CANCER RESEARCH UK | CAMBRIDGE INSTITUTE

CRUK CI cluster introduction

Using the Cambridge Institute's High Performance Computing Facilities

UNIVERSITY OF CAMBRIDGE

Overview

This brief course will give you two things:

1. A refresher on unix and an introduction to cluster computing

2. Basic instruction on using our scheduler

3. Some performance hints

It *won't* make you an expert on parallel computing and HPC, but will let you get to work.

This course has a practical component, for which you will need an ssh client and cluster account.

Session I

1	Unix refresher
2	Cluster introduction
3	Practical – unix processes

Session II

4	Using the scheduler
5	Practical – job submission

Session III

6	Some performance hints

CANCER RESEARCH UK | CAMBRIDGE INSTITUTE

Unix refresher

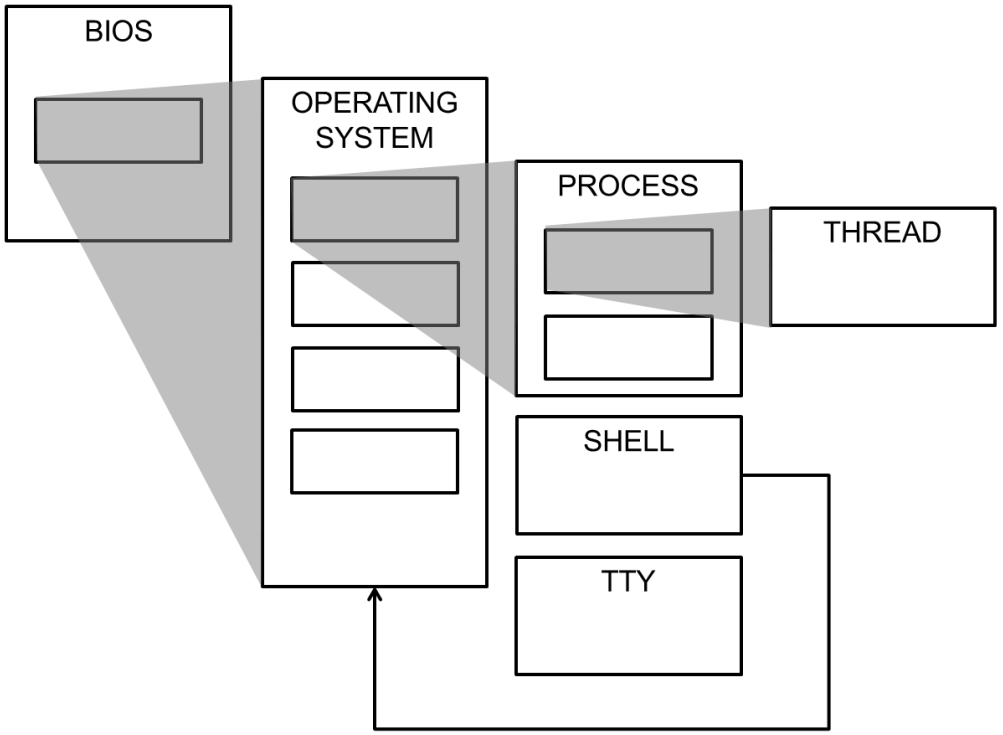
(we have a course if this is all new…)

UNIVERSITY OF CAMBRIDGE

Operating Systems and Processes

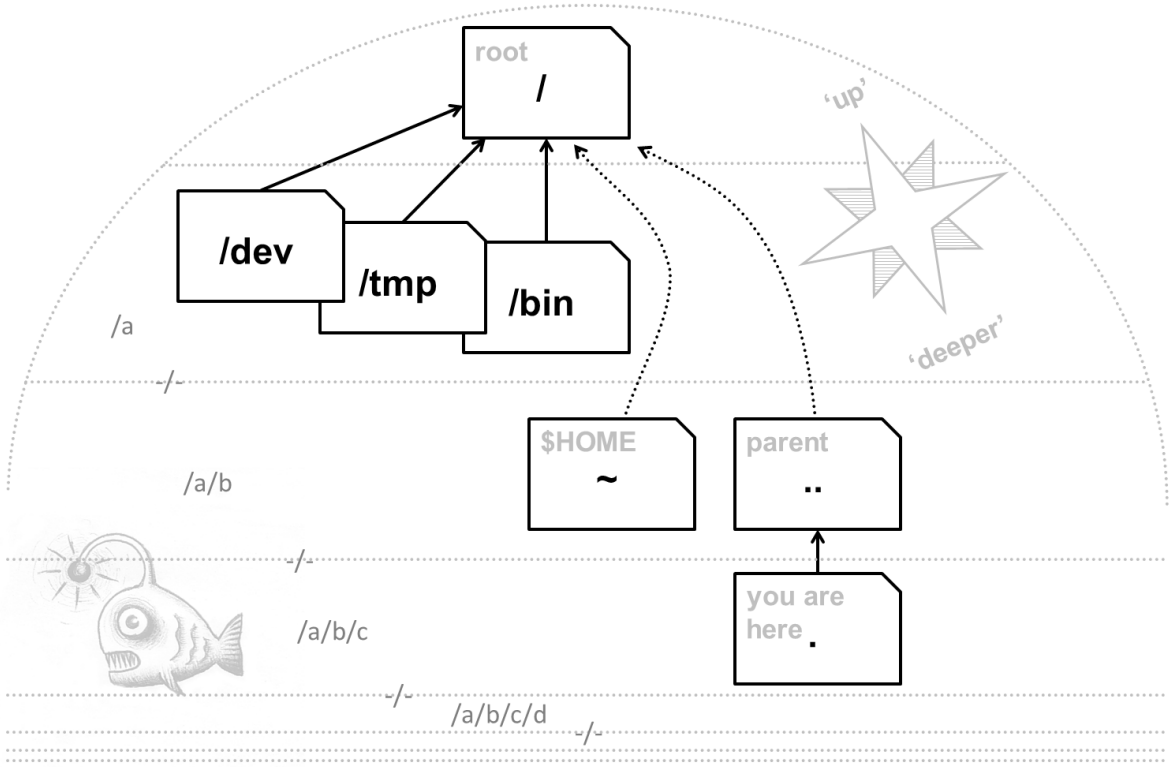
'Unix' or 'linux' (or 'UNIX') is our *operating system* – the program that controls the processes and their access to the network, screen, etc.

The shell is a *process* – it happens to be one that can see its own OS, which is one of the reasons it's so useful.



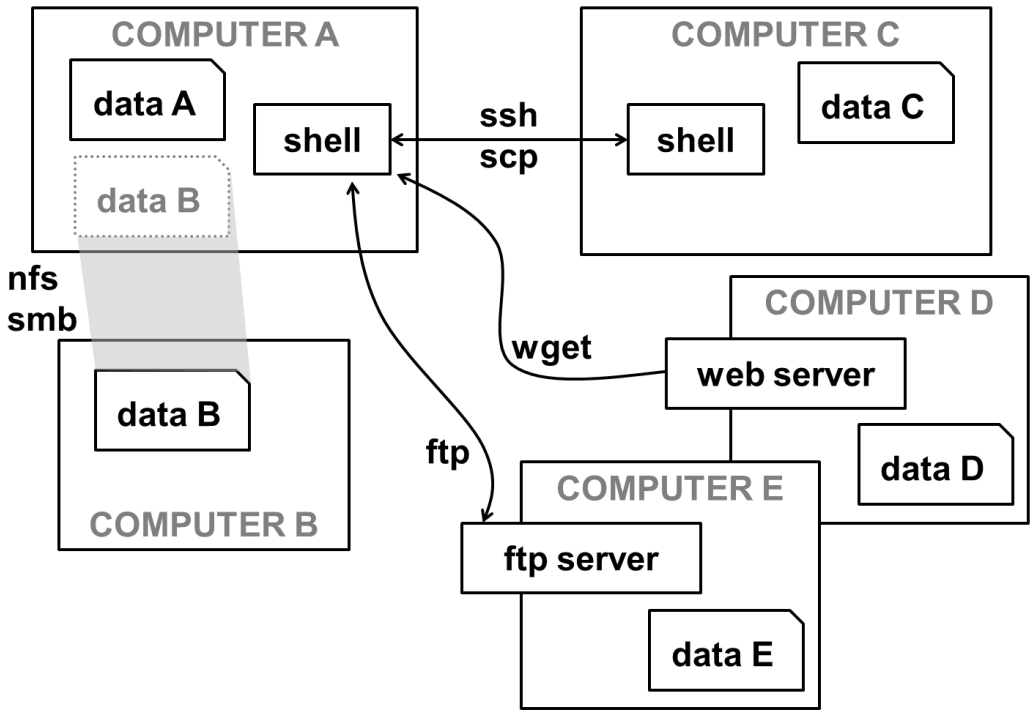
Navigation concepts

You need to be able to navigate without a GUI.
Fortunately some things are always in the same place.
Unix file systems are trees, *with the roots at the top*.



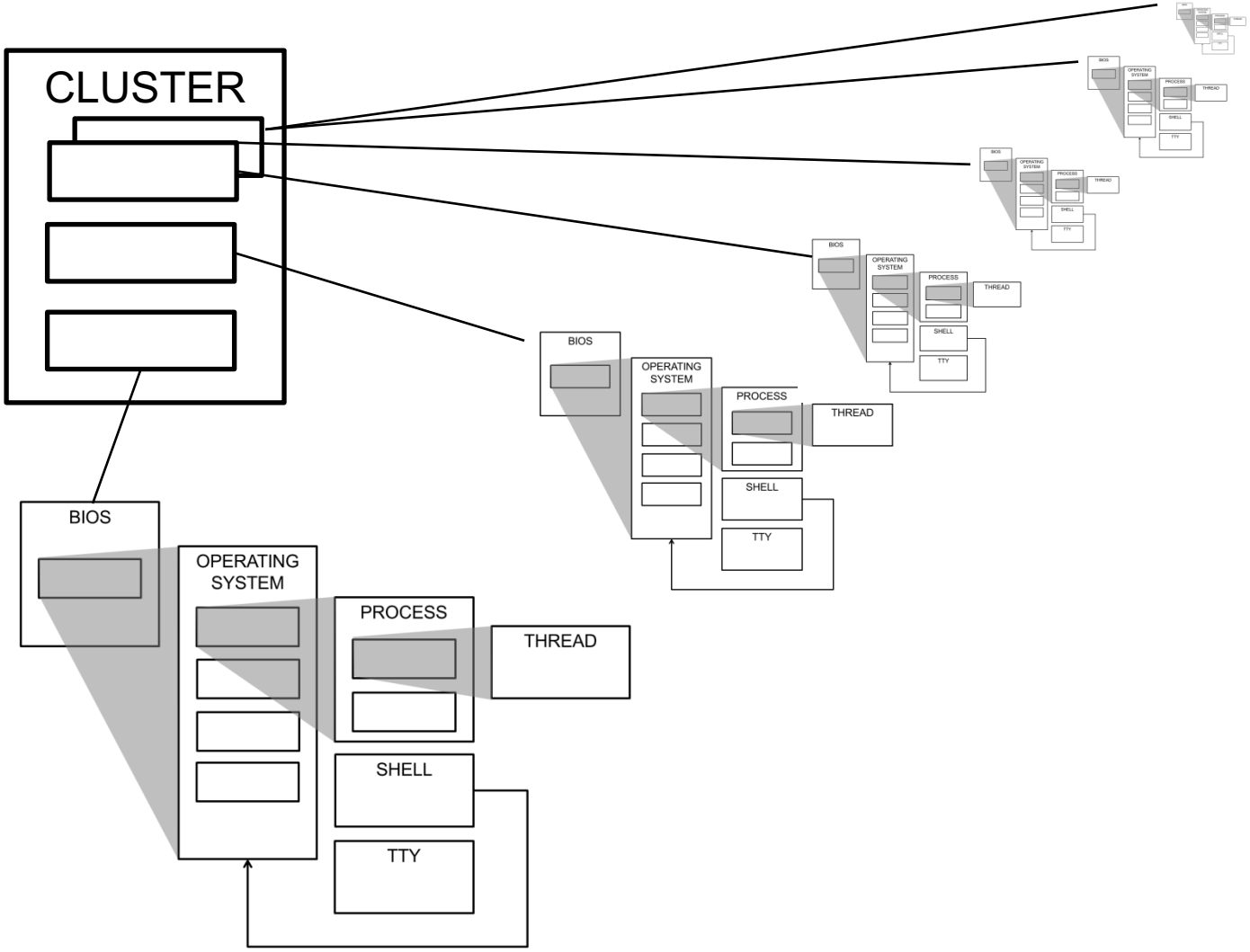
Moving data, or yourself

Most of the ways of moving data around the internet were developed for Unix first. You also have the option of going to where the data is, with a remote shell.



A cluster is just many computers together…

Each one with its own OS, processes, and shell environments.



CANCER RESEARCH UK | CAMBRIDGE INSTITUTE

CRUK CI HPC clusters

Key Cluster-related Staff

Guy Robinson (Head of IT & SC)

Marc O'Brien (Technical Architect)

Nigel Berryman (IT Operations Manager)

Charles Thomson (IT Specialist)

Jon Marshall (System Administor)

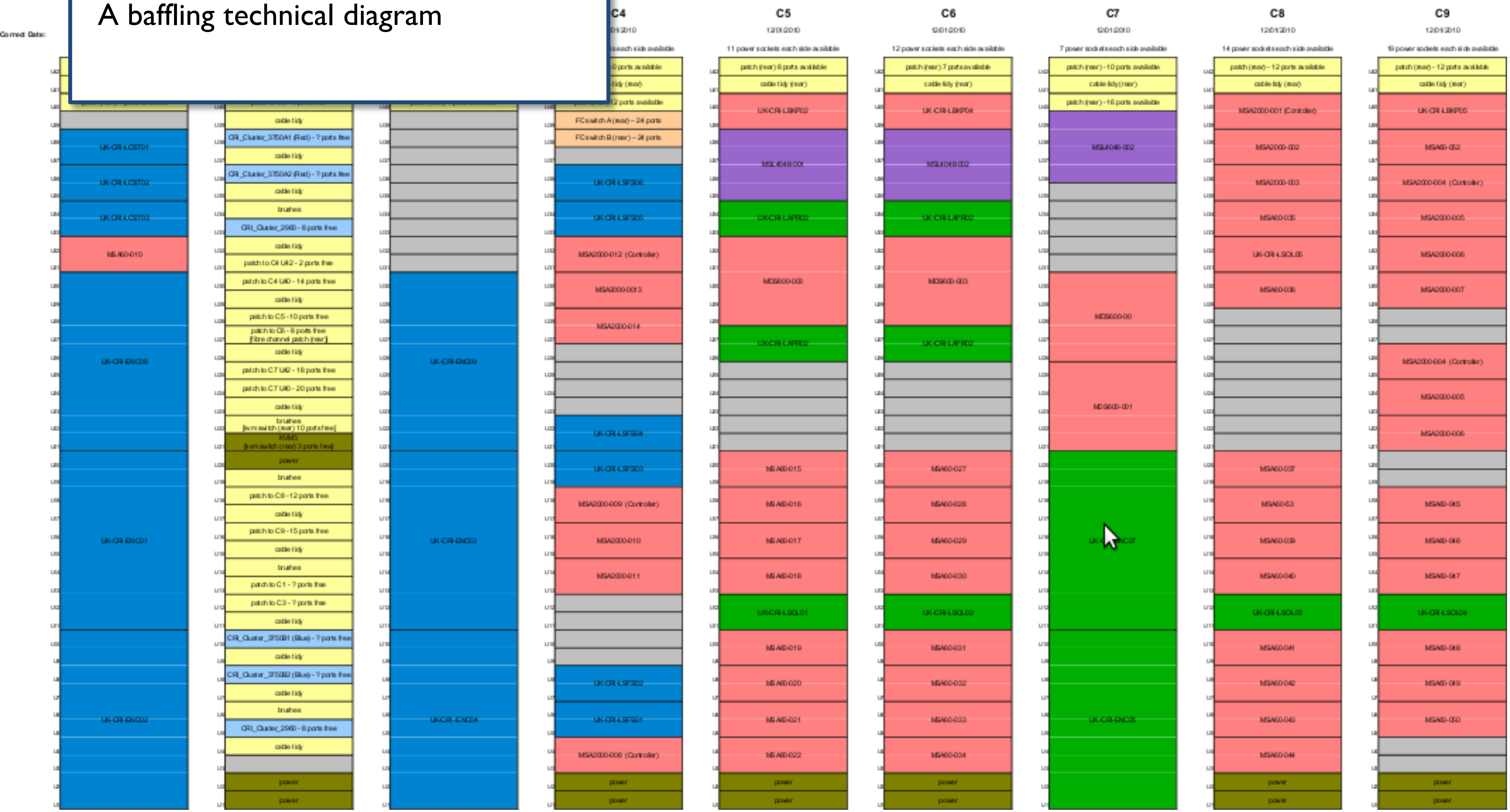
CRUK CI IT & SC Help Desk

Use the usual helpdesk route for day-to-day problems – if an issue is affecting you it may be affecting many people.

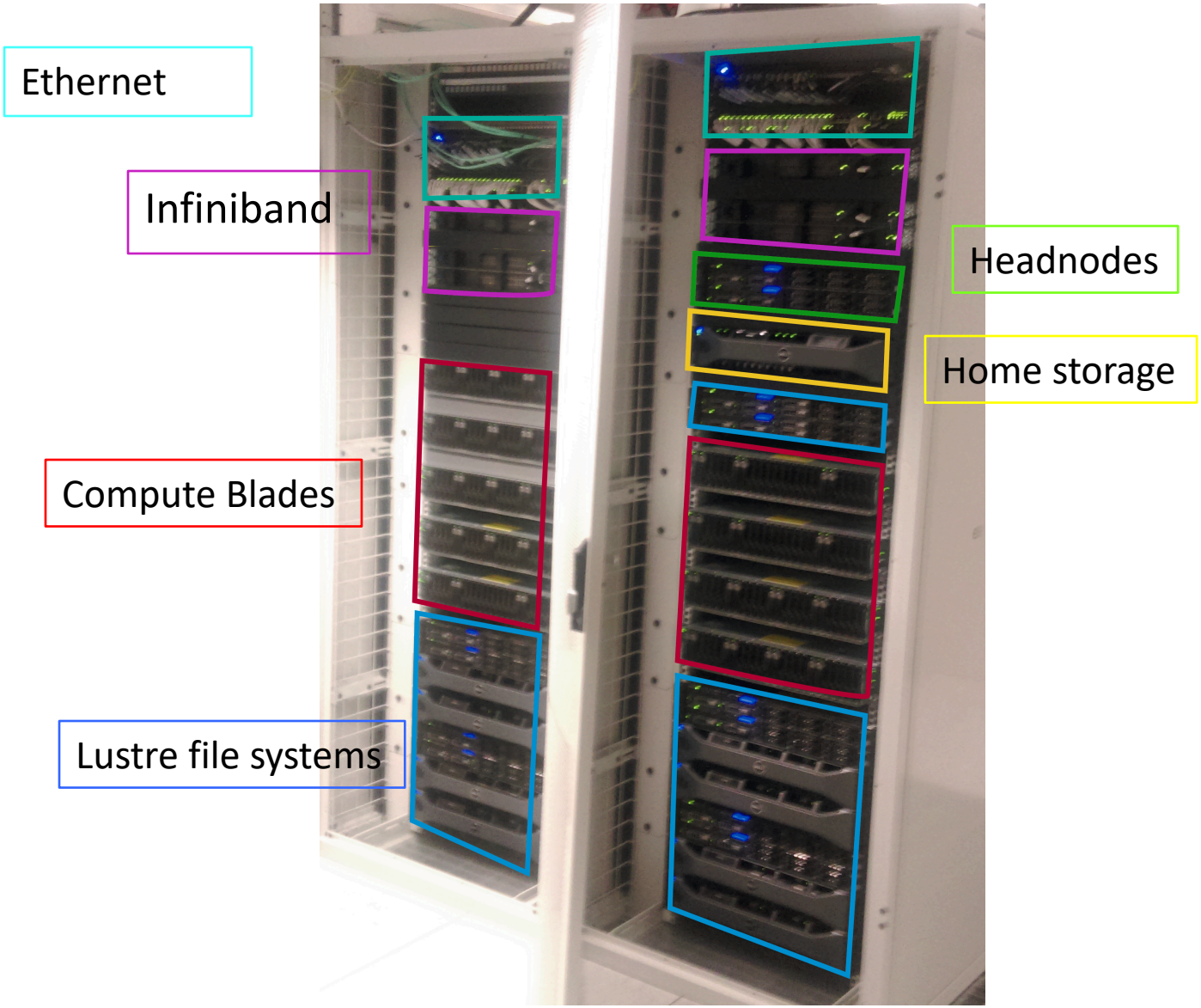
helpdesk-it@cruk.cam.ac.uk

01223 769600

A baffling technical diagram



Cluster1


Ethernet
Infiniband
Compute Blades
Lustre file systems
Headnodes
Home storage

CRUK CI HPC Cluster Specifications

- 33 compute nodes
- 2 x 20 core Intel Broadwell CPUs
- 1320 cores
- 320 GB RAM per node
- 2 x 196TB Lustre parallel file-systems
- Job scheduler (SLURM)
- CentOS 7.3

Other ways of computing: Dedicated Servers

Database servers
- BioInformatics database servers
- Many 10s TB storage

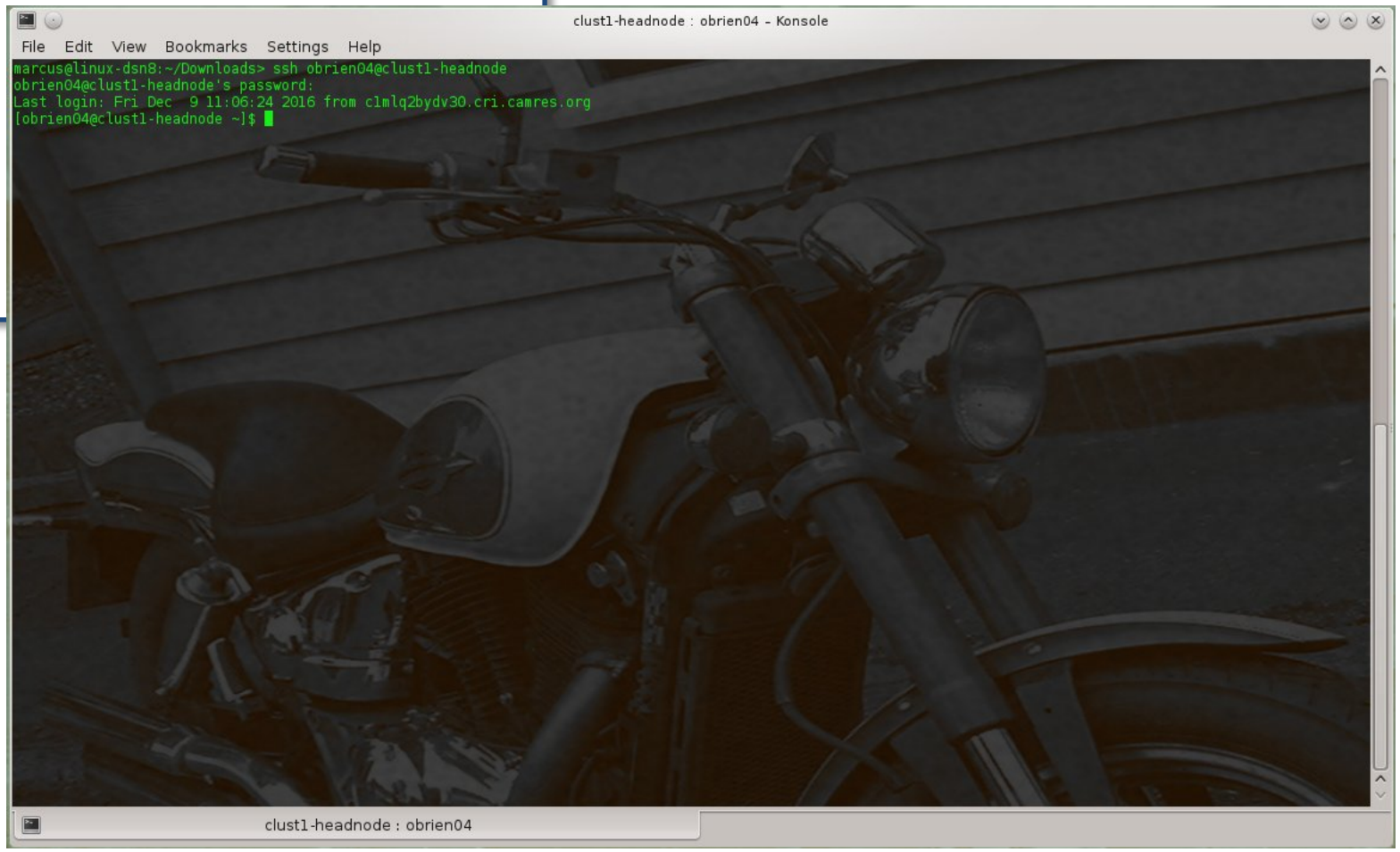
Group Servers
- Dedicated server for individual research groups

Virtual Servers
- VMware cluster (virtual servers)
- Ideal for non-CPU intensive applications, such as web servers

Accessing the HPC cluster

Login to the head node via secure shell (ssh)

[user@computer ~] ssh user@clust1-headnode



/home Directory

- 19TB file system shared over NFS4
- Intended for source code and common user applications & libraries
- Group quota
- Mounted on head node and all compute nodes
- Daily backed up

```
[user@clust1-headnode~ ]$ xfs-quota -c "quota -g"
Disk quotas for Group computing (1048)
Filesystem              Blocks      Quota      Limit  Warn/Time
Mounted on
/dev/mapper/mpatha         244          0          0   00 [--------]
/home
```

Transfer data into and out of the cluster

Use SSH scp (SCoPy) to transfer data

Bulk transfers using rsync (delta-transfer algorithm)

```
user@laptop:~> scp -r data/ clust1-headnode:/scratchb/xxlab/
file01.dat 100% 100KB 100.0KB/s 00:00
file02.dat 100% 100KB 100.0KB/s 00:00
file03.dat 100% 100KB 100.0KB/s 00:00

user@laptop:~> rsync -av data/ clust1-headnode:/scratchb/xxlab/data
sending incremental file list
file01.dat
file02.dat
file03.dat
sent 307464 bytes received 72 bytes 615072.00 bytes/sec
total size is 307200 speedup is 1.00
```

Cluster Storage

1. What's wrong with NFS or CIFS?
2. The Lustre parallel file-system
3. Working with Lustre

The Problem

Extreme I/O demand on storage

- HPC cluster can have 10s to 1000s compute nodes
- x Many users
- x 1000s jobs
- + Millions of small and large files

Required shared filesystem
Breaks most filesystems !!!

Lustre: Parallel Filesystem

Lustre is a massively parallel distributed file system

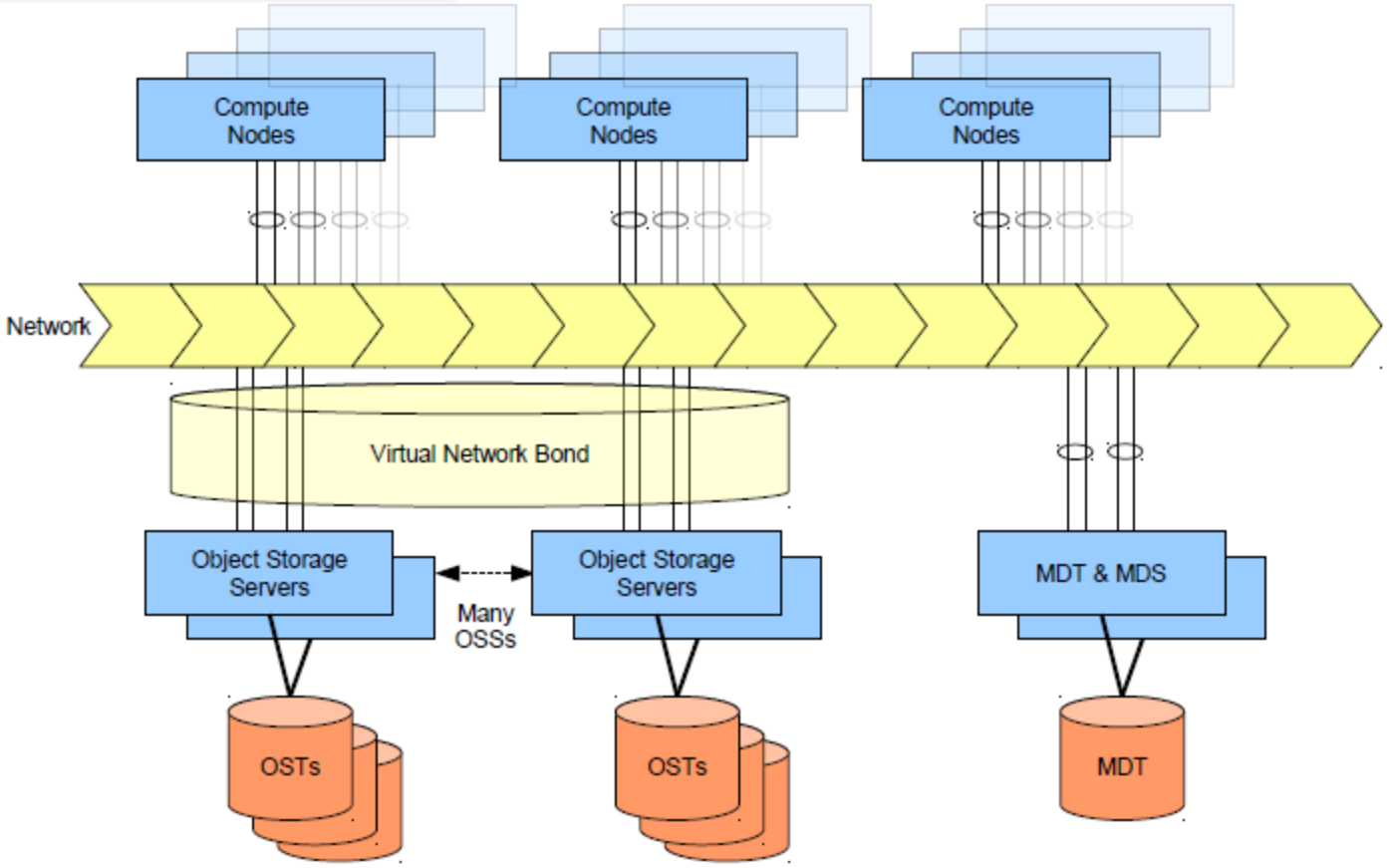
- Deployed in 7 out of 10 most powerful supercomputers
- POSIX compliant

Lustre design paradigm concepts

- Separation of file meta-data and storage allocation
- Scalable data serving through parallel data striping
- Aggregate network bandwidth
- Distributed operation

.

Lustre Architecture



Lustre Quotas

- Lustre file system quotas are applied to each group
- Units in kilobytes (oddity)
- "quota" & "limit" of zero signifies no quota (as shown for user)

```
clust1-headnode ~ $ lfs quota /mnt/scratchb
Disk quotas for user user321 (uid 442255):
Filesystem kbytes quota limit grace files quota limit grace
/mnt/scratchb 275648748 0 0 - 984 0 0 -
Disk quotas for group xxlab (gid 987):
Filesystem kbytes quota limit grace files quota limit grace
/mnt/scratchb 3471466428 3600000000 4000000000 - 681541 0 0 -
```

Lustre Health Check

Check the status of each lustre
Display the usage of each distributed Object Storage Target component

```
clust1-headnode ~ $ lfs check servers
lfs check servers
scratcha-MDT0000-mdc-ffff81041158cc00 active.
scratcha-OST0000-osc-ffff81041158cc00 active.
...
scratchb-OST0007-osc-ffff81041158cc00 active.

clust1-headnode ~ $ lfs df -h
UUID bytes Used Available Use% Mounted on
scratcha-MDT0000_UUID 239.0G 856.7M 224.5G 0% /mnt/scratcha[MDT:0]
scratcha-OST0000_UUID 5.4T 2.2T 2.9T 40% /mnt/scratcha[OST:0]
...
scratchb-OST0007_UUID 5.4T 2.1T 3.0T 39% /mnt/scratchb[OST:7]
filesystem summary: 42.9T 16.3T 26.6T 37% /mnt/scratcha
```