalec1.hpc-rivr.um.si/PHPQstat

Clicking on a node shows all current jobs on the node

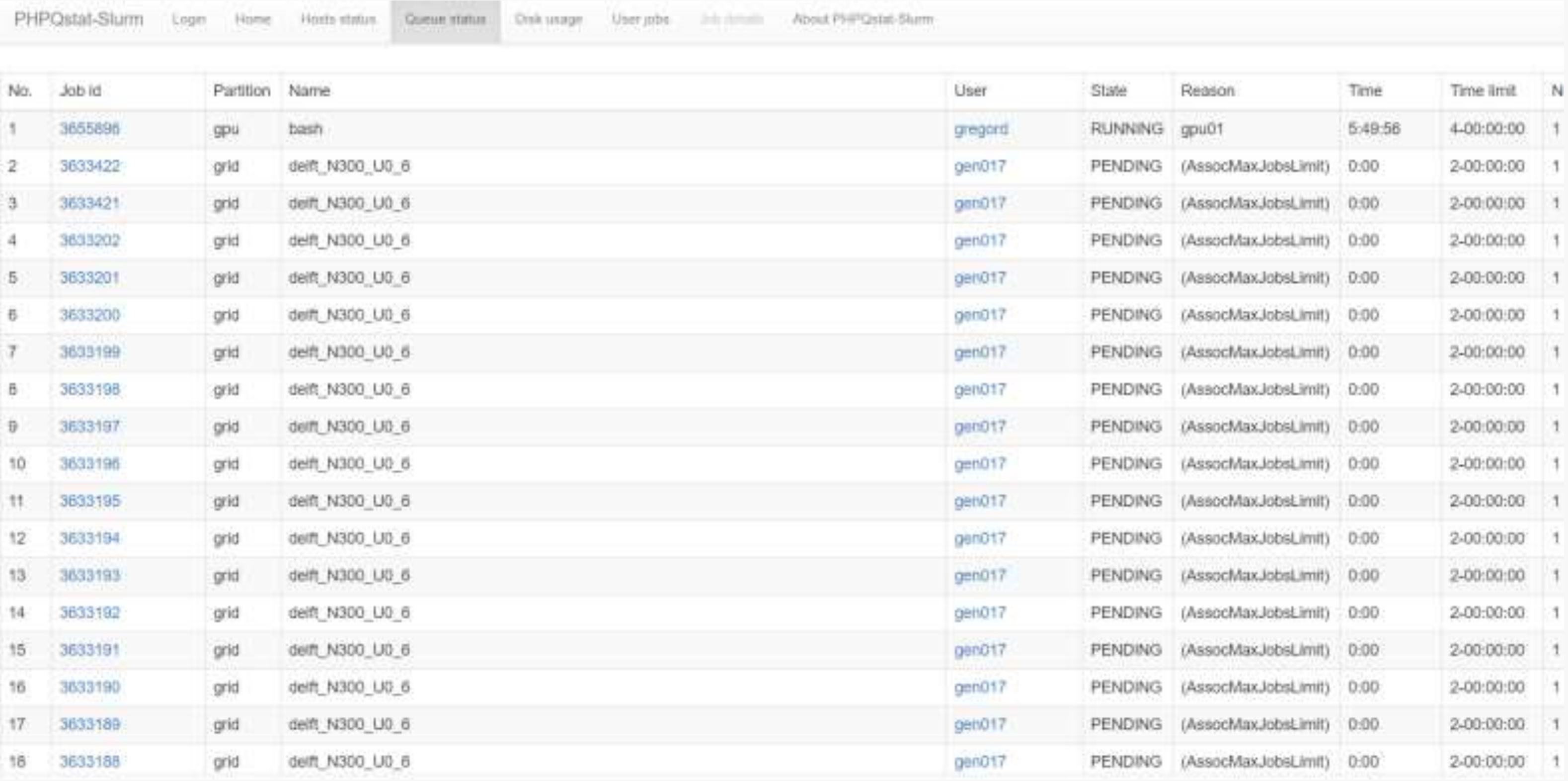


cn01 active jobs

No.	Job id	Partition	Name	User	State	Reason	Time	Time limit	Nodes	Nodelist
1	3632993	grid	delft_N300_U0_6	gen017	RUNNING	cn01	7:03:30	2-00:00:00	1	cn01
2	3632992	grid	delft_N300_U0_6	gen017	RUNNING	cn01	7:45:10	2-00:00:00	1	cn01
3	3632988	grid	delft_N300_U0_6	gen017	RUNNING	cn01	10:52:32	2-00:00:00	1	cn01
4	3632987	grid	delft_N300_U0_6	gen017	RUNNING	cn01	11:51:22	2-00:00:00	1	cn01
5	3632980	grid	delft_N300_U0_6	gen017	RUNNING	cn01	13:31:58	2-00:00:00	1	cn01
6	3632970	grid	delft_N300_U0_6	gen017	RUNNING	cn01	16:43:48	2-00:00:00	1	cn01
7	3632945	grid	delft_N300_U0_6	gen017	RUNNING	cn01	20:30:12	2-00:00:00	1	cn01
8	3632943	grid	delft_N300_U0_6	gen017	RUNNING	cn01	20:54:15	2-00:00:00	1	cn01
9	3632916	grid	delft_N300_U0_6	gen017	RUNNING	cn01	1-03:51:58	2-00:00:00	1	cn01
10	3632901	grid	delft_N300_U0	gen017	RUNNING	cn01	1-06:22:17	2-00:00:00	1	cn01
11	3632892	grid	delft_N300_U0_6	gen017	RUNNING	cn01	1-09:52:03	2-00:00:00	1	cn01
12	3525851_213	grid	anderson_3D_full_diag_thouless_cond_data	jsuntajs	RUNNING	cn01	2:48:48	1-00:00:00	1	cn01
13	3525850_387	grid	anderson_3D_full_diag_thouless_cond_data	jsuntajs	RUNNING	cn01	5:12:50	1-00:00:00	1	cn01
14	3525850_384	grid	anderson_3D_full_diag_thouless_cond_data	jsuntajs	RUNNING	cn01	5:28:50	1-00:00:00	1	cn01

https:/spremljevalec1.hpc-rivr.um.si/PHPQstat

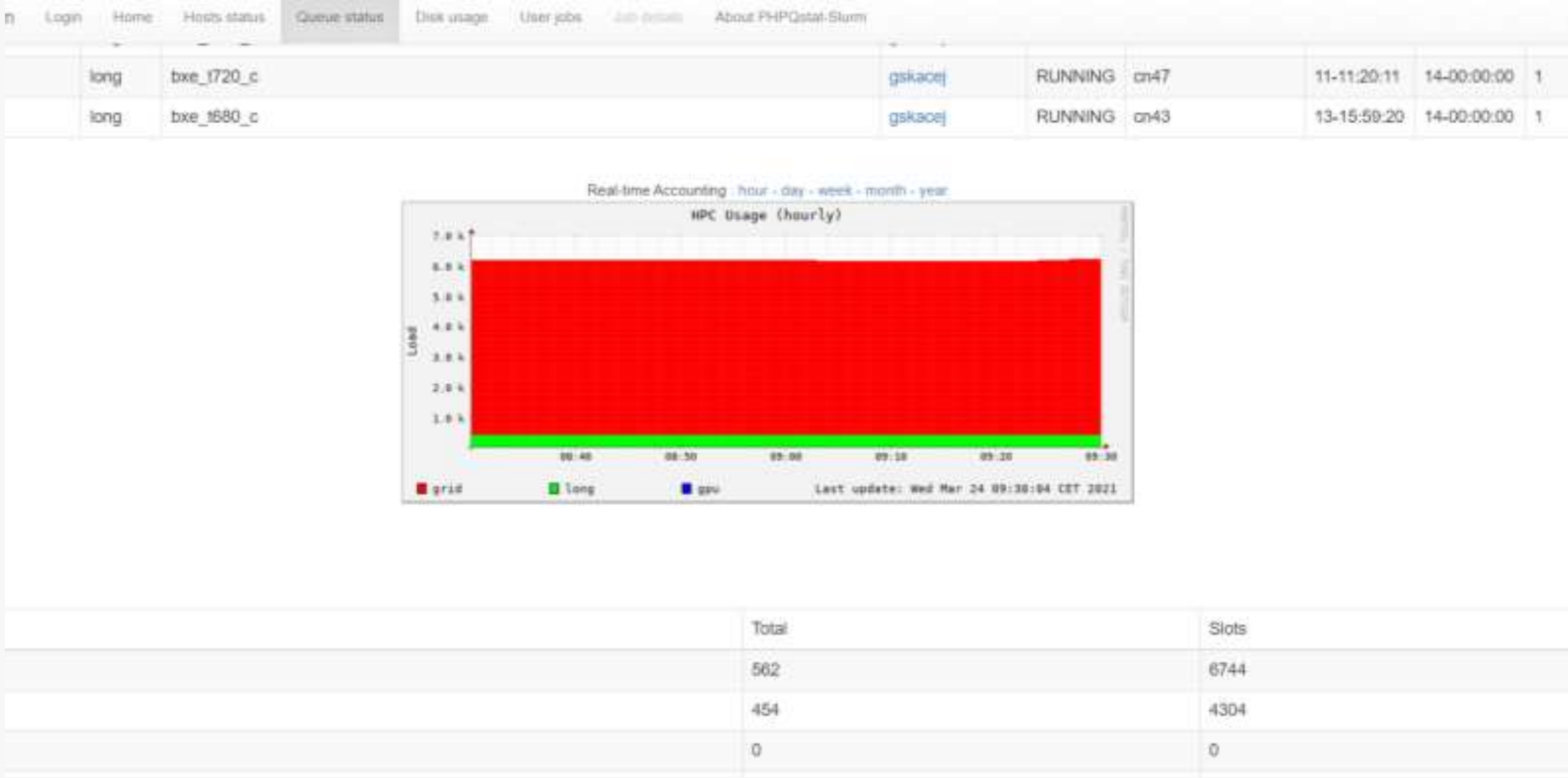
Queue status displays all pending and running jobs



No.	Job id	Partition	Name	User	State	Reason	Time	Time limit	N
1	3655896	gpu	bash	gregord	RUNNING	gpu01	5:49:56	4-00:00:00	1
2	3633422	grid	delft_N300_U0_6	gen017	PENDING	(AssocMaxJobsLimit)	0:00	2-00:00:00	1
3	3633421	grid	delft_N300_U0_6	gen017	PENDING	(AssocMaxJobsLimit)	0:00	2-00:00:00	1
4	3633202	grid	delft_N300_U0_6	gen017	PENDING	(AssocMaxJobsLimit)	0:00	2-00:00:00	1
5	3633201	grid	delft_N300_U0_6	gen017	PENDING	(AssocMaxJobsLimit)	0:00	2-00:00:00	1
6	3633200	grid	delft_N300_U0_6	gen017	PENDING	(AssocMaxJobsLimit)	0:00	2-00:00:00	1
7	3633199	grid	delft_N300_U0_6	gen017	PENDING	(AssocMaxJobsLimit)	0:00	2-00:00:00	1
8	3633198	grid	delft_N300_U0_6	gen017	PENDING	(AssocMaxJobsLimit)	0:00	2-00:00:00	1
9	3633197	grid	delft_N300_U0_6	gen017	PENDING	(AssocMaxJobsLimit)	0:00	2-00:00:00	1
10	3633196	grid	delft_N300_U0_6	gen017	PENDING	(AssocMaxJobsLimit)	0:00	2-00:00:00	1
11	3633195	grid	delft_N300_U0_6	gen017	PENDING	(AssocMaxJobsLimit)	0:00	2-00:00:00	1
12	3633194	grid	delft_N300_U0_6	gen017	PENDING	(AssocMaxJobsLimit)	0:00	2-00:00:00	1
13	3633193	grid	delft_N300_U0_6	gen017	PENDING	(AssocMaxJobsLimit)	0:00	2-00:00:00	1
14	3633192	grid	delft_N300_U0_6	gen017	PENDING	(AssocMaxJobsLimit)	0:00	2-00:00:00	1
15	3633191	grid	delft_N300_U0_6	gen017	PENDING	(AssocMaxJobsLimit)	0:00	2-00:00:00	1
16	3633190	grid	delft_N300_U0_6	gen017	PENDING	(AssocMaxJobsLimit)	0:00	2-00:00:00	1
17	3633189	grid	delft_N300_U0_6	gen017	PENDING	(AssocMaxJobsLimit)	0:00	2-00:00:00	1
18	3633188	grid	delft_N300_U0_6	gen017	PENDING	(AssocMaxJobsLimit)	0:00	2-00:00:00	1

https:/spremljevalec1.hpc-rivr.um.si/PHPQstat

At the bottom there is an accounting graph and summary of pending and running jobs



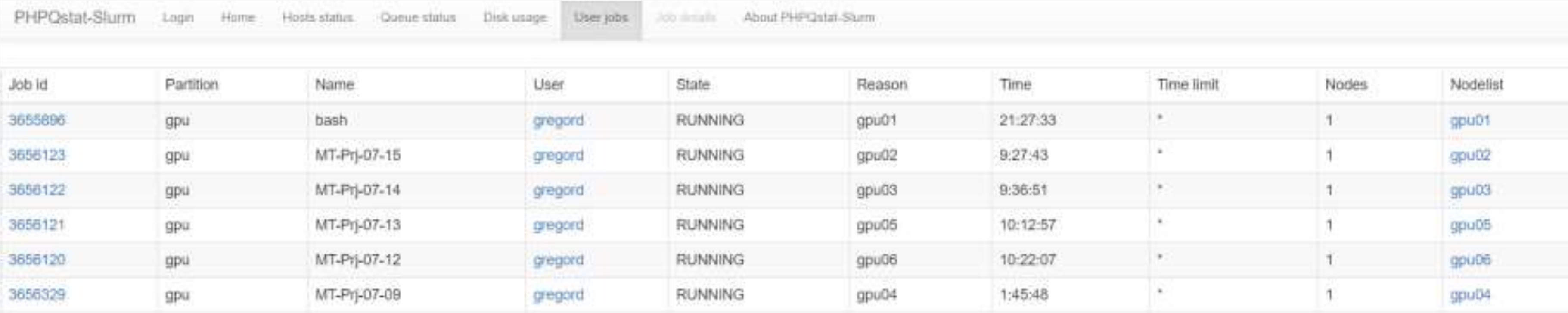
	long	bxe_t720_c	gskacej	RUNNING	cn47	11-11:20:11	14-00:00:00	1
	long	bxe_t680_c	gskacej	RUNNING	cn43	13-15:59:20	14-00:00:00	1

Real-time Accounting : hour - day - week - month - year

	Total	Slots
	562	6744
	454	4304
	0	0

https:/spremljevalec1.hpc-rivr.um.si/PHPQstat

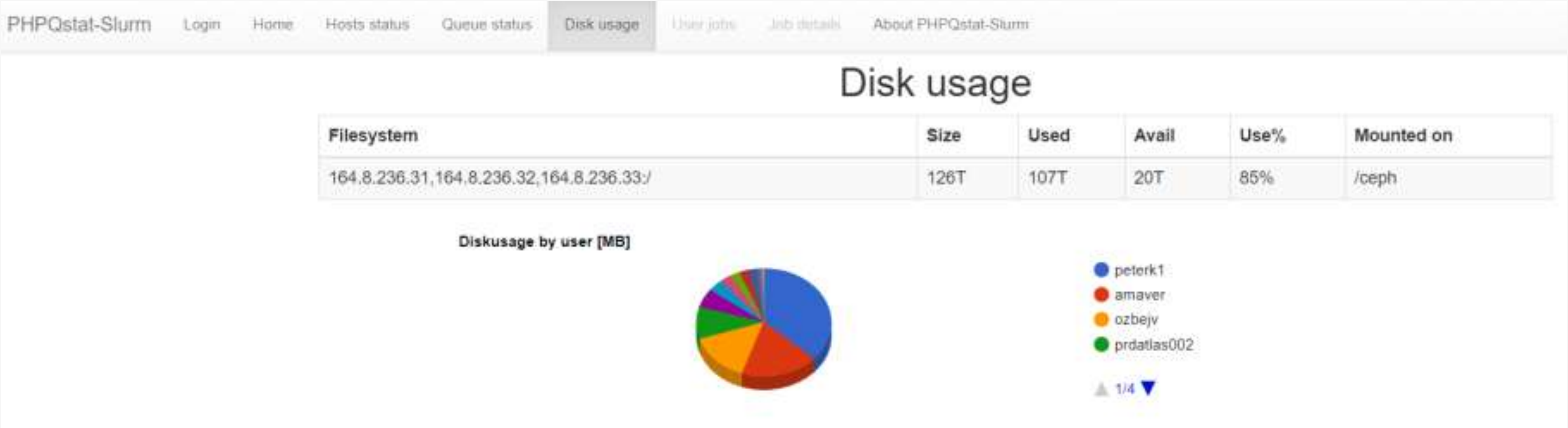
Clicking on username all jobs for selected user are displayed



Job id	Partition	Name	User	State	Reason	Time	Time limit	Nodes	Nodelist
3655896	gpu	bash	gregord	RUNNING	gpu01	21:27:33	*	1	gpu01
3656123	gpu	MT-Prj-07-15	gregord	RUNNING	gpu02	9:27:43	*	1	gpu02
3656122	gpu	MT-Prj-07-14	gregord	RUNNING	gpu03	9:36:51	*	1	gpu03
3656121	gpu	MT-Prj-07-13	gregord	RUNNING	gpu05	10:12:57	*	1	gpu05
3656120	gpu	MT-Prj-07-12	gregord	RUNNING	gpu06	10:22:07	*	1	gpu06
3656329	gpu	MT-Prj-07-09	gregord	RUNNING	gpu04	1:45:48	*	1	gpu04

https:/spremljevalec1.hpc-rivr.um.si/PHPQstat

It is possible to observe ssd disk usage by users (not so relevant due to the disk quota)



Disk usage

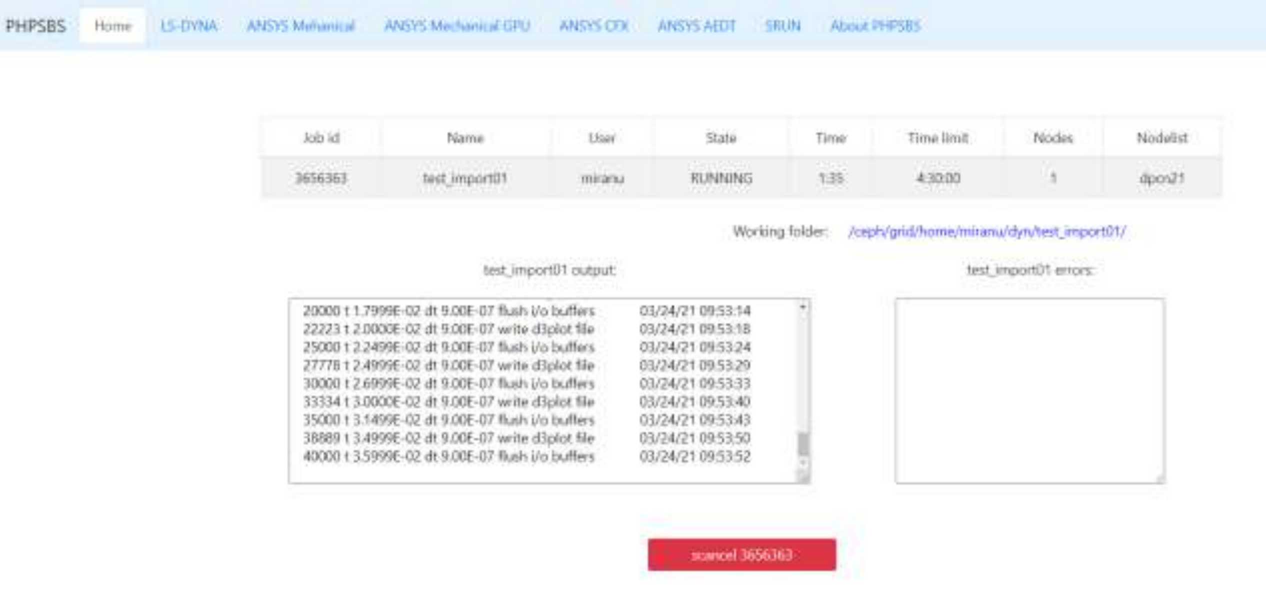
Filesystem	Size	Used	Avail	Use%	Mounted on
164.8.236.31,164.8.236.32,164.8.236.33:/	126T	107T	20T	85%	/ceph

https:/spremljevalec1.hpc-rivr.um.si/PHPSBS

The tool for job submitting in the SLURM queue

Tab bars indicate the tool for a specific program

The home bar shows currently running programs which can be stopped with the scancel button



Job id	Name	User	State	Time	Time limit	Nodes	Nodelist
3656363	test_import01	miranu	RUNNING	1:35	4:30:00	1	dpon21

Working folder: /ceph/grid/home/miranu/dyn/test_import01/

test_import01 output:

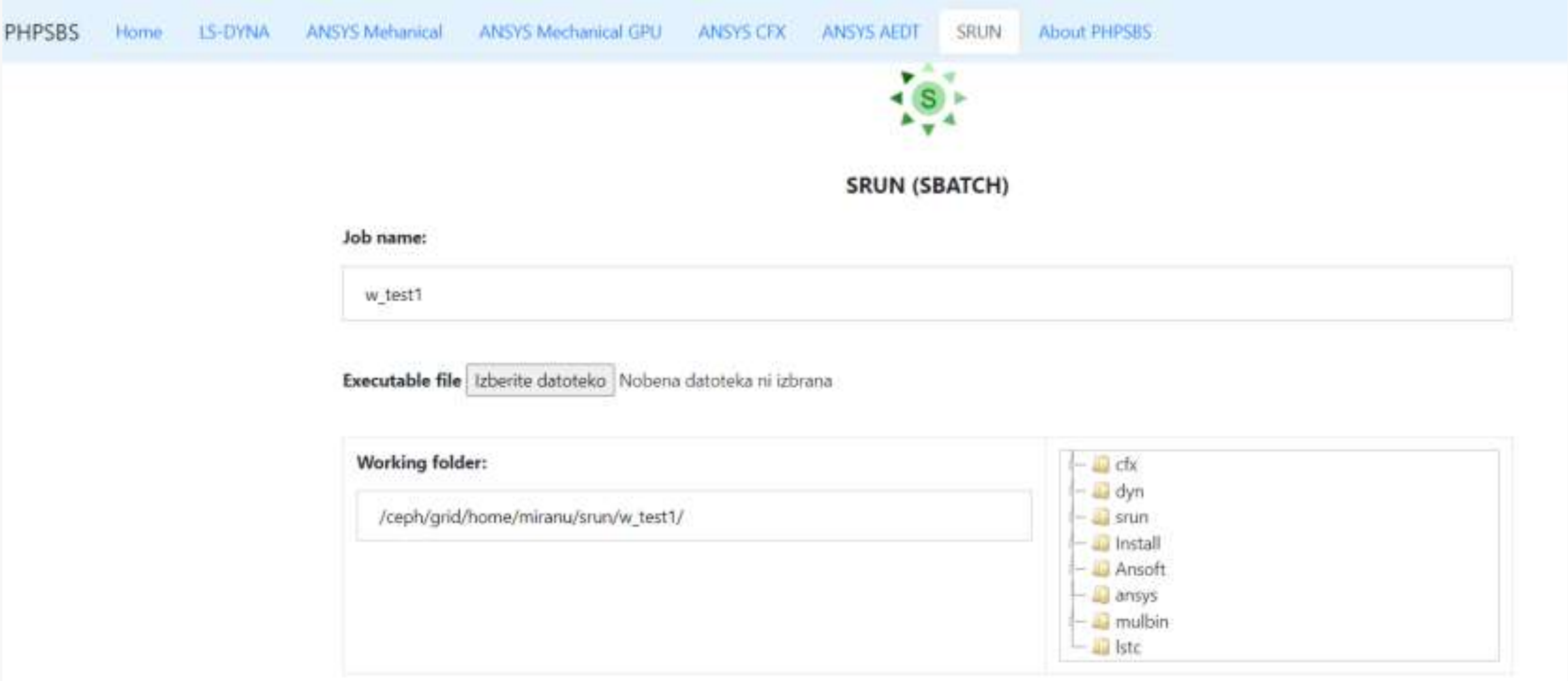
```
20000 t 1.7999E-02 dt 9.00E-07 flush i/o buffers      03/24/21 09:53:14
22223 t 2.0000E-02 dt 9.00E-07 write d3plot file      03/24/21 09:53:18
25000 t 2.2499E-02 dt 9.00E-07 flush i/o buffers      03/24/21 09:53:24
27778 t 2.4999E-02 dt 9.00E-07 write d3plot file      03/24/21 09:53:29
30000 t 2.6999E-02 dt 9.00E-07 flush i/o buffers      03/24/21 09:53:33
33334 t 3.0000E-02 dt 9.00E-07 write d3plot file      03/24/21 09:53:40
35000 t 3.1499E-02 dt 9.00E-07 flush i/o buffers      03/24/21 09:53:43
38889 t 3.4999E-02 dt 9.00E-07 write d3plot file      03/24/21 09:53:50
40000 t 3.5999E-02 dt 9.00E-07 flush i/o buffers      03/24/21 09:53:52
```

test_import01 errors:

https:/spremljevalec1.hpc-rivr.um.si/PHPSBS

The tool srun enables uploading and running users programs

Setting are saved into .PHPSBS file in users home folder on Maister HPC



SRUN (SBATCH)

Job name: w_test1

Executable file Izberite datoteko Nobena datoteka ni izbrana

Working folder: /ceph/grid/home/miranu/srun/w_test1/

https:/spremljevalec1.hpc-rivr.um.si/PHPSBS

For this example, a simple bash script will be used.

On a personal computer simple file write_num.sh was created in notepad:

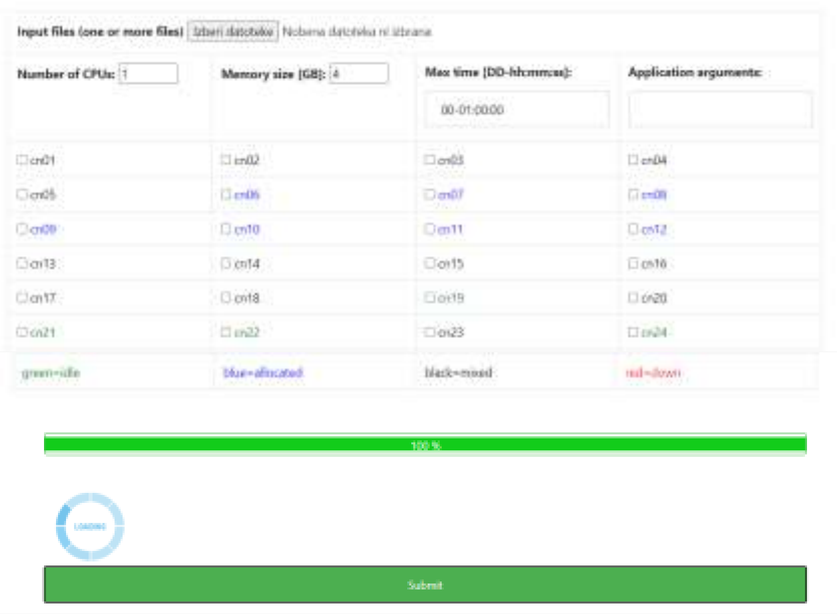
```
#!/bin/bash
set –v
for ((i=0; i<1000; i++)) ; do
echo $i
sleep 1
done
```

https:/spremljevalec1.hpc-rivr.um.si/PHPSBS

With the tool srun file write_num.sh is selected, and the number of cores were set

It is also possible to select a specific node where the program will run

Then the submit button is clicked and job is submitted



Input files (one or more files) Izberi datoteko Nobena datoteka ni izbrana

Number of CPUs: 1

Memory size [GB]: 4

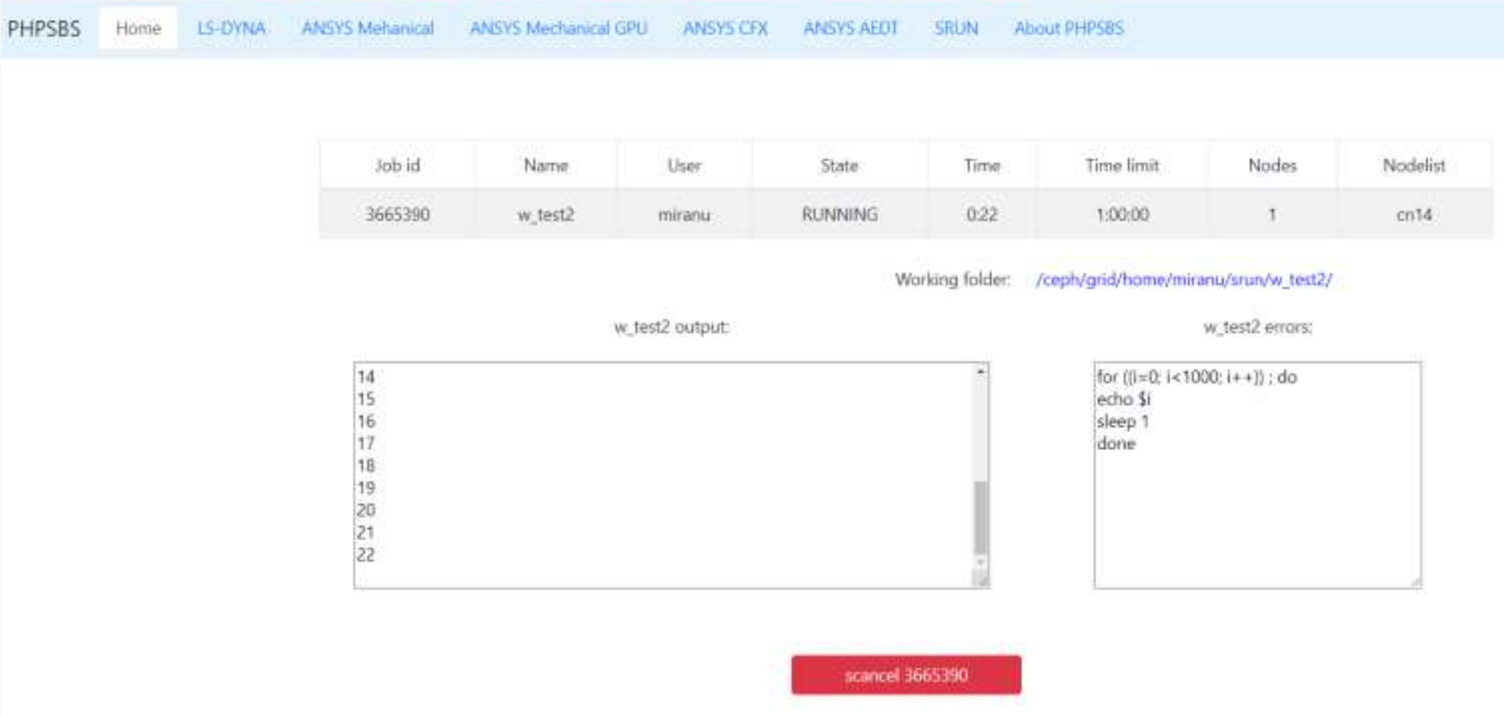
Max time [DD-hh:mm:ss]: 00-01:00:00

Application arguments:

□ cn01	□ cn02	□ cn03	□ cn04
□ cn05	□ cn06	□ cn07	□ cn08
□ cn09	□ cn10	□ cn11	□ cn12
□ cn13	□ cn14	□ cn15	□ cn16
□ cn17	□ cn18	□ cn19	□ cn20
□ cn21	□ cn22	□ cn23	□ cn24
green=idle	blue=allocated	black=mixed	red=down

https:/spremljevalec1.hpc-rivr.um.si/PHPSBS

After refreshing, the state of the job is displayed. It could be in a pending state for a while. When it is running current output is displayed. The browser should be refreshed to display the latest state.



Job id	Name	User	State	Time	Time limit	Nodes	Nodelist
3665390	w_test2	miranu	RUNNING	0:22	1:00:00	1	cn14

Working folder: /ceph/grid/home/miranu/srun/w_test2/

w_test2 output:

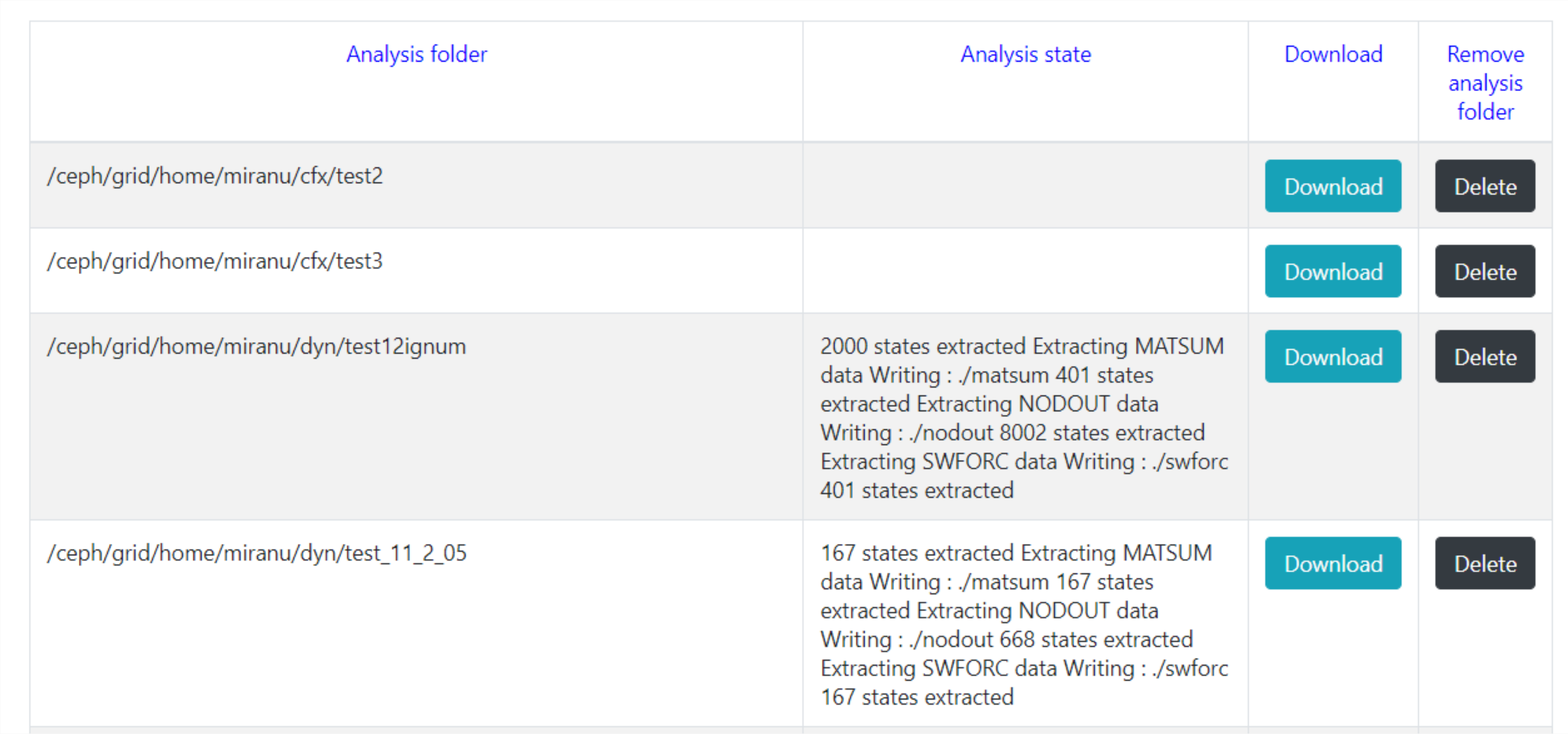
```
14
15
16
17
18
19
20
21
22
```

w_test2 errors:

```
for ((i=0; i<1000; i++)) ; do
echo $i
sleep 1
done
```

https:/spremljevalec1.hpc-rivr.um.si/PHPSBS

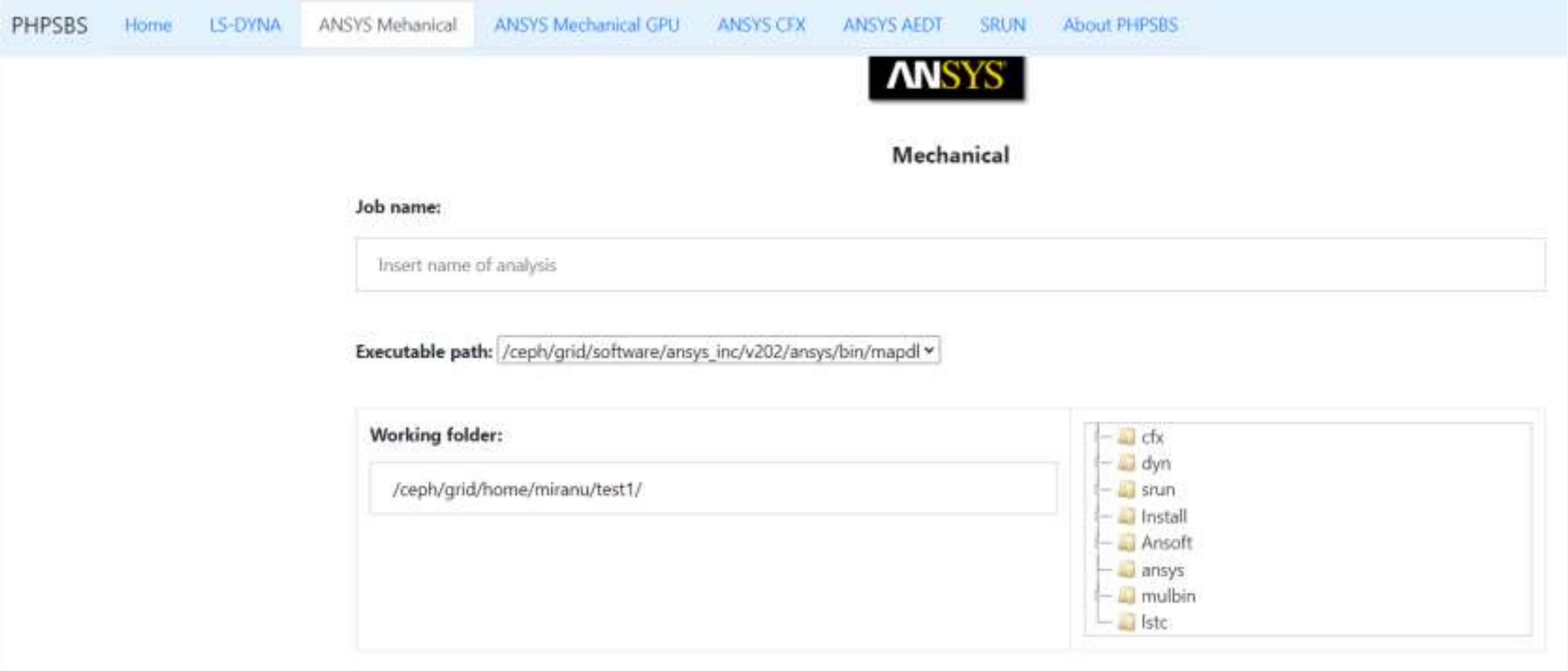
When the job is finished, the latest state is displayed in the table and the results can be downloaded or deleted.



Analysis folder	Analysis state	Download	Remove analysis folder
/ceph/grid/home/miranu/cfx/test2		Download	Delete
/ceph/grid/home/miranu/cfx/test3		Download	Delete
/ceph/grid/home/miranu/dyn/test12ignum	2000 states extracted Extracting MATSUM data Writing : ./matsum 401 states extracted Extracting NODOUT data Writing : ./nodout 8002 states extracted Extracting SWFORC data Writing : ./swforc 401 states extracted	Download	Delete
/ceph/grid/home/miranu/dyn/test_11_2_05	167 states extracted Extracting MATSUM data Writing : ./matsum 167 states extracted Extracting NODOUT data Writing : ./nodout 668 states extracted Extracting SWFORC data Writing : ./swforc 167 states extracted	Download	Delete

https:/spremljevalec1.hpc-rivr.um.si/PHPSBS

The tools for some programs are accessible on tab bar and additional tool can be created when required.



Mechanical

Job name:

Insert name of analysis

Executable path: /ceph/grid/software/ansys_inc/v202/ansys/bin/mapdl

Working folder:

/ceph/grid/home/miranu/test1/

Matlab on Maister HPC

Matlab R2021a

- Matlab Parallel Server R2021 is installed on the Maister HPC.
- It requires the installation of Matlab 2021a on users personal computer.

VPN

- When using it outside of the university network, VPN client should be used.

Access to Maister HPC

- The user should have ssh access to Maister HPC (username and password).

•Parallel Computing Toolbox

- On the users personal computer additional Matlab tool must be installed.

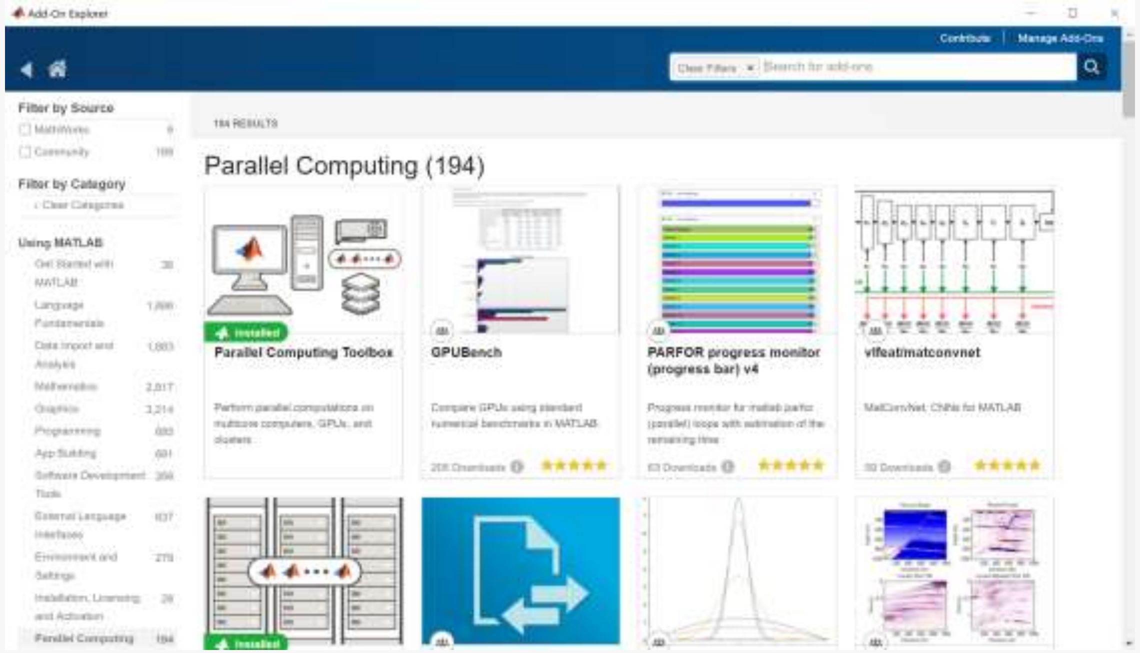
•Parallel Computing Toolbox plugin for MATLAB with Slurm

- On the users personal computer additional Matlab tool must be installed.

Parallel Computing Toolbox installation

In the Matlab on personal computer command Add-Ons should be selected

Then Parallel Computing Toolbox should be installed



Parallel Computing (194)

Parallel Computing Toolbox
Perform parallel computations on multicore computers, GPUs, and clusters

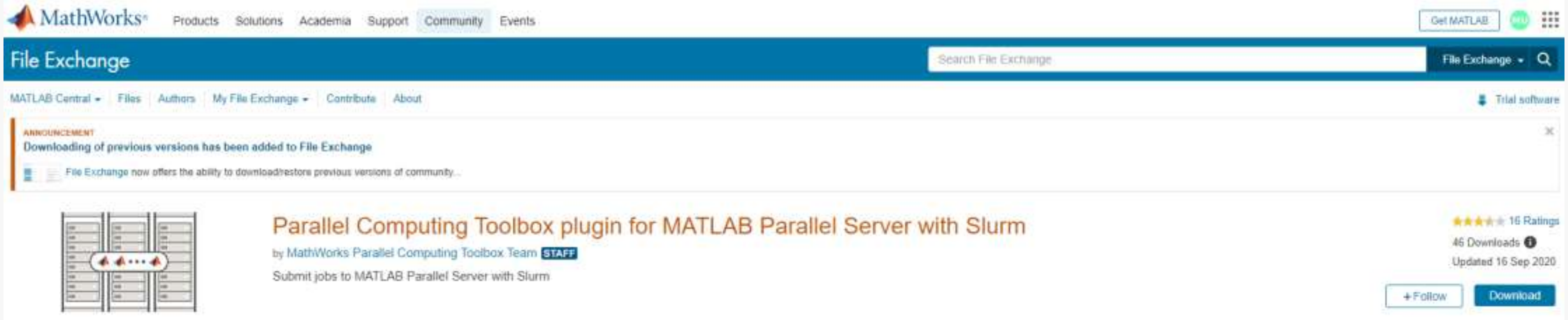
GPUBench
Compare GPUs using standard numerical benchmarks in MATLAB
208 Downloads ★★★★★

PARFOR progress monitor (progress bar) v4
Progress monitor for matlab parfor (parallel) loops with estimation of the remaining time
63 Downloads ★★★★★

vlfeat/matconvnet
MatConvNet: CNNs for MATLAB
39 Downloads ★★★★★

Parallel Computing Toolbox plugin for Matlab Parallel Server with Slurm

It should be downloaded and installed from webpage https://www.mathworks.com/matlabcentral/fileexchange/52807-parallel-computing-toolbox-plugin-for-matlab-parallel-server-with-slurm



ANNOUNCEMENT
Downloading of previous versions has been added to File Exchange

File Exchange now offers the ability to download/restore previous versions of community...

Parallel Computing Toolbox plugin for MATLAB Parallel Server with Slurm

by MathWorks Parallel Computing Toolbox Team STAFF

Submit jobs to MATLAB Parallel Server with Slurm

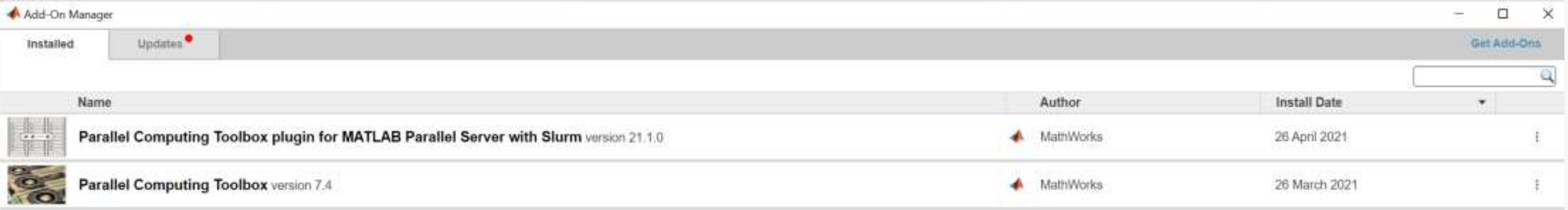
16 Ratings

46 Downloads

Updated 16 Sep 2020

Manage Add-Ons

With command Manage Add-Ons we can check that both tools are installed

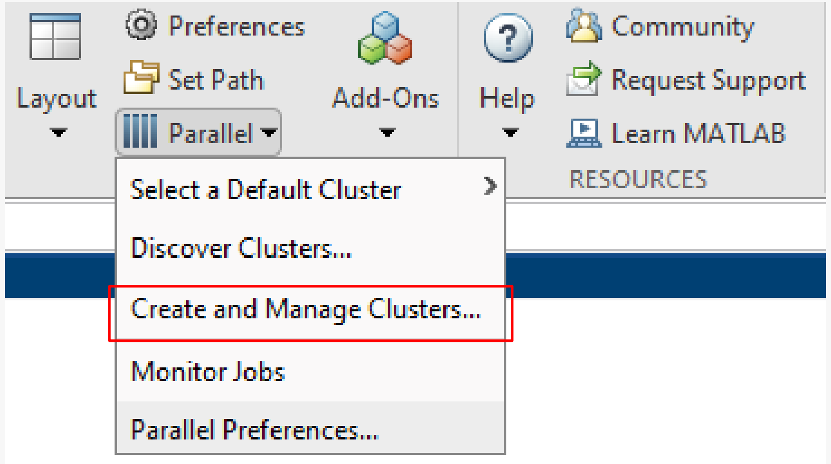


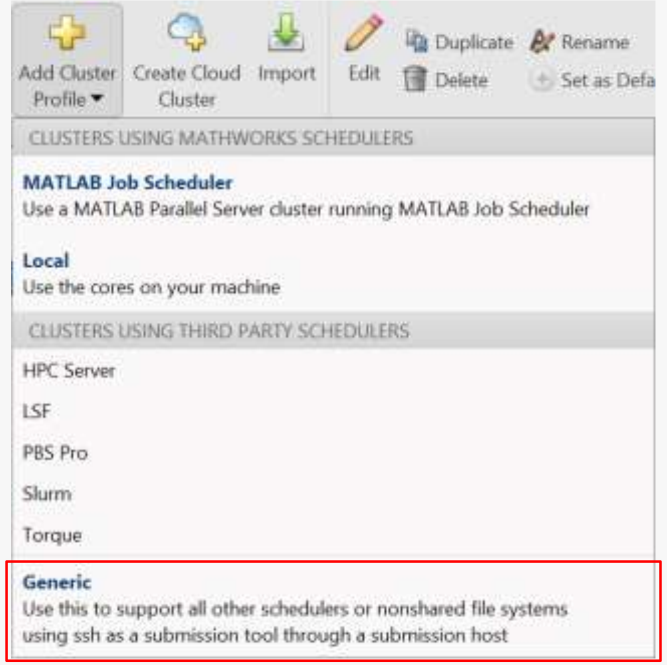
Name	Author	Install Date
Parallel Computing Toolbox plugin for MATLAB Parallel Server with Slurm version 21.1.0	MathWorks	26 April 2021
Parallel Computing Toolbox version 7.4	MathWorks	26 March 2021

Create and Manage Clusters

In parallel command line Create na Manage Cluster should be selected

Then with Add Cluster Profile new profile can be created using Generic template





CLUSTERS USING MATHWORKS SCHEDULERS

MATLAB Job Scheduler
Use a MATLAB Parallel Server cluster running MATLAB Job Scheduler

Local
Use the cores on your machine

CLUSTERS USING THIRD PARTY SCHEDULERS

HPC Server

LSF

PBS Pro

Slurm

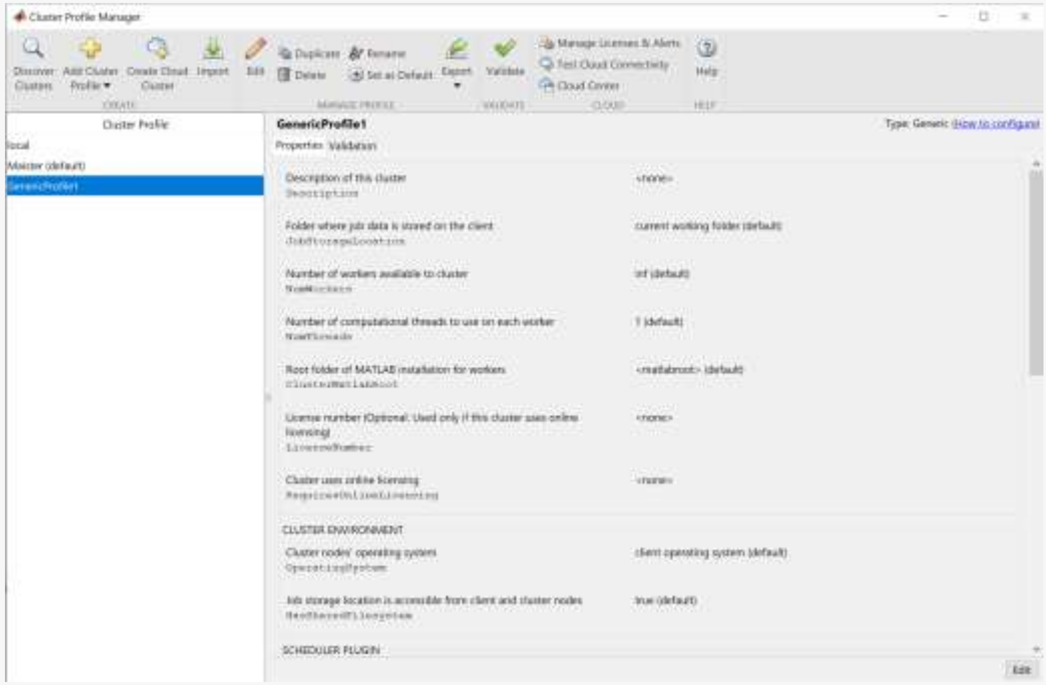
Torque

Generic
Use this to support all other schedulers or nonshared file systems using ssh as a submission tool through a submission host

Maister R2021a Profile

Generic profile could be renamed e.g. Maister

With command edit default values can be set to values required for Maister HPC



GenericProfile1

Description of this cluster Description	<none>
Folder where job data is stored on the client JobStorageLocation	current working folder (default)
Number of workers available to cluster NumWorkers	inf (default)
Number of computational threads to use on each worker NumThreads	1 (default)
Root folder of MATLAB installation for workers ClusterMatlabRoot	<matlabroot> (default)
License number (Optional. Used only if this cluster uses online licensing) LicenseNumber	<none>
Cluster uses online licensing RequiresOnlineLicensing	<none>

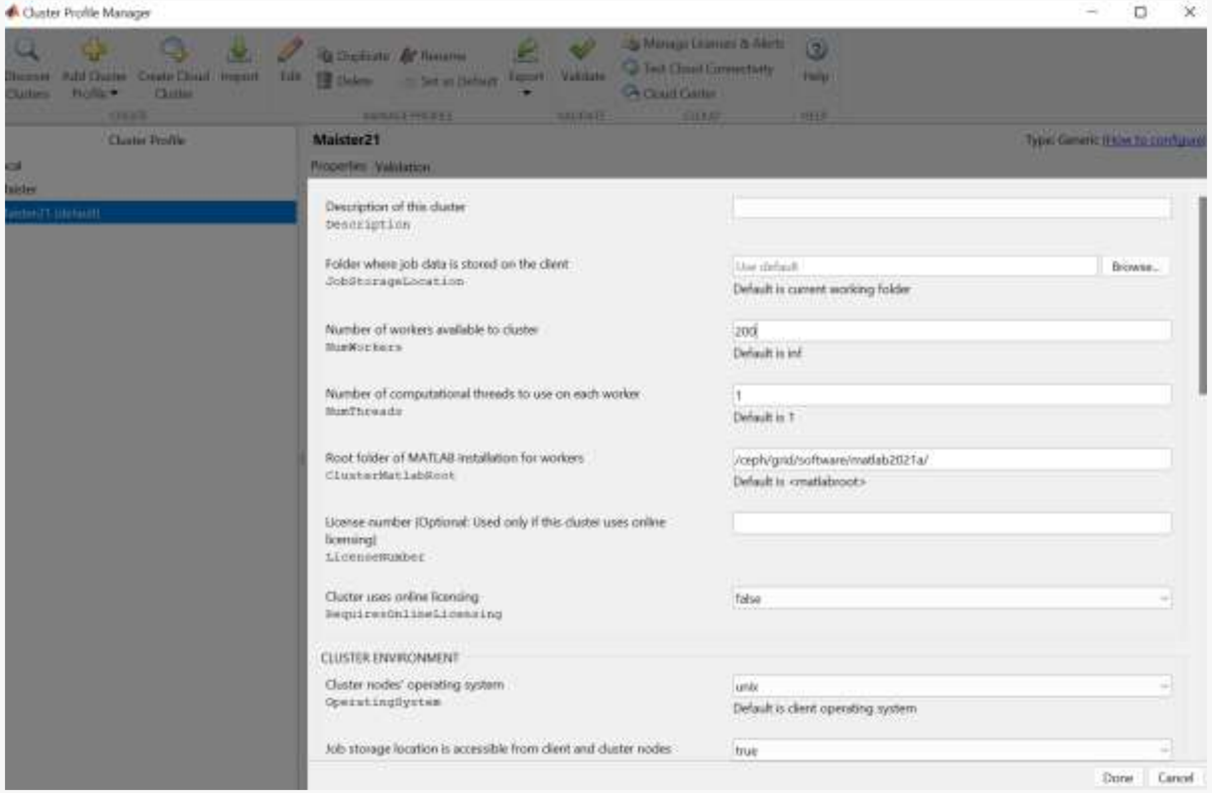
CLUSTER ENVIRONMENT

Cluster nodes' operating system OperatingSystem	client operating system (default)
Job storage location is accessible from client and cluster nodes HasSharedFilesystem	true (default)

SCHEDULER PLUGIN

Maister R2021a Profile

- Description = Maister
- NumWokers = number of cores. Default is infinite but 200 is quite large value for Maister HPC.
- NumThreads=2 (every core have two threads)
- ClusterMatlabRoot=/ceph/grid/software/matlab2021a



Maister21

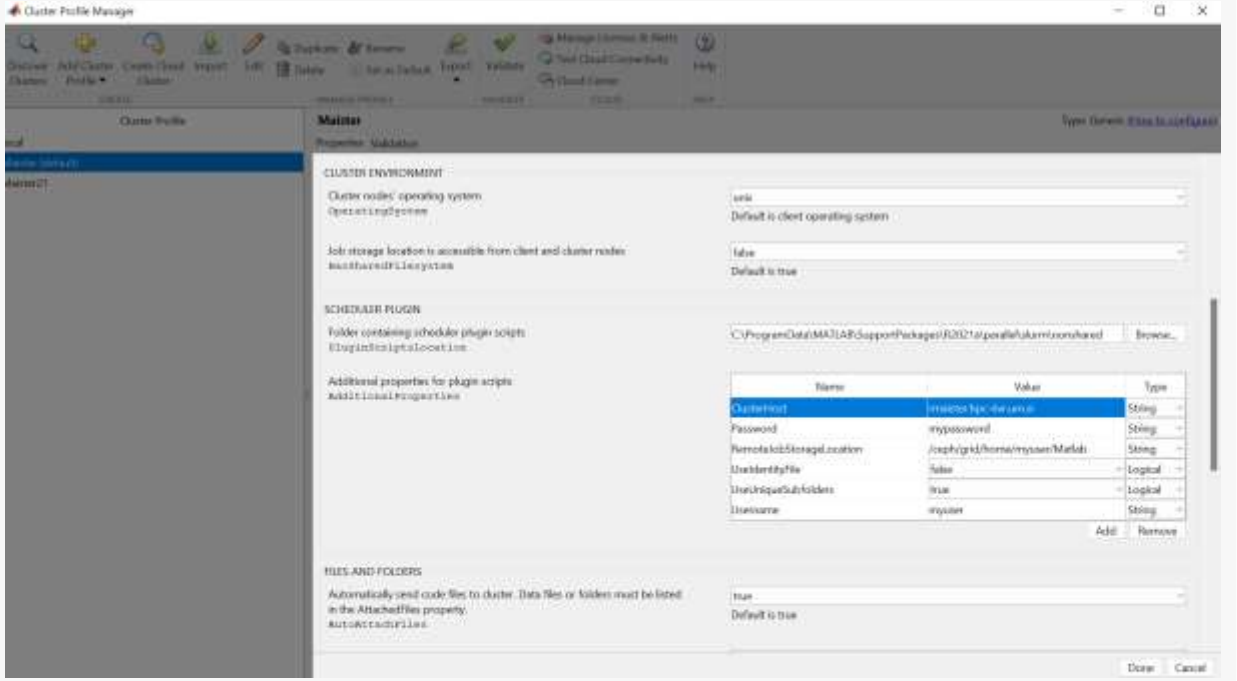
Description of this cluster Description	
Folder where job data is stored on the client JobStorageLocation	Use default Default is current working folder
Number of workers available to cluster NumWorkers	200 Default is inf
Number of computational threads to use on each worker NumThreads	1 Default is 1
Root folder of MATLAB installation for workers ClusterMatlabRoot	/ceph/grid/software/matlab2021a/ Default is <matlabroot>
License number (Optional: Used only if this cluster uses online licensing) LicenseNumber	
Cluster uses online licensing RequiresOnlineLicensing	false

CLUSTER ENVIRONMENT

Cluster nodes' operating system OperatingSystem	unix Default is client operating system
Job storage location is accessible from client and cluster nodes	true

Maister R2021a Profile

- PluginScriptLocation= folder on personal computer with Matlab data
- AdditionalProperties = hostname, username and password for ssh access
- RemoteJobStorageLocation = subfolder for Matlab temporary files



Maister

Type: Generic

CLUSTER ENVIRONMENT

Cluster nodes' operating system
`OperatingSystem`
unix
Default is client operating system

Job storage location is accessible from client and cluster nodes
`HasSharedFilesystem`
false
Default is true

SCHEDULER PLUGIN

Folder containing scheduler plugin scripts
`PluginScriptsLocation`
C:\ProgramData\MATLAB\SupportPackages\R2021a\parallel\slurm\nonshared

Additional properties for plugin scripts
`AdditionalProperties`

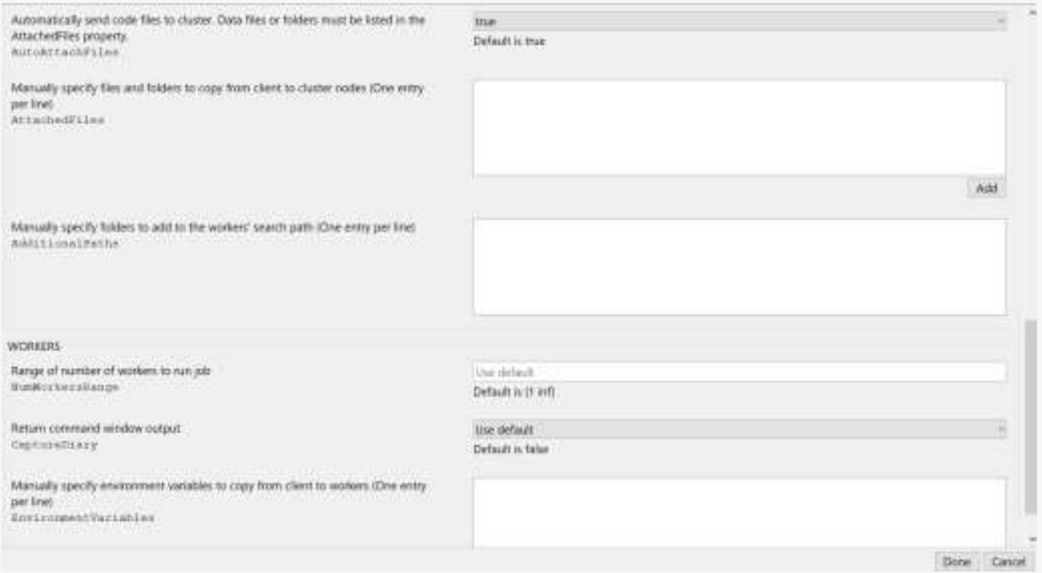
Name	Value	Type
ClusterHost	maister.hpc-rivr.um.si	String
Password	mypassword	String
RemoteJobStorageLocation	/ceph/grid/home/myuser/Matlab	String
UseIdentityFile	false	Logical
UseUniqueSubfolders	true	Logical
Username	myuser	String

FILES AND FOLDERS

Automatically send code files to cluster. Data files or folders must be listed in the AttachedFiles property.
`AutoAttachFiles`
true
Default is true

Maister R2021a Profile

- Remaining fields can have default values
- Profile is saved with button Done



Automatically send code files to cluster. Data files or folders must be listed in the AttachedFiles property.
AutoAttachFiles

true

Default is true

Manually specify files and folders to copy from client to cluster nodes (One entry per line)
AttachedFiles

Manually specify folders to add to the workers' search path (One entry per line)
AdditionalPaths

WORKERS

Range of number of workers to run job
NumWorkersRange

Use default

Default is [1 inf]

Return command window output
CaptureDiary

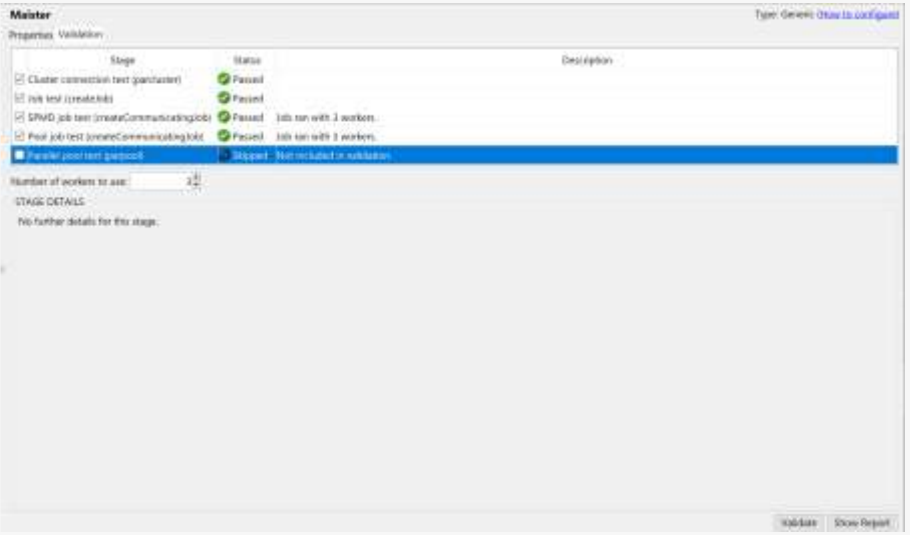
Use default

Default is false

Manually specify environment variables to copy from client to workers (One entry per line)
EnvironmentVariables

Maister R2021a Profile validation

- Select tab Validation
- Choose small number of workers e.g. 3
- Parallel Pool test does not work
- Lately also Job test does not work



Maister

Type: Generic (How to configure)

Properties Validation

Stage	Status	Description
☑ Cluster connection test (parcluster)	✓ Passed	
☑ Job test (createJob)	✓ Passed	
☑ SPMD job test (createCommunicatingJob)	✓ Passed	Job ran with 3 workers.
☑ Pool job test (createCommunicatingJob)	✓ Passed	Job ran with 3 workers.
□ Parallel pool test (parpool)	Skipped	Not included in validation

Number of workers to use: 3

STAGE DETAILS

No further details for this stage.

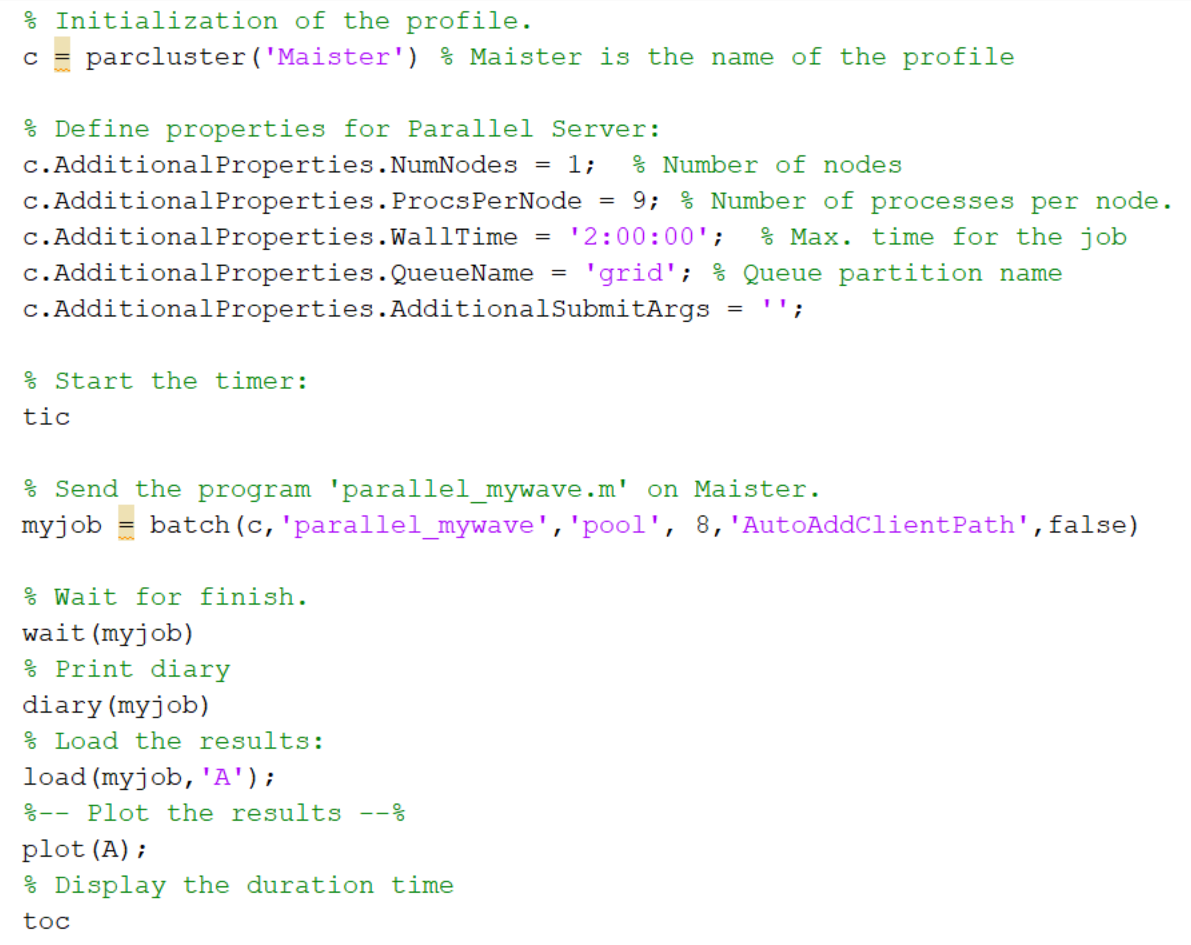
Maister R2021a Example

- https://www.hpc.iastate.edu/guides/using-matlab-parallel-server#ex2_parallel
- Parallel job
- Parallel command in the Matlab (e.g. parfor) enables parallel execution
- **parallel_mywave.m** :

```
parfor i=1:10000000
    A(i)=sin(i*2*pi/2500000);
end
```

Maister R2021a Example

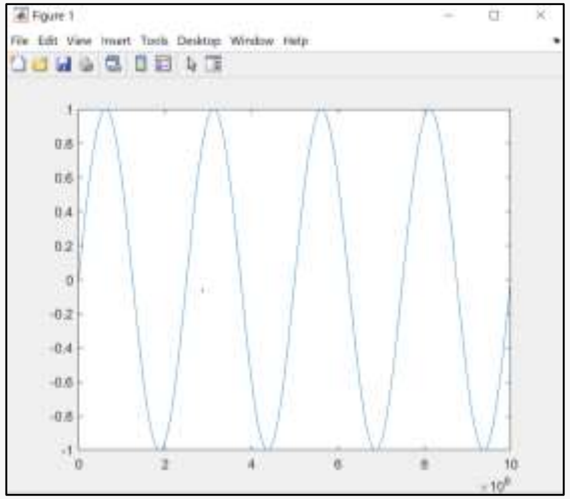
- Program for parallel execution on the Maister HPC
- Total number of processes is 9 while 8 processes are dedicated to the job and one for job control

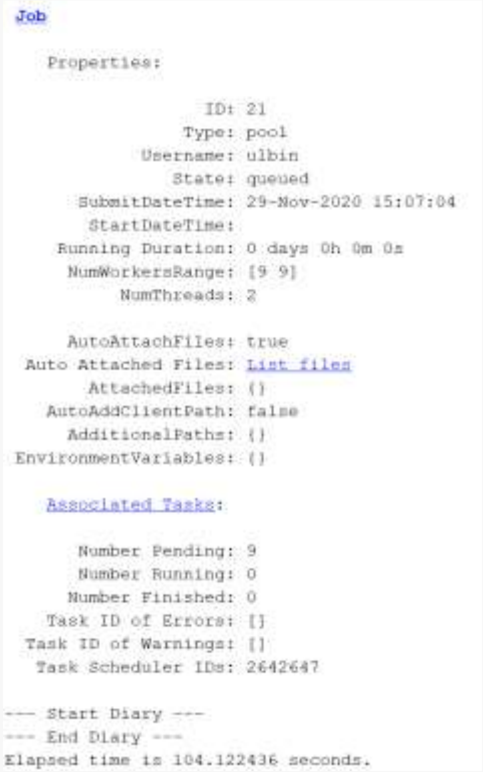


```
% Initialization of the profile.
c = parcluster('Maister') % Maister is the name of the profile

% Define properties for Parallel Server:
c.AdditionalProperties.NumNodes = 1;  % Number of nodes
c.AdditionalProperties.ProcsPerNode = 9; % Number of processes per node.
c.AdditionalProperties.WallTime = '2:00:00';  % Max. time for the job
c.AdditionalProperties.QueueName = 'grid'; % Queue partition name
c.AdditionalProperties.AdditionalSubmitArgs = '';

% Start the timer:
tic

% Send the program 'parallel_mywave.m' on Maister.
myjob = batch(c,'parallel_mywave','pool', 8,'AutoAddClientPath',false)

% Wait for finish.
wait(myjob)
% Print diary
diary(myjob)
% Load the results:
load(myjob,'A');
%-- Plot the results --%
plot(A);
% Display the duration time
toc
```

Maister R2021a Example

- Program is waiting on the personal computer
- At the end diary is printed and for this example plot is drawn





```
Job

   Properties:

                   ID: 21
                 Type: pool
             Username: ulbin
                State: queued
       SubmitDateTime: 29-Nov-2020 15:07:04
        StartDateTime:
     Running Duration: 0 days 0h 0m 0s
      NumWorkersRange: [9 9]
           NumThreads: 2

      AutoAttachFiles: true
  Auto Attached Files: List files
        AttachedFiles: {}
    AutoAddClientPath: false
      AdditionalPaths: {}
 EnvironmentVariables: {}

   Associated Tasks:

       Number Pending: 9
       Number Running: 0
      Number Finished: 0
    Task ID of Errors: []
  Task ID of Warnings: []
   Task Scheduler IDs: 2642647

--- Start Diary ---
--- End Diary ---
Elapsed time is 104.122436 seconds.
```

Graphical access to Maister HPC

Powerful GPU nodes

- Sometimes GPUs could be used for interactive work.
- Only one node is dedicated to interactive work.

X – windows on personal computer

- For display the personal computer should be used.
- It is required that is running X-Windows subsystem.

Access to Maister HPC

- The user should have ssh access to Maister HPC (username and password).
- The ssh terminal aplication should support X11 forwarding.

•Running graphical programs on Maister node

- Running the program on dedicated node displays the window on personal computer.

Summary

- Maister HPC is powerful supercomputer
- The user can run their own programs, open source programs or commercial programs with academic license
- There are several ways for accessing and running jobs on Maister HPC
- Some tools enables usage without knowledge of special operating system or queue management

THANK YOU

Miran Ulbin

miran.ulbin@um.si