Minerva Infrastructure Meeting

December 2, 2010

Participants

Jason Allen (CD/FEF), Gabe Perdue, Jaewon Park, Dave Schmitz, Marc Mengel, Kevin McFairland, Don Gustafson (CD/networking) , Margaret Votava, Rick Snider, John Urish, Lee Lueking, Igor Mandrichenko (and a few others I forgot)

Agenda

The service level requirements for, and strategy for deploying and maintaining, the Minerva near-line machines.

Overview

There are currently two machines that are used for almost-real-time processing of Minerva detector data. The output of this processing is used to monitor the experiment operation and detect problems to which the shifters need to respond. There are two categories of processing being done, 1) fast turn-around low statistics of the current run, and 2) high statistics jobs that take longer.

The discussion was prompted by the recent power outage in Feynman. Both machines are located in FCC2 and were unavailable for more nearly 36 hours. During this time the experiment was "flying blind", although fortunately no problems emerged. The power outage also illuminated the vulnerability of this monitoring to outages of the central services being used. These services include networking and central BlueArc storage.

The experiment acknowledges that it was understood in the original agreement with CD that the machines would receive regular working hours (8/5) support from FEF. This implies that if either of the machines goes down during off hours it would not be attended to until the next business day. It was assumed that the second machine would be the fail over. They had requested assurance that the network between WH12 CR and the machines have 24/7 support.

Requirements

Several requirements were enumerated in the discussion:

1. The short turn-around monitoring is critical and should never be unavailable for more than 2 hours (actual words were "2 or 3 hours").
2. The long turn-around monitoring could experience outages of 24 hours (actual words "a day or so") without serious consequences. However, having at least one server always available is highly desirable.
3. Information from the short turn-around monitoring must always be available to the CR for shifter review.

Strategy and future plan

With these requirements in mind, it was decided that the only place where power and networking are assured at all times when the detector is running, is in the detector hall. Therefore, a machine will be found that can be located in a newly acquired rack Minerva owns in the NUMI Hall. This will be available even if networking and/or power in WH and/or FCC are down. Minerva will maintain this machine (system administration, kernel upgrades, etc. ) similar to their online machines. This will be on the same private network, behind an ACL protected firewall, as the online. The experiment will work out the details of accessing it in the CR and elsewhere as needed.

When the new FCC3 computer room is ready it will have its own UPS independent of the FCC2 computer room. Networking will have networking equipment in both computer rooms and this will provide redundancy to avoid future networking outages. Minerva is purchasing 2 new servers similar to the 2 existing mnvnearline machines. These new nodes will be deployed in the FCC3 computer room thus providing redundancy for the servers as well, in the event of power loss in either of the computer rooms. A procedure needs to be established to move the Condor submission node in the event the node it is on goes down for an extended period.

System maintenance for the machines in Feynman will be done by FEF on the normal Intensity Frontier maintenance days (Third Thursday of the month). Experience has shown that reboots for kernel upgrades are needed only once every two months, so this is roughly the interval between reboots. Jason said he would look into the possibility for a variance to extend the time between reboots even more, but procedures for any machine with interactive logins will need proper review by computer security. Jason discussed a “rolling reboot” procedure used for CDF as an option to consider, although all of the nodes could be rebooted at the same time if the down time is fairly short. However the reboots are performed, the Condor batch jobs for each worker need to be emptied before it is rebooted. The submission node may be ok if it is rebooted quickly as it should recover the queue from where it left off.

Minerva relies on the central BlueArc service for sharing data among CR shifters and others needing to examine it. The data stored here includes detector data and processing results from the nearline machines. In order to circumvent the single point of failure caused by the NFS mounted bluearc, a small dedicated NAS server that would be a “shadow” server was discussed. Data would be written in parallel to both the bluarc, and the shadow storage. In the event that the bluearc failed, the system would fail over to the shadow box. Other options may be available upon further reflection. The current bluearc mounts should be reviewed to ensure that the nearline machines are unaffected if the service is unavailable.