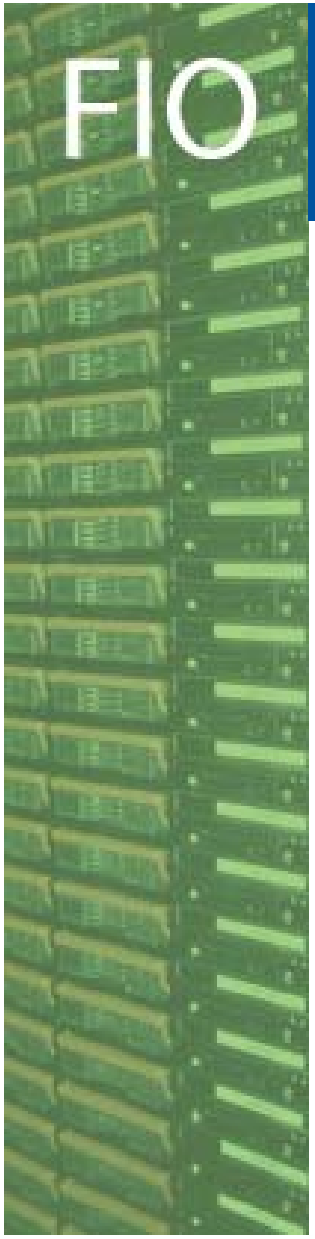


Fabric Infrastructure and Operations

CERN IT Department

CASTOR Operational experiences

HEPiX Taiwan Oct. 2008

Miguel Coelho dos Santos

CERN - IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it

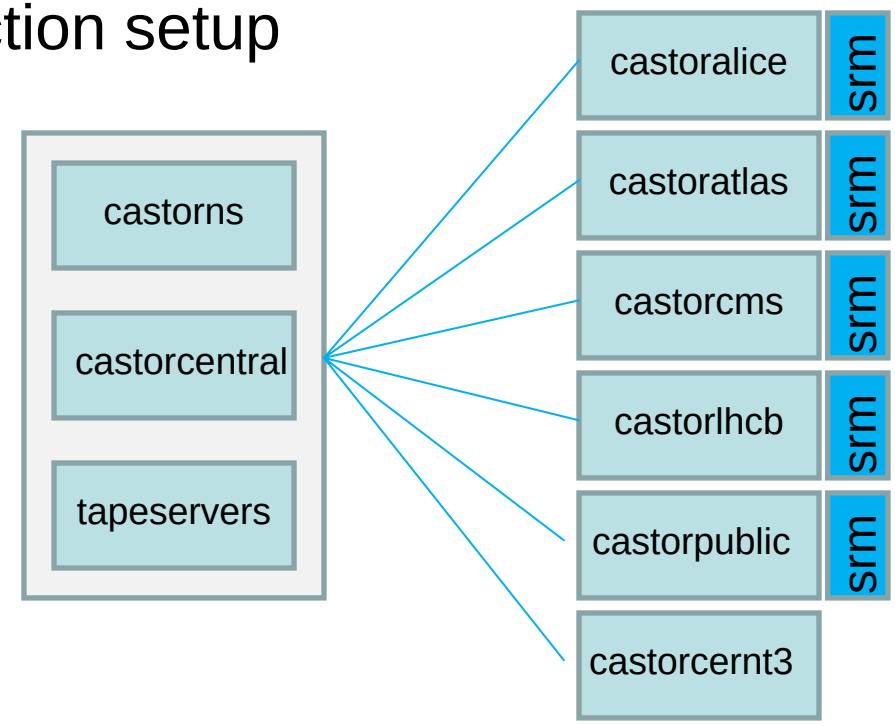
Summary

- CASTOR Setup at CERN
- Recent activities
- Operations
- Current Issues
- Future upgrades and activities

CERN - IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it

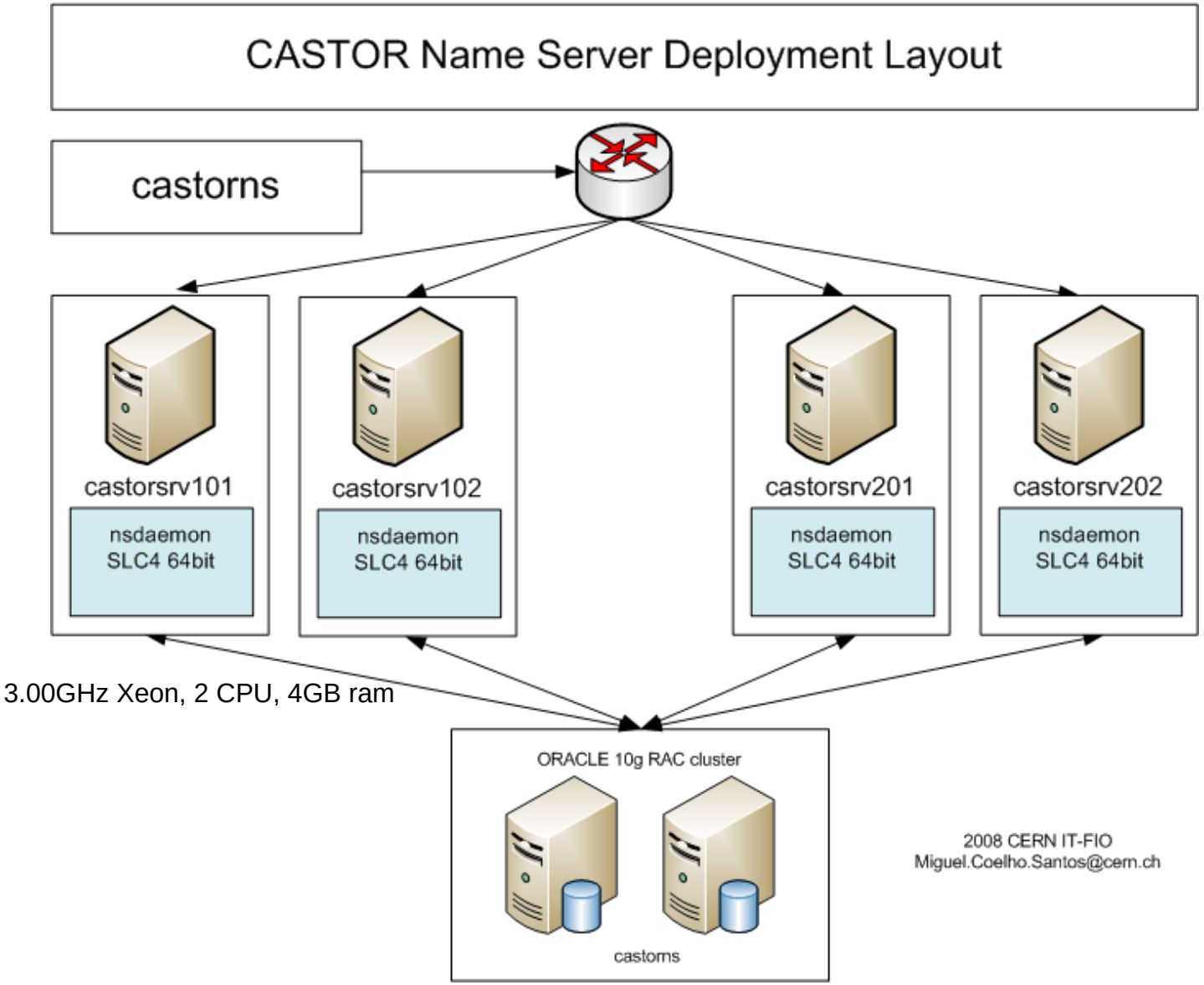
CASTOR Setup at CERN

- Some numbers regarding the CASTOR HSM system at CERN.
  - 5 tape robots, 120 drives
  - 35K tape cartridges, 21 PB
  - 900 disk servers, 28K disks, 5 PB
  - 108M files in name space, 13M copies on disk
  - 15PB of data on tape
- Production setup



CERN - IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it

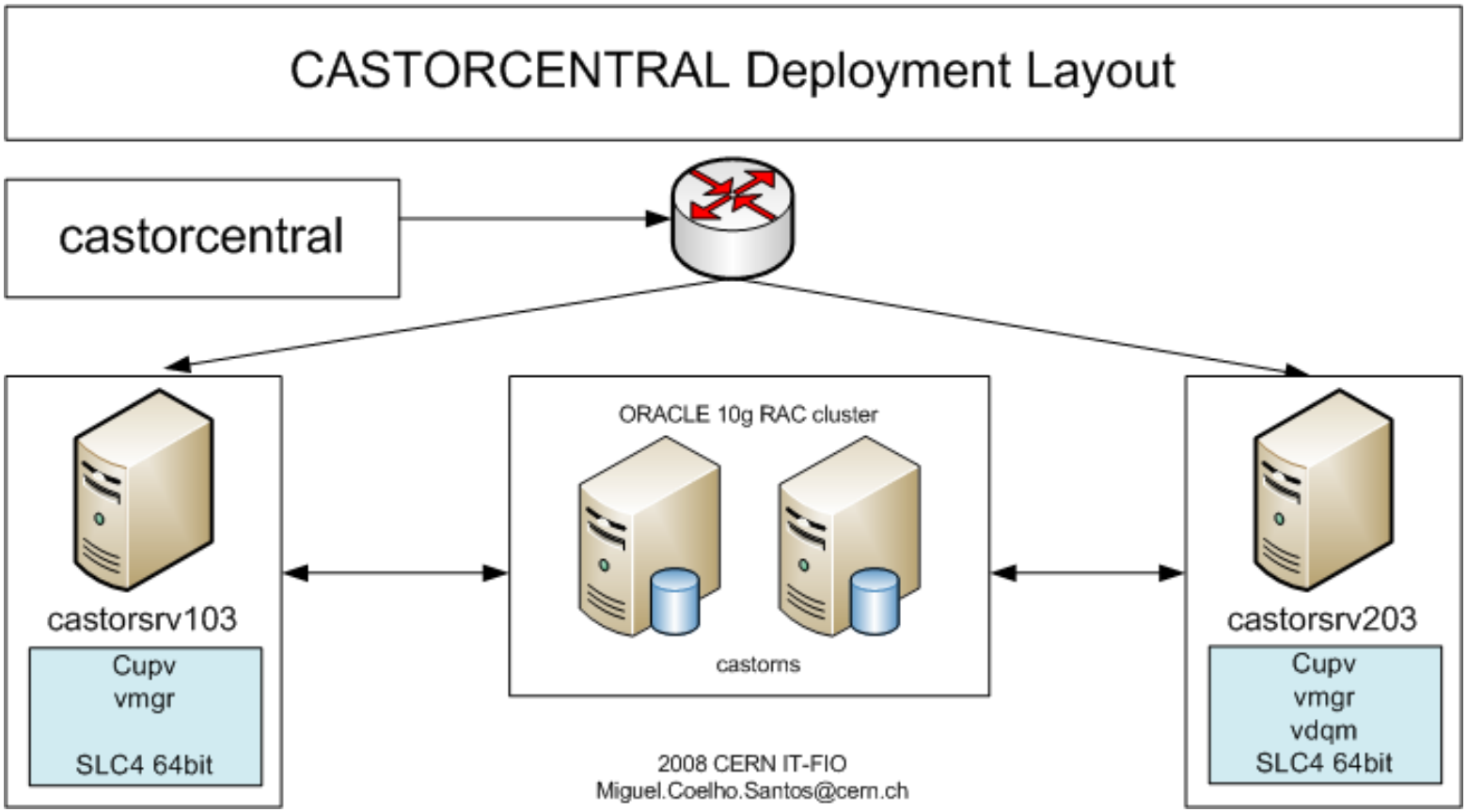
CASTORNS setup



2008 CERN IT-FIO
Miguel.Coelho.Santos@cern.ch

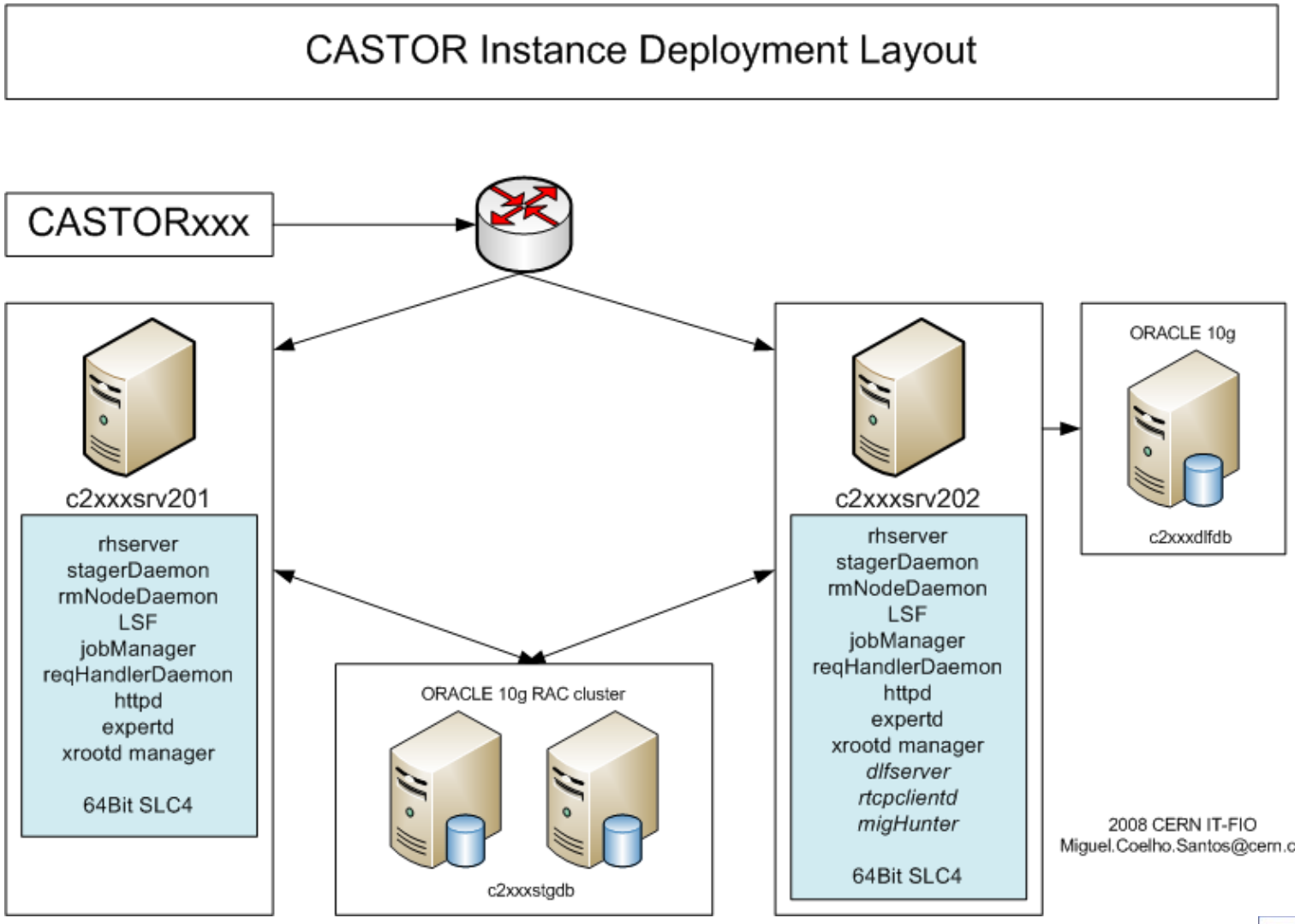
CERN - IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it

CASTORCENTRAL setup



3.00GHz Xeon, 2 CPU, 4GB ram

Instance setup (cernt3)



CPU E5410 @ 2.33GHz, 8 cores, 16G

On High-Availability

- High-available setup achieved by running more than one daemon of each kind
- Strong preference for Active-Active setups, i.e. load-balancing servers using DNS
- Remaining single points of failure:
  - vdqm
  - dlfserver
  - rtcpclientd
  - mighunter
- Deployment of CERNT3 model to other production instances depends on deployment of new SRM version

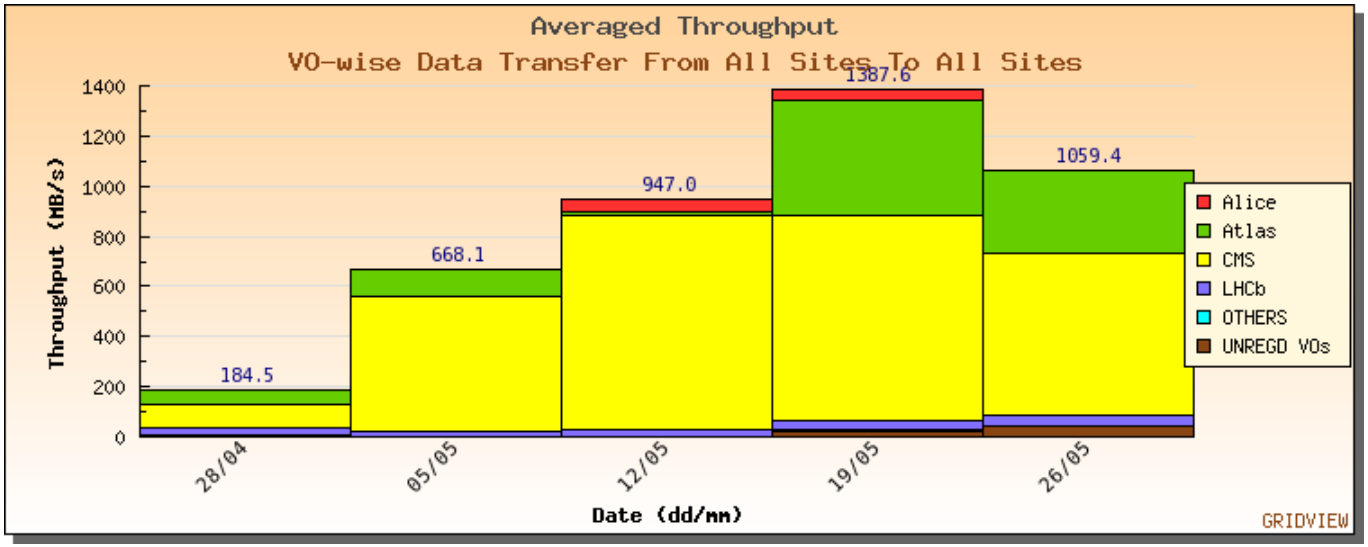
Recent activities

- CCRC-08 was a successful test
- Introducing data services piquet and debugging alarm lists
- Ongoing cosmic data taking
- Created REPACK instance
  - Needed a separate instance to exercise repack at larger scale
  - Problems are still being found and fixed
- Created CERNT3 instance (local analysis)
  - New monitoring (daemons and service)
  - Running 2.1.8 and new xroot redirector
  - Improved HA deployment

CERN - IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it

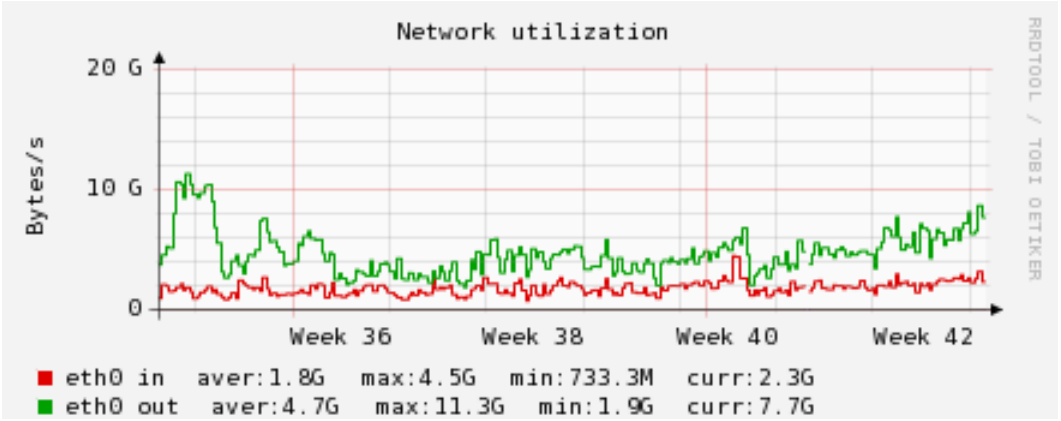
CCRC-08

- Large number of files moved through the system
- Performance (speed, time to tape, etc) and reliability met targets => Tier-0 works OK
- SRM was the most problematic component: SRMserver deadlock (no HA possible), DB contention, gcWeight, crashes.

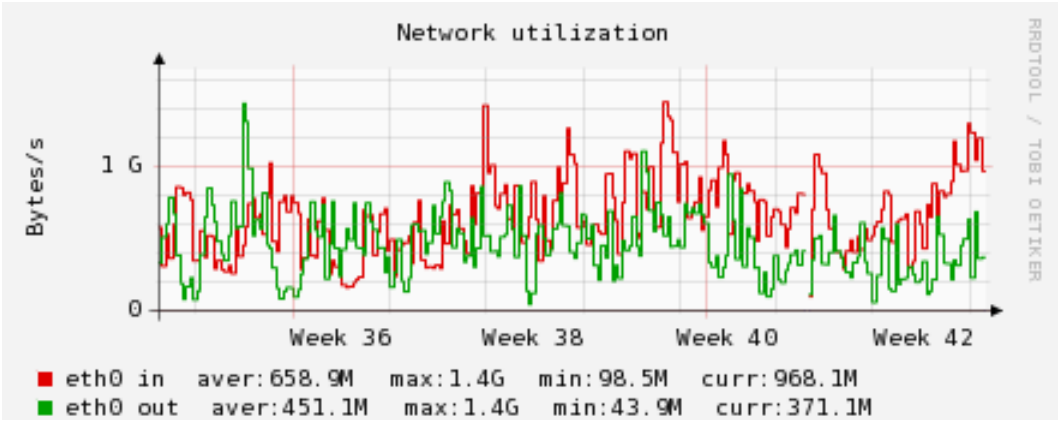


Current data activity

- No accelerator... still there is data moving
- Disk Layer



- Tape Layer



CERN - IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it

Data Services piquet

- In preparation for data taking, besides making the service more resilient to failures (HA deployment), the Service Manager on Duty rota was extended for 24/7 coverage of critical data services => Data Services Piquet
- The Services covered are: CASTOR, SRM, FTS and LFC.
- Manpower goes up to increase coverage but:
  - The number of admins that can do changes goes up
  - The average service specific experience of admins doing changes goes down

Improve documentation
Make admin Processes more robust (and formal)

CERN - IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it

Operations

- Some processes have been reviewed recently
- Change management
  - Application upgrades
  - OS upgrade
  - Configuration change or emergencies
- Procedure documentation

DISCLAIMER

ITIL covers extensively some of these topics. The following slides are a summary of a few of our internal processes. They originate from day to day experience, some ITIL guidelines are followed but as ITIL points out, implementation depends heavily on specific environment. Our processes might not apply to other organizations.

CERN - IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it

Change Management (I)

- Application Upgrades, for example to deploy castor patch 2.1.7-19
- Follow simple rules
  - Exercise the upgrade on pre-production setup, identical as possible to production setup
  - Document the upgrade procedure
  - Reduce concurrent changes
  - Announce with enough lead-time, O(days)
  - "*No upgrades on Friday*"
  - Don't upgrade multiple production setups at the same time

CERN - IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it

Change Management (II)

- OS Upgrades
- Monthly OS upgrade, 5K servers
  - Test machines are incrementally upgraded every day
  - Test is frozen into Pre-production setup
    - All relevant (specially *mission critical*) services should (must!) have a pre-prod instance, castorpps for CASTOR
  - 1 week to verify mission critical services
    - internal verification
  - 1 week to verify the rest of the services
    - external verification (several *clients,* for example voboxes*)*
  - Changes to production are deployed on a Monday (D-day)
  - D-day is widely announced since D-19 (=7+7+5) days

CERN - IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it

Change Management (III)

- Change requests by customer
  - Handled case by case
  - It is generally a documented, exercised, standard change => low risk
- Emergency changes
  - Service already degraded
  - Not change management but incident management!
- Starting to work on measuring change management success (and failures) in order to keep on improving the process

CERN - IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it

Procedure documentation

- Document day-to-day operations
- Procedure writer (expert)
- Procedure executer
  - Implementing service changes
  - Handling service incidents
- Executer validates procedure each time it executes it successfully, otherwise the expert is called and the procedure updated
- Simple/Standard incidents are now being handled by recently arrived team member without any special CASTOR experience

CERN - IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it

Current Issues (I)

- File access contention
  - LSF scheduling is not scaling as initially expected
  - Most disk movers are memory intensive: gridftp, rfio, root.
  - So far xrootd (scala) seems more scalable and redirector should be faster
- Hardware replacement
  - A lot of disk server movement (retirements+arrivals). Moving data is very hard. Need drain disk server tools.
- File corruption
  - Not common but happened recently that a raid array started to go bad without alarm from fabric monitoring
  - Some files on disk were corrupted.
  - Need to calculate and verify checksum of every file on every access.
  - 2.1.7 has basic RFIO only check summing on file creation (update bug to be fixed, removes checksum)
  - No checksum on replicas
  - 2.1.8 has replica checksum and checksum before PUT

Current issues (II)

- Hotspots
  - Description
    - Without any hardware/software failure, users are not able to access some files on disk within acceptable time
  - Causes

1. Mover contention

Not enough resources (memory) on disk server to start more movers. Problem is specially relevant in gridftp and rfio transfers**.**
xroot should solve it for local access. Grid access needs to be addressed.

2. File location

Although more movers can be started on the disk server, other disk server resources are being starved, for example network or disk IO. Can be caused by:
1.Multiple high speed accesses to the same file (few hot files)
2.Multiple high speed accesses to different files (various hot files)
There is currently no way to relocate hot files or to cope with peaks by temporarily having more copies of hot files.

CERN - IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it

Current issues (III)

- Small files
  - Small files expose overheads, i.e. initiating/finishing transfers, seeking, meta data writing (for example tape labels, catalogue entries and POSIX information on disk (xfs))
  - Currently tape label writing seems to be the biggest issue...
  - More monitoring should be put in place to understand better the various contributions from the different overheads
- Monitoring
  - Current monitoring paradigm relies on parsing log messages ☹
  - CASTOR would benefit from metric data collection at daemon level

Future upgrades

- Upgrade FTS
  - It will allow to start using internal gridftp! (lower memory footprint)
- Upgrade SRM
  - A lot of instabilities in recent versions ☹ ☹ ☹
  - Getting SRM stable is critical
  - It will allow to run load balanced stagers everywhere!
- 2.1.8
  - New xrootd
  - Complete file check summing
  - Etc (see talk by S. Ponce)
- Monitoring
  - New fabric monitoring sensor and more lemon metrics are in the pipeline

CERN - IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it

Future activities

- Data taking
- Local analysis support
- Disk server retirements
  - ~65 disk servers to retire 1Q 2009
  - More throughout the year

FIO

Questions?

CERN IT Department

CERN - IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it