Information Technologies

Mills End-of-Life Plan and Policies

The Mills cluster was purchased in late 2011 and was turned over to investors at the start of 2012. The cluster has a proposed five year lifespan, putting its end-of-life in the January 2017 time frame. At this time, that leaves approximately one year of "life" left in Mills.

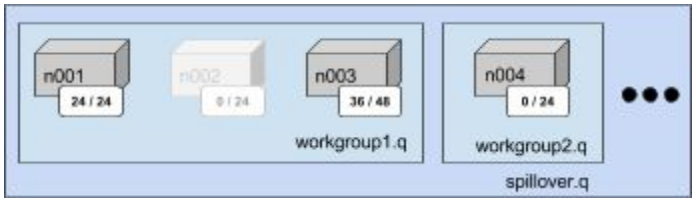
In December of 2015, the vendor from whom Mills was purchased informed Central IT that they would not be able to purchase any hardware support for the year 2016. In short, the hardware is now old enough that the vendor does not expect to be able to secure adequate replacement components. Since node failures are bound to occur, an alternate plan is necessary.

Keeping Mills Online

Ganglia monitoring of Mills shows that the average usage of the cluster (by core) over the last year is 32%. Peak usage tends to stay below the 50% mark. With upwards of 3000 cores unused on average, Mills should have plenty of capacity to handle the typical job load in spite of node failures. The one caveat is the organization of the job scheduler: each investor's queues are serviced by a fixed set of nodes that are "owned" by that investor. If any of those nodes fail, the capacity of that investor's queue to service jobs is diminished.

With no hardware support contract possible for 2016, this job scheduling mechanism has no predictable effects on investors' resources: hardware can fail in various ways for an equally varied number of reasons. The *standby queues* on Mills encompass all nodes of the cluster, so any investor losing "owned" nodes does still have that option for augmenting diminished capacity; however, the hard runtime limits on jobs that are run in standby queues can make that option infeasible as a replacement for "owned" resources for which no runtime limits apply.

A solution to this problem was implemented on the Farber community cluster when it was constructed in 2014. Like Mills, the nodes in Farber are marked as "owned" by specific investors and have "owned" queues that feed jobs to them. In addition, a *spillover queue* that encompasses all nodes exists secondary to all "owned" queues. By preference, regular jobs (jobs that do not run in a standby queue) will fill the "owned" queue for a workgroup. Should an "owned" node be busy or offline, resources from the spillover queue will be employed in lieu of waiting for the "owned" node to become available. Consider the following figure depicting a four-node cluster:



Workgroup1 owns 96 cores spanning n001, n002, and n003 and has jobs occupying 60 cores. Since n002 is offline that workgroup's "owned" queue has a capacity of just 72 cores. A job requesting 24 cores will not be eligible to execute solely in "workgroup1.q." But the node "owned" by workgroup2 is idle at the moment, so between "workgroup1.q" and "spillover.q" there are actually 36 cores available and the 24-core job is still able to execute.

Based on the usage statistics from Ganglia, with the addition of a spillover queue on Mills, the cluster could remain functional even under the failure of a large percentage of the nodes.

Hardware Failures — Policies

A spillover queue will keep Mills functional despite node failures—up to a point. Enough nodes must remain online to service the workload without producing a sizable backlog of jobs. ***If at any time in 2016 greater than 40% of the cores in Mills have failed and are offline, the cluster will be considered "dead" and its shutdown schedule will be accelerated.*** A compute node is deemed "failed" when any of the following conditions exist:

- ***Node does not power on or does not pass POST (power-on self test)***
- ***Failed network adapter (Infiniband or ethernet)***
- ***Failed memory***
- ***Failed processor***
- ***Failed hard disk***

When a node has failed, its owner reserves the right to take possession of it. ***IT will notify a node's owner of the failure and—if the owner does not take possession—reserves the right to reuse any functional components of the node in the repair of other nodes in the cluster.*** Thus, as nodes fail IT can accrue an inventory of spare parts to mitigate other node failures. ***Best efforts will be made to repair failed nodes with replacement parts on-hand.***

The head node of Mills is a special case: without it, the cluster cannot function. ***If the head node fails efforts will be made to repair or replace it.***

Besides the compute nodes, the cluster also includes networking and storage hardware that will likewise no longer be under a support contract:

- As readily-available commodity hardware, ***all attempts will be made to replace failed ethernet switches within the cluster.***

- The Infiniband network in Mills is integral to multi-node parallelism in users' jobs as well as being absolutely necessary for the high-speed Lustre filesystems. ***The service contract for the QLogic 12000-120 Infiniband switch has been extended through the year 2016.***

- The `/lustre-scratch` filesystem exists on hardware for which there is no support contract. ***Failure of one MDS/OSS server within each high-availability pair can be sustained with degraded performance. Failure of both servers in a high-availability pair will leave the filesystem unusable for the remainder of Mills' lifespan.***

    - The Lustre object storage targets (OSTs) are backed by RAID6 volumes which can survive with up to two failed hard disks. ***Failed hard disks will be replaced with hard disks reaped from previously-failed RAID6 sets on the Lustre cluster.***
    - The full failure of any RAID6 set marks the loss of an OST, making the `/lustre-scratch` filesystem unusable. ***When an OST fails, the filesystem will be reformatted without that Lustre storage target.*** Since it is a scratch filesystem, such data loss is not considered terminal to its usefulness.
    - ***The failure of the RAID6 set that backs the Lustre metadata target (MDT) constitutes a terminal failure of the filesystem.***
- The `/archive` filesystem exists on hardware for which there is no support contract.
    - ***Failure of the production server will necessitate the move of its disaster-recovery partner into a production role, with no further offsite replication of the data.***
    - ***Failure of the disaster-recovery server will end offsite replication of the data.***
    - ***Best efforts will be made to replace failed component hard disks during the year 2016 to prevent data loss and overall failure of the filesystem.***

With these policies in place, IT is optimistic that Mills will remain online through its remaining year of service.

Shutdown Schedule

In January 2017 (or under a failure that marks the cluster "dead") IT will cease to offer support for Mills:

- ***No further updates to the operating system***
- ***No further updates to any software packages maintained by IT***
- ***No new user accounts will be created***
- ***Support will occur at the discretion of IT staff for diagnosis of user issues***

Should Mills remain functional under all policies above through January 2017, ***IT will entertain the possibility of allowing the cluster to remain online in this capacity for an additional period (to be determined at that time).***

At the end of any extension period (or in January if no extension is granted) the job scheduler on Mills will be terminated and all compute nodes will be powered-down. ***At this time, investors may claim their nodes and remove them from Mills if they wish to keep the hardware.***

After investors have been given the chance to claim their hardware, IT will begin to dismantle the cluster and dispose of the hardware still in its possession. The `/lustre-scratch` filesystem (if still online) will be

shut down immediately as it constitutes scratch (temporary) storage. The head node will remain online for up to six months to allow ongoing access to data resident in:

- User home directories
- `/lustre/work` workgroup directories
- `/archive` workgroup directories

Important!

Investors are urged to begin planning now with respect to their data resident on filesystems in the Mills cluster. IT has no plans to mass-migrate any of the data present in `/lustre/work` or `/archive` to longer-term storage systems. It is up to individuals and/or workgroups to determine:

- What data is expendable versus what data must be retained
- Where to store the data that must be retained

IT staff will gladly meet with workgroups to discuss options for archival of the necessary data as well as mechanisms by which the data can be moved into whatever longer-term solution is chosen.