CERN site report

CASTOR F2F'09 @ RAL

Olof Barring

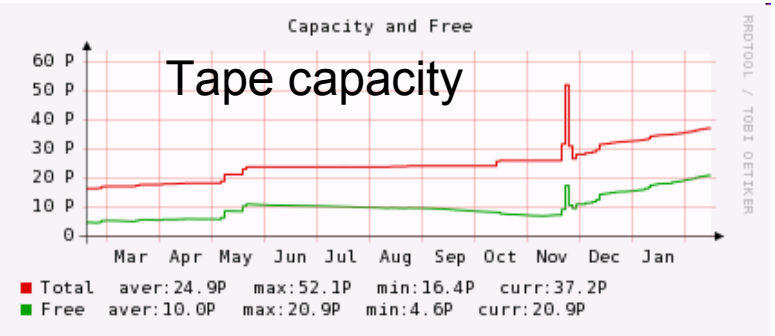
Summary

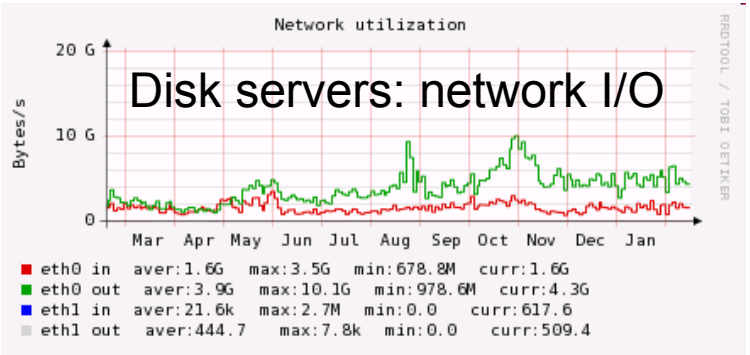
- CASTOR operations at CERN
  - CASTOR@CERN in numbers
  - Production setups
  - Staffing
- Ongoing activities
  - Preparations for 2.1.8 deployment
  - Monitoring: wLCG disk cache metrics
  - Disk server h/w retirements
- Common incidents
  - Hotspots
  - Networks
  - Changes

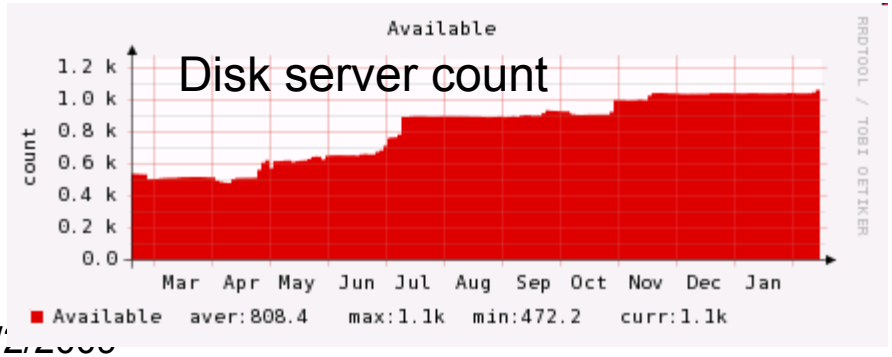
CASTOR in numbers

- Not much changed since Miguel's talk at the Taiwan workshop in October

5 tape robots, 130 drives (soon all 1TB) 37 PB total, 20 PB free
1050 disk servers, 30K disks, 6 PB
116M files in name space, 16M copies on disk



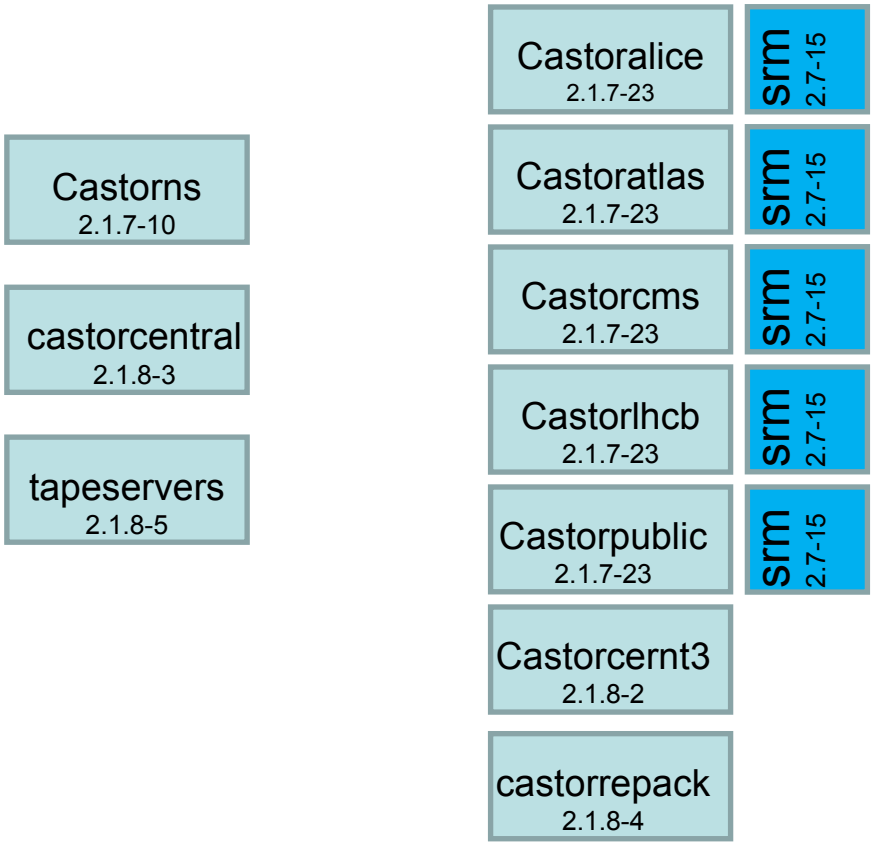




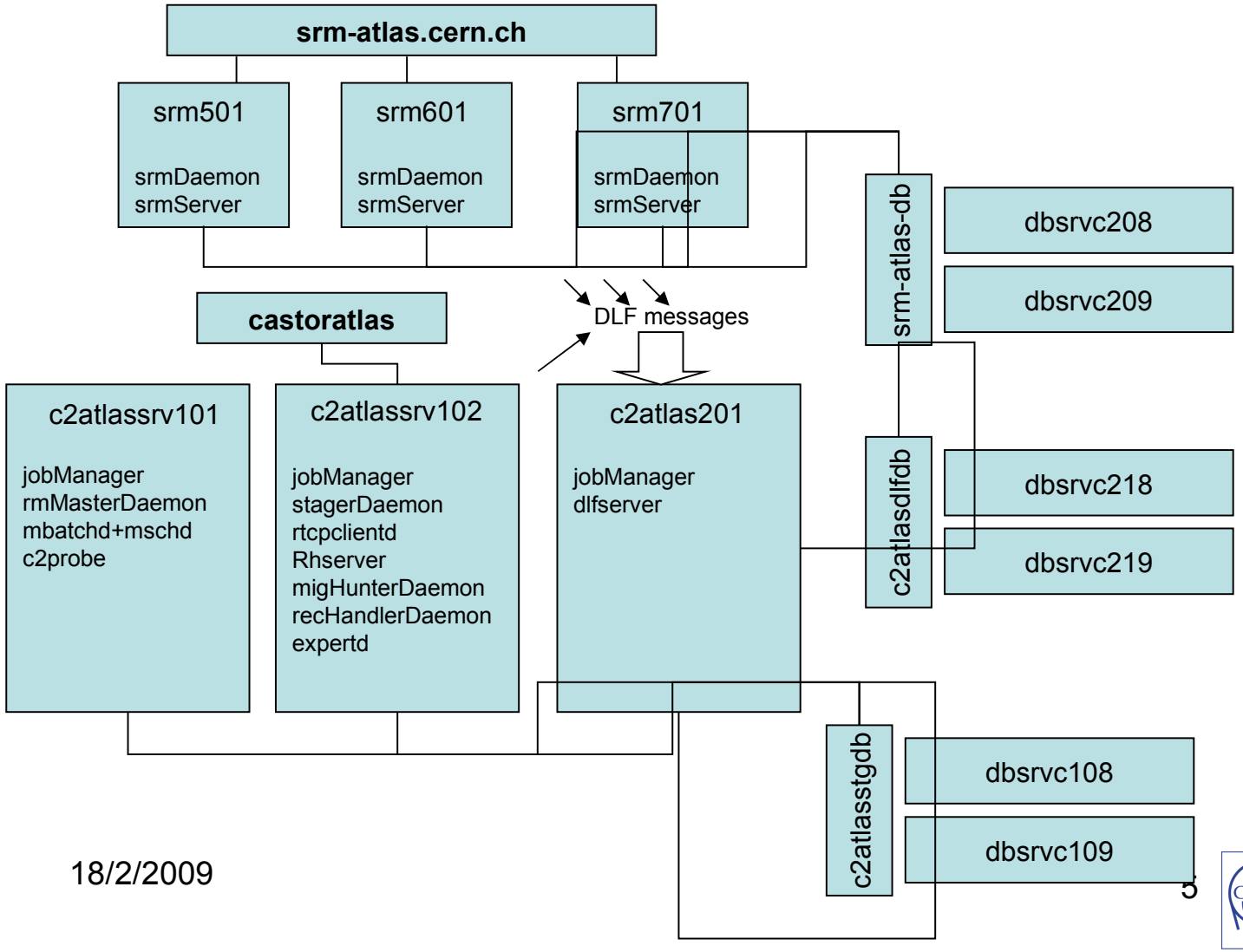
CERN - IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it

CASTOR Setup at CERN

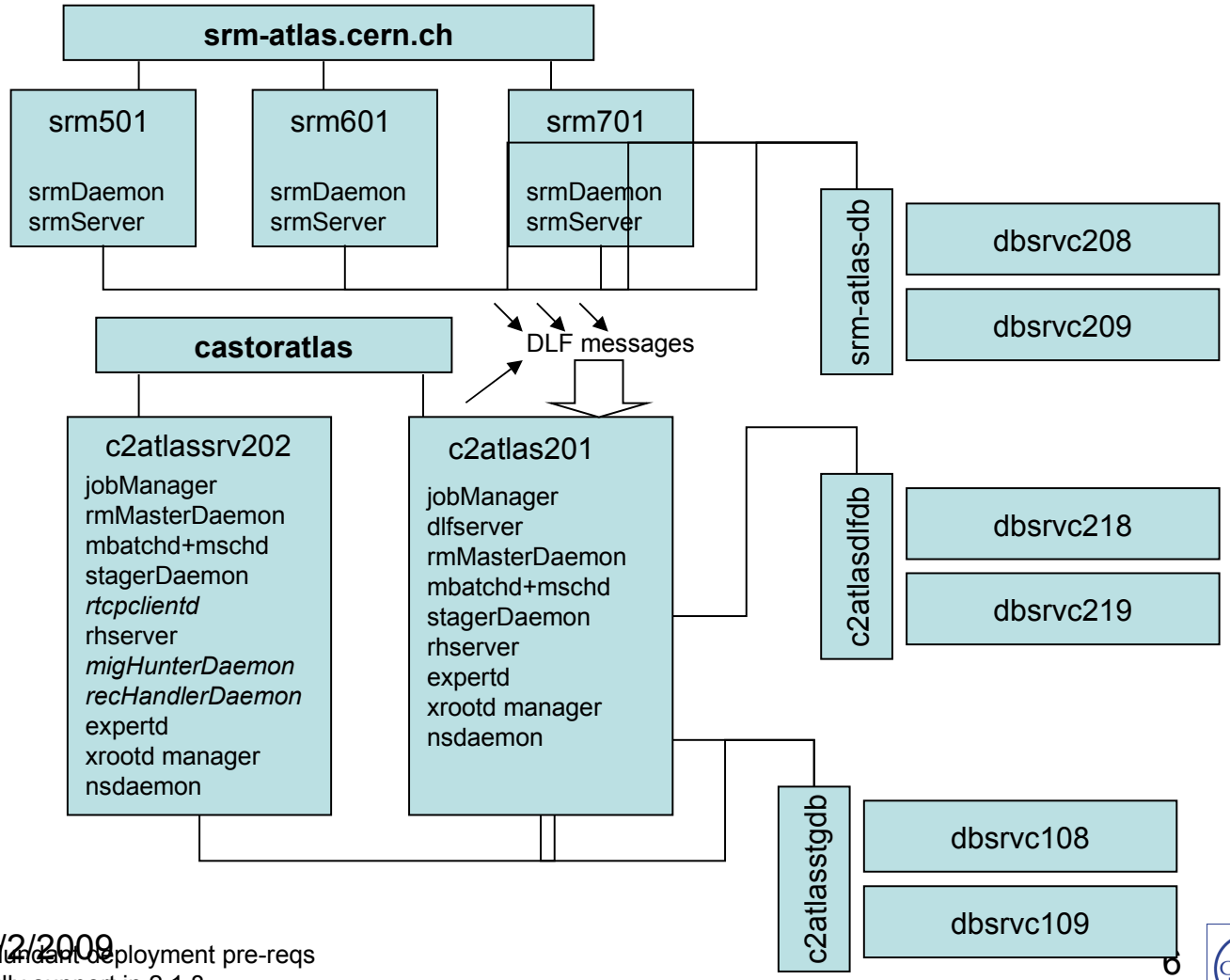
- Production setup 17/2/2009



Service deployment, today



Service deployment, tomorrow



Redundant deployment pre-reqs
- Fully support in 2.1.8
- No pre-2.1.7 clients

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

Service deployment

- SRM
  - 1.3 release was a deployment headache
    - Old castor client (2.1.3) prevented us from going ahead with redundant deployment of stager
    - Backend daemon could not be loadbalanced
    - Trusted rfio access for passing credential delegation
    - External gridftp
  - 2.7 - what a relief! ☺
    - Deployed just before Christmas
    - SRM now loadbalanced, redundant, native gridftp (except LHCb) and secure(r)
    - We can start the loadbalanced deployment of stager
    - Logging is better but there are still some room for improvements
- CASTOR
  - 2.1.7 has been running stable for 6 months but…
  - 2.1.8 features are eagerly awaited
    - Full xrootd support: pressure from VOs (particular ALICE but also the others)
    - User and group accounting: pressure from ATLAS and CMS
    - Redundant deployment of (almost) all daemons → reduced vulnerability to network 'glitches', which were a recurring problem in 2008
    - Automated, robust and natively supported interfaces for draining diskservers → industrial scale server retirements

Staffing

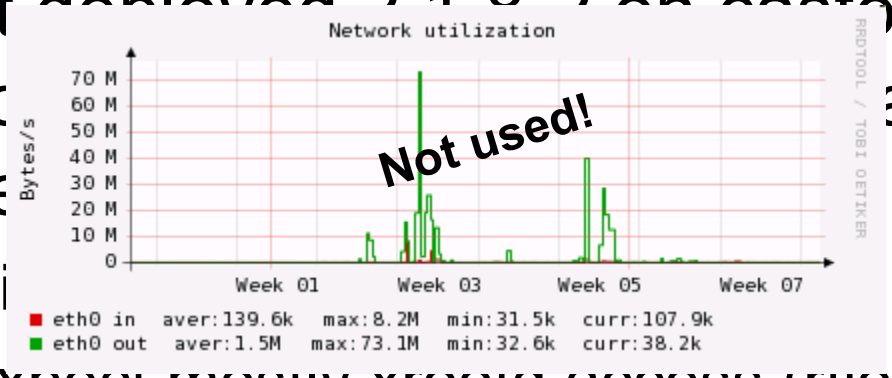
- CASTOR2 stager + SRM operation
  - ~3 FTEs spread over Ignacio, Jan, Miguel + myself
  - ~2FTEs from 1/3/2009 when Jan is moving to new responsibilities (leading the linux support team)
- castor-external-op meeting chair
  - Taken over by Ignacio
  - Many thanks to Jan who has chaired the meeting for the last year and organised this workshop!
- Also staff reductions in the tape team: 4 → 2FTEs (see Tim's talk)

Ongoing activities: 2.1.8

- First deployed 2.1.8-2 on castorcernt3 in October 2008, to meet the “Analysis” use-case
  - Disk-only (no tape) access
  - Expect mostly xrootd access (rfio/rootd allowed but discouraged)
- Agreed with developers and tape op to exercise the tape side on castorrepack
  - 2.1.8-3 in November
  - 2.1.8-4 just before Xmas

Ongoing activities: 2.1.8

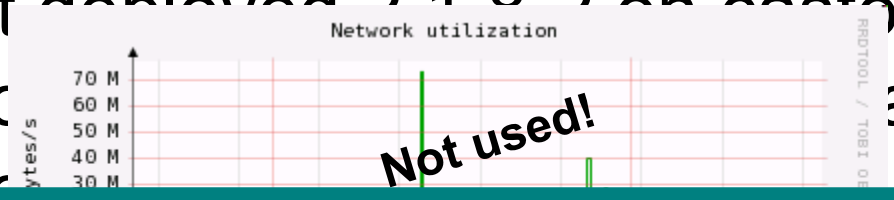
- First deployed 2.1.8-2 on castorcernt3 in Octo… …alysis" use-case
  - Di…
  - Expect mostly xrootd access (rfio/rootd allowed but discouraged)



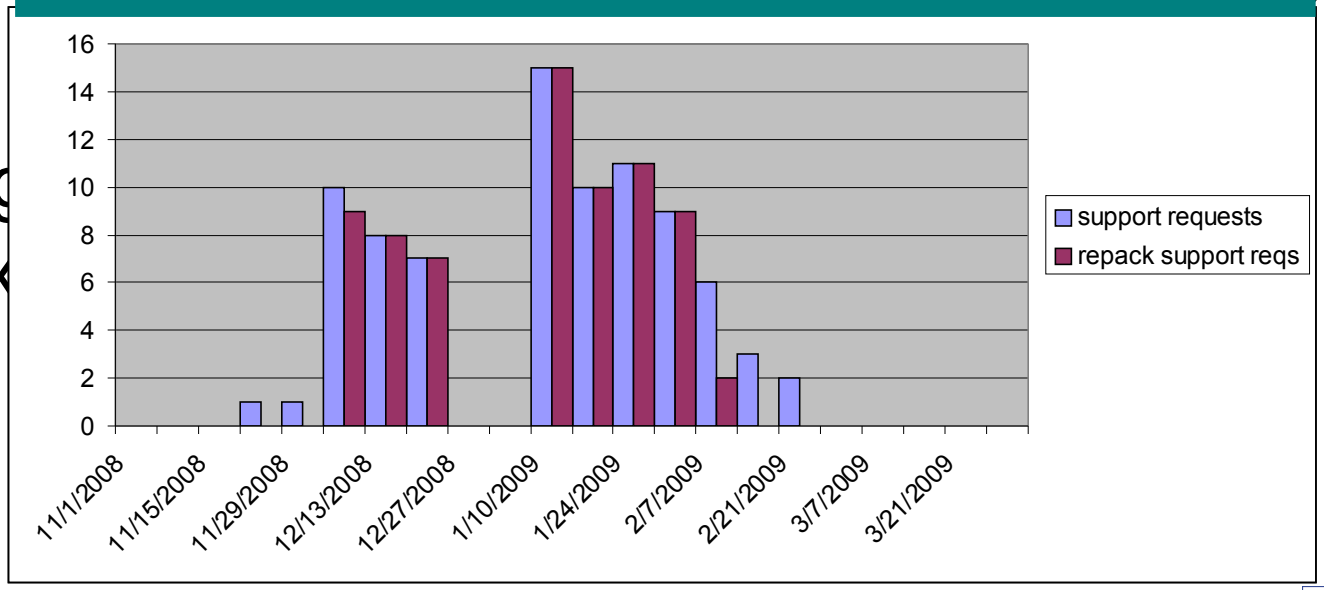
- Agreed with developers and tape op to exercise the tape side on castorrepack
  - 2.1.8-3 in November
  - 2.1.8-4 just before Xmas

Ongoing activities: 2.1.8

- First deployed 2.1.8-2 on castorcernt3 in Octo… …alysis" use-cas…
  - …
  - …
- Ag… ex…
  - …
  - …



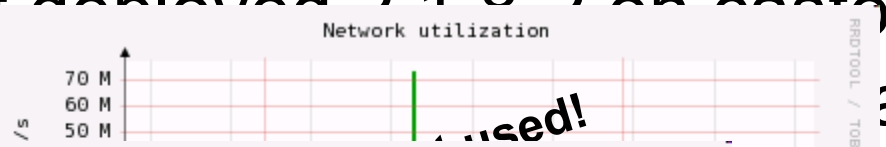
Use of savannah support requests has allowed for a rapid fix turnover while keeping the release and deployment process under control.

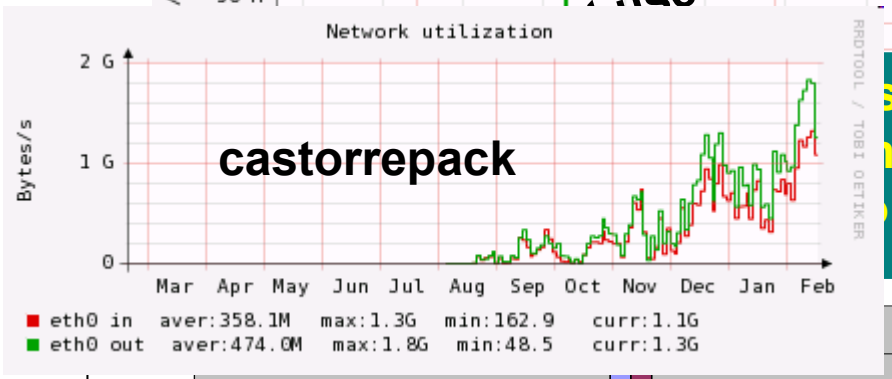


CERN - IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it

Ongoing activities: 2.1.8

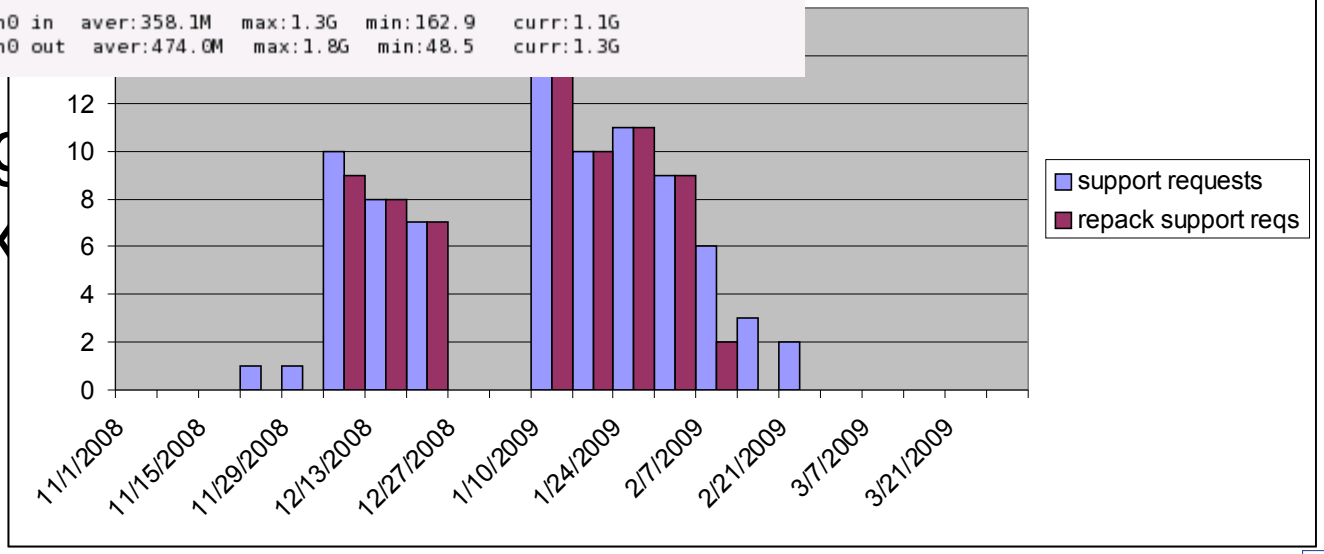
- First deployed 2.1.8-2 on castorcernt3 in Octo… …alysis" use-





…s has allowed for a …e release and …l.

- Ag… ex…



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

Monitoring: wLCG disk cache metrics

- Common disk cache metrics proposed to wLCG MB a year ago
- Storage system independent metrics
- Three main categories (http://indico.cern.ch/conferenceDisplay.py?confId=27421)
  - Usage & Usage Patterns
    - Average file size
    - Average data transferred per tape mount
    - Average daily mounts per tape volume
    - Requests/second
    - Percentage of requests for non-disk resident files
  - Performance
    - Network utilisation per pool
      - where relevant, with comparison to target
    - # simultaneous active transfers
      - with breakdown by read/write and access mode (streaming/random)
    - Average I/O performance of tape system
  - Responsiveness
    - Stager
      - Time between request and availability of first byte
        - » Distinguish cases where file in requested pool, in another pool, on tape
      - Queued transfer requests
      - Average & maximum disk server busy time
    - SRM
      - Time between initial request and return of TURL
    - Tape Layer
      - Time between CASTOR request and drive allocation

Monitoring for operation

- Service related metrics are useful if they can be correlated with system metrics, e.g.
  - File open time vs CPU load or IOwait (hot servers)
  - Failed request rate vs network timeouts
  - Average transfer slot occupancy vs CPU load
- CERN uses Lemon for system metrics other sites use Nagios or other systems
- See more details in Miguel's talk

CluMan - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://cluman.web.cern.ch/cluman/#p|c2cms|d|state

Most Visited | AIS Applications Home... | CASTOR (Cern Advan... | CERN -- Administrativ... | Communications Syste... | Home | LES PAGES BLANCHES | S-E-Banken Internetk... | Storage System Stand... | Stugor på Västkusten! | CASTORLHCB - Distrib...

PostMortemMay24 < FIOgroup < TWiki | CASTOR - Support: Browse Items [LCG... | CluMan | LCG-LHCC Mini Review (16 February 20... | Agnosticism - Wikipedia, the free encyc... | LCG-LHCC Referees Meeting (19 Febru...

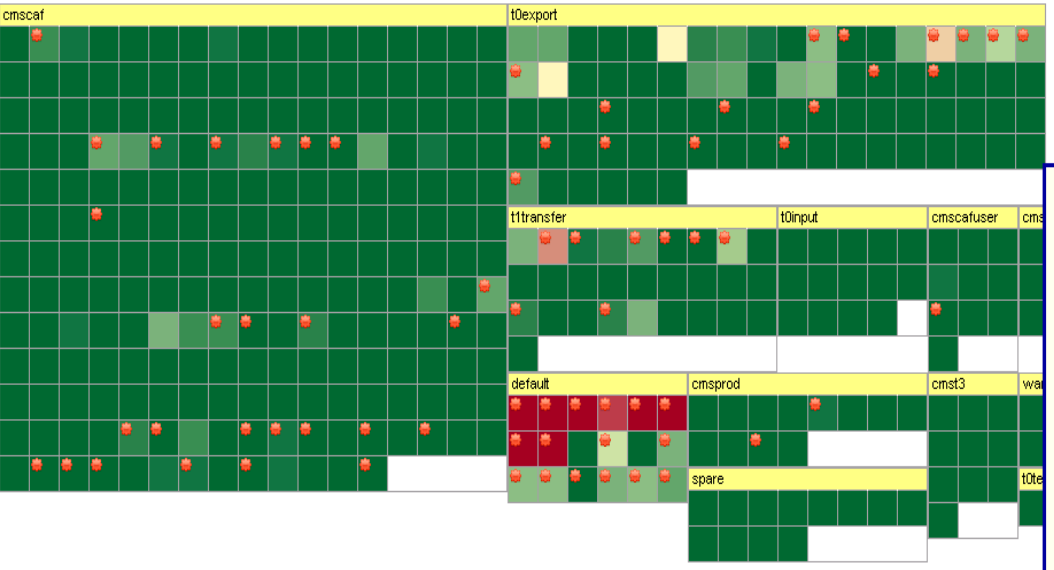
Project | Refresh | Reload

Project *c2cms*

Fetched 417 entities. Took 11.413 seconds.

clear

Map View | Tree View



Color cells by:

- state
- CPU Load
- IO wait
- Slot usage
- rfio network timeouts
- Migrator
- Recaller Streams
- temperature
- er of users
- e in seconds
- access warning
- g kernel
- U load
- duction

Name: **lxfsrl1804**
Type: **host**

CPU Load: **.17**
Recaller Streams: **0**
state: **production**
CPU Model: **Intel(R)_Xeon(R)_CPU_____E5405__@_2.00GHz**
Uptime in seconds: **5919032**
rfio network timeouts: **0**
CPU load: **.17**
Network traffic (in): **0**
Slot usage: **30**
Migrator: **0**
Network traffic (out): **0**
State: **production**
CPU temperature: **26.5**
Kernel: **2.6.9-78.0.1.EL.cernsmp**
Number of users: **3**
IO wait: **.01**

CluMan - the cluster management tool - developed by CERN IT/FIO (more)

released: 10:50, Thursday 04 Dec 2008

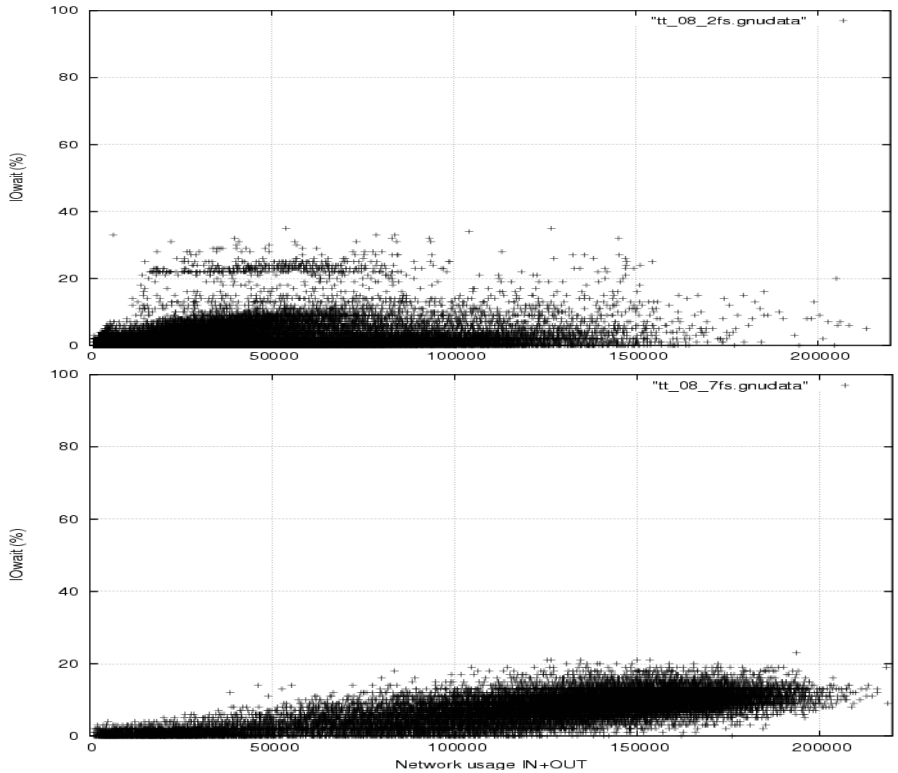
Done

start | presentations | C2f2f_CERN_sit... | Tier0Status_20... | CastorMetrics4L... | CluMan - Mozilla... | Downloads | sr | Inbox - Microso... | RE: disk cache ... | CASTOR metric... | 10:53 PM

Disk cache hardware retirements

- CERN procures with 3 years warranty
  - 1050 servers → average 1 server runs out of warranty every day
  - Draining mode depends on service class
    - Disk1Tape0: full draining - all data must be copied
    - DiskXTape1: only unique copies of CANBEMIGR files must be copied
  - At least 50% of our capacity is Tape0! → one server every two days require full draining
  - 1 server holds at least 5TB of data → 12hrs for full drain at GigE speed
  - Average ~0.25 (12hrs/24hrs * 50% * 1 server/day) full draining is running all the time!
- In reality because we procure in bulk we also retire in bulk and the activity is pretty spiky
- Process must be planned and automated
  - Wheel in replacement well in time: requires that the replacement exists
  - Put old server in standby (draining) and launch the active draining: major exercise in 2.1.7 but there is hope for 2.1.8

Common incidents: hotspots



- Signature: server goes into high IO wait and delivers (almost) no data
- Can only recover by killing off all RUNning requests (i.e. set server state to DISABLED)
- Observed correlations to specific RAID configurations
- See more details in Ignacio's talk

Common incidents: network

- Problems usually detected as service degradation (SAM tests, c2probe)
- Can be difficult to distinguish from other potential problem sources
  - Service changes
  - Database problems
- Redundant routing makes it difficult to localize the problem

- Preventive/debugging measures
  - Monitor application logs (e.g. rfio) for network timeouts or dropped oracle sessions
  - Monitor packet losses or retransmits
  - Be prepared to run tcpdump with increasing bandwidth (-b)

Common incidents: changes

- It's working, don't touch it!
- Changes are normal
  - Security, Linux upgrades
  - Configuration updates
  - CASTOR releases
  - SRM
- Changes has been a common source for incidents
  - A strict and disciplined and certification process is the only remedy
- But can't test everything
  - Every major upgrade ideally should come with a rollback plan
- Recent examples:
  - SRM mkdir on-the-fly spec change

Incidents

Date (of incident)	Brief description
26/01/2009	Backwards incompatible change on SRM affected ATLAS / LHCb
23/01/2009	FTS/SRM/CASTOR transfer problems for Atlas
17/01/2009	FTS transfer problems for Atlas
8/1/2009	many user jobs killed on lxbatch due to memory problems
5/12/2008	srm-atlas degraded due to BringOnline? requests
25/11/2008	Batch service downtime following cooling incident
20/11/2008	CASTORATLAS degradation for 1.5 hours
17/11/2008	Database connections blocked due to IPMI reconfiguration
11/11/2008	WN misconfiguration due to bug in yaim module affecting the python path settings
24/10/2008	FTS service degradation
14/10/2008	unfair public fair share for LHCb
12/10/2008	CASTORLHCB degradation
8/10/2008	Silent data corruption in CASTORCMS/CMSCAF
10/9/2008	ATLAS transfers/access of first beam data stuck
28/08/2008	SRM CMS post-mortem - 28 August 2008
25/08/2008	Network post mortem - Friday 22 through Monday 25 August 2008
29/07/2008	FTS-T0-EXPORT post mortem - 28 and 29 July 2008
1/7/2008	Problems with myproxy service on px201
5/6/2008	Post-mortem of 1.3-26 problems
24/05/2008	Postmortem of CERN-PROD SRM v22 blockage on Saturday morning 24/5/2008
9/5/2008	Problem accessing data in CASTOR @ CERN
14/02/2008	Postmortem of Tier-0 transfer problems Feb 14 - 18 2008