Instructions for Parkes disk cleaning

Parkes has two VLBI xraids, `pkxserve1` and `pkxserve2`. These reside in the 4th cabinet from the left in the Northern plinth (near the pulsar digital filterbanks), with disks currently living under the North window.

The xraids are connected to `pkvsi1` and `pkvsi2` respectively (only on the internal network, use `pkvsi1-ext` from outside and via `venice`).

Things that may need to be done are:
- Redetect devices to refresh the number of slices and disk size.
- Change the number of slices on a set of disks.
- Put a new partition or partition table on some disks (after reslicing, or because there was a problem).
- Reformat disks so they are clean for the upcoming run.
- Rebuild a degraded array.
The most common, of course, is simply reformatting.

redetect **Redetecting devices** needs to be done whenever you change disks so that you have more slices, or larger or smaller disks, to ensure that queries give sensible feedback. The simplest way to do this is simply to restart the computer (`pkvsi1` or `pkvsi2` according to which xraid the different disks are in). However it may not be possible to reboot the machine if other people are currently logged onto the computer - check this using the command `'w'`. If you know who is logged on and you believe nothing critical will happen use the command "`shutdown -r 5`" - this will send a message to all terminals announcing a reboot in 5 minutes (or some other number as appropriate).
If it is not possible to restart the computer at the time and based on the output of `parted` you believe it is necessary to run a redetection, you can force it manually, but unless you're extremely confident about what you're doing, ***don't***! We may be able to modify Curtin's scripts to run on other machines, but as this doesn't seem to have ever been a problem in the past (maybe older kernels didn't make so many assumptions so it didn't used to happen but now it does...), if it does seem to be necessary, send Claire an email and someone at Curtin will reload the scsi devices.

reslice **Adding or removing slices on an xraid set** requires use of the xraid admin tools. Current version is 1.5.1. Adding slices is unlikely to be necessary anymore as we are moving toward using only single partition 3TB disk sets for VLBI. However, in the past disk sets were sliced into two (or occasionally three) 1.5TB devices. There is no longer a reason to prefer one way to the other, however the advantage to observers and PIs is that without the slicing, fewer disk changes are needed mid-experiment, meaning less for the observers to worry about, and less time and thus data lost in transitions.
To reslice devices, should it be necessary (unlikely but not impossible), do the following from any computer at Parkes:

```
hostnode> ssh -X vlbi@pkvsi1
pkvsi1:~> cd xserve-raid/
pkvsi1:~/xserve-raid> cd RAID-Admin-1.5.1/
pkvsi1:~/xserve-raid/RAID-Admin-1.5.1> java -jar RAID_Admin.jar
```

For screenshots of what to do with the RAID tools, see the Curtin wiki page on this.

Once the admin GUI has opened, you can select the xraid you're interested in, and look at the "Arrays

and Drives" tab, where you can see the size of each set and how many slices it has.
If you need to change the slice settings, select the "Advanced" button from the tool bar. You'll need to enter a password - ask Chris or Claire.
You can then select the set of drives you want to slice/unslice, confirm that you know what you're doing, and go ahead. Slicing doesn't take long. However once you have resliced, you'll then need to repartition the disks before formatting for use.

repartition **A disk may need to have a partition applied** if the disk set has been resliced, or the partition table is broken for some other reason. As we use disks which are >2TB, `fdisk` does not work, and instead we need to use `parted`.
Chris has written scripts for partitioning and formatting: Replace '1' with '2' according to which machine the disk set in question is mounted in.

```
hostname> ssh vlbi@pkvsi1
pkvsi1:~> su -
Password:
```

%color=#006600% Root password for vsi machines
<code>pkvsi1:~# df -h </code> %color=#006600% **Check if the disks are currently mounted - MUST unmount to partition**

```
Filesystem          Size  Used Avail Use% Mounted on
/dev/sda1           9.2G  7.8G 1007M  89% /
tmpfs               764M     0  764M   0% /lib/init/rw
udev                 10M  112K  9.9M   2% /dev
tmpfs               764M     0  764M   0% /dev/shm
/dev/md1            723G  112G  612G  16% /data/internal
/dev/md0            1.6T  1.2T  437G  73% /data/removable
/dev/sdc1           2.7T  202M  2.7T   1% /data/xraid0
/dev/sdd1           2.7T  202M  2.7T   1% /data/xraid1
```

%color=#006600% Make note of device names - /dev/sdc1 is /data/xraid0
%color=#006600% **Make sure you're going to change the disk set you want to**

```
pkvsi1:~# umount /data/xraid0
pkvsi1:~# e2label /dev/sdc1
ATNF V017A
```

%color=#006600% Finally, run the partitioning script on device sdc
<code>pkvsi1:~# partitiondisk.pl /dev/sdc </code> **We now have a freshly partitioned disk. Can check that did what you want by running** `parted` **and simply type the command** `print all` **- this shows all disks attached to the host. If that didn't work for some reason, try formatting manually using Parted.** <code» parted /dev/sdc (parted) help</code>
%color=#006600%**View the available commands.**
We'll make a GPT disk, with a primary partition that fills the whole disk.

```
(parted) mklabel gpt
(parted) mkpart primary 0 -0
```

```
(parted) quit
```

As with any other hard disk, after partitioning the disk needs to be formatted.

reformat **Reformatting a disk set** is the most commonly required task, and hopefully the only one you'll ever need to do! :)
Chris has written a script which will do this, it lives on the pkvsi machines.
Substitute appropriate machine numbers and device names below - check with Chris if you don't have the vlbi or root passwords.

```
hostname> ssh vlbi@pkvsi1
pkvsi1:~> su -
Password:
```

%color=#006600%Root password for vsi machines.
Check if the disks are currently mounted - must unmount to format.
<code>pkvsi1:~# df -h Filesystem Size Used Avail Use% Mounted on /dev/sda1 9.2G 7.8G 1007M 89% / tmpfs 764M 0 764M 0% /lib/init/rw udev 10M 112K 9.9M 2% /dev tmpfs 764M 0 764M 0% /dev/shm /dev/md1 723G 112G 612G 16% /data/internal /dev/md0 1.6T 1.2T 437G 73% /data/removable /dev/sdc1 2.7T 202M 2.7T 1% /data/xraid0 /dev/sdd1 2.7T 202M 2.7T 1% /data/xraid1</code> %color=#006600%Make note of device names - /dev/sdc1 is /data/xraid0.**

```
pkvsi1:~# umount /data/xraid0
```

%color=#006600%Check that the disk sets already contain labels (which ideally should be correct, you can check that they are during formatting and change afterward with e2label if necessary.
<code>pkvsi1:~# e2label /dev/sdc1 ATNF V017A </code> %color=#006600%Run the formatting script for VLBI disks**

```
pkvsi1:~# formatdisk.pl /dev/sdc1
pkvsi1:/~# mount /dev/sdc1 /data/xraid0
```

%color=#006600%Mount the disk so that you can access it again

If the script doesn't work for some reason, the manual command is this:
mke2fs -T largefile4 -j -m 0 -L "ATNF V017A" /dev/sdc1
followed by mounting the disks and running chown vlbi:vlbi /dev/sdc1 - This means we use ext2 format, but the -j means Journalling, so it's actually ext3 format. This isn't the most efficient format, but the journalling is good if accidents happen with disk removal! - m 0 means don't reserve space for root, and -L allows us to specify the disk volume label, and we want the disks to be owned by the group VLBI.

Repeat for as many disk sets as necessary - can do up to 4 at a time with two chassis (two on pkvsi1 and two on pkvsi2). When all 4 are done (or as many as necessary), run 'df -h' again to check that your devices are mounted, are ~2.7TB and have ~202MB / 1% used, the rest available.

rebuild Rebuilding a degraded array **needs to occur whenever a disk loaded in an array comes**

up with a red light, and removing and reinserting it does not fix the problem. Sometimes you may also need to change disks for an orange light, too, though usually they can be fixed by "making available" (explained below).

If you insert a spare disk (at the moment Curtin supplies these when needed), the rebuilding process will occur automatically, but it may be soluble without a new disk, and it's wise to monitor that it's going okay even with a fresh disk; this can be done using the xraid tools.

For instructions on how to start the xraid admin gui, see "reslicing" instructions above.

The disks come in sets of 7, which are linked in a RAID 5 array, meaning that if a disk is lost, all data is still recoverable, so a disk failure is nothing to panic about. However the situation should be rectified as soon as possible, as the loss of a second disk will result in total loss of data!!

Similarly, you must be careful to always load disks in the correct order, and if a spare disk is required, replace it in the position of the failed disk, don't move any of the other disks.

%color=#ff9900%*For an orange light failure:* This indicates that the chassis can see the disk but doesn't think it belongs to this set. Most likely cause is that the disk was not correctly seated in the chassis when the xraid was powered up. A failed disk can be removed without shutting down the xraid chassis (but only one disk, and only the failed one!).

In this case, carefully remove the disk with a warning (you may need to press the warning button on the chassis to stop the alarm noise), being careful not to release any of the green-light disks. Check the contacts on the disk, and try reinserting the disk, being sure to push it right in until it makes the "thunk" noise. Wait a minute or two and see what colour the light goes. If the disk is now accepted and the chassis thinks it's okay, the array will automatically start to rebuild, there will be lots of flashing blue lights!

If the disk is still in an error state it will remain orange or turn red. In this case, start the xraid Admin tools, and look at the state of the individual disk. The xraid may recognise the disk as valid, but think it belongs to a different array (even if it clearly doesn't!). If so, click the "Utilities" button on the toolbar (password required), and select "Make a disk available". Select the failed disk, and proceed as the tools guide you. Once the disk is available, the rebuild will start automatically, and you can monitor it by selecting the "Arrays and Drives" tab on the front page, clicking the "arrays" radio button and selecting the set you're interested in.

In one case that I've encountered the disk lost it's SMART capability - not quite sure what that is, but I think it's not fatal, just means it's lost some of its self-monitoring capability. My feeling is to just ignore the warning.

If for some reason the disk can't be made available and the set rebuilt with the same disk, a spare will be needed - contact Claire or someone else at Curtin, as ATNF has run out of spares.

For a red light failure: This indicates that the chassis cannot see the disk at all. This is either because the contacts are very poor, or because the disk has failed - either physically, or it has bad sectors that can't be avoided. In virtually all cases the only remedy is to get a replacement disk (see above) and replace it. A failed disk can be hot-swapped, so you can continue to use the disk set without redundancy, and when the spare disk arrives you can carefully remove the failed disk without powering down, and replace with the new disk, then check the rebuilding as described previously.

Note: The time disks seem most prone to failure is when they are removed, inserted or transported - xraids are made to be swappable, but not really to the extent that we use them for that purpose!! Consequently if you DO have a failed disk set, leave it loaded until it's fixed!

If any of this doesn't make sense, call Chris or Claire (who is at least 2 hours behind in WA).

From:
<https://www.atnf.csiro.au/vlbi/dokuwiki/> - **ATNF VLBI Wiki**

Permanent link:
https://www.atnf.csiro.au/vlbi/dokuwiki/doku.php/lbaops/lbaobservingnotes/parkesclean

Last update: **2015/12/18 16:38**