Minos Grid Experience – NuComp 2009/10

- minos_jobsub – remote access, good defaults
- Admin – proxies, authorization, configuration
- File access – code, dcache, Bluearc, OSG security
- Monitoring – Minos plots, bluwatch
- Parrot – AFS available globally, performance
- Scaling issues

Minos Grid – NuCOMP 2009/10

minos_jobsub

- Available on all Minos Cluster interactive nodes

  - http://www-numi.fnal.gov/condor/

- Submits and controls local and GlideinWMS jobs

- Helps manage input and output files

- Regulates rate of condor_rm due to SAZ limits

- minos_q shows global summary of jobs

Minos Grid Experience - NuComp 2009/10

Administration

- User registration
    - Log into minos25 master node to setup crontab to keep a Kerberos proxy alive for use by Condor
    - VOMS registration requires KCA proxy in brower
- Condor configuration -
    - Minos provides config files
    - FEF must push out with cfengine
    - Minos can clean up working files under condor.

Minos Grid Experience – NuComp 2009/10

File Access

- Code
  - Base releases in AFS (moving to /grid/fermiapp)
  - User test releases in /minos/scratch
- Dcache – raw data and reconstructed files – 30 TB cache
- Bluearc – Analysis and personal working areas – 80+ TB
  - Many small files are common
- OSG security – presume a malicious root user on workers
  - OSE (/grid/data) writeable, noexe
  - GCE (/grid/fermiapp) readonly, exe

Minos Grid Experience – NuComp 2009/10

Monitoring

- Minos Condor plots by rbpatter -

    – http://www-numi.fnal.gov/computing/condor_monitoring/index_day_all.html

- Bluearc

- Check file open for success and speed every minute from several hosts

    – http://www-numi.fnal.gov/computing/dh/bluwatch/

- Log data rates from NexSan/Satabeast and HDS every minute from 1 host

    – http://www-numi.fnal.gov/computing/dh/bluwatch/rate/2009/10/14/minos27.txt

Minos Grid Experience – NuComp 2009/10

Bluearc performance

- d0ora2 was primary cause of problems

    - Scheduled 18:00 Sat/Sun/Tue/Thu

    - special hardware and software, misconfigured

    - resolved Jun 25 to Jul 13, then returned.

    - retired, upgraded to d0ora3 on Aug 25

- LOCKING - /grid/fermiapp/minos/scripts/cpn

    - Allows 'n' copies, presently 10, polls optimally

    - The old cp1 locked 340K times since April 27

    - New cpn has locked 200K times since Aug 21

Minos Grid Experience - NuComp 2009/10

Bluearc performance

- Shifted D0 to separate disks - they are causing sustained overloads

- Shifting Minos to separate disks now

- See monitoring plots, such as

    - http://www-numi.fnal.gov/computing/dh/bluearc/rates/minos27/NOW.png - current day

    - http://www-numi.fnal.gov/computing/dh/bluearc/rates/minos27/WEEK.png - current week

    - http://www-numi.fnal.gov/computing/dh/bluearc/rates/minos27/ - historical

    - http://www-numi.fnal.gov/computing/dh/bluearc/rates/minos-sam04/20090629.week.png - scratch

- Failures are very rare !

    - We have seen periods of nearly 1 year with no failures to open a file.

Minos Grid Experience - NuComp 2009/10

PARROT deployment on Fermigrid

- PARROT has problems with SAM projects, etc
    - Problems with I/O intensive activity (like cp1 or cpn)
    - Parrot is designed for CPU intensive applications
- We plan to rsync the parrot-accessed files to /grid/fermiapp and /minos/data

Minos Grid Experience - NuComp 2009/10

Scaling issues

- Minos has run as many as 1900 jobs on Fermigrid, goal is 5000

- Why not buy them for dedicated use ? Fine, if we have $1M.

- We run all the our Condor processes on one node (unusual)

    - This seems to handle the load well, with modern hardware.

- Bluearc/general file access considerations.

    - CDF and D0 are moving .2 to .8 GB/sec, peaks over 1 GB/sec
        - Data served by thousands of disks, via SAM/DCache
    - /grid/data has only about 100 physical disks
        - NFS has no scheduling, so we do it ourselves (cpn)
    - How do we track usage ? This is hard with NFS at present.