Slurm at UPPMAX

How to submit jobs with our queueing system

Jessica Nettelblad
sysadmin at UPPMAX

Slurm at UPPMAX

- Intro
- **Queueing with Slurm**
  - **How to submit jobs**
- Testing
  - How to test your scripts before submission
- Monitoring
  - How to keep track on how the job is doing
- Examples

Free!

Watch!
Futurama S2 Ep.4
Fry and the Slurm factory

Simple Linux Utility
for Resource
Management

slurm
workload manager

Open source!
https://github.com/SchedMD/slurm

Popular!
Used at many universities all
over the world

Submit jobs

Tell Slurm what the job needs

Which script?

- Syntax:
    - sbatch filename

```
sbatch startjob.sh
```

Input after file name

- Syntax:
  - sbatch filename input arguments

```
sbatch startjob.sh input1 input2
```

Flags before filename

- Syntax:
  - sbatch flags filename input arguments

```
sbatch -A g2017016 -p core -n 1 -t 10:00 startjob.sh input1 input2
```

Flags in bash script

- Syntax:

  - sbatch filename input arguments

```
sbatch startjob.sh input1 input2

cat startjob.sh
#!/bin/bash
#SBATCH –A g2017016
#SBATCH –p core
#SBATCH –n 1
#SBATCH –t 10:00


start
Do this
Do that
end
```

Important flags

- **-A**
    - **Project name**
    - `projinfo`
    - **Example: -A g2017016**

- **-t**
    - **Run time**
    - **days-hours:minutes:seconds**
    - **Max: 10 days, 10-00:00:00**
    - **Example: -t 3-10:00:00**

- -p
    - Partition
    - One or more cores: core, devcore
    - One or more nodes: node, devel
    - Example: -p core

- -n
    - Number of tasks (cores)
    - Enough cores for threads: Defined by the job script or program used
    - Enough cores for memory: Test! Start low.
    - Example: 1

```
sbatch –A g2017016 –p core –n 1 –t 20:00 startjob.sh
```

Submit startjob.sh to Slurm, account it to project g2017016, let it run on 1 core for 20 minutes.

More flags

- **Job name**
  - -J testjob

- **--res=g2017016_WED**
  - Use reserved nodes

- Email notifications
  - --mail-type=FAIL
  - --mail-type=TIME_LIMIT_80
  - [--mail-user=jessica.nettelblad@it.uu.se](mailto:jessica.nettelblad@it.uu.se)

- Output redirections
  - Default: work directory
  - --output=/proj/g2017016/nobackup/private/jessine/testjob/output/

- https://slurm.schedmd.com/sbatch.html
  - Most, but not all options are available at every center
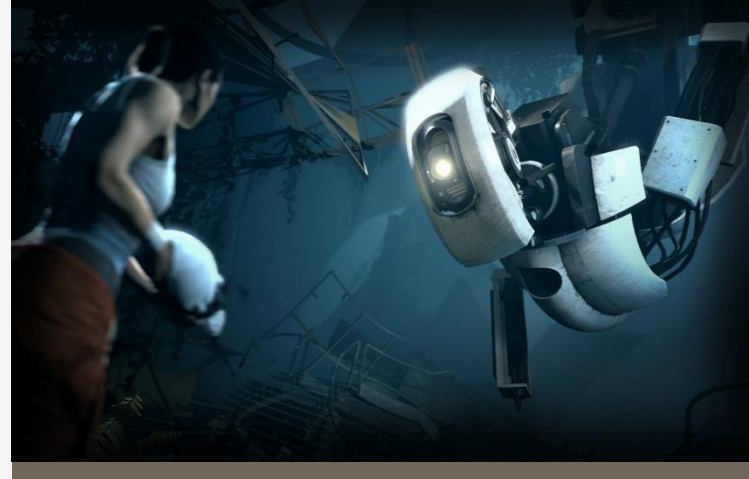
Monitoring

Keep track of the job
- In queue
- While running
- When finished

Check queue

- **jobinfo**
  - Shows all jobs in queue. Modified squeue.
  - https://slurm.schedmd.com/squeue.html

- How many jobs are running?
  - `jobinfo | less`
    - Type q to exit
- When are my jobs estimated to start?
  - `jobinfo –u jessine`

Understanding the queue

- Priority
  - FAQ: http://www.uppmax.uu.se/support/faq/running-jobs-faq/your-priority-in-the-waiting-job-queue/

- Bonus jobs
  - Most projects has a quota of 2000 hours per 30 day
  - More than quota? Bonus job!
  - Can still run, no limits – but lower priority
  - How much have your project used? `projinfo`

- Dependency
  - Jobs won't start until another has finished successfully
  - Set with --dependency

Check progress

- **`uppmax_jobstats`**
    - less /sw/share/slurm/*/uppmax_jobstats/*/<job id>
    - Shows memory and core usage

- `scontrol show job <job id>`

- ssh to the compute node
    - `top`
    - `tail -f` (on result file)

Check finished job

- **`slurm.out`**
  - Check it for every job
  - Look for error messages
- **`uppmax_jobstats`**
- `finishedjobinfo –s today`

Cancel job

- Cancel all my jobs
  - `Scancel –u jessine`
- Cancel all my jobs named testjob
  - `scancel –j testjob`
- Cancel job with job id 4456
  - `scancel 4456`
  - Find job id? `jobinfo –u jessine`

Testing

Test using the
- interactive command
- dev partition
- fast lane

Testing in interactive mode

- `interactive` **instead of** `sbatch`
- No script needed
- `interactive -A g2017016 -t 15:00`

- Example:
  - A job script didn't work. I start an interactive job and submit line for line.

Testing in devel partition

- `-p devcore –n 4 –t 60`
  - Normal sbatch job: sbatch flags jobscript input
  - Run on four cores for 60 minutes
  - Note: Max one job submitted
- Job starts quickly!

- Example:
  - Before submitting the real job, I want to make sure the script and submit works as intended.

Testing in a fast lane

- `--qos=short`
  - Up to 15 minutes, up to four nodes

- `--qos=interact`
  - Up to 12 hours, up to one node

- Example:
  - My job is shorter than 15 minutes or 12 hours. I add a qos flag, and the job will get higher priority.
  - Note: Only one running job with qos – the rest will stay in queue.

Examples

Basic example

```
#!/bin/bash -l
#SBATCH -A g2016011
#SBATCH -p core
#SBATCH -n 1
#SBATCH -t 10:00:00
#SBATCH -J jour2

module load bioinfo-tools samtools/0.1.19 bwa
export SRCDIR=$HOME/baz/run3

cp $SRCDIR/foo.pl $SRCDIR/bar.txt $SNIC_TMP/
cd $SNIC_TMP

./foo.pl bar.txt

cp *.out $SRCDIR/out2
```

Basic example explained

- `#!/bin/bash`
    – starts the bash interpreter

- `#SBATCH -A g2016011`
    – "#" starts a comment that bash ignores
    – "#SBATCH" is a special signal to SLURM
    – "-A" specifies which account = project will be "charged".

- `#SBATCH -p core`
    – sets the partition to core, for jobs that uses less than one node.

- `#SBATCH -n 1`
    – requests one task = one core

Basic example explained

- `#SBATCH -t 10:00:00`
    - Time requested: 10 hours.
- `#SBATCH -J jour2`
    - jour2 is the name for this job
    - mainly for your convenience
- `module load bioinfo-tools samtools/0.1.19 bwa`
    - bioinfo-tools, samtools version 0.1.19 and bwa is loaded.
    - can specify versions or use default (risky)
- `export SRCDIR=$HOME/baz/run3`
    - Environment variable SRCDIR is defined

Basic example explained

- `cp $SRCDIR/foo.pl $SRCDIR/bar.txt $SNIC_TMP/`
- `cd $SNIC_TMP`
    - Copy foo.pl and bar.txt to $SNIC_TMP, then go there.
    - $SNIC_TMP is a job specific directory on the compute nodes.
    - Recommended! Can be much faster than home.
- `./foo.pl bar.txt`
    – Actual script with code to do something.
    – Call one command, or a long list of actions with if-then, etc.
- `cp *.out $SRCDIR/out2`
    - $SNIC_TMP is a temporary folder. It's deleted when job is finished.
    - Remember to copy back any results you need!

Group jobs

```
#!/bin/bash -l
#SBATCH -A g2016011
#SBATCH -p core
#SBATCH -n 4
#SBATCH -t 2-00:00:00
#SBATCH -J br_para_02

B=/sw/data/uppnex/igenomes/
S=Annotation/Genes/genes.gtf

grep AT2G48160 $B/Arabidopsis_thaliana/Ensembl/TAIR9/$S &
grep NM_030937 $B/Homo_sapiens/NCBI/GRCh38/$S &
grep NM_025773 $B/Mus_musculus/NCBI/GRCm38/$S &
grep XM_008758627.1 $B/Rattus_norvegicus/NCBI/Rnor_6.0/$S &
wait
```

Split jobs

```
#!/bin/bash

TOOL=z_tools/3
IGEN=/sw/data/uppnex/igenomes

cd $IGEN

for v in [[:upper:]]*/*/*
do
  echo $v
  cd $TOOL
  sbatch star_index.job $v
  cd $IGEN
  sleep 1
done
```

We're here to help!

- If you run into problems after this course? Just ask someone for help!
  - Check userguides and FAQ on uppmax.uu.se
  - Ask your colleagues
  - Ask UPPMAX support: support@uppmax.uu.se