ASGC Site Report

CASTOR F2F meeting

HungChe Jen (hung-che.jen@cern.ch)
OPS Team
Academia Sinica Grid Computing

Outline

- 1. Current architecture of Castor
  - Service overview
  - Hardware resource
  - Storage classes
  - Monitoring
- 2. Issues/problem/Event
- 3. Plans for 2009

Overview (1)

Production Castor 2

- Core service (Stager, NS, DLF...)
  - Version: 2.1.7-19-2
  - Running on SLC4/64bit
  - Intel Xeon 5150 2.66 GHz Dual Core x2, 4GB memory
  - LSF 7 update 3
  - 3 nodes Oracle RAC
- SRM v2.2
  - Version : 2.7-12
  - srmServer x3, srmDaemon x3
  - Running on SLC4 64bit
  - 2 nodes oracle RAC

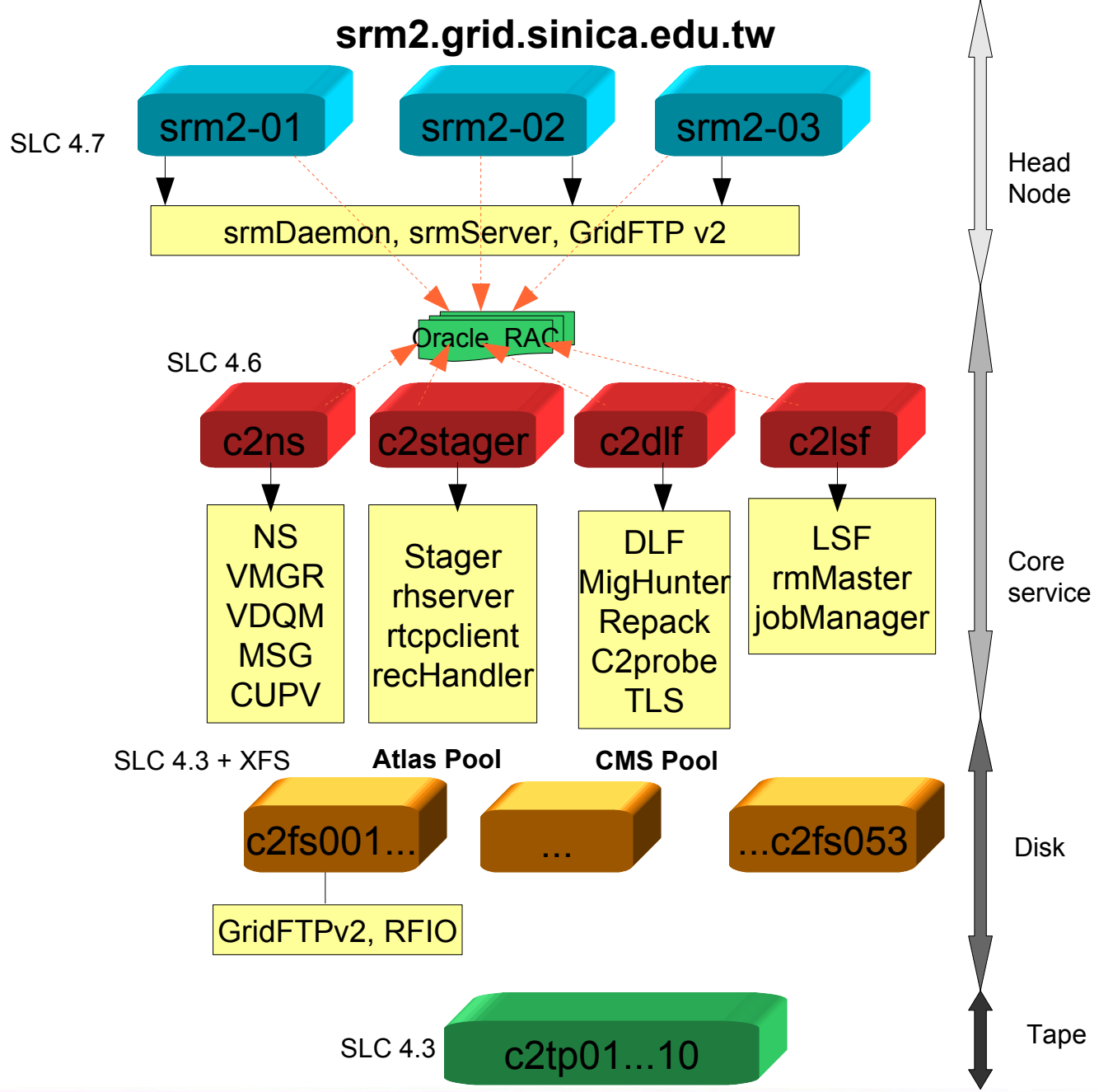
Overview (2)

- 53 Disk servers
  - SLC4, 64Bit machine
  - 8GB memory
  - XFS file system
- 8 Tape servers
  - SLC4, 32Bit machine
  - 4GB memory

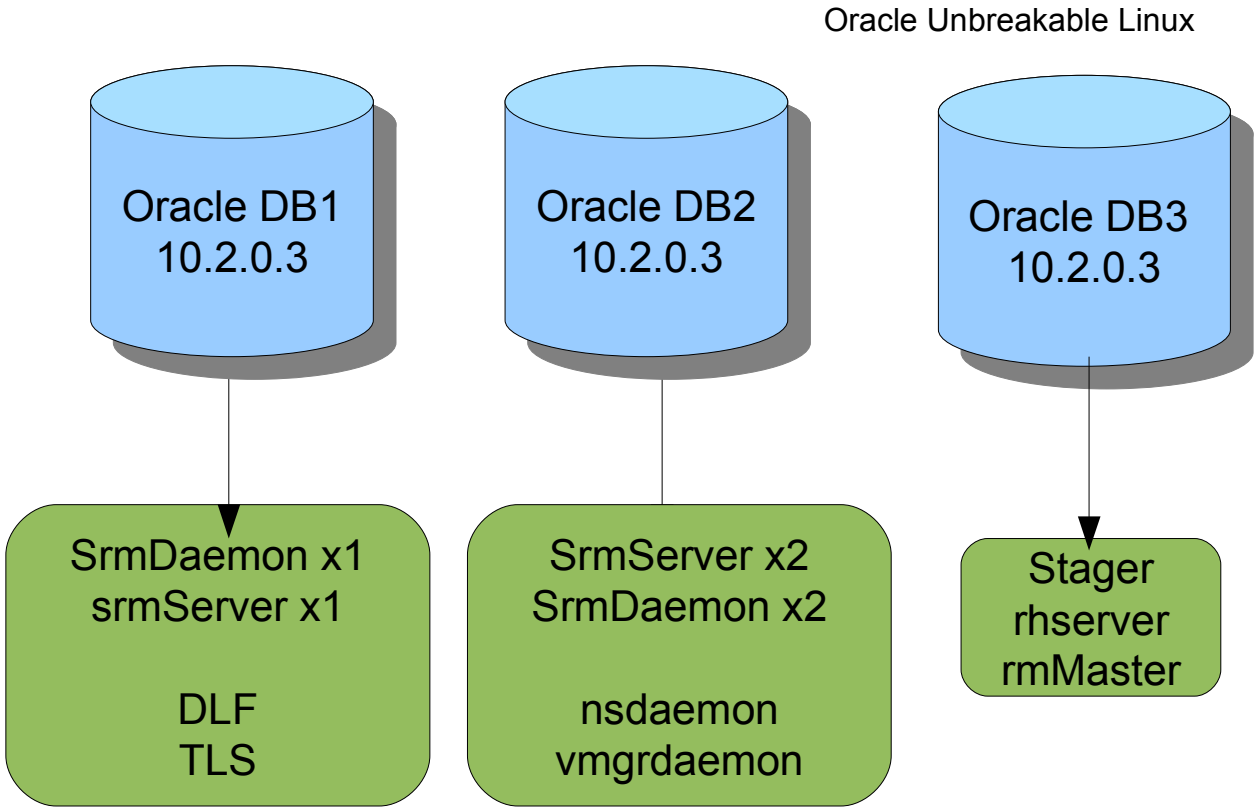
Overview (3)

Pre-production Castor 2

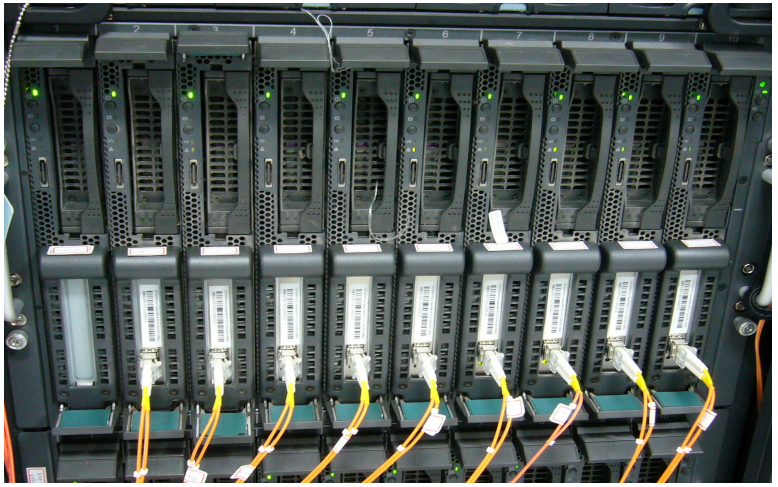
- Core service (Stager, NS, DLF...)
  - Version: 2.1.7-24
  - Running on SLC4 64bit
  - LSF 7 update 3
  - Oracle single instance
- 2 tape server with LTO4 tape drive
- SRM v2.2
  - Version : 2.7-12
  - srmServer and srmDaemon x1
  - Running on SLC4 64bit
  - Oracle single instance


srm2.grid.sinica.edu.tw
SLC 4.7
srm2-01
srm2-02
srm2-03
srmDaemon, srmServer, GridFTP v2
Oracle RAC
SLC 4.6
c2ns
c2stager
c2dlf
c2lsf
NS
VMGR
VDQM
MSG
CUPV
Stager
rhserver
rtcpclient
recHandler
DLF
MigHunter
Repack
C2probe
TLS
LSF
rmMaster
jobManager
SLC 4.3 + XFS
Atlas Pool
CMS Pool
c2fs001...
...
...c2fs053
GridFTPv2, RFIO
SLC 4.3
c2tp01...10
Head Node
Core service
Disk
Tape

Oracle RAC for Castor/SRM



Hardware (1)



- All server use Blade system
  - SMP Xeon 3.0, 8GB ECC Memory
  - Support remote power control
  - Quanta S72A + LSI FC card
  - IBM HS21 Blade system + Qlogic FC card

- Disk Array (Total Capacity: ~1 PB)
  - SilverStor TN-6224S-FFG
  - Single controller
  - 3.5” SATA -II 750GB/1TB 7,200R/16MB*24pcs
  - Cache RAM : 512MB DDR
  - RAID 6 + 1 hot spare

Hardware (2)

- Tape System 1
  - IBM TotalStorage 3584
  - 8 LTO3 Tape Drives
  - 2489 Slots, 2000 LTO3 cartridges
  - Total capacity:800TB
- Tape System 2
  - IBM TotalStorage 3584
  - 4 LTO4 Tape Drives
  - 727 Slots, 660 LTO4 cartridges
  - Total capacity: 528TB

New Hardware Resources

- SilverStor TN-6224S-FFG (x26)
  - Single controller
  - 3.5” SATA -II 1.5 TB 7,200R/16MB * 24pcs
  - Cache RAM : 1G DDR
  - RAID6 + 1 hot spare
  - 1 Disk array: 31.5TB, Total capacity : ~820TB

- Nexsan SATABeast (x16)
  - Single controller
  - 3.5” SATA -II 1TB 7,200R/32MB * 36pcs
  - Cache RAM : 1G DDR
  - RAID6 + 1 hot spare
  - 1 Disk array: 33TB, Total capacity :~528 TB

Atlas & CMS Storage Classes

storage class	disk servers	Disk(TB)	Tape(TB)	Job Slot
atlasPrdD0T1	2	30.01	117.19	200
atlasPrdD1T0	15	278.12		810
atlasGROUPDISK	1	19.09		40
atlasMCDISK	5	95.43		500
atlasMCTAPE	2	38.17	39.06	80
Total	25	460.82	156.25	1630

storage class	disk servers	Disk(TB)	Tape(TB)	Job Slot
cmsLTD0T1	9	139.11	454.3 7.81 3.91	495
cmsPrdD1T0	7	119.31		455
cmsWANOUT	4	72.26		220
Total	20	330.68	466.02	1170

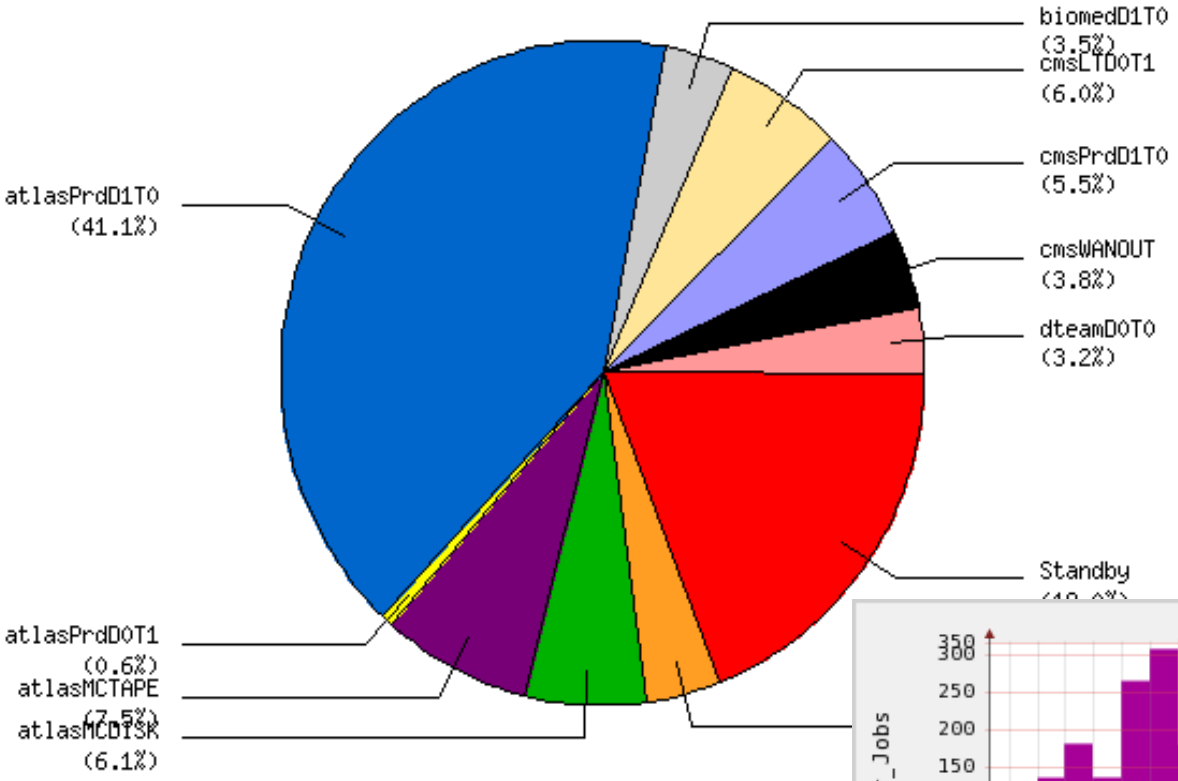
Biomed, dteam and Standby disk

storage class	disk servers	Disk(TB)	Tape(TB)	Job Slot
dteamD0T0	1	19		45
biomedD1T0	1	19		100
Standby	7	110		
Total	9	148	0	145

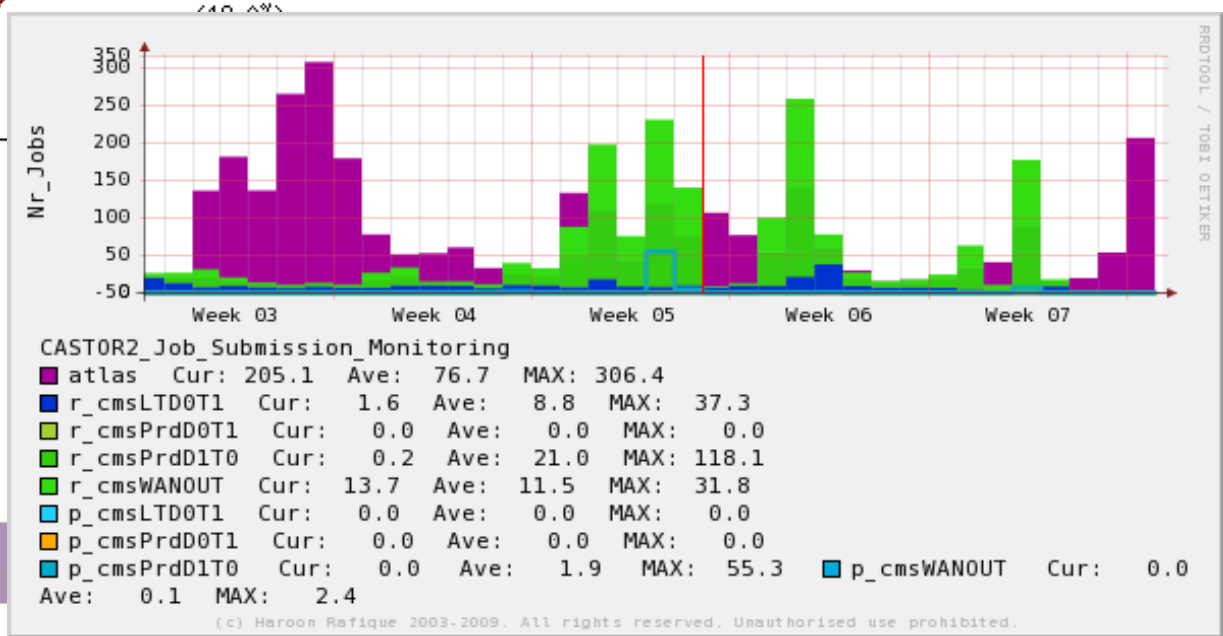
Monitoring (1)

- Nagios
  - Plug-in to help monitoring disk servers status and service availability (gridftp , SRM).
  - E-mail , SMS notification
- Ganglia
- RRD base monitoring
- C2probe
  - Availability monitoring
- TLS (Tape logging System)
  - Tape accounting

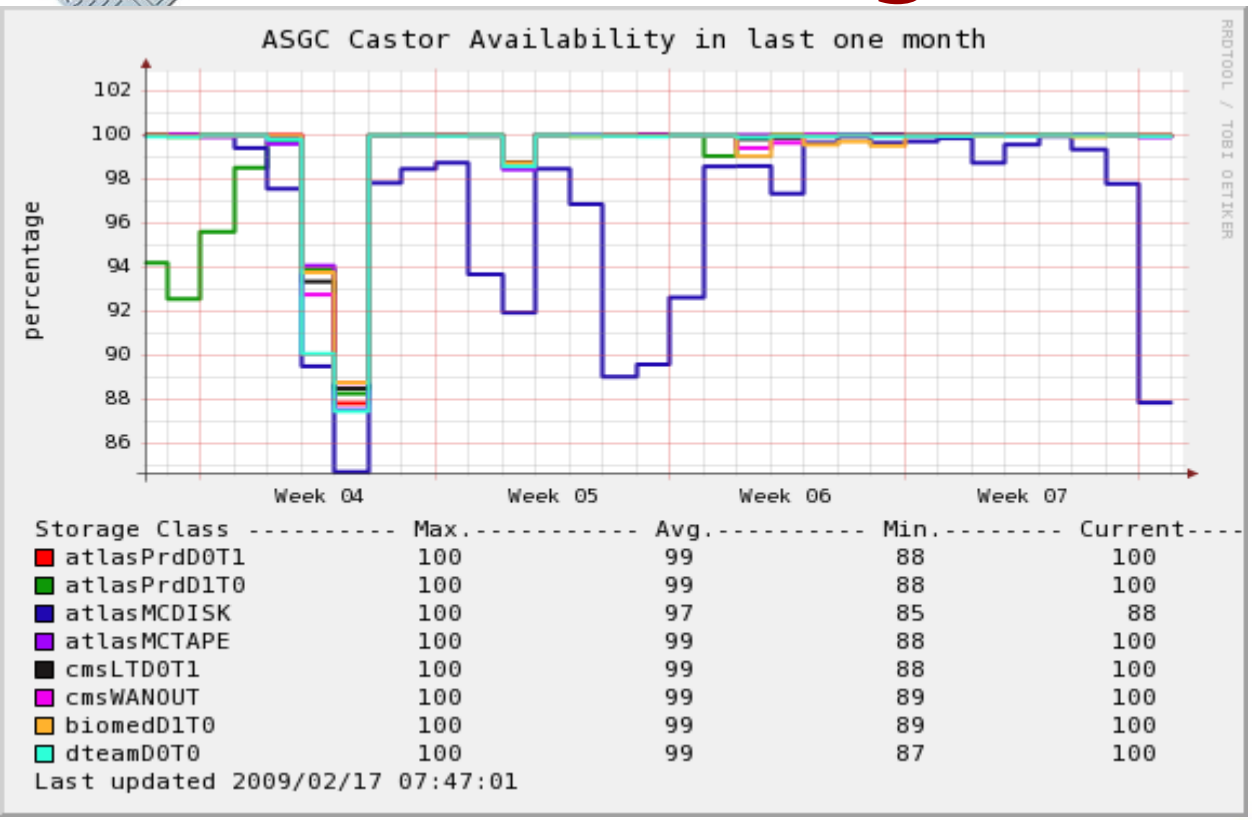
Monitoring (2) – RRD



- Show statistics of Disk/ Tape usage
- LSF jobs monitoring



Monitoring (3) – c2probe



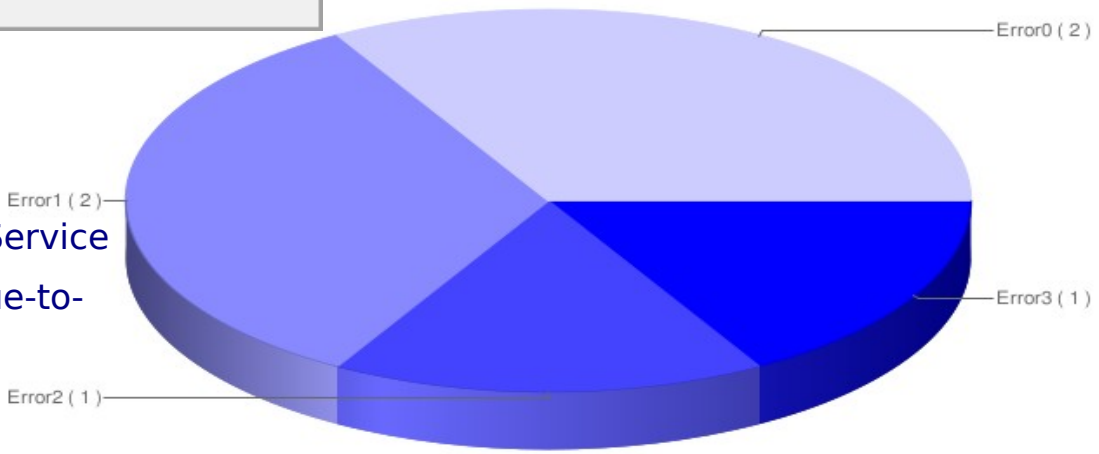
Storage Class	Max.	Avg.	Min.	Current
atlasPrdD0T1	100	99	88	100
atlasPrdD1T0	100	99	88	100
atlasMCDISK	100	97	85	88
atlasMCTAPE	100	99	88	100
cmsLTD0T1	100	99	88	100
cmsWANOUT	100	99	89	100
biomedD1T0	100	99	89	100
dteamD0T0	100	99	87	100

Last updated 2009/02/17 07:47:01

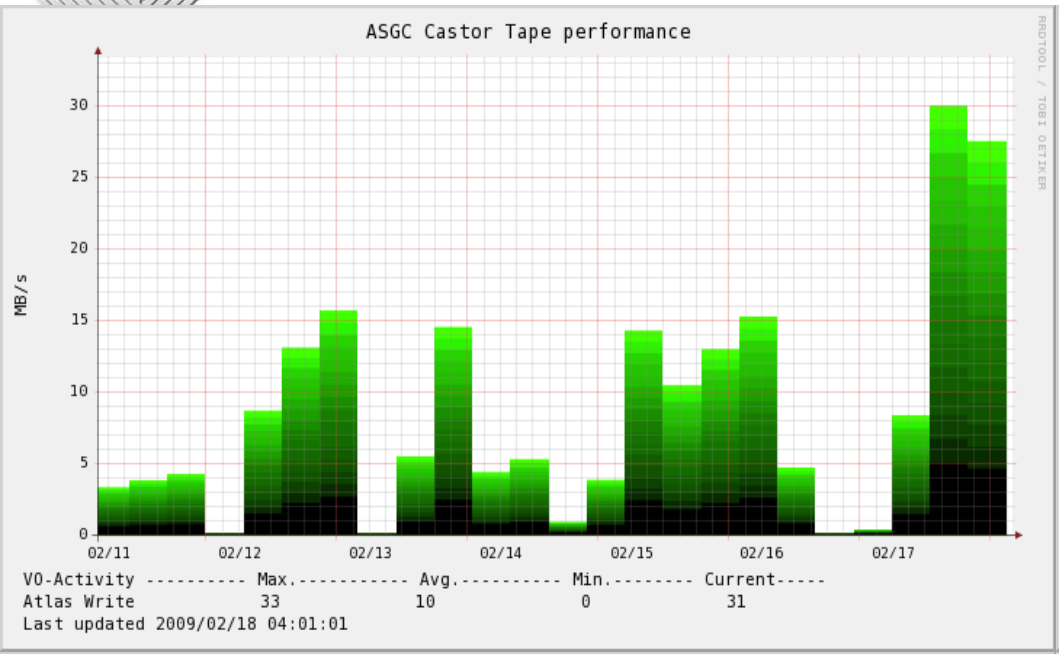
- Availability plot

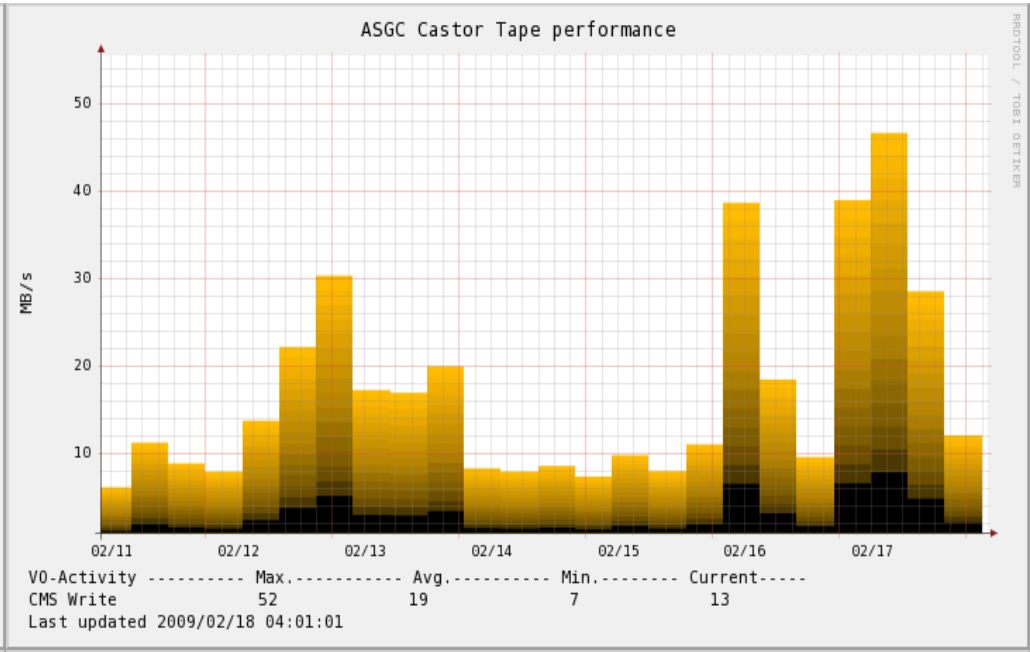
- Error Pie chart

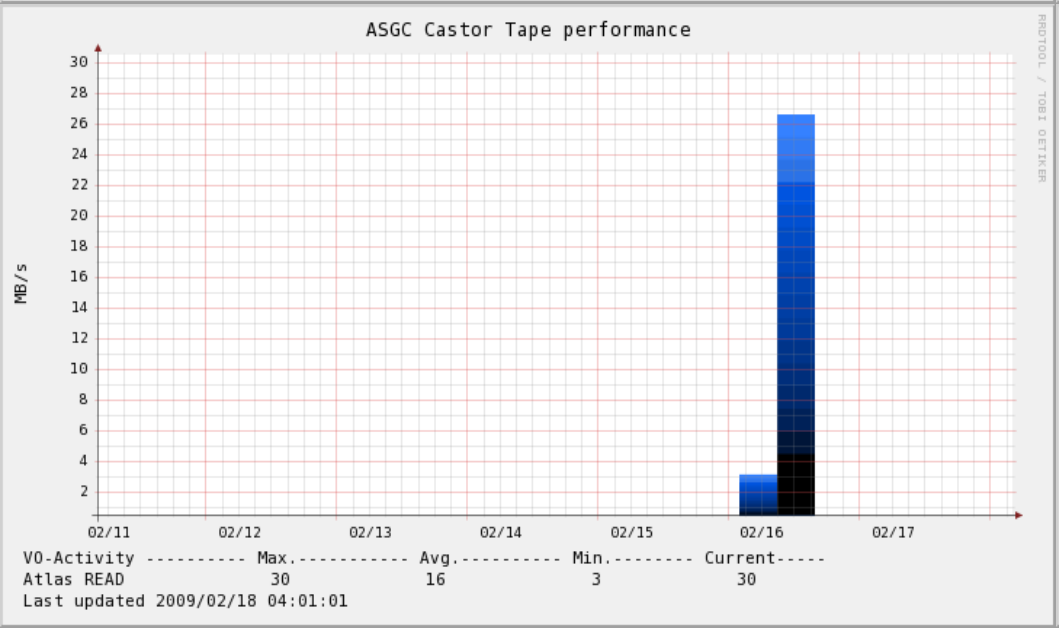
Error (count)	Message
Error0 (2)	job: error
Error1 (2)	migrator: Copy-failed
Error2 (1)	Stager: Impossible-to-get-the-Service
Error3 (1)	JobManager: Job-terminated-due-to-scheduling-error

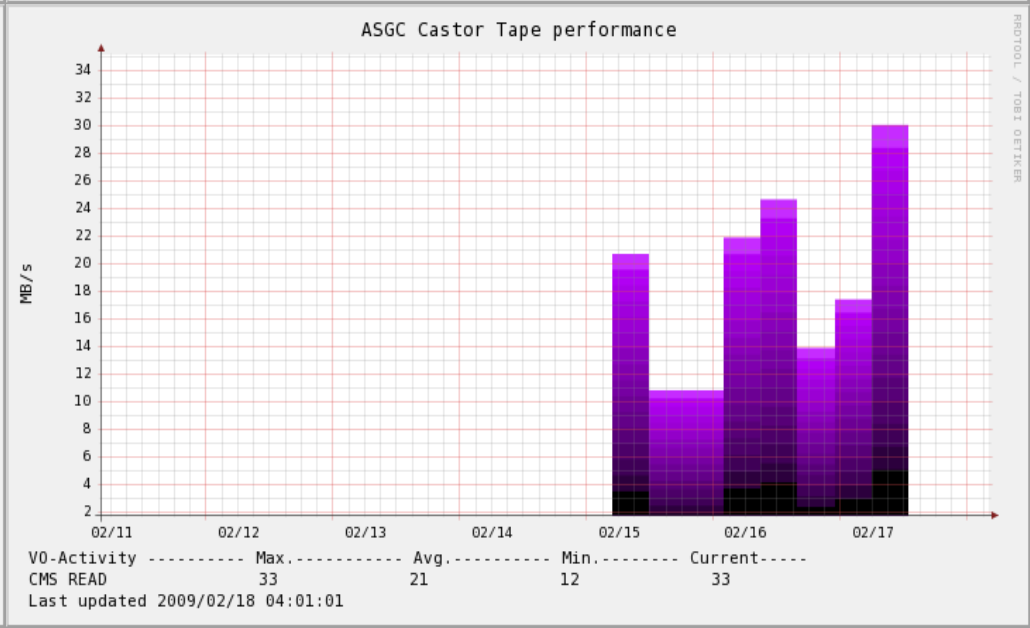


Monitoring (4) – TLS (Tape Logging System)









Issue / Problem / Event (1)

- One Blade chassis lost the input power from sockets. Castor was not function since all core services are located in the this chassis !!

- Consider migrating part of the core services into different chassis? Or have second core services on different location to reduce the risk.

Issue / Problem / Event (2)

- Oracle RAC stability affect Castor core service
  - Data block corruption
  - One RAC node unstable
  - ORA-00600 error INTERNAL ERROR CODE, ARGUMENTS:, [KTSPFREDO-4]
  - etc...

Issue / Problem / Event (3)

- "Big ids" problem have found in both Stager and SRM databases. e.g:

```
SQL> select count(*) from id2type where
id>10E18;


COUNT(*)
----------
65691
```

- "No type found for id"
  - The daemon is keep trying to process the request in status=0 which don't have Id2Type.
  - Fixing all missing id and removing old request.
- Looking forward the long term solution.

Plans for 2009

- Schedule downtime in March
  - Castor upgrade to 2.1.7-24
  - Disk array firmware upgrade
- 4 LTO4 tape drive become production service.
- Improving Tape accounting/monitoring
- Setup a dedicated disk pool for Repack activity.

Plans for 2009 (2)

- High Availability and Load Balancing Core service (April)
  - Needs second Stager, NS, LSF on another Blade chassis.
  - Attach different power, network...
- Oracle RAC needs consolidation
  - One more RAC node
  - Moving datafile to ASM from OCFS2?
- New disk servers online