We will be having our regular bi-weekly Infrastructure WG telecon on Monday, February 1, at 12 Noon CT.

Agenda

- Existing Resource Usage Update
    - ♦ TeraGrid Resources (Startup Allocation)
        - ◊ Service Units
            - · Allocated: 30K SUs on abe; 30K SUs on lincoln (nVidia Tesla GPUs)
            - · Remaining: ~29.4 SUs (abe); 30K SUs (lincoln) as of Feb1
        - ◊ Disk Storage
            - · **/cfs/projects/lsst**
            - · Allocated: 5TB
            - · Remaining: 5TB
        - ◊ Tape Storage
            - · Allocated: 40TB
            - · Remaining: 40TB

- Cost Sheet Update
    - ♦ Baseline verion is v45
        - ◊ ?https://www.lsstcorp.org/docushare/dsweb/Get/Version-12185/Infrastructure-Costs-v45.xls
        - ◊ caveats apply: v45 does not *exactly* match PMCS
    - ♦ Current version now v73
        - ◊ ?https://www.lsstcorp.org/docushare/dsweb/Get/Document-6284/Infrastructure-Costs-v73.xls
        - ◊ ?https://www.lsstcorp.org/docushare/dsweb/Get/Document-8189/CostSummaryWithBaseline-
    - ♦ Summary of Changes
        - ◊ New floorspace model (-2550K/-11336K) (**ArchSite 620->871 (peak 920); BaseSite 480->843 (peak 864)**)
            - · ?LSST-50 Floorspace tab: Floorspace calculation does not take into account the increase in drive capacities over time
            - · ?LSST-69 New Model for Floorspace (Lease Costs) at ArchSite
            - · ?LSST-70 Lease Costs at ArchSite: How are the Floorspace rates expected to change over time?
        - ◊ ?LSST-83 Track number of cores as a primary technical metric
            - · Highlights the need to pay attention to ?Amdahl's Law
        - ◊ ?LSST-89 Fix bad assumption regarding hardware tape compression - eliminate 2-to-1 (+461K/+1147K)
    - ♦ Questions & Notes
        - ◊ Ramp up: One of the things in the cost sheet that I wonder about is our "ramp up", i.e. we're currently planning on buying 1/3 of the hardware 3 years early, 2/3 two years early, etc. I wonder if 3 years early is a little too soon.
    - ♦ Upcoming Changes
        - ◊ **Priority is updating the Power & Cooling estimates**
            - · ?LSST-10 Update Power & Cooling at Base Site (info already received from RonL)
            - · ?LSST-47 Power Costs at BaseSite: Use Historical Data to Model Future Power Prices
            - · ?LSST-36 Update Power & Cooling at ArchSite
            - · ?LSST-36 P&C and Floorspace at PCF (rates, payment approach, green features of PCF)

- ◊ ?LSST-78 Move the 3% CPU spare from document 2116 "CPU Sizing" to document 6284 "Cost Estimate"
- ◊ ?LSST-79 Add tape library replacement to ArchAOS and BaseAOS
- ◊ ?LSST-28 Optimal CPU Replacement Policy
- ◊ ?LSST-14 Processor Sizing Update (Doc2116 LSST CPU Sizing)
- ◊ ?LSST-37 Missing controller costs for disk

♦ Next steps with cost sheet
- ◊ Full review each of the elements of the cost sheet (boxes of the mapping document)
  - · More readable description of the formulas being used
  - · Identification and documentation of assumptions
  - · Identification and documentation of external data input
- ◊ Serves two significant purposes
  - · Allows for better internal reviews (validation of models and information used)
  - · Provides justifications for external reviews
- ◊ Results in an updated (or replacement of) Document-1684 and related documents ("Explanation of Cost Estimates")

• DC3b Infrastructure Options/Costs? for the Performance Tests

♦ ?http://dev.lsstcorp.org/trac/wiki/DC3bHardwareRequirements

♦ ?LSST-11 DC3b Hardware

♦ Compute
- ◊ Research Proposal submitted to TeraGrid for allocations starting Apr1
- ◊ ?https://www.lsstcorp.org/docushare/dsweb/Get/Document-8529/LSST-TeraGrid-Proposal.pd
- ◊ should hear back in early March

♦ ImSim data (both catalog and image files)
- ◊ 47TB image files; 15TB database
- ◊ **Is the InfraWG responsible for protecting this data against loss?**

♦ Scratch Disk
- ◊ Project space (est 20TB) plus available scratch (50TB) is ~70TB (scratch on abe is 100TB, but shared and variable)
- ◊ Intermediate and output data is 140TB for PT1, 200TB for PT2, 300TB for PT3
- ◊ The **scratch gap is 70TB** for PT1, 130TB for PT2, 230TB for PT3

♦ Database Disk
- ◊ Can use existing SAN
- ◊ ?LSST-73 Add additional storage from our existing SAN allocation to lsst10 /scr
  - · Update: we are going ahead and adding our remaining SAN allocation to lsst10. The total space available for MySQL data will be 14.7 TB. See ?LSST-73 for more details.
- ◊ The **database gap is 3TB** for PT1, 4TB for PT2, 6TB for PT3

♦ Tape Storage
- ◊ **Does the catalog data (the database) need to be backed up?**
- ◊ Total tape needed is (162TB, 250TB, 400TB)
- ◊ The **tape gap is 62TB for PT1**, 150TB for PT2, 300TB for PT3
- ◊ Pricing
  - · $62/TB (for single copy) [$25/tape=400GB]; for 300TB is ~$19K (LTO-3)
  - · $31/TB (for single copy) [$50/tape=1.6TB]; for 300TB is ~$9K (LTO-5) [plus faster bandwidth than LTO-3]
  - · Note1: Working with PI for possible purchase strategies, which include PI subsidizing during tape usage and LSST subsidizing LTO-5 tapes, to avoid the need to buy LTO-3 (old technology) tapes.

Additional notes
    - ◊ Note2: Mass storage will exist post-2010 (ongoing talks with NCSA PI); New system by Oct1 (estimated)
    - ◊ Note3: Estimated DC3b data loss from tape failures published on mailing list
  - ♦ Timelines
    - ◊ Lead time for acquisition, installation, etc.

- DC3b Infrastructure Options/Costs? for Data Serving
  - ♦ ?http://dev.lsstcorp.org/trac/wiki/DC3bDataServingRequirements
  - ♦ ?LSST-54 Connections Speeds between lsst10 and the SAN Storage. We need 300 MB/s. What are our options?
    - ◊ Adapter slots on lsst10 will not support 8Gb HBA, **getting price estimate for a new database server**
  - ♦ Image Retrieval needs are **unspecified** (servers? spinning disk?); software? (ftp?); See next agenda item also (REDDnet).

- Distributed File Management (REDDNET/Lstore/iRods/DataNet)
  - ♦ What is **the role of REDDnet** during DC3b?

- Update on LSST Database Performance Tests Using SSDs (Arun/Jacek?)

- Database access from Pipeline processing during DC3b
  - ♦ discuss details, specifically, is access occuring:
    - ◊ during pipeline startup?
    - ◊ during stage pre/post process?
    - ◊ during stage process()?
  - ♦ effects on performance, scalability, etc.

- Miscellaneous
  - ♦ Possibility of partnering with BW on GPFS. Nothing concrete yet, but could lead to shared expertise and special licensing.

- Upcoming Conferences of Interest
  - ♦ 11th LCI International Conference on High-Performance Clustered Computing
    - ◊ March 8-11, 2010 Pittsburgh
    - ◊ ?http://www.linuxclustersinstitute.org/conferences/

- InfraWG Ticket Update

Notes

Attendees: DavidG, K-T, JeffK, Jacek, Daniel, Ray, AndyC, MikeF

- Action items reflected in JIRA tickets.

- only 5TB for ImSim catalog (not 15TB)
- will store all ImSim data received from AndyC on dual-tape
  - ♦ 47TB of image files
  - ♦ 5TB of catalog data (flat files used to import into mysql)

    - 10TB of ImSim input data (gzip files used as input to ImSim processing)
- will need ~7TB of spinning disk on the database (5TB of data plus ~2TB for indexes) for ImSim catalog
    - total database disk now 15TB (PT3)
- database gap is now gone with these revisions
- database backups will be on single-copy tape
- scratch gap is not an issue
    - pipelines will write and read from tape during processing
- we may not need faster bandwidth to the SAN storage from the database server during PT1
- REDDnet
    - pilot project will explore it as a data distribution mechanism for providing data access to collaborators
    - can REDDnet replicate the catalog database?

Useful Links

- InfraWG Home Page
    - ?http://dev.lsstcorp.org/trac/wiki/InfrastructureWG
- InfraWG Tickets (in priority order)
    - ?All InfraWG Tickets