High Performance Computing for Genomics

Part I: High Performance Computing

Overview

- Login
- Storage
- Credit system
- Module system
- Hardware
- PBS
- Job Submission

Login

Login is possible through various methods.

Command line example (Linux/Mac):

```
$ ssh -X vsc3XXXX@login1-tier2.hpc.kuleuven.be
```

Or via putty (Windows)

List of Login nodes

Node	URL
Genius	login1-tier2.hpc.kuleuven.be
Genius	login2-tier2.hpc.kuleuven.be
Genius (NX Client)	login3-tier2.hpc.kuleuven.be
Genius (NX Client)	login4-tier2.hpc.kuleuven.be

ssh -X vsc3XXXX@login1-tier2.hpc.kuleuven.be

If you are not logged in yet,

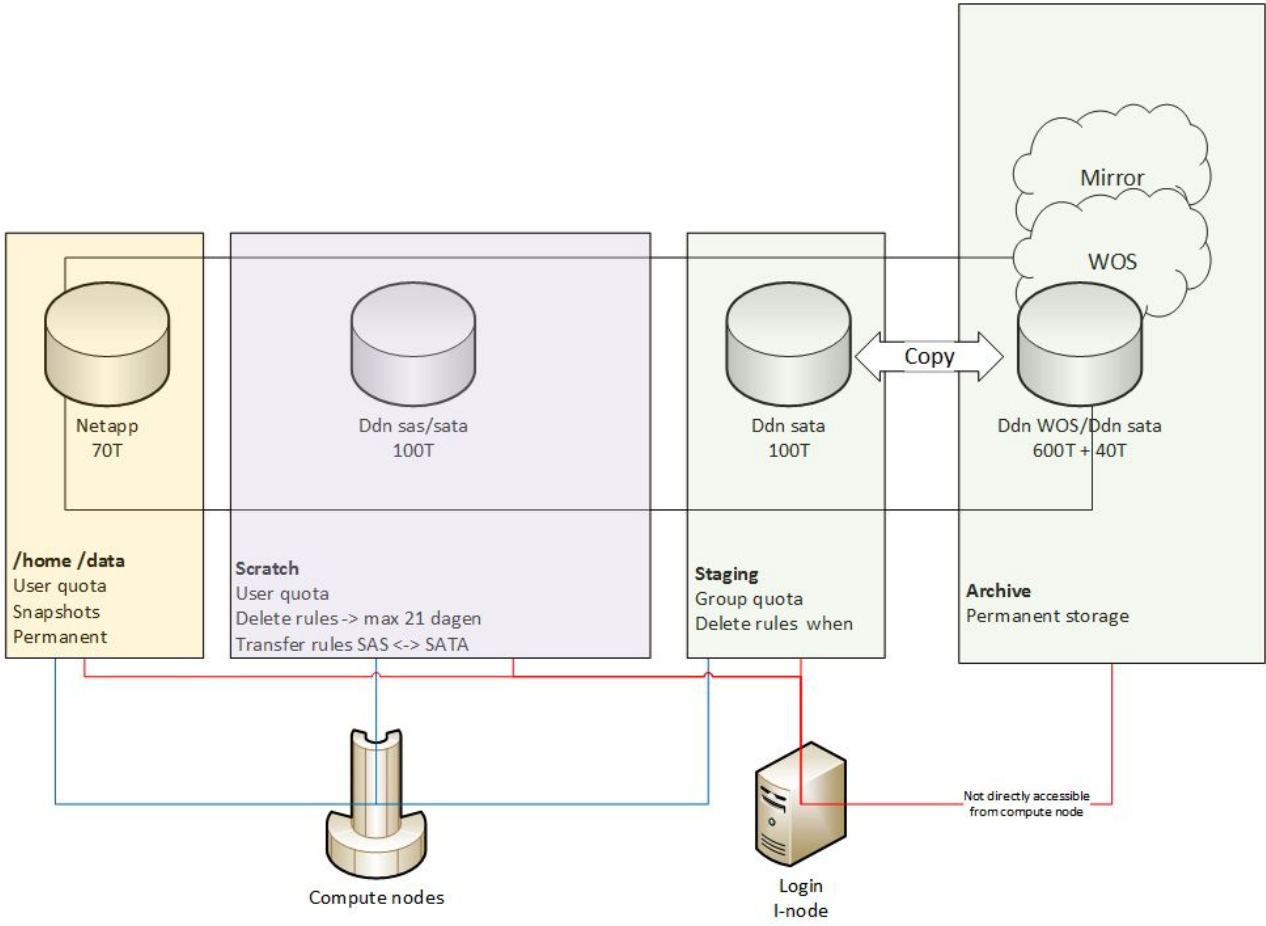
Now is the moment

Overview

- Login
- Storage
- Credit system
- Module system
- Hardware
- PBS
- Job Submission

Storage

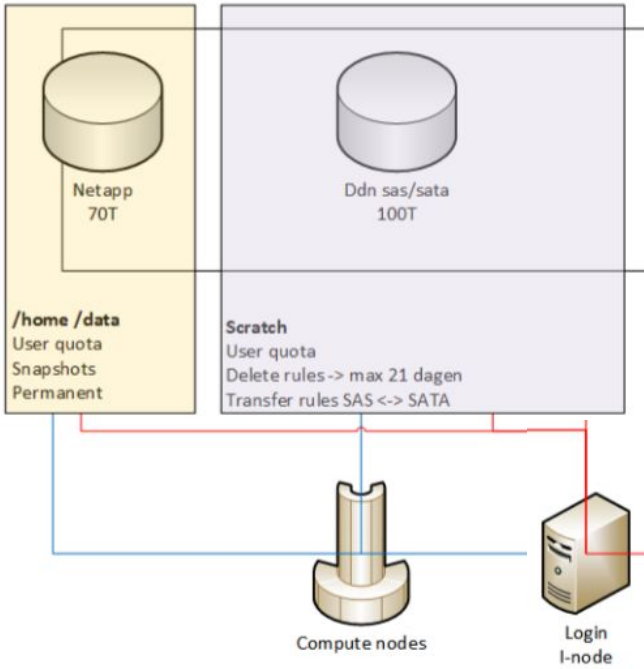
- NAS storage, fully back-up with snapshots for /home and /data
- Scratch storage, fast parallel filesystem
- Staging storage, to store current working projects
- Archive storage, to store large amounts of data for long time



Personal Storage

Personal Storage, only accessible by the owner
Fully Backup

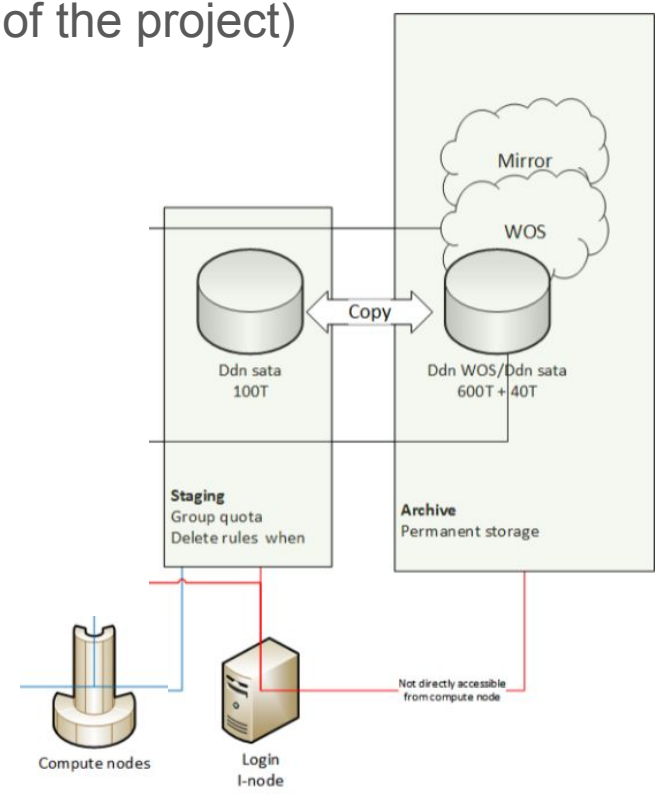
- Home Directory
  - $VSC_HOME or /user/leuven/3XX/vsc3XXXX
  - 25GB
  - Important data, configuration files
- Data Directory
  - $VSC_DATA or /data/leuven/3XX/vsc3XXXX
  - 75GB
  - Important data, biger then Home
- Scratch Space
  - $VSC_SCRATCH or /scratch/leuven/3XX/vsc3XXXX
  - 100GB
  - Temporary date, will be deleted within 21 days



Project Storage

Permissions are managed per group (usually the group of the project)

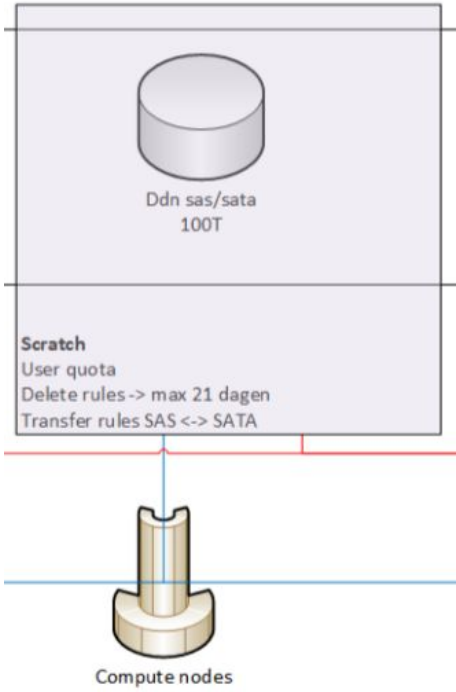
- Staging Space
  - Storage for project files (while working at the project)
  - Size is project dependable (per 1TB)
- Archive Storage
  - Long term storage and backup for project files (or staging)
  - Size is project dependable (per 1TB)
  - Only accessible from the login node



Temporary Storage

Scratch: fast storage, used during the processing of the data

- Node Scratch
  - $VSC_SCRATCH_NODE
  - Size is machine dependable, min 150 GB
  - Is NOT accessible from the login node
- Site Scratch
  - $VSC_SCRATCH_SITE
  - Similar as $VSC_SCRATCH_NODE
- Personal Scratch
  - $VSC_SCRATCH
  - Personal scratch, 100GB
  - Accessible from the login node
  - Data will be deleted within 21 days



Exercise 1

Find the path for the following locations:

- $VSC_HOME
- $VSC_DATA
- $VSC_SCRATCH
- $VSC_SCRATCH_NODE

Copy Data

```
$ rsync -ahr --progress /staging/leuven/stg_00019/workshop/* .
```

rsync --progress -ahr filename targetDirectory

--progress	Show the progress of the copy
-a	Preserve almost everything
-h	Output numbers must be Human readable
-r	Copy directories recursively

Exercise 1b

Go to your data directory

Clone the github repository:

```
$ git clone https://github.com/GenomicsCoreLeuven/vsc_ngs_workshop.git
```

Overview

- Login
- Storage
- Credit system
- Module system
- Hardware
- PBS
- Job Submission

Credit System

VSC ~ regular bank

Each job costs a certain number of credits

Each job belongs to a project

Each project account has its own credits

Users have access to one or multiple project accounts
(multiple users can have access to the same project account)

Credit System: balance

Load the accounting module

```
$ module load accounting
```

Check the amount of credits at your disposal

```
$ mam-balance
```

Credit System: statement

Get an overview of the transactions

```
$ mam-statement
```

Get an overview of the transactions of a certain project, over a certain time

```
$ mam-statement -g lp_projectname -s 2015-09-01 -e 2015-09-30
```

-g	groupname
-u	username
-s	start
-e	end

Credits System: quotes

Estimating the cost of a job is possible by requesting a quote:

```
$ gquote -l nodes=3:ppn=4:ivybridge,pmem=2gb,walltime=48:00:00
```

- nodes=3:ppn=4:ivybridge
  3 nodes, with 4 processors per node of the ivybridge type
- pmem=2gb
  2GB of memory (RAM)
- walltime=48:00:00
  The job will run for maximum 48 hours

Job Cost Calculation

The effective cost of the job is calculated using this formula:

```
(0.000278*nodes*walltime + startup)*nodetype
```

nodes	The number of nodes that were reserved
walltime	The effective duration of the job (in seconds)
startup	0.1, it is added for the overhead of scheduling
nodetype	A representation of the node type's performance

Exercise 2

- Get a quote for 5 minutes computing on 1 node, 5 processors of the ivybridge type
- Get a quote for 20 minutes computing on 2 nodes, 4 processors of the haswell type, using 100Gb of memory
- Get a quote for 1 hour for both the ivybridge and the haswell processor

Overview

- Login
- Storage
- Credit system
- Module system
- Hardware
- PBS
- Job Submission

Module System

For many programs, multiple versions are installed, and each version requires specific libraries.

Toolchains consists of compilers and libraries, together with software depending on those libraries.

The module system inside the toolchain is used to manage the environment variables, and all dependencies to resolve possible conflicts.

Module System: Module av

Many software packages are installed as modules:

```
$ module av
```

```
------------------------------ /apps/leuven/skylake/2018a/modules/all ------------------------------
   7zip/16.02-GCCcore-6.4.0
   ABINIT/8.4.4-intel-2018a
   ABINIT/8.8.3-intel-2018a                                        (D)
   ABySS/2.0.2-foss-2018a
   ADMIXTOOLS/5.1-foss-2018a
   ANTLR/2.7.7-intel-2018a-Python-3.6.4
   ANTs/2.3.1-foss-2018a-Python-2.7.14
   ATK/2.28.1-foss-2018a
```

Module System: load and unload

Load a module

```
$ module load BEAST

$ module load  BEAST/2.1.2

$ module load  BEAST/2.1.2  zlib/1.2.8-foss-2014a
```

Unload a module

```
$ module unload BEAST
```

Unload all modules

```
$ module purge
```

Module System: list

View all loaded modules

```
$ module list
Currently Loaded Modulefiles:
  1) /thinking/2014a
  2) Java/1.7.0_51
  3) icc/2013.5.192
  4) ifort/2013.5.192
  5) impi/4.1.3.045
  6) imkl/11.1.1.106
  7) intel/2014a
  8) beagle-lib/20140304-intel-2014a
  9) BEAST/2.1.2
 10) GCC/4.8.2
 11) OpenMPI/1.6.5-GCC-4.8.2
 12) gompi/2014a
 13) OpenBLAS/0.2.8-gompi-2014a-LAPACK-3.5.0
 14) FFTW/3.3.3-gompi-2014a
 15) ScaLAPACK/2.0.2-gompi-2014a-OpenBLAS-0.2.8-LAPACK-3.5.0
 16) foss/2014a
 17) zlib/1.2.8-foss-2014a
```

Overview

- Login
- Storage
- Credit system
- Module system
- Hardware
- PBS
- Job Submission

Hardware

- Genius
  Thin node cluster
- Superdome
  Shared Memory Processing
- Accelerators
  Like GPUs
- Storage

Hardware: Genius

- Genius
  - Main Cluster
  - 3 types of nodes:
    - Thin nodes (Skylake)
      86 nodes, 2x18 cores/node, 192GB RAM
    - Big nodes (Skylake)
      10 nodes, 2x18 cores/node, 768 GB RAM
      Partition: bigmem
    - GPGPU (Skylake)
      20 nodes, 2x18 cores/node, 192GB RAM
      4 NVIDEA P100, 16GB GDDR
      Partition: gpu
- Superdome

Hardware: Genius

- Genius
- Superdome
  - 1 type of node:
    - Shared Memory nodes (Skylake)
      8 nodes, 14 cores/node, 750GB RAM
      Partition: superdome
  - Up to 6TB RAM!

Hardware: Overview

Cluster	Partition	#cores (threads) per node	Usable Memory	#Credits/hour
Genius		36	192GB	10
Genius	bigmem	36	768GB	10
Genius	gpu	36 + 4 GPU	192GB	
Genius	superdome	14	Shared 750GB	10

The Operating System also needs some RAM, good practice is to reserve 4GB.

Hardware: which task on which cluster?

Cluster	Partition	Task Description	Task Example
Genius	Regular And gpu	Memory low jobs, with lots of I/O	• Alignment • Read Mapping • Variant Calling • Read Counting • ...
Genius	bigmem	High memory jobs	• Full Genome (elprep, ...) • Small De Novo Assemblies • Reference-based Assemblies • ...
Genius	superdome	Really High memory jobs, computing power less important	• Large De Novo Assemblies • ...

Overview

- Login
- Storage
- Credit system
- Module system
- Hardware
- PBS
- Job Submission

Portable Batch System (PBS)

Portable Batch System (or simply PBS) is the name of computer software that performs job scheduling. Its primary task is to allocate computational tasks, i.e., batch jobs, among the available computing resources. It is often used in conjunction with UNIX cluster environments. PBS is supported as a job scheduler mechanism by several meta schedulers ...

-- by Wikipedia

PBS script

Every shell script (.sh) can be turned into a PBS script (.pbs).
A typical scripts will need the following adjustments:

1. Include PBS headers
2. Load the software modules
3. Optimize I/O using a scratch node
4. Check if results are copied/stored in a data or staging directory

PBS header

Headers contain the run parameters

Each line starts with #PBS then the parameter and the value

```
#!/bin/bash -l
#PBS -l walltime=12:00:00
#PBS -l mem=100gb
#PBS -l nodes=1:ppn=20
#PBS -M mail@mail.com
#PBS -m aeb
#PBS -N jobname
#PBS -A lp_projectname
```

PBS header: job description

Walltime: the maximum time the job can run

```
#PBS -l walltime=12:00:00
```

Mem: the maximum memory available for the job

```
#PBS -l mem=100gb
```

Number of nodes, processors per node (ppn) and the processor type

```
#PBS -l nodes=1:ppn=20
```

PBS header: notifications

You can send a mail @ certain conditions

-M specifies the mail adres

```
#PBS -M mail@mail.com
#PBS -m aeb
```

m the conditions when a mail has to be send:

b	When the job begins
e	When the job ends
a	When the job is aborted

PBS header: billing

The name of the job, this will be visible for other users

```
#PBS -N jobname
```

The project that will be used for the billing

```
#PBS -A lp_projectname
```

PBS script

Every shell script (.sh) can be turned into a PBS script (.pbs).
A typical scripts will need the following adjustments:

1. Include PBS headers
2. **Load the software modules**

```
module load  BEAST/2.1.2
```

3. Optimize I/O using a scratch node
4. Check if results are copied/stored in a data or staging directory

PBS script

Every shell script (.sh) can be turned into a PBS script (.pbs).
A typical scripts will need the following adjustments:

1. Include PBS headers
2. Load the software modules
3. **Optimize I/O using a scratch node**
   $VSC_SCRATCH, $VSC_SCRATCH_NODE
   Scratch is fast temporary storage
4. Check if results are copied/stored in a data or staging directory

PBS script

Every shell script (.sh) can be turned into a PBS script (.pbs).
A typical scripts will need the following adjustments:

1. Include PBS headers
2. Load the software modules
3. Optimize I/O using a scratch node
4. **Check if results are copied/stored in a data or staging directory**
   Scratch is temporary, when the job is finished, data will be lost

Overview

- Login
- Storage
- Credit system
- Module system
- Hardware
- PBS
- Job Submission

Job submission

On all partitions:

```
$ qsub run-job.pbs
```

You have to specify the partition in the PBS script (superdome request an extra line):

```
#PBS -L tasks=1:lprocs=14:place=numanode:memory=700gb
#PBS -l partition=superdome
```

When a job is submitted, the job id is returned.
jobIDs are unique.

Jobs progress

An overview of all your jobs:

```
$ qstat -u vsc3XXXX
hpc-p-svcs-5:

Req'd       Req'd       Elap
Job ID                  Username    Queue    Jobname
SessID  NDS   TSK   Memory       Time     S   Time
----------------------- ----------- -------- ----------------
------ ----- ------ --------- --------- - ---------
30017323.hpc-p-svcs-5   vsc3XXXX    q21d     jobname
197071     1     10       1gb 480:00:00 R  00:04:16
```

Job estimated start

Get the estimated start of a job:

```
$ showstart 30017323.hpc-p-svcs-5
job 30017323 requires 10 procs for 20:00:00:00
Estimated Rsv based start in                 00:00:00 on Mon Apr
18 16:29:10
Estimated Rsv based completion in          20:00:00:00 on Sun May
8 16:29:10
Best Partition: smp1
NOTE:  job is running
```

Job overview

An overview of every parameter and all used resources can be requested:

```
$ checkjob 30017323.hpc-p-svcs-5
job 30017323
AName: jobname
State: Running
Creds:  user:vsc3XXXX  group:vsc3XXXX  account:lp_projectname
class:q21d  qos:normal
WallTime:   00:02:43 of 20:00:00:00
SubmitTime: Mon Apr 18 16:28:56
  (Time Queued  Total: 00:00:14  Eligible: 00:00:14)
StartTime: Mon Apr 18 16:29:10
TemplateSets:  DEFAULT
Total Requested Tasks: 10
Req[0]  TaskCount: 10  Partition: smp1
Memory >= 1024M  Disk >= 0  Swap >= 0
Dedicated Resources Per Task: PROCS: 1  MEM: 24G
```

Job stop/kill/delete

Sometimes a job has to be killed.

```
qdel 30017323.hpc-p-svcs-5
```

Exercise 3

1. Open paths.pbs
   Edit the PBS header (change the mail, and accounting information)
2. Ask a quote based on the PBS header
3. Start the job on thinking
4. Check the start time and status of the job
5. Check the output of the job

Coffee break