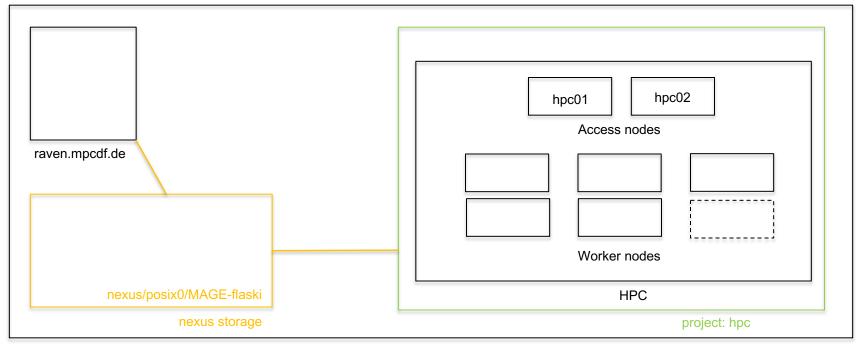


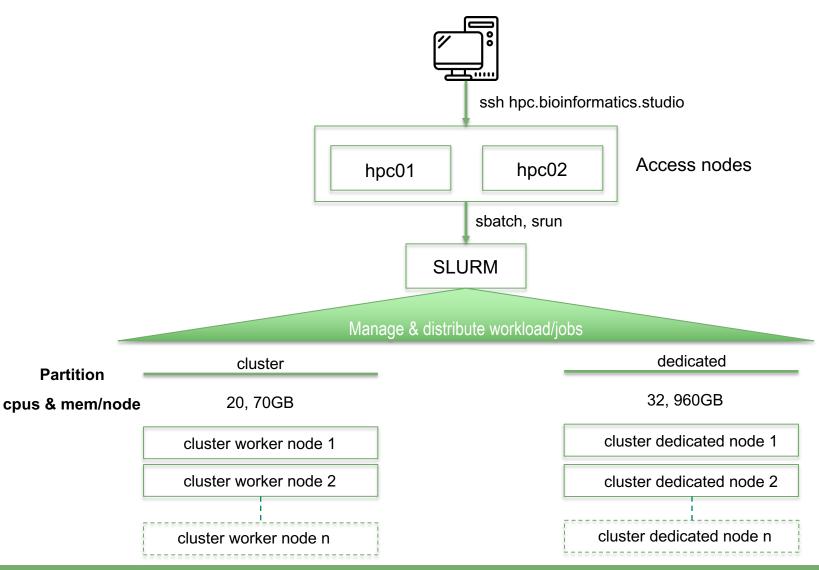
Introduction to hpc.bioinformatics.studio

Cloud Infrastructure



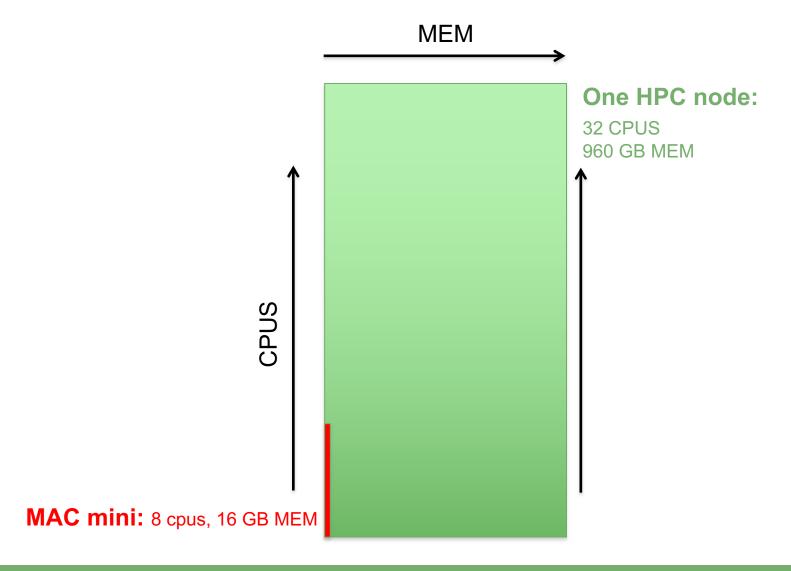
mpcdf@Garching

Workflow Architecture

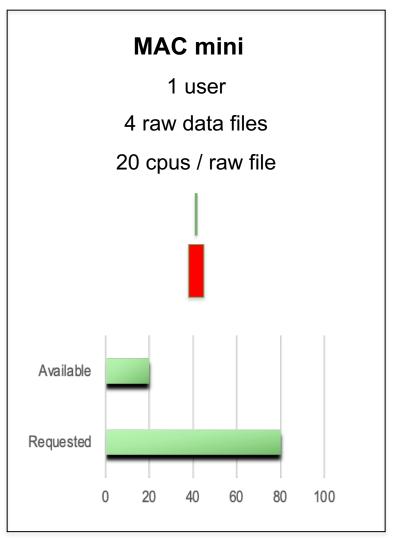


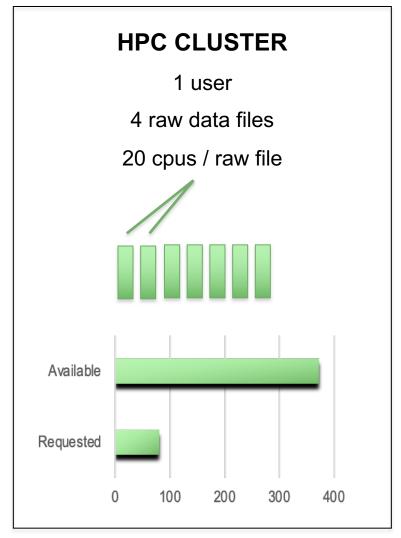
MAX PLANCK INSTITUTE FOR **BIOLOGY OF AGEING** | BIOINFORMATICS

Why a HPC cluster?



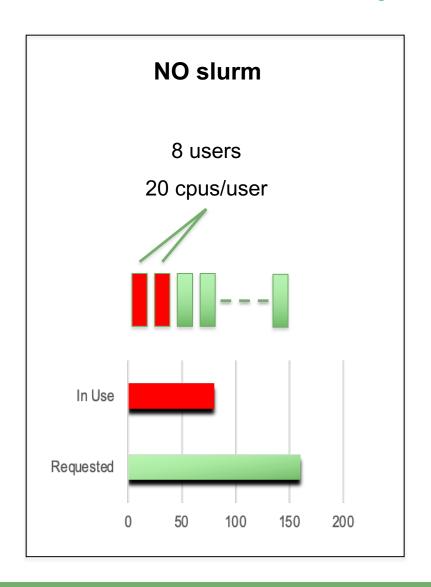
Why a HPC cluster?

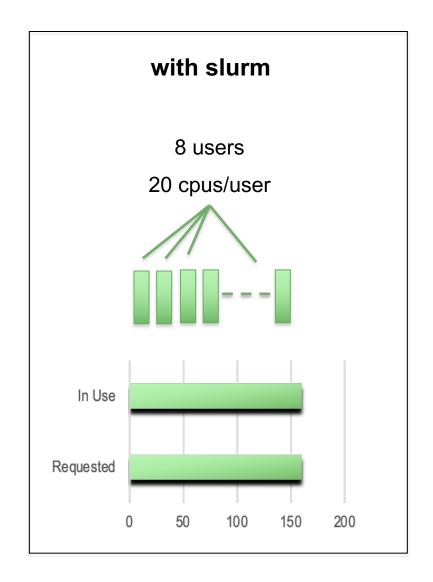




^{***} The resources shown here are to provide an overview, the numbers are not exact as the hpc.bioinformatics.studio

Why SLURM?





How SLURM?

NO slurm

> bwa mem sample1.fastq

with slurm

> srun bwa mem sample1.fastq

How SLURM?

NO slurm

> bwa mem -T 18 sample1.fastq

with slurm

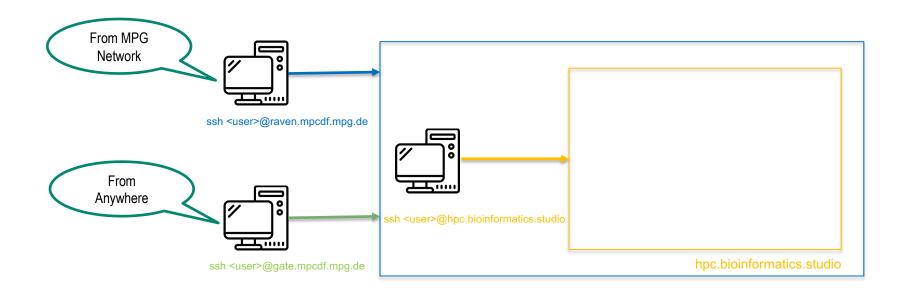
```
> srun --cpus-per-task=18 \
--mem=15gb \
--time=5-24 \
--partition=cluster
bwa mem -T 18 sample1.fastq

( 5 days and 24 hours = 6 days alternative: 144:00:00;
maximum = 14 days )
```

Getting Started

Prerequisite

- A user account at MPCDF (MPCDF Registration)
- A hpc.bioinformatics.studio account (contact: <u>bioinformatics@age.mpg.de</u>)



SLURM Commands: cluster info

```
hpc02:~$ # View partition information
hpc02:~$ sinfo
PARTITION AVAIL TIMELIMIT NODES STATE NODELIST
            up 14-00:00:0
                              9 idle slurm-worker-cluster-[0-8]
cluster*
dedicated
            up 14-00:00:0
                                   mix slurm-worker-dedicated-[2-3]
            up 14-00:00:0
                              2 alloc slurm-worker-dedicated-[0-1]
dedicated
dedicated up 14-00:00:0 2 idle slurm-worker-dedicated-[4-5]
hpc02:~$ # Show information of nodes
hpc02:~$ sinfo -N
NODELIST
                         NODES PARTITION STATE
slurm-worker-cluster-0
                            1 cluster* idle
slurm-worker-cluster-1
                            1 cluster* idle
                            1 cluster* idle
slurm-worker-cluster-2
                            1 cluster* idle
slurm-worker-cluster-3
slurm-worker-cluster-4
                            1 cluster* idle
                            1 cluster* idle
slurm-worker-cluster-5
slurm-worker-cluster-6
                            1 cluster* idle
slurm-worker-cluster-7
                            1 cluster* idle
slurm-worker-cluster-8
                            1 cluster* idle
slurm-worker-dedicated-0
                            1 dedicated mix
slurm-worker-dedicated-1
                            1 dedicated alloc
slurm-worker-dedicated-2
                            1 dedicated mix
slurm-worker-dedicated-3
                            1 dedicated mix
slurm-worker-dedicated-4
                            1 dedicated idle
slurm-worker-dedicated-5
                            1 dedicated idle
```

SLURM Commands: cluster info

```
hpc02:~$ # Show information of nodes
hpc02:~$ sinfo -N -O partitionname, nodehost, cpus, cpusload, freemem, memory
                    HOSTNAMES
                                                              CPU_LOAD
PARTITION
                                         CPUS
                                                                                   FREE_MEM
                                                                                                       MEMORY
                    slurm-worker-cluster20
                                                                                  69565
                                                                                                       70365
cluster
                                                              0.00
cluster
                    slurm-worker-cluster20
                                                              0.00
                                                                                  69609
                                                                                                       70365
cluster
                    slurm-worker-cluster20
                                                              0.00
                                                                                  67129
                                                                                                       70365
                    slurm-worker-cluster20
cluster
                                                              0.00
                                                                                   66891
                                                                                                       70365
                    slurm-worker-cluster20
                                                             0.00
                                                                                  66889
cluster
                                                                                                       70365
                                                                                  66894
cluster
                    slurm-worker-cluster20
                                                              0.00
                                                                                                       70365
                                                              0.00
cluster
                    slurm-worker-cluster20
                                                                                   66905
                                                                                                       70365
cluster
                    slurm-worker-cluster20
                                                              0.00
                                                                                  66933
                                                                                                       70365
                    slurm-worker-cluster20
                                                             0.00
                                                                                                       70365
cluster
                                                                                   66911
dedicated
                                                             15.88
                                                                                  769817
                                                                                                       967194
                    slurm-worker-dedicat32
dedicated
                                                             15.20
                                                                                  679470
                                                                                                       967194
                    slurm-worker-dedicat32
                    slurm-worker-dedicat32
                                                             14.98
                                                                                  738568
dedicated
                                                                                                       967194
dedicated
                    slurm-worker-dedicat32
                                                              3.79
                                                                                   929163
                                                                                                       967194
dedicated
                    slurm-worker-dedicat32
                                                              0.00
                                                                                   964917
                                                                                                       967194
dedicated
                    slurm-worker-dedicat32
                                                             0.01
                                                                                   964890
                                                                                                       967194
hpc02:~$ # Show information about nodes with a specific state (e.g., idle, alloc, mix, etc.)
hpc02:~$ sinfo -t alloc
PARTITION AVAIL TIMELIMIT NODES
                                   STATE NODELIST
             up 14-00:00:0
cluster*
                                      n/a
                                      mix slurm-worker-dedicated-[2-3]
dedicated
             up 14-00:00:0
             up 14-00:00:0
                                 2 alloc slurm-worker-dedicated-[0-1]
dedicated
```

SLURM Commands: submitting jobs

```
hpc02:~$ # Submit jobs to SLURM
hpc02:~$ # sbatch [options] script.sh
hpc02:~$ # Where script.sh is the shell script containing the commands you want to execute
hpc02:~$ sbatch script.sh
Submitted batch job 7540
hpc02:~$ # Common options:
hpc02:~\$ # -p < partition>: Specify the partition/queue for the job.
hpc02:~$ # -n <tasks>: Number of tasks in the job.
hpc02:~$ # --cpus-per-task=<cores>: Specify the number of CPU cores per task.
hpc02:~$ # --mem=<memory>: Request memory for the job
hpc02:~$ sbatch -p cluster --cpus-per-task=1 --mem=1gb -t 00:10:00 test.sh
Submitted batch job 7541
hpc02:~$ # Check running jobs
hpc02:~$ squeue ==
            JOBID PARTITION
                                NAME
                                         USER ST
                                                       TIME NODES NODELIST(REASON)
             7540 cluster test_job
                                        hamin R
                                                       1:22
                                                                1 slurm-worker-cluster-0
                                                                 1 slurm-worker-cluster-0
             7541 cluster test_job
                                                       0:34
                                        hamin R
```

SLURM Commands: submitting jobs

```
hpc02:~$ # Submissions wihtout arguments specifications will result in -p "cluster" and a time limit of "2 weeks"
hpc02:~$ # You can also include SLURM parameters inside the script ie.
hpc02:~$ cat script.sh 🌰
#!/bin/bash
#SBATCH --job-name=test_job
                                     # Job name
#SBATCH -p cluster
                                     # Assign to a cluster
#SBATCH --cpus-per-task=1
                                     # Number of CPUs for task
#SBATCH --mem=1gb
                                     # Job memory request
#SBATCH --time=00:05:00
                                     # Time limit hrs:min:sec
#SBATCH -o test_output.log
                                     # Define the output path
pwd; hostname; date; sleep 300
hpc02:~$ sbatch script.sh 
Submitted batch job 7542
hpc02:~$ # Check the running jobs
hpc02:~$ squeue
            JOBID PARTITION
                                NAME
                                         USER ST
                                                       TIME NODES NODELIST(REASON)
                                                                 1 slurm-worker-cluster-0
             7542
                    cluster test_job
                                        hamin R
                                                       0:19
```

SLURM Commands: checking queue status

```
hpc02:~$ # Show the queue information
hpc02:~$ squeue
            JOBID PARTITION
                                 NAME
                                          USER ST
                                                        TIME
                                                              NODES NODELIST(REASON)
                                                                  1 slurm-worker-cluster-0
                    cluster script.s
                                         hamin R
                                                        5:05
              7548
                    cluster script.s
                                                        5:02
                                                                  1 slurm-worker-cluster-0
             7549
                                         hamin R
             7550
                    cluster script.s
                                         hamin R
                                                        5:02
                                                                  1 slurm-worker-cluster-0
                                         hamin
                                                                  1 slurm-worker-cluster-0
             7551
                    cluster script.s
                                                        2:51
             7552 dedicated script.s
                                         hamin
                                                        0:27
                                                                  1 slurm-worker-dedicated-0
             7553 dedicated script.s
                                                        0:24
                                                                  1 slurm-worker-dedicated-0
                                         hamin
             7554 dedicated script.s
                                         hamin R
                                                        0:21
                                                                  1 slurm-worker-dedicated-0
hpc02:~$ # Queue infromation with option
hpc02:~$ # -u <username>: Show jobs for a specific user
hpc02:~$ # -p <partition>: Show jobs in a specific partition
hpc02:~$ squeue -p dedicated
            JOBID PARTITION
                                          USER ST
                                                        TIME
                                                              NODES NODELIST(REASON)
                                 NAME
                                                                  1 slurm-worker-dedicated-0
             7552 dedicated script.s
                                         hamin R
                                                        1:39
             7553 dedicated script.s
                                         hamin
                                                        1:36
                                                                  1 slurm-worker-dedicated-0
             7554 dedicated script.s
                                                        1:33
                                                                  1 slurm-worker-dedicated-0
                                         hamin R
```

SLURM Commands: controlling jobs

```
hpc02:~$ # Show detailed information about a job
hpc02:~$ # scontrol show job <job_id>
hpc02:~$ scontrol show job 7552 🤙
JobId=7552 JobName=script.sh
  UserId=hamin(60571) GroupId=mage(17600) MCS_label=N/A
  Priority=4294894274 Nice=0 Account=bioinformatics QOS=normal
  JobState=COMPLETED Reason=None Dependency=(null)
  Requeue=1 Restarts=0 BatchFlag=1 Reboot=0 ExitCode=0:0
  RunTime=00:10:01 TimeLimit=00:15:00 TimeMin=N/A
  SubmitTime=2023-09-28T12:12:04 EligibleTime=2023-09-28T12:12:04
  AccrueTime=2023-09-28T12:12:04
  StartTime=2023-09-28T12:12:05 EndTime=2023-09-28T12:22:06 Deadline=N/A
  SuspendTime=None SecsPreSuspend=0 LastSchedEval=2023-09-28T12:12:05
  Partition=dedicated AllocNode:Sid=hpc-login:1634788
  ReqNodeList=(null) ExcNodeList=(null)
  NodeList=slurm-worker-dedicated-0
  BatchHost=slurm-worker-dedicated-0
  NumNodes=1 NumCPUs=2 NumTasks=0 CPUs/Task=2 ReqB:S:C:T=0:0:*:*
  TRES=cpu=2, node=1, billing=2
  Socks/Node=* NtasksPerN:B:S:C=0:0:*:* CoreSpec=*
  MinCPUsNode=2 MinMemoryNode=2G MinTmpDiskNode=0
  Features=(null) DelayBoot=00:00:00
  OverSubscribe=OK Contiguous=0 Licenses=(null) Network=(null)
  Command=/nexus/posix0/MAGE-flaski/service/hpc/home/hamin/script.sh
  WorkDir=/nexus/posix0/MAGE-flaski/service/hpc/home/hamin
  StdErr=/nexus/posix0/MAGE-flaski/service/hpc/home/hamin/slurm-7552.out
  StdIn=/dev/null
  StdOut=/nexus/posix0/MAGE-flaski/service/hpc/home/hamin/slurm-7552.out
   Power=
```

SLURM Commands: controlling jobs

```
hpc02:~$ # Show information about a partition
hpc02:~$ # scontrol show partition <parition_name>
hpc02:~$ # scontrol show partition <parition_name>
hpc02:~$ scontrol show partition dedicated

PartitionName=dedicated
AllowGroups=ALL AllowAccounts=ALL AllowQos=ALL
AllocNodes=ALL Default=NO QoS=N/A
DefaultTime=NONE DisableRootJobs=NO ExclusiveUser=NO GraceTime=0 Hidden=NO
MaxNodes=UNLIMITED MaxTime=14-00:00:00 MinNodes=0 LLN=NO MaxCPUsPerNode=UNLIMITED
Nodes=slurm-worker-dedicated-1,slurm-worker-dedicated-2,slurm-worker-dedicated-0,slurm-worker-dedicated-3,slurm-worker-dedicated-5,slurm-worker-dedicated-4
PriorityJobFactor=1 PriorityTier=1 RootOnly=NO ReqResv=NO OverSubscribe=NO
OverTimeLimit=NONE PreemptMode=OFF
State=UP TotalCPUs=192 TotalNodes=6 SelectTypeParameters=NONE
JobDefaults=(null)
DefMemPerNode=UNLIMITED MaxMemPerNode=UNLIMITED
```

SLURM Commands: controlling jobs

```
hpc02:~$ # Show information about a node
hpc02:~$ # scontrol show node <node_name>
hpc02:~$ scontrol show node slurm-worker-dedicated-2 '
NodeName=slurm-worker-dedicated-2 Arch=x86_64 CoresPerSocket=32
   CPUAlloc=32 CPUTot=32 CPULoad=8.49
   AvailableFeatures=(null)
   ActiveFeatures=(null)
   Gres=(null)
   NodeAddr=192.168.42.6 NodeHostName=slurm-worker-dedicated-2 Version=19.05.5
   OS=Linux 5.4.0-117-generic #132-Ubuntu SMP Thu Jun 2 00:39:06 UTC 2022
   RealMemory=967194 AllocMem=0 FreeMem=176841 Sockets=1 Boards=1
   State=ALLOCATED ThreadsPerCore=1 TmpDisk=0 Weight=1 Owner=N/A MCS_label=N/A
   Partitions=dedicated
   BootTime=2023-08-04T14:12:48 SlurmdStartTime=2023-08-24T10:29:46
   CfgTRES=cpu=32,mem=967194M,billing=32
   AllocTRES=cpu=32
   CapWatts=n/a
   CurrentWatts=0 AveWatts=0
   ExtSensorsJoules=n/s ExtSensorsWatts=0 ExtSensorsTemp=n/s
```

SLURM Commands: cancel a job

```
hpc02:~$ squeue -u hamin <
            JOBID PARTITION
                                NAME
                                         USER ST
                                                       TIME NODES NODELIST(REASON)
             7710
                    cluster script.s
                                        hamin R
                                                       8:34
                                                                 1 slurm-worker-cluster-0
                                                                1 slurm-worker-cluster-0
             7709 cluster script.s
                                        hamin R
                                                       8:37
hpc02:~$ # Cancel a job
hpc02:~$ # scancel <job_id>
hpc02:~$ scancel 7709
hpc02:~$ squeue -u hamin
            JOBID PARTITION
                                NAME
                                         USER ST
                                                       TIME NODES NODELIST(REASON)
                                                                 1 slurm-worker-cluster-0
             7710
                    cluster script.s
                                        hamin R
                                                       8:59
```

Procedures: prevent overloading the cluster

```
#!/bin/bash
# Limit the number of simultaneously running jobs
cd ~/project/raw data
for f in $(ls *.fastq);
  do rm ~/project/slurm logs/${f}.*.out
  # wait if the running jobs exceed the limit
  while [ 'squeue -u username | wc -l' -gt "500"];
   do echo "sleeping"; sleep 300
  done
  sbatch --cpus-per-task=18 --mem=15gb --time=5-24 \
  -p cluster -o ~/project/slurm logs/${f}.%j.out ~/project/tmp/${f}.sh << EOF
  #!/bin/bash
  # necessary operations
  EOF
done
exit
```

SLURM Concepts: at a glance

An open-source workload manager and job scheduler

Manages and schedules tasks across a cluster of computers

Job: Unit of work submitted to SLURM for execution

Partitions: Groups of nodes with similar characteristics

Nodes: Individual computers in the cluster that execute jobs

Resources: Includes CPU cores, memory, and other hardware resources

Singularity/Apptainer

Why Singularity?

- Container engine for HPC
- No root privilege required
- Reproducible, portable, distributable container
- Easy conversion from docker to singularity container

Singularity: example

```
hpc02:~$ # Build a singularity container from Docker image
hpc02:~$ singularity pull my_container.sif docker://ubuntu:latest
INFO: Converting OCI blobs to SIF format
INFO: Starting build...
Getting image source signatures
Copying blob 445a6a12be2b done
Copying config c6b84b685f done
Writing manifest to image destination
```

```
hpc02:~$ # Execute the built container
hpc02:~$ singularity exec my_container.sif /bin/bash <==</pre>
Apptainer> ls /
bin
                 etc lib
                           lib64
                                     media nexus
                                                             singularity sys usr
     dev
                                                  proc run
boot environment home lib32 libx32 mnt
                                            opt
                                                  root sbin srv
                                                                          tmp
                                                                              var
Apptainer>
```

```
hpc02:~$ # Execute bioinformatics_software container  hpc02:~$ singularity exec /nexus/posix0/MAGE-flaski/service/images/bioinformatics_software.v4.0.3.sif /bin/bashcontainer:..laski/service/hpc/home/hamin$
```

The Modules System

- Our mpgagebioinformatics/bioinformatics_software make use of the modules system to load and unload required software
- The modules system loads software (version of choice) and changes environment variables (eg. LD_LIBRARY_PATH)

```
container:..laski/service/hpc/home/hamin$ # Shows available modules
container:..laski/service/hpc/home/hamin$ module avail 🤝
             ------/modules/modulefiles/general
jdk/18.0.2(default) jupyterhub/2.3.1(default) perl/5.32.1(default) python/3.9.13(default) rlang/4.2.1(default)
                                         -----/modules/modulefiles/libs -----
bzip2/1.0.8(default) gsl/2.7.1(default) htslib/1.16(default) imagemagick/7.1.0-47(default) openblas/0.3.21(default) xz/5.2.5(default)
                                       ------ /modules/modulefiles/bioinformatics ------
abismal/3.0.0(default)
                        emboss/6.6.0(default)
                                               kallisto/0.48.0(default) rsem/1.3.3(default)
                                                                                                subread/2.0.3(default)
bamutil/1.0.15(default)
                       epiteome/1.0.0(default) kenttools/435(default)
                                                                       samtools/1.15.1(default)
                                                                                                tophat/2.1.1(default)
bedtools/2.30.0(default) expat/2.4.8(default)
                                               lofreq/2.1.5(default)
                                                                       segemehl/0.3.4(default)
                                                                                                trimgalore/0.6.7(default)
bismark/0.24.0(default)
                        fastac/0.11.9(default)
                                              meme/5.4.0(default)
                                                                       seatk/1.3.0(default)
                                                                                                 trimmomatic/0.39(default)
blast/2.13.0(default)
                        flexbar/3.5.0(default)
                                              mitools/1.5.0(default)
                                                                       skewer/0.2.2(default)
                                                                                                 vcftools/0.1.16(default)
bowtie/1.3.1
                        gatk/4.2.6.1(default)
                                               near/1.0.0(default)
                                                                       snpeff/4.3.t(default)
                                                                                                 vdjtools/1.2.1(default)
bowtie/2.4.5(default)
                       gsea/4.3.2(default)
                                              ngsutils/0.5.9(default)
                                                                      spades/3.15.4(default)
                                                                                                walt/1.1.0(default)
```

The Modules System

```
container:..laski/service/hpc/home/hamin$ # Shows a description of the SAMtools module
container:..laski/service/hpc/home/hamin$ module whatis samtools 🛑
 ------/modulefiles/bioinformatics --
    samtools/1.15.1: Version 1.15.1 of samtools
container:..laski/service/hpc/home/hamin$ # Show environment changes for SAMtools
container:..laski/service/hpc/home/hamin$ module show samtools 🛑
/modules/modulefiles/bioinformatics/samtools/1.15.1:
module-whatis
               {Version 1.15.1 of samtools}
conflict
               samtools
prepend-path
              PATH /modules/software/samtools/1.15.1/bin
              MANPATH /modules/software/samtools/1.15.1/share/man
prepend-path
prepend-path
               INFODIR /modules/software/samtools/1.15.1/share/man
container:..laski/service/hpc/home/hamin$ # Load SAMtools
container:..laski/service/hpc/home/hamin$ module load samtools 🦃
container:..laski/service/hpc/home/hamin$ # List all loaded modules
container:..laski/service/hpc/home/hamin$ module list
Currently Loaded Modulefiles:
1) samtools/1.15.1(default)
container:..laski/service/hpc/home/hamin$ # Unload the SAMtools module
container:..laski/service/hpc/home/hamin$ module unload samtools 🛑
container:..laski/service/hpc/home/hamin$ # Unload all loaded modules
container:..laski/service/hpc/home/hamin$ module purge 
container:..laski/service/hpc/home/hamin$ # Check again the module list
container:..laski/service/hpc/home/hamin$ module list 
No Modulefiles Currently Loaded.
```

RAVEN & hpc.bioinformatics.studio

Where to Run Your Jobs?

You can run your slurm jobs in both Raven & hpc.bioinformatics.studio

RAVEN	hpc.bioinformatics.studio
For all MPI institutesMore resourcesMore usersLess time-limit	 For MPG users only Less resources Less users Higher time-limit (14 days)

Use Bioinformatics Containers

Useful Bioinformatics software containers are available to use in:

/nexus/posix0/MAGE-flaski/service/images/

Raven

```
hamin@raven02:~> # Load singularity module
hamin@raven02:~> module load singularity

In 2021 the Singularity open source project split into two projects called Apptainer and SingularityCE.
On this cluster, this Singularity module will just load the Apptainer module.
Please, consider adjusting your scripts accordingly.

Loading singularity/link2apptainer

Loading requirement: apptainer/1.1.7
hamin@raven02:~> # Use Bioinformatics Software image to launch container
hamin@raven02:~> singularity exec /nexus/posix0/MAGE-flaski/service/images/bioinformatics_software.v4.0.3.sif /bin/bash

INFO: fuse2fs not found, will not be able to mount EXT3 filesystems
container:~$
```

Use Bioinformatics Containers

Useful Bioinformatics software containers are available to use in:

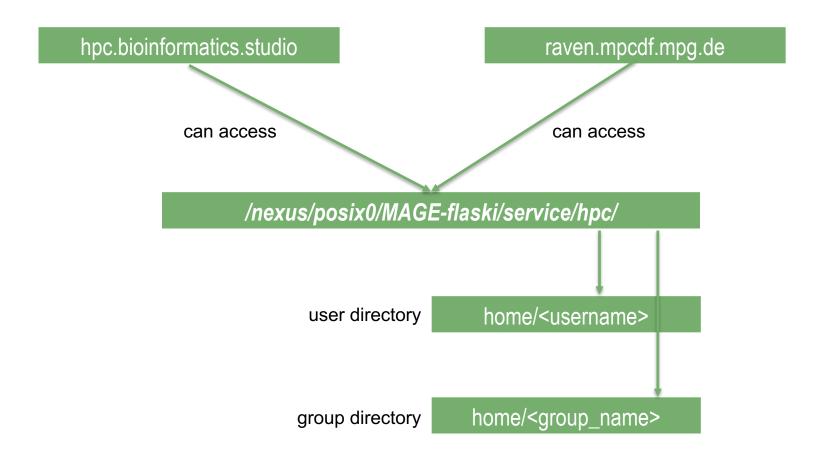
/nexus/posix0/MAGE-flaski/service/images/

hpc.bioinformatics.studio

```
hpc01:~$ # Singularity module is loaded by default
hpc01:~$ # Use Bioinformatics Software image to launch container
hpc01:~$ singularity exec /nexus/posix0/MAGE-flaski/service/images/bioinformatics_software.v4.0.3.sif /bin/bash
container:..laski/service/hpc/home/hamin$ # Check the available modules in Bioinformatics Software container
container:..laski/service/hpc/home/hamin$ module avail
                                    jdk/18.0.2(default) jupyterhub/2.3.1(default) perl/5.32.1(default) python/3.9.13(default) rlang/4.2.1(default)
                             bzip2/1.0.8(default) gsl/2.7.1(default) htslib/1.16(default) imagemagick/7.1.0-47(default) openblas/0.3.21(default) xz/5.2.5(default)
                                              -----/modules/modulefiles/bioinformatics
abismal/3.0.0(default)
                       cufflinks/2.2.1(default) hisat/2.1.0(default)
                                                                     near/1.0.0(default)
                                                                                            seatk/1.3.0(default)
                                                                                                                      tophat/2.1.1(default)
bamutil/1.0.15(default)
                       cytoscape/3.9.1(default) homer/4.11.0(default)
                                                                     ngsutils/0.5.9(default)
                                                                                            skewer/0.2.2(default)
                                                                                                                     trimgalore/0.6.7(default)
bedtools/2.30.0(default) emboss/6.6.0(default)
                                              igrec/3.1.1(default)
                                                                     nlopt/2.7.1(default)
                                                                                            snpeff/4.3.t(default)
                                                                                                                     trimmomatic/0.39(default)
                       epiteome/1.0.0(default)
                                                                                            spades/3.15.4(default)
bismark/0.24.0(default)
                                              iseerna/1.2.2(default)
                                                                     picard/2.27.4(default)
                                                                                                                      vcftools/0.1.16(default)
                                                                                            sratoolkit/2.11.3(default) vdjtools/1.2.1(default)
blast/2.13.0(default)
                       expat/2.4.8(default)
                                              kallisto/0.48.0(default) primer3/2.6.1(default)
bowtie/1.3.1
                       fastac/0.11.9(default)
                                              kenttools/435(default)
                                                                     quast/5.2.0(default)
                                                                                            star/2.7.10a(default)
                                                                                                                      walt/1.1.0(default)
                                                                     rsem/1.3.3(default)
bowtie/2.4.5(default)
                       flexbar/3.5.0(default)
                                              lofreq/2.1.5(default)
                                                                                            star/2.7.10b
bwa/0.7.17(default)
                       gatk/4.2.6.1(default)
                                              meme/5.4.0(default)
                                                                     samtools/1.15.1(default) stringtie/2.2.1(default)
bwtool/face601(default)
                      qsea/4.3.2(default)
                                              mitools/1.5.0(default)
                                                                     segemehl/0.3.4(default)
                                                                                            subread/2.0.3(default)
```

Data Handling

Data Access



Data Limit

user directory

home/<username>

Data limit: 100 GB per user

Close to threshold (80 GB+): warning email

Cross limit (100 GB+): notify to limit size

Failed to limit: data cleaned up

User not logged in for 3 months: tar data & deactivate user

group directory

home/<group_name>

File not accessed in 6 weeks: delete file

Backup Data

From <u>raven.mpcdf.mpg.de</u> transfer data to <u>/raven/ptmp</u>

rsync -rtvh /nexus/posix0/MAGE-flaski/service/hpc/group/<group>/<folder> /raven/ptmp/<user>/

```
hamin@raven01:~> rsync -rtvh /nexus/posix0/MAGE-flaski/service/hpc/group/Bioinformatics/fol1 /raven/ptmp/hamin/
sending incremental file list
fol1/
fol1/f1
fol1/f2
sent 176 bytes received 58 bytes 468.00 bytes/sec
total size is 0 speedup is 0.00
hamin@raven01:~> ls /raven/ptmp/hamin/fol1
f1 f2
```

Or to archive file system <u>/r</u>

rsync -rtvh /nexus/posix0/MAGE-flaski/service/hpc/group/<group>/<folder> /r/<username first letter>/<user>/

```
hamin@raven01:~> rsync -rtvh /nexus/posix0/MAGE-flaski/service/hpc/group/Bioinformatics/fol1 /r/h/hamin/
sending incremental file list
fol1/
fol1/f1
fol1/f2
sent 176 bytes received 58 bytes 468.00 bytes/sec
total size is 0 speedup is 0.00
hamin@raven01:~> ls /r/h/hamin/fol1
f1 f2
```

Thank You!

Documentation: https://github.com/mpg-age-bioinformatics/cluster_first_steps

Support: bioinformatics@age.mpg.de

Website: https://bioinformatics.age.mpg.de/