Maxwell for CMS

The CMS partitions serve for Machine/Deep Learning applications of CMS groups at DESY.

For details about the setup of these partitions please have a look at the pages on hardware, limits and constraints .

Partitions

The CMS resources are "divided" into three partitions:

- The **cms-uhh** partition. It's a high priority partition. Jobs in this partition will immediately **cancel any jobs in the all or cms partition** using the same nodes!
- The **cms-desy** partition. It's a high priority partition. Jobs in this partition will immediately **cancel any jobs in the all or cms partition** using the same nodes!
- The **cms** partition. It's a regular priority partition combining all cms-uhh and cms-desy resources. Jobs in this partition will terminate jobs in the all partition (like on any other partition).

For access to the cms-uhh partition please get in touch with admins for Gregor Kasieczka's group. For access to the cms-desy partition please get in touch with CMS DESY admins. Access to either cms-uhh or cms-desy implies access to the cms partition as well.

Interactive Login Nodes

There are no login nodes associated with the CMS resources in Maxwell.

- **ssh max-display.desy.de**: will connect you to one of the display nodes. FastX might your better choice. Please have a look at the Remote Login and the FastX documentation.
- **ssh max-wgs:** will connect you to the generic login node.
- **Please note:** max-display.desy.de is directly accessible from outside. All other login nodes can only be reached by first connecting to bastion.desy.de.

Login nodes are always shared resources sometimes used by a large number of concurrent users. Don't run compute or memory intense jobs on the login nodes, use a batch job instead!

The XFEL Batch resource in Maxwell

As a first step login to one of the login nodes and check which Maxwell resources are available for your account using the my-partitions command:

my-partitions

```
[@max-wgs ~]$ my-partitions

      Partition   Access   Allowed
groups
-------------------------------------------------------------------------------------------------------------
----
            all      yes   all                                  <------- will be available if any of the
resources below is "yes!
           cfel       no   cfel-wgs-users
            cms      yes   max-cms-uhh-users,max-cms-desy-users  <------- look for this one as a member of
CMS UHH or CMS DESY
       cms-desy       no   max-cms-desy-users                   <------- look for this one as a member of
CMS DESY
        cms-uhh       no   max-cms-uhh-users                    <------- look for this one as a member of
CMS UHH
           cssb       no   max-cssb-users
      epyc-eval       no   all
          exfel      yes   exfel-wgs-users
      exfel-spb       no   exfel-theory-users,school-users
       exfel-th       no   exfel-theory-users
   exfel-theory       no   exfel-theory-users
     exfel-wp72       no   exfel-theory-users
        fspetra       no   max-fspetra-users
           grid       no   max-grid-users
           jhub       no   all
        maxwell      yes   maxwell-users,school-users            <------- might be granted if you have
suitable applications
            p06       no   max-p06-users
         petra4       no   p4_sim
             ps       no   max-ps2-users
            psx       no   max-psx2-users
            uke       no   max-uke-users
           upex       no   upex-users,school-users
     xfel-guest       no   max-xfel-guest-users,p4_sim
        xfel-op       no   max-xfel-op-users
       xfel-sim       no   max-xfel-sim-users
```

If it says "yes" for partition "cms" you are ready to go. If so you will also see a "yes" at least for partition "all". Let's assume that you've got the cms-uhh resources. Same rules apply to the other cms partitions/resources.

If you have an application, which is started by a script called my-application, and doesn't require a GUI, you can simply submit the script as a batch-job:

sbatch

```
[@max-wgs ~]$ sbatch --partition=cms-uhh --time=12:00:00 --nodes=1 my-application
Submitted batch job 1613895

# the job might already be running
[@max-wgs ~]$ squeue -u $USER
             JOBID PARTITION     NAME     USER ST       TIME  NODES NODELIST(REASON)
           1614464       cms   my-app     user  R       0:06      1 max-cmsg001
# Status of the job                           R: running. PD: pending
```

This works for any application smart enough not to strictly require an X-environment, matlab, comsol, ansys, mathematica, idl and many others can be executed as batch jobs. To make it more convenient you can add the SLURM directives directly into the script:

sbatch script

```
[@max-wgs ~]$ cat my-application
#!/bin/bash
#SBATCH --partition=cms-uhh
#SBATCH --time=1-12:00:00      # request 1 day and 12 hours
#SBATCH --mail-type=END,FAIL   # send mail when the job has finished or failed
#SBATCH --nodes=1              # number of nodes
#SBATCH --output=%x-%N-%j.out  # per default slurm writes output to slurm-<jobid>.out. There are a number
of options to customize the job
[...] # the actual script.
```

The email-notification will be sent to <user-id>@mail.desy.de. That should always work, so you don't actually need to specify an email-address. If you do, please make sure it's a valid address. For further examples and instructions please read Running Jobs on Maxwell.

If you think that it's much to complicated to write job-scripts or if you can't afford to invest the time to look into it: we are happy to assist. Please drop a message to maxwell.service@desy.de, we'll try our best.

Running interactive batch jobs

If you absolutely need an interactive environment, X-windows features like a GUI, there are options to do that in the batch environment. For example:

salloc

```
# request one node for 8 hours:
[@max-wgs ~]$ salloc --nodes=1 --time=08:00:00 --partition=cms
salloc: Pending job allocation 1618422
salloc: job 1618422 queued and waiting for resources
salloc: job 1618422 has been allocated resources
salloc: Granted job allocation 1618422
salloc: Waiting for resource configuration
salloc: Nodes max-cmsg001 are ready for job


# now you got a node allocated. So you can ssh into the node
[@max-wgs ~]$ ssh max-cmsg001
[@max-cmsg001 ~]$ # run your application!
[@max-cmsg001 ~]$ exit # this terminates the ssh session, it does NOT terminate the allocation
logout
Connection to max-cmsg001 closed.
[@max-wgs ~]$ exit
exit
salloc: Relinquishing job allocation 1618422
# now your allocation is finished. If in doubt use squeue -u $USER or sview to check for running sessions!
```

There are a few things to consider:

- Interactive jobs with salloc easily get forgotten, leaving precious resources idle. We do accounting and monitoring!
- Keep the time short: there is hardly a good reason to run an interactive jobs for more than the working hours. Use a batch job instead.
- Terminate allocations as soon as the job is done!

Other Maxwell Resources

Being member of CMS-UHH and maybe having access to the cms-uhh partition doesn't need to be the end of the story. If you have parallelized applications suitable for the Maxwell cluster you can apply for the Maxwell resource like everyone else on campus. Please send a message to maxwell.service@desy.de briefly explaining your use case. You can easily distribute your job over the partitions:

multiple partitions

```
[@max-wgs ~]$ cat my-application
#!/bin/bash
#SBATCH --partition=cms-uhh,cms,maxwell,all
#SBATCH --time=1-12:00:00      # request 1 day and 12 hours
#SBATCH --mail-type=END,FAIL   # send mail when the job has finished or failed
#SBATCH --nodes=1              # number of nodes
#SBATCH --output=%x-%N-%j.out  # per default slurm writes output to slurm-<jobid>.out. There are a number
of options to customize the job
#SBATCH --constraint=P100      # make sure that the same type of GPU is available in all & maxwell
partitions!
 [...] # the actual script.
```

The partition will be selected from cms-uhh OR maxwell OR cms OR all starting with the highest priority partition. So your job will run on the cms-uhh partition if nodes are available, on the maxwell partition if nodes are available and finally on the all partition if none of the other partitions specified have free nodes. Keep in mind that you should however select the partition according to the type of work you are doing. And a job can never combine nodes from different partitions, so check the limits applying to partitions.

To check availability of nodes and characteristics use sinfo (https://slurm.schedmd.com/sinfo.html)

sinfo

```
[@max-wgs ~]$ sinfo -p cms -o "%10P %.6D %8c %8L %12l %8m %30f %N"
PARTITION   NODES CPUS     DEFAULTT TIMELIMIT    MEMORY   AVAIL_FEATURES                 NODELIST
cms            10 40       12:00:00 1-00:00:00   256000   INTEL,V4,E5-2640,GPU,P100,GPUx max-cmsg[001-010]

[@max-wgs ~]$ sinfo -p cms -o "%10P %.6D %10s"
PARTITION   NODES JOB_SIZE
cms            10 1-8
```