

Packrat, a CERI Halloween Treat

Robert Thomas DeBula (rdebula)

Fri 10/31/2014 12:25 PM

Inbox

To: Center for Earthquake Research and Information - Faculty <ceri_faculty@memphis.edu>; Center for Earthquake Research and Information - Staff <ceri_staff@memphis.edu>; Center for Earthquake Research and Information - Students <ceri_students@memphis.edu>;

Announcing Packrat (no, not me, and not my office)

Our previous, Veritas tape robot based, backup system for CERI essential data was rendered inoperable due to hardware failure of the (obsolete and unsupported) Sun tape robot. It was not practical to replace it due to the costs of hardware repair and the cost of the Veritas software. I proposed that for far less money, using drives we had (and repurposed to backups) it would be possible to add some server hardware and create an online network accessible disk based backup system to replace it (and more). Chuck approved the necessary funding for server and support hardware to provide the service proposed to him as a replacement. This new system is known as "Packrat" (you'll never access it directly by that name) because, well, you have to call it something and the name seems to fit. Of course, as always, final responsibility for the backup of all of a persons' data is still theirs (flash keys are cheap, so are external disks). This has been the policy at CERI all along (you ultimately bear responsibility for backing up your own data).

That said, Packrat is an archival data storage backup system accessible to CERI users (plugged in to the wired the ethernet connections) in the same manner as the rest of /gaia space (if you can access /gaia space from your machine, you can access the backed up archives on Packrat). In fact, for the CERI community, this backup space appears as part of /gaia space, and resides in "/gaia/archive". The contents of /gaia/archive are read only, you cannot write to it (it's backup, it's history, it should not be altered as someone might delete a bug or something in past data and change our present and CERI would be overrun by virtual data dinosaurs). Within /gaia/archive you will (currently) see three relevant directories: **data**, **machome**, and **sunhome**. In /gaia/archive/data are backups of most of the shared data common to CERI, including, but not limited to:

ag	compendium	dunedain	gps	nele	nmcd	scratch	thesis
bc	dbe	fdl	msf	nga	nvt	seisnet	workdisk

Access permissions (read only) are the same as they are on the original data (if you cannot access it in normal /gaia space, you shouldn't be able to through /gaia/archive either). Within /gaia/archive/sunhome are the Sun home directories for everyone at CERI (and relicts of the past). Within /gaia/archive/macspace are the Mac lab home directories of everyone at CERI. The other /gaia/archive directories you will see are for special purposes and wouldn't be accessed by most folks at CERI normally.

If, from the Unix command line on a machine from which you can access

/gaia space, you:

cd /gaia/archive/data (or **sunhome** or **machome**)

you will see listings of the available archived directories. There are daily archives (kept for the past 31 days), weekly archives (kept for 12 weeks), and monthly archives (kept for 60 months). If we start running out of space, those numbers may need adjusted downward in the future, but that is how they are currently set.

Currently (today), doing an "ls /gaia/archive/data" you would see a listing of the form:

```
daily-1392139640_2014.09.25.23.45.01
weekly-1392139726_2014.03.06.09.50.02
monthly-1392139869_2014.03.04.10.26.39
```

It should be easy to decipher what period each directory name listed represents. If you pick the date (or closest to it) you want, and cd into it, you will see the contents are as they were precisely at the point in time when the backup "snapshot" was taken. The backups are gathered between 6 and 10PM (it will eventually run longer) each evening, so the "snapshot" is from the evening of the day in question, some time generally between 6PM and midnight. Things behind the curtain are not that simple, but using the /gaia space structure, and a lot of background techno-fiddling, they have been made to look that way.

The beauty of this is multifold: It backs up a lot more (sadly, not the PC sabaki Windows machine data as yet, unless we can figure out a way to add it eventually) data than the Veritas tape backups ever did. Access to that backup data is essentially just as fast as access to the regular /gaia space data over the network. The data is all in (transparent to you) compressed form to save on space (the processing power of Packrat using the ZFS filesystem autocompression can compress and decompress data generally faster than the network can carry it). And, lastly, there is no need to call anyone to access or restore backed up data, you can access it directly yourself, online, all the time (except for maintenance, outages, etc.). The server with all of this /gaia/archive backup data is **not** housed in the CERI machine room, or even in the same building, so it is nearby, but still "offsite" from House 1 (so if ravenous rat hordes invade the machine room and chew up all the cables, Packrat will be safe, at least until they're hungry again and begin foraging in other houses).

The Packrat server has additional expansion capability, we'd just need to add more disks. Again, disks fail (Packrat is a RAID based system and can survive 2 disks failing out of each VDEV pool of 10, and believe me, that has already saved the archives during development more than once, with an array of that many disks hardware failures do occur), computers fail (Packrat has dual power supplies and a big UPS, but no generator for long outages), and it is up to each of **you** to provide that final level of backup for anything you consider important, just as it always has been. This should just make recovering something you accidentally deleted last week much, much, easier (it has actually already been used a few times, once by me to recover a messed up Mac home directory for a student, and once to recover some Sun information by a faculty testing pioneer).

Hopefully, you reached this point and did not fall asleep when my run-on sentences overwhelmed your optical cortex.

Some (likely to be) **FAQs**:

Q: I created a file Wednesday morning and then I accidentally deleted it Wednesday afternoon, can I recover it from /gaia/archive?

A: No, the daily backups only take place some time between 6PM and Midnight. If data is created and destroyed before the daily backups take place, it won't be in /gaia/archive.

Q: Are some things excluded from the backups and are not saved in /gaia/archive?

A: Yes, some things are excluded, such as trashcans, cache files, iTunes audio files, etc., and the exclusion list will probably grow to include other things as well. The current exclusions are:

```
Library/Caches/*
Library/PubSub/*
Music/iTunes/*
.Trash/*
.Trashes/*
.Spotlight-V100/*
.fsevents/*
```

Do not depend upon the contents of scratch or workdisk being backed up either, as by their nature they are supposed to contain *temporary* files.