Duke Atlas Tier 3 Site

Doug Benjamin

(Duke University)

Duke Atlas Group

Personnel

- 4 Tenured Faculty

- 1 Research Professor, 2 Research Scientists

- 3 PostDocs (located at CERN)

- 1 PostDoc (who will eventually transition from CDF at the Tevatron to Atlas)

- Several students

- Thus our Tier 3 has to be configured with to handle ***4 separate*** analyses at once

Duke Tier 3 site (current HW configuration)

- 3 TB Fileserver (Centos 4.6 64 bit OS)
  - Data and user home areas exported to other nodes in cluster
  - Grid Storage Element (Bestman-gateway with local FS)
- 2 interactive nodes

  2 dual core AMD , SL4.4 (32bit) 500 GB local storage

- One worker node class machine (32 bit SL 4.4)

  - Atlas code server
  - Bestman-gateway/Xrootdfs/Fuse
- Xen hypervisor SL5.2 with two virtual machines
  - Gums server (up and running but not really used)
  - Condor Master machine

HW config continued

- ## 2 worker nodes (Dell 1950 1U)

  - 8 cores ,2 GB/core, SL4.7 (64bit) 300 GB local storage

  - 8 Condor slots – in separate partitions (16 GB/each)

  - Xrootd data server

- ## 1 worker node (assembled from parts)

  - 8 core, 2 GB/core, SL 4.7 (64 bit), 2.7 TB local disk (3)

  - 8 Condor slots – separate partitions (16 GB/each)

  - Xrootd data server

- ## Xen hypervisor w/ 3 virtual machines (assembled from parts)

  - 1 worker node. 4 cores, 2GB/core, SL 4.7 (64bit), 300 GB local storage, 4 Condor slots, Xrootd data server

  - Xrootd director (also Proof master if Proof is added to cluster)

  - Extra VM (though to be Proof master initially)

Duke Tier 3 Software/Support Personnel

- Software installed
  - Several recent Atlas releases from kits – done manually
  - Condor , WLCG client
  - Bestman-Xrootd, Bestman gateway, gridftp
- Support Personnel
  - Duke Physics department system managers maintain OS and interactive user accounts
  - Duke HEP Research scientist (me) Installs/maintains Atlas, OSG and Condor software
    - 0.10 to 0.25 FTE available for these activities
    - Implies the code must be simple to install, debug and operate

Services at Duke Tier 3

- Condor cluster with 28 slots
  - Use basic condor install/configuration with small changes
  - Configuration files all on common disk
- Xrootd data system (with Bestman-gateway SRM)
- NFS data storage with Bestman-gateway SRM
  - *Bestman-gateway SRM   Easy to install*
- Atlas code releases from kits
  - Atlas env. similar to LXPlus (cmthome/ requirements file)
  - differs from how things are done at ANL
- Wiki visible to all
  - Important for communication amongst support Personnel
  - Source of training materials for group members

Cluster Evolution/ Design Constraints

- Last year after OSG meeting in NC – setup a OSG CE with help of RENCI. Now concluding that OSG CE total cost of ownership is too high (effort involved to maintain) for benefits

- Previous experience lets me to conclude that I will want managed storage sooner than later (hence work w/ Xrootd)

- Need to fit within constrains of Department computer room

  - Rack mounted hardware

  - Mirrored system disks (reduce sys admin efforts)

  - Standard configurations. / Use virtual machines when appropriate

- Need to reduce load on Department system admins.

  (they are providing best effort support)

- Adapt system to needs/wishes of the users

Upgrade plans

Short term (within 1-2 months)

- Replacement of interactive login machines
  - 8 core , 2 GB/core, 500 GB local scratch space
- Additional Disk -
  - 16 TB standalone file server (assembled from parts ~ 6K$)
- Managed switch

Longer term (minimal survival mode) (within 1 year)

(assumes no new sources of computing funds)

- 64 cores (2 GB/core) (~10 machines)
- ~ 20 TB disk storage (on worker nodes)

Add Local Panda submission to site

- Allow users to test their jobs before going to grid
- Needs http interface to local file catalog

Our current analysis model

Based on a small Tier 3

- Heavily depends on outside computing

- Fraction of data copied to local storage

- Test programs with local machines

- Send to grid - larger scale jobs

- Resulting histogram files returned to local machines

- Disadvantage -

- Very reliant on availability on outside CPU and storage at the Tier1 and Tier2 sites

Upgrade Plan (ideal model)

- ~25 8 core machines

- ~ 100 TB disk space

  - This much space requires some space management

    - Typical file size << 1 GB/file

- Analysis model

  - Copy most of DPD data locally

  - Analyze locally except for the very large jobs

  - Would allow us quickly analyze the data as it is available

Challenges

- Interaction with Duke Computing infrastructure
    - Campus Firewalls / Local Firewalls
        - open ports normally closed - slow process
        - Determine required ports by services open
    - Campus networking
        - "Last mile problem" - Duke U is on Internet 2
        - Working with Campus networking people to solve problem
- Getting files to our site

    - Using dq2-ls and dq2-get to identify and copy files to Duke

    - dq2-ls and dq2-get often fail at Duke – silent failures – return nothing.   Can see no pattern in failures
- Interface to Condor batch system

    - my job has many parts – not a bunch of single Condor jobs

    - ARCON scripts form ANL should help – but need to be extended for MC type jobs (no input files but output files)

Doug Benjamin

Challenges cont.

- <u>Setting up Bestman-Xrootd</u>
    - VDT documentation – confusing at best in current form
    - Use other sources (Proof documentation, Scalla documentation, e-mail from Wei Yang – direct and through OSG storage)
    - Excellent help form Wei Yang
- <u>Registering Duke Tier 3 as part of OSG</u>
    - I am first of these types of Tier 3 to register (SE only)

        provide storage for local users only
- <u>Funding for Computing</u>
    - In past 3 years - received no funds from DOE for computing
    - All computing comes from a small grant from Duke U.
    - Need to find additional resources if we are to grow to the size we envision we need to be effective
- Longer term - <u>Local Data management</u>

Conclusions

- Duke Tier 3 is an example of a small tier site being designed and configured to provide enough resources to allow group members to analyze the Atlas data effectively on a limited budget

- To reduce the support load excellent documentation is critical.

- Tier 3's will need some support – need to have a robust Tier 3 support group.

- Without sufficient funding to Tier 3 sites, it implicitly assumes that sufficient resources will be made available to the general users at the Tier1 and Tier 2 sites.