Welcome

Virtual tutorial starts at 15.00 BST
ARCHER FileSystems
Adrian Jackson
adrianj@epcc.ed.ac.uk
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
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

<table>
<thead>
<tr>
<th>obdidx</th>
<th>objid</th>
<th>objid</th>
<th>group</th>
</tr>
</thead>
<tbody>
<tr>
<td>13</td>
<td>14246234</td>
<td>0xd9615a</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>14271068</td>
<td>0xd9c25c</td>
<td>0</td>
</tr>
<tr>
<td>21</td>
<td>14245673</td>
<td>0xd95f29</td>
<td>0</td>
</tr>
<tr>
<td>42</td>
<td>13982337</td>
<td>0xd55a81</td>
<td>0</td>
</tr>
</tbody>
</table>
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
stripes_count:   48
stripes_size:    1048576
stripes_offset:  -1

<table>
<thead>
<tr>
<th>obdidx</th>
<th>objid</th>
<th>objid</th>
<th>group</th>
</tr>
</thead>
<tbody>
<tr>
<td>36</td>
<td>13975788</td>
<td>0x540ec</td>
<td>0</td>
</tr>
<tr>
<td>44</td>
<td>13984603</td>
<td>0x5635b</td>
<td>0</td>
</tr>
<tr>
<td>12</td>
<td>14242013</td>
<td>0x950dd</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>14270476</td>
<td>0x8b08c</td>
<td>0</td>
</tr>
<tr>
<td>20</td>
<td>14251397</td>
<td>0x97565</td>
<td>0</td>
</tr>
<tr>
<td>41</td>
<td>13977735</td>
<td>0x548f7</td>
<td>0</td>
</tr>
<tr>
<td>33</td>
<td>13973984</td>
<td>0x35850</td>
<td>0</td>
</tr>
<tr>
<td>25</td>
<td>13987172</td>
<td>0x56df4</td>
<td>0</td>
</tr>
<tr>
<td>17</td>
<td>14248714</td>
<td>0x96hoa</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>14269848</td>
<td>0x96d88</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>14249421</td>
<td>0x96dcd</td>
<td>0</td>
</tr>
<tr>
<td>29</td>
<td>13993927</td>
<td>0x587c7</td>
<td>0</td>
</tr>
<tr>
<td>37</td>
<td>13983370</td>
<td>0x55e8a</td>
<td>0</td>
</tr>
<tr>
<td>45</td>
<td>13981347</td>
<td>0x55e8e</td>
<td>0</td>
</tr>
<tr>
<td>13</td>
<td>14246343</td>
<td>0x961c7</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>14271177</td>
<td>0x92c90</td>
<td>0</td>
</tr>
<tr>
<td>21</td>
<td>14245782</td>
<td>0x95e96</td>
<td>0</td>
</tr>
<tr>
<td>42</td>
<td>13982446</td>
<td>0x55eae</td>
<td>0</td>
</tr>
<tr>
<td>34</td>
<td>13982005</td>
<td>0x55935</td>
<td>0</td>
</tr>
<tr>
<td>26</td>
<td>14002872</td>
<td>0x96ab8</td>
<td>0</td>
</tr>
<tr>
<td>18</td>
<td>14251198</td>
<td>0x974ee</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>14264184</td>
<td>0x9a778</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>14246114</td>
<td>0x96e02</td>
<td>0</td>
</tr>
<tr>
<td>30</td>
<td>14007080</td>
<td>0x5bb28</td>
<td>0</td>
</tr>
<tr>
<td>38</td>
<td>13971500</td>
<td>0x5302c</td>
<td>0</td>
</tr>
<tr>
<td>46</td>
<td>13984976</td>
<td>0x55440</td>
<td>0</td>
</tr>
<tr>
<td>14</td>
<td>14249207</td>
<td>0x96c27</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>14261862</td>
<td>0x99e66</td>
<td>0</td>
</tr>
<tr>
<td>22</td>
<td>14242738</td>
<td>0x95e62</td>
<td>0</td>
</tr>
<tr>
<td>43</td>
<td>13982681</td>
<td>0x55bd9</td>
<td>0</td>
</tr>
<tr>
<td>35</td>
<td>13980468</td>
<td>0x55334</td>
<td>0</td>
</tr>
<tr>
<td>27</td>
<td>13987067</td>
<td>0x555eb</td>
<td>0</td>
</tr>
<tr>
<td>19</td>
<td>14250207</td>
<td>0x970df</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>14259427</td>
<td>0x994e3</td>
<td>0</td>
</tr>
<tr>
<td>11</td>
<td>14241050</td>
<td>0x94d1a</td>
<td>0</td>
</tr>
<tr>
<td>31</td>
<td>13993329</td>
<td>0x58571</td>
<td>0</td>
</tr>
<tr>
<td>39</td>
<td>13982395</td>
<td>0x55aab</td>
<td>0</td>
</tr>
<tr>
<td>47</td>
<td>13979769</td>
<td>0x55079</td>
<td>0</td>
</tr>
<tr>
<td>15</td>
<td>14244193</td>
<td>0x959f1</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>14270901</td>
<td>0x9c1b5</td>
<td>0</td>
</tr>
<tr>
<td>23</td>
<td>14250475</td>
<td>0x971ab</td>
<td>0</td>
</tr>
<tr>
<td>40</td>
<td>13979889</td>
<td>0x55021</td>
<td>0</td>
</tr>
<tr>
<td>32</td>
<td>13978890</td>
<td>0x54e0a</td>
<td>0</td>
</tr>
<tr>
<td>24</td>
<td>13969092</td>
<td>0x574e4</td>
<td>0</td>
</tr>
<tr>
<td>16</td>
<td>14242977</td>
<td>0x954a1</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>14261946</td>
<td>0x99e0a</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>14240286</td>
<td>0x94eae</td>
<td>0</td>
</tr>
<tr>
<td>28</td>
<td>13969584</td>
<td>0x57640</td>
<td>0</td>
</tr>
</tbody>
</table>
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;
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
    - /ncerc
    - /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
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.

- Homespase 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/