Welcome

Virtual tutorial starts at 15.00 BST

archer

epcc

THE UNIVERSITY OF EDINBURGH

NATURAL ENVIRONMENT RESEARCH COUNCIL

EPSRC

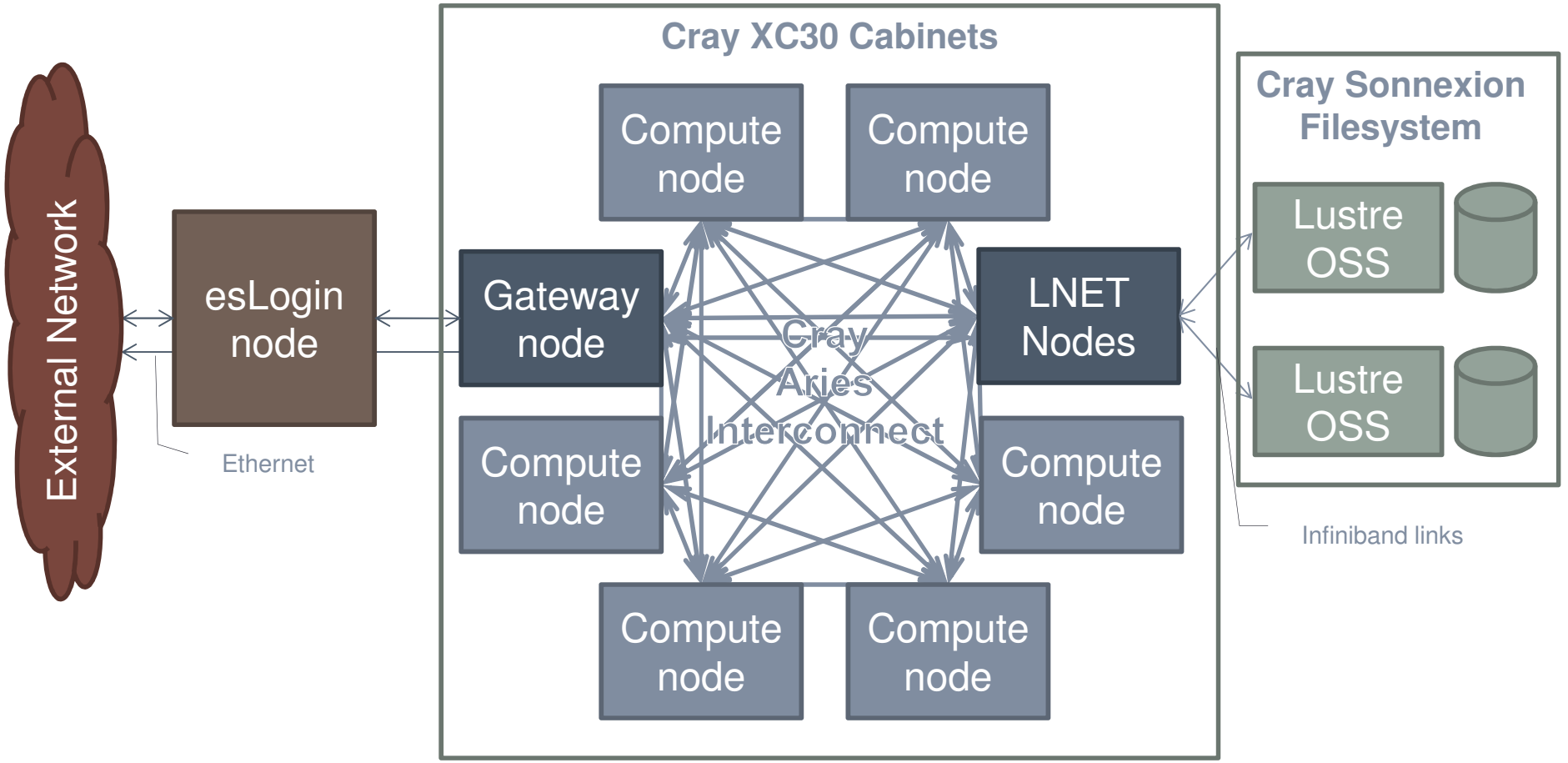
ARCHER FileSystems

Adrian Jackson

adrianj@epcc.ed.ac.uk

epcc

THE UNIVERSITY OF EDINBURGH


Cray XC30 Cabinets
Compute node
Compute node
Gateway node
Cray Aries Interconnect
LNET Nodes
Compute node
Compute node
Compute node
Compute node
External Network
esLogin node
Ethernet
Cray Sonnexion Filesystem
Lustre OSS
Lustre OSS
Infiniband links

Filesystems

- /home – NFS, not accessible on compute nodes
  - For source code and critical files
  - Backed up
  - > 200 TB total
- /work – Lustre, accessible on all nodes
  - High-performance parallel filesystem
  - Not backed-up
  - > 4PB total
- RDF – GPFS, not accessible on compute nodes
  - Long term data storage

/home

- **Note: /home is not mounted on the compute nodes so all files required for your calculations must be available on the /work filesystem.**

- The home directory for each user is located at:
  - /home/[project code]/[group code]/[username] where:
  - [project code] is the code for your project (e.g., x01);
  - [group code] is the code for your project group, if your project has groups, (e.g. x01-a) or the same as the project code, if not;
  - [username] is your login name.
  - Each project is allocated a portion of the total storage available,
    - project PI can able to sub-divide this quota among the groups and users within the project
- Environment variable $HOME is automatically set to point to your home directory.
- Backed-up
  - first to a second set of hard disks
  - second to tape.

/work

- /work is high-performance, parallel Lustre filesystems.
  - Each project will be assigned space on a particular Lustre partition with the assignments chosen to balance the load across the available infrastructure.
  - /work should be used for reading and writing during simulations.

- **Not backed-up**

- Links from the /home filesystem to directories or files on /work are strongly discouraged.

Disk Quotas

- /work
  - Lustre `lfs quota` command can be used to get more detailed quota information than is available on the SAFE.

- To check the quota for your project group:
  - `lfs quota -g [project code] /work/[project code]`
    Information on the disk usage for an individual can be checked with
  - `lfs quota -u [username] /work/[project code]`

/work performance

- Lustre filesystem has a number of I/O servers
  - By default each file is assigned to 4 I/O servers and split across them in 1MB chunks: **striping**
  - ARCHER has 48 virtual I/O servers (**OSTs**)
- Programs using parallel I/O and writing/read large amounts of data can benefit from changing default behaviour
- Increasing file striping allows program to exploit all I/O servers
  - Stripe can be set per file or per directory
  - Set per directory, anything created within that directory inherits the directory lustre configuration

```
adrianj@eslogin004:~> lfs getstripe /work/z01/z01/adrianj/temp
/work/z01/z01/adrianj/temp
stripe_count:   4 stripe_size:    1048576 stripe_offset:  -1

adrianj@eslogin004:~> touch /work/z01/z01/adrianj/temp/test.dat
adrianj@eslogin004:~> lfs getstripe /work/z01/z01/adrianj/temp
/work/z01/z01/adrianj/temp
stripe_count:   4 stripe_size:    1048576 stripe_offset:  -1
/work/z01/z01/adrianj/temp/test.dat
lmm_stripe_count:   4
lmm_stripe_size:    1048576
lmm_layout_gen:     0
lmm_stripe_offset:  13
        obdidx           objid           objid           group
            13        14246234       0xd9615a               0
             5        14271068       0xd9c25c               0
            21        14245673       0xd95f29               0
            42        13982337       0xd55a81               0
```

adrianj@eslogin004:~> lfs setstripe -c -1 /work/z01/z01/adrianj/temp

```
adrianj@eslogin004:~> touch /work/z01/z01/adrianj/temp/test.dat
adrianj@eslogin004:~> lfs getstripe /work/z01/z01/adrianj/temp
/work/z01/z01/adrianj/temp
stripe_count:   -1 stripe_size:    1048576 stripe_offset:  -1
/work/z01/z01/adrianj/temp/test.dat
lmm_stripe_count:   48
lmm_stripe_size:    1048576
lmm_layout_gen:     0
lmm_stripe_offset:  36
        obdidx           objid           objid           group
            36        13975788       0xd540ec               0
            44        13984603       0xd5635b               0
            12        14242013       0xd950dd               0
             4        14270476       0xd9c00c               0
            20        14251397       0xd97585               0
            41        13977735       0xd54887               0
            33        13973584       0xd53850               0
            25        13987172       0xd56d64               0
            17        14248714       0xd96b0a               0
             1        14269848       0xd9bd98               0
             9        14249421       0xd96dcd               0
            29        13993927       0xd587c7               0
            37        13983370       0xd55e8a               0
            45        13981347       0xd556a3               0
            13        14246343       0xd961c7               0
             5        14271177       0xd9c2c9               0
            21        14245782       0xd95f96               0
            42        13982446       0xd55aee               0
            34        13982005       0xd55935               0
            26        14002872       0xd5aab8               0
            18        14251198       0xd974be               0
             2        14264184       0xd9a778               0
            10        14246114       0xd960e2               0
            30        14007080       0xd5bb28               0
            38        13971500       0xd5302c               0
            46        13984976       0xd564d0               0
            14        14249207       0xd96cf7               0
             6        14261862       0xd99e66               0
            22        14242738       0xd953b2               0
            43        13982681       0xd55bd9               0
            35        13980468       0xd55334               0
            27        13987067       0xd56cfb               0
            19        14250207       0xd970df               0
             3        14259427       0xd994e3               0
            11        14241050       0xd94d1a               0
            31        13993329       0xd58571               0
            39        13982395       0xd55abb               0
            47        13979769       0xd55079               0
            15        14244193       0xd95961               0
             7        14270901       0xd9c1b5               0
            23        14250475       0xd971eb               0
            40        13979889       0xd550f1               0
            32        13978890       0xd54d0a               0
            24        13989092       0xd574e4               0
            16        14242977       0xd954a1               0
             0        14261946       0xd99eba               0
             8        14240286       0xd94a1e               0
            28        13989584       0xd576d0               0
```

Filesystems

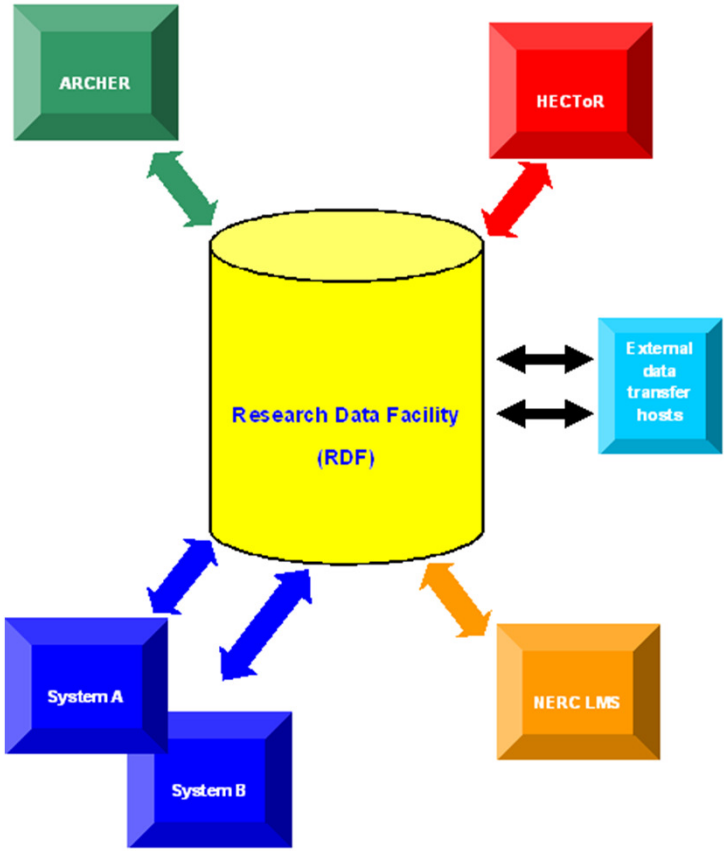
- No /tmp on backend nodes
  - GNU Fortran, file OPEN statements with STATUS='SCRATCH‘
  - export GFORTRAN_TMPDIR=/work/[project]/[group]/[username]/tmp
- Group permissions setup per project
  - Possible to access files on group permissions with projects but beyond a project would need world readable files
- Sharing data
  - Within projects
    - /work/projectcode/projectcode/shared
  - Between projects
    - /work/projectcode/shared

HAS BEEN SETUP TODAY (14/05/14)

Research Data Facility (RDF)

- RDF is designed for long term data storage

- RDF consists of
  - 12.3 PB disk
  - 30 PB backup tape
  - Provide a high capacity robust file store;
  - Persistent infrastructure - will last beyond any one national service;



archer

epcc

THE UNIVERSITY OF EDINBURGH

RDF

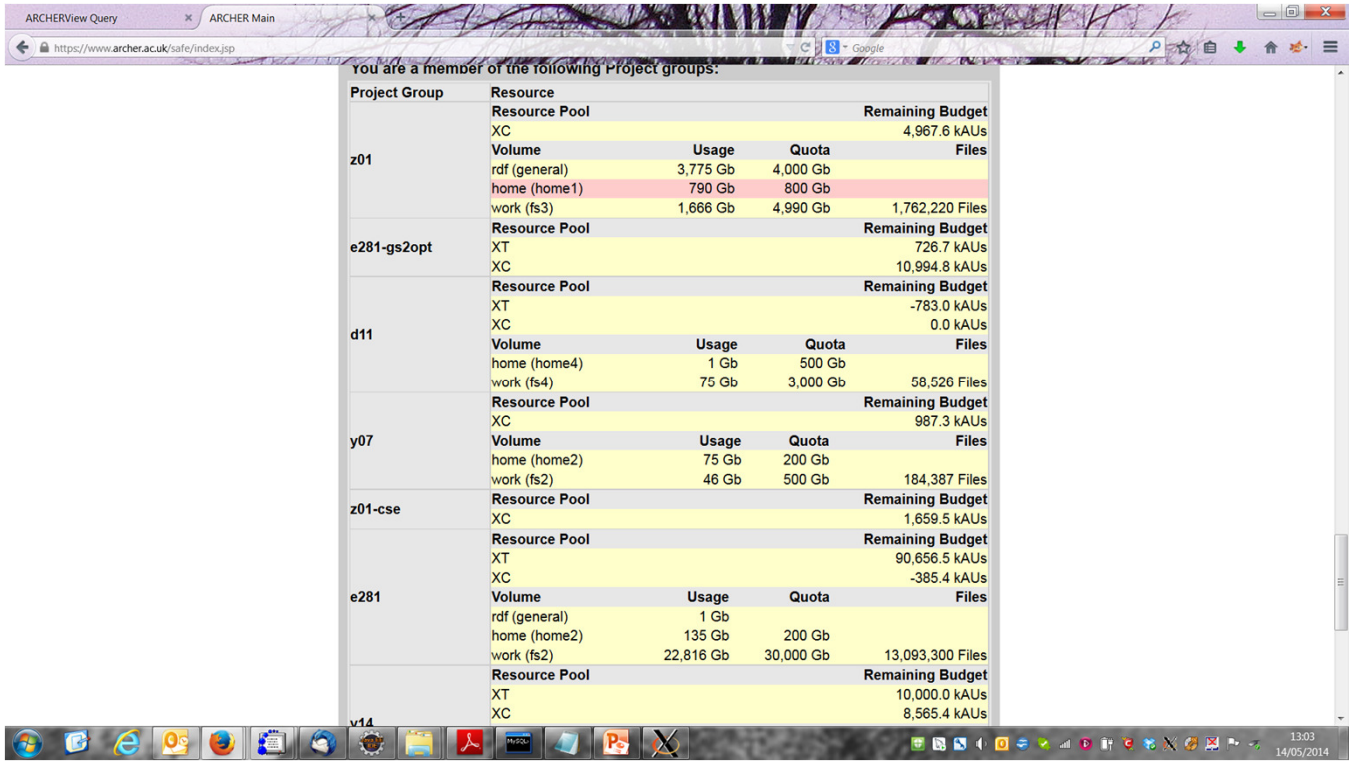
- RDF directly mounted from ARCHER
  - The name of the filesystem will depend on your funding body. At present three filesystems have been created:
    - /epsrc
    - /nerc
    - /general
  - These filesystems are only visible on the ARCHER login nodes.
  - cp command gives the best performance on transferring data from ARCHER filesystems to the RDF.

- Users moving large volumes of data via rsync etc. are recommended to use the serial batch queues. Large transfer jobs running on the login nodes may be terminated.

- External access to RDF
  - Through Data Mover Nodes dtn01.hector.ac.uk – dtn04.hector.ac.uk
  - GridFTP setup on dtn01 and dtn02

SAFE disk quota

- View disk quotas
  - Values for disk use are updated four times a day



Project Group	Resource			
z01	Resource Pool			Remaining Budget
	XC			4,967.6 kAUs
	Volume	Usage	Quota	Files
	rdf (general)	3,775 Gb	4,000 Gb	
	home (home1)	790 Gb	800 Gb	
	work (fs3)	1,666 Gb	4,990 Gb	1,762,220 Files
e281-gs2opt	Resource Pool			Remaining Budget
	XT			726.7 kAUs
	XC			10,994.8 kAUs
d11	Resource Pool			Remaining Budget
	XT			-783.0 kAUs
	XC			0.0 kAUs
	Volume	Usage	Quota	Files
	home (home4)	1 Gb	500 Gb	
	work (fs4)	75 Gb	3,000 Gb	58,526 Files
y07	Resource Pool			Remaining Budget
	XC			987.3 kAUs
	Volume	Usage	Quota	Files
	home (home2)	75 Gb	200 Gb	
	work (fs2)	46 Gb	500 Gb	184,387 Files
z01-cse	Resource Pool			Remaining Budget
	XC			1,659.5 kAUs
e281	Resource Pool			Remaining Budget
	XT			90,656.5 kAUs
	XC			-385.4 kAUs
	Volume	Usage	Quota	Files
	rdf (general)	1 Gb		
	home (home2)	135 Gb	200 Gb	
	work (fs2)	22,816 Gb	30,000 Gb	13,093,300 Files
y14	Resource Pool			Remaining Budget
	XT			10,000.0 kAUs
	XC			8,565.4 kAUs

SAFE disk quota

- Two types of space in SAFE (like time):
  - *general group*
    - same code as the project
    - Includes every member of the project, so everyone can use this quota.
  - *reserve group*
    - *projectcode-reserve*
    - No members, so no one can use the disk space which is in its quotas.
- Homespace and workspace are administered separately
  - Each have overall quota
- Can also have quotas for the project groups which you create

SAFE **group** disk quota management

- Assigning disk quota to project groups in SAFE creates new directories for that disk quota
  - i.e. project t01, creates a group t01-a, with some time and some disk quota on /home. This creates a new directory:
  - /home/t01/t01-a
  - If you add a user to that group it will also create a directory for that user in the group directory, i.e.:
  - /home/t01/t01-a/username
  - Files created in this directory will count towards the group quota, files created in the normal project directory (i.e. /home/t01/t01/username) count against the general project quota
  - Really, files assigned to quotas by the file group they are created under (can check using `ls -l` can change using the `chown` command)

SAFE **user** disk quota management

- User disk quotas are completely separate from project quotas.
  - Simply putting a limit on the amount of disk space a user can use in a project's /home or /work file space
  - Can have total user limits that exceed project disk quota

www.archer.ac.uk/documentation/safe-guide/safe-guide-pi.php

Goodbye

Virtual tutorial has finished

Please check here for future tutorials and training

http://www.archer.ac.uk/training

http://www.archer.ac.uk/training/virtual/

archer

epcc

THE UNIVERSITY OF EDINBURGH