HPC

Introduction to Unix and HPC

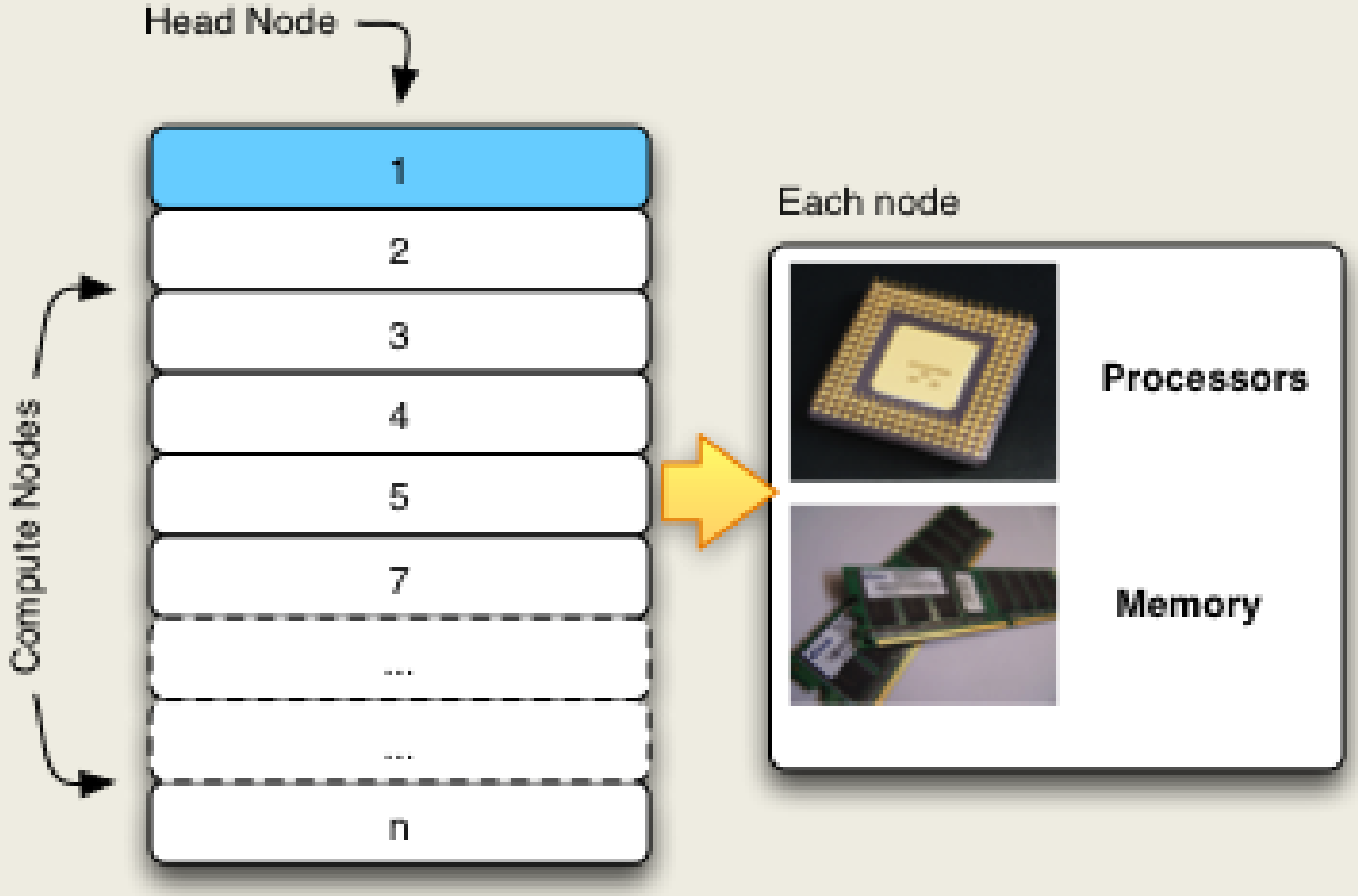
HPC machines

System	Memory Arch	Cores	Nodes	Memory
Octane (training)	Distributed	48	3	48GB
Orange	**Distributed**	**1,600**	**100**	**8TB**
NCI - current	**Distributed**	**11,936**	**1492**	**37TB**
NCI - Q1 2013	**Distributed**	**57,472**	**?**	**158TB**

The typical HPC workflow

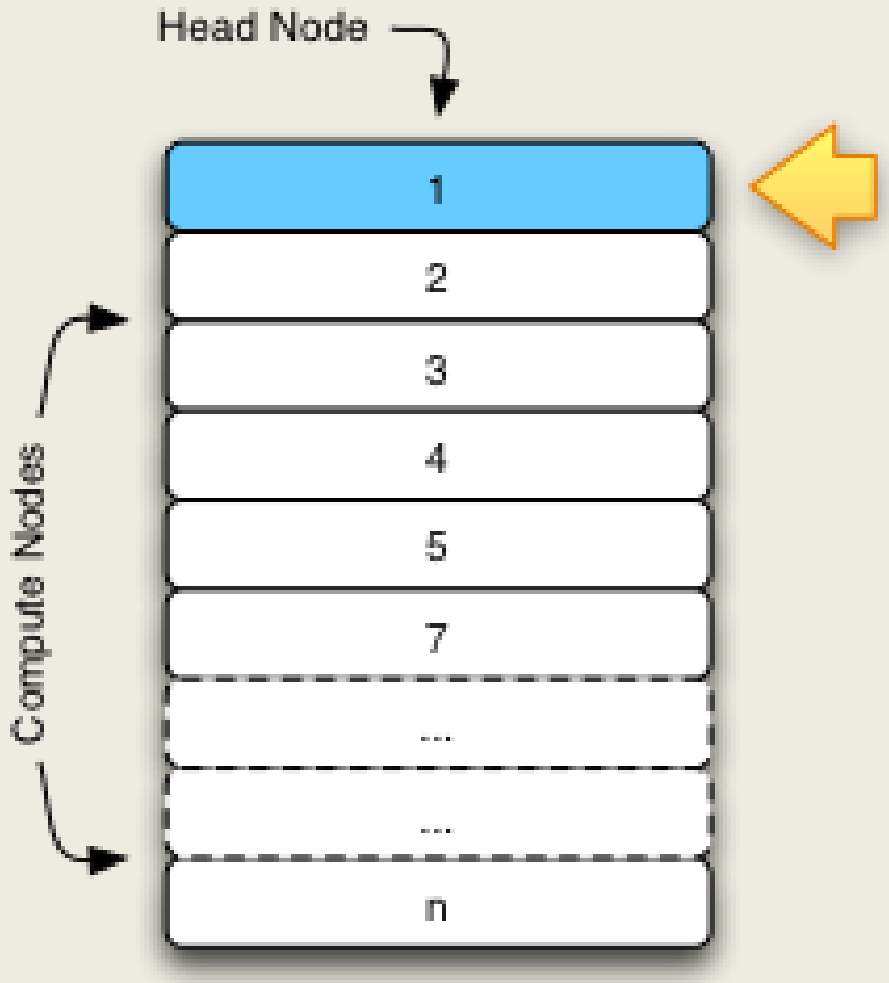
- In HPC we talk about **jobs**, these are simply commands we wish to run.
- They are generally time consuming and resource intensive.
- Jobs are run **non-interactively**.
- We add our jobs to a **queue**.
- When the machine has free resources the jobs run.
- Once jobs complete, we can inspect their output.

The HPC "Cluster"



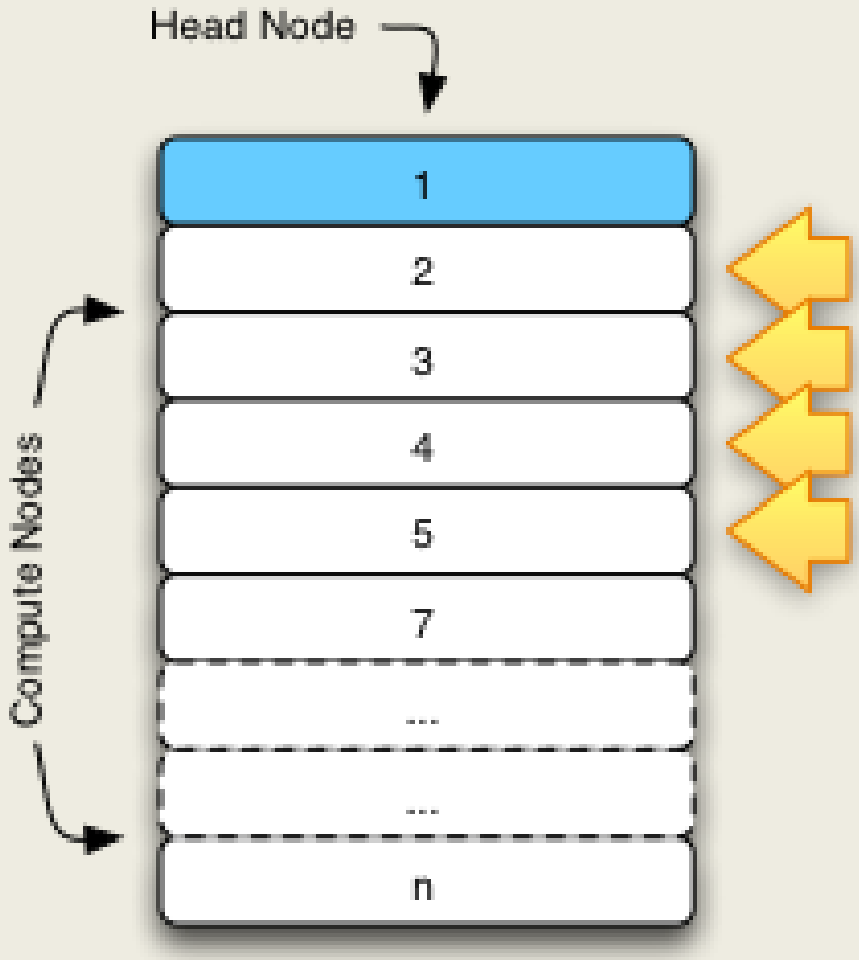
INTERSECT

The Head Node



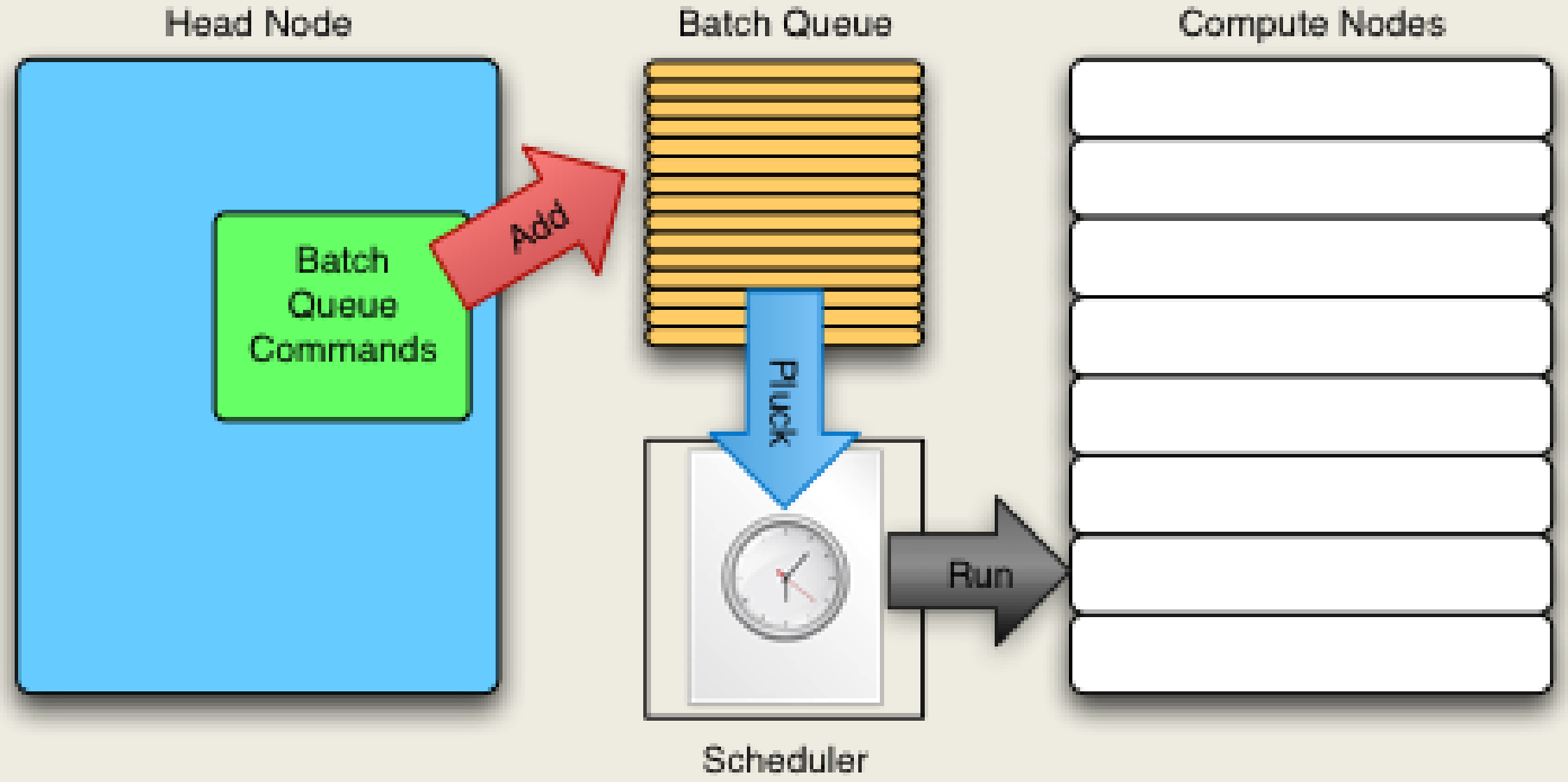
- Interactive programs
- SSH sessions
- Testing
- Compiling
- Queuing jobs

Compute Nodes



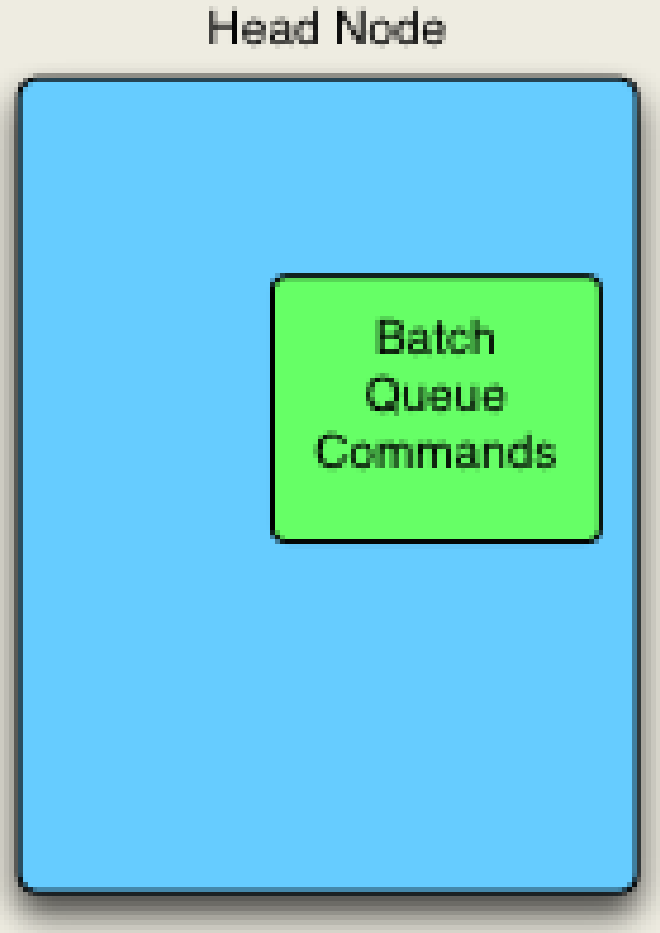
- These run your jobs
- Managed by the **scheduler**
- You never interact with them directly

The Batch Queuing System



INTERSECT

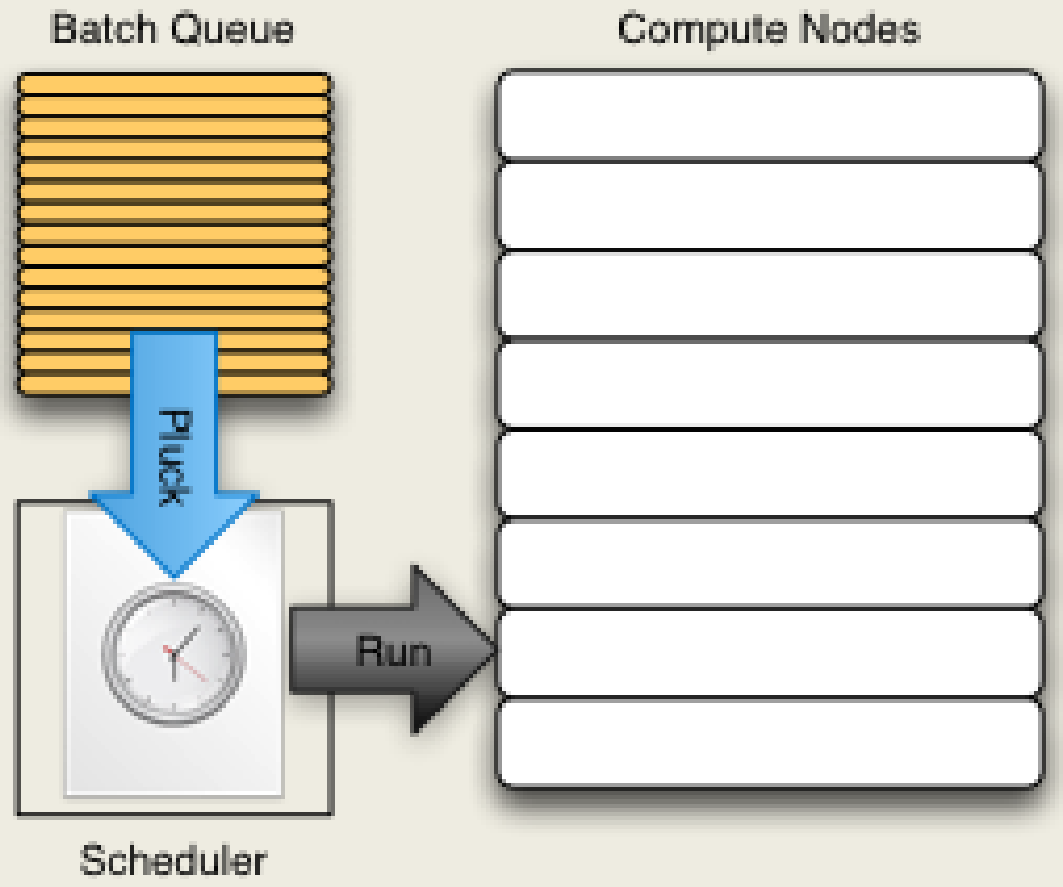
Batch Queuing component



- The batch system is a normal program
- Lets you add and remove jobs from the queue and monitor the queue
- Script/command line driven
- **ANUPBS on NCI**
- **PBSPro on Orange**

The Scheduler component

- Allocates jobs to compute nodes
- Optimizes usage of resources
- "Optimize" can mean many things
- Non-trivial
- Never interact with directly



INTERSECT

PBSPro Batch System

In order to use the batch system productively, we need to know how to perform three actions:

Add a job to the queue
Remove a job from the queue
See where our job is in the queue

Command	Description
qsub <job-script>	Submit a job (add to queue) Returns a <job-number>
qdel <job-number>	Delete job (remove from the queue)
qstat <job-number>	Monitor jobs

Monitoring the queue

Command	Description
qstat -a	List all jobs in the queue
qstat -u <username>	List all jobs of a particular user
qstat -f <job-number>	Show detailed information about a job

Exercise 1: Monitoring the queue with qstat

Add a job to the queue

- To add a job to the queue, we write a **job script**.
- The job script is a simply a BASH script.
- It has some special comments that pass information to PBSPro.
- When we want to queue the job, we pass its filename as a parameter to **qsub**:
- **qsub <job-script>**
- The batch queuing system will return a number that uniquely identifies the job.

A sample job script

```
#!/bin/bash
# Request resources
# * 10 minutes wall time to run
#PBS -l walltime=00:10:00
# * 1 node, 1 processor
#PBS -l nodes=1:ppn=1
# * 100 megabytes physical memory allocated to job
#PBS -l mem=100mb
# Specify a project code (for accounting)
#PBS -P a40
# Move to directory job was submitted in
cd $PBS_O_WORKDIR
# Specify the job to be done
date
sleep 10
date
```

You've got mail!

```
# Set email address
#PBS -M fred@intersect.org.au
# Send an email when jobs
# begins (b), gets aborted (a)
# and ends (e)
#PBS -m abe
```

Exercise 2: Submitting a sample job

Useful Environment Variables

- These are available in the context of your job script.

Command	Description
PBS_O_WORKDIR	The directory the job was submitted from
PBS_JOBID	The job number given when the job was submitted

Job limits on Orange

- 200 hours of **walltime**
- 96 **CPU cores**
- If requesting over 16 cores, requests must be in multiples of 16
- 64GB of **memory** per 16 cores on a standard node. e.g. 128GB for 32 cores
- 128GB of **memory** per 16 cores for the large memory nodes

Priorities of Jobs

- Resources available to the project
- Walltime
- Number of cores
- Memory
- Number of jobs (fair share)

Best strategy

- Submit jobs constantly/daily
- Have about 10-20 jobs in the machine
- Be realistic with walltime
- Don’t ask for resources you don’t need!

Disk Partitions - Orange

/home	
Mounted under:	/home/username
Size:	60GB default
Backed up:	Yes
Speed:	Fast disk (parallel file system)
Life time:	Permanent

Disk Partitions - Orange

/home/project-name	
Mounted under:	/home/project-name
Size:	no default size
Backed up:	No
Speed:	Medium speed
Life time:	6 months - merit allocation period

There will also be some “repository space” for large datasets, such as bioinformatics databases

Disk Partitions - Orange

/scratch/2weeks	
Mounted under:	/scratch/2weeks
Size:	Limit of disk - 5TB
Backed up:	No
Speed:	Medium speed
Life time:	6 months - merit allocation period

Warning: Shared among users, so important consider file permissions!

Disk Partitions

You can find out more about the partitions on the HPC machine using the **df** command.

Command	Description
df -h	Show disk free space for all partitions in human readable format

You can find out more about current disk usage, using the **du** command.

Command	Description
du -hs .	Show disk usage of current directory in human readable format

Quotas

- You can use the **quota** command to find out about disk usage quotas in force.
- There is no quota on /scratch for performance reasons.
- The quota on /home/project-name depends on your allocation.

Command	Description
quota	Shows quotas on disk usage

Conclusion

- In this course we have covered the basics of the Unix command line, transferring data, and the specifics of our HPC machine.
- Next steps:
  - Apply for a start up account on Orange or NCI
  - Complete our survey
- Any questions?