



# RAL CASTOR Update

CASTOR delta Review

6-7 Dec 2006

Shaun de Witt

Bonny Strong, Tim Folkes, and Chris Kruk

Bonny Strong  
RAL

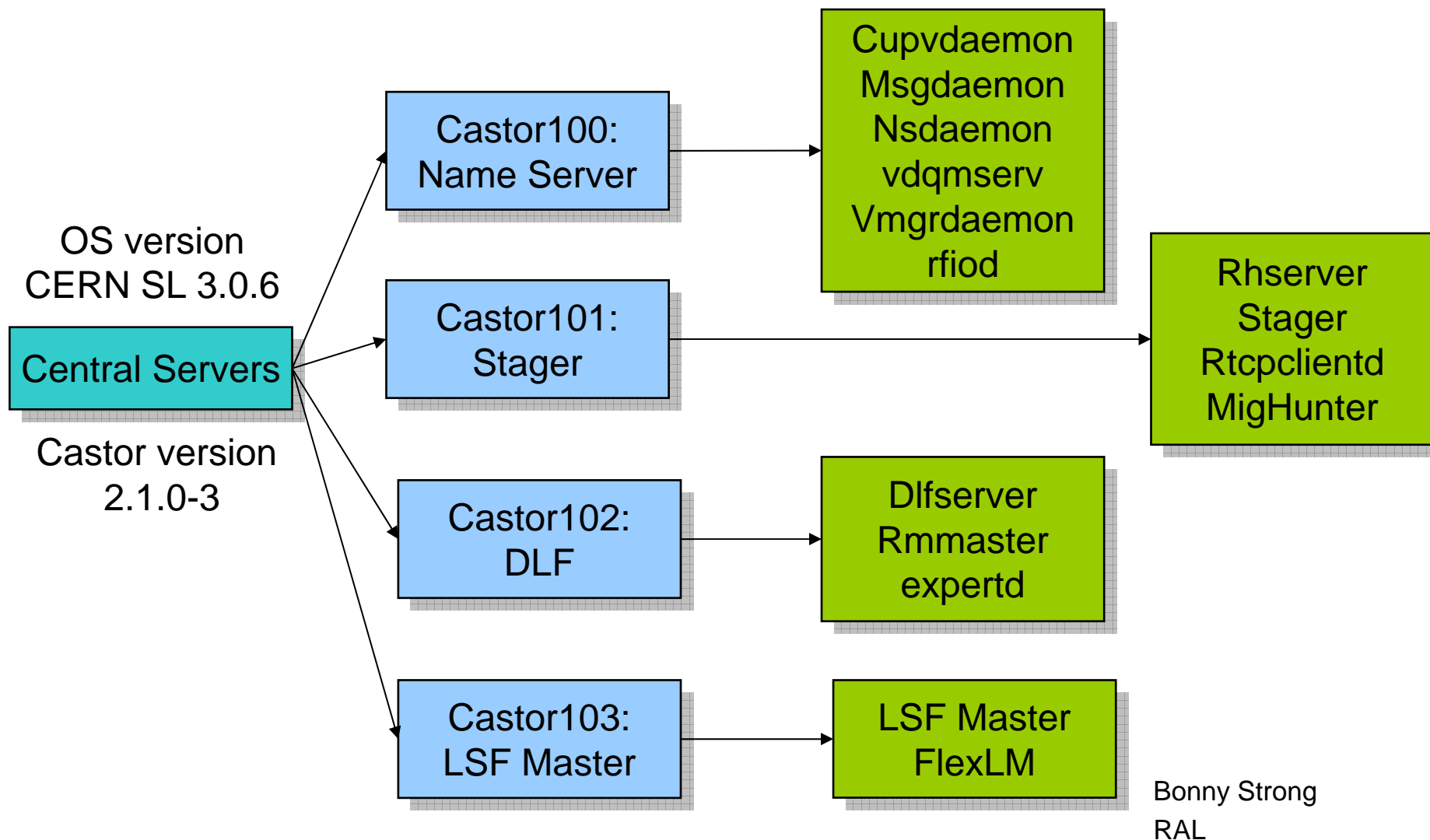
1. Current Status
2. Tape Systems: Status/Issues
3. Experience with CMS CSA06
4. Developing Monitoring at RAL
5. Issues/Problems
6. Goals for Next 6 Months

- Production platform

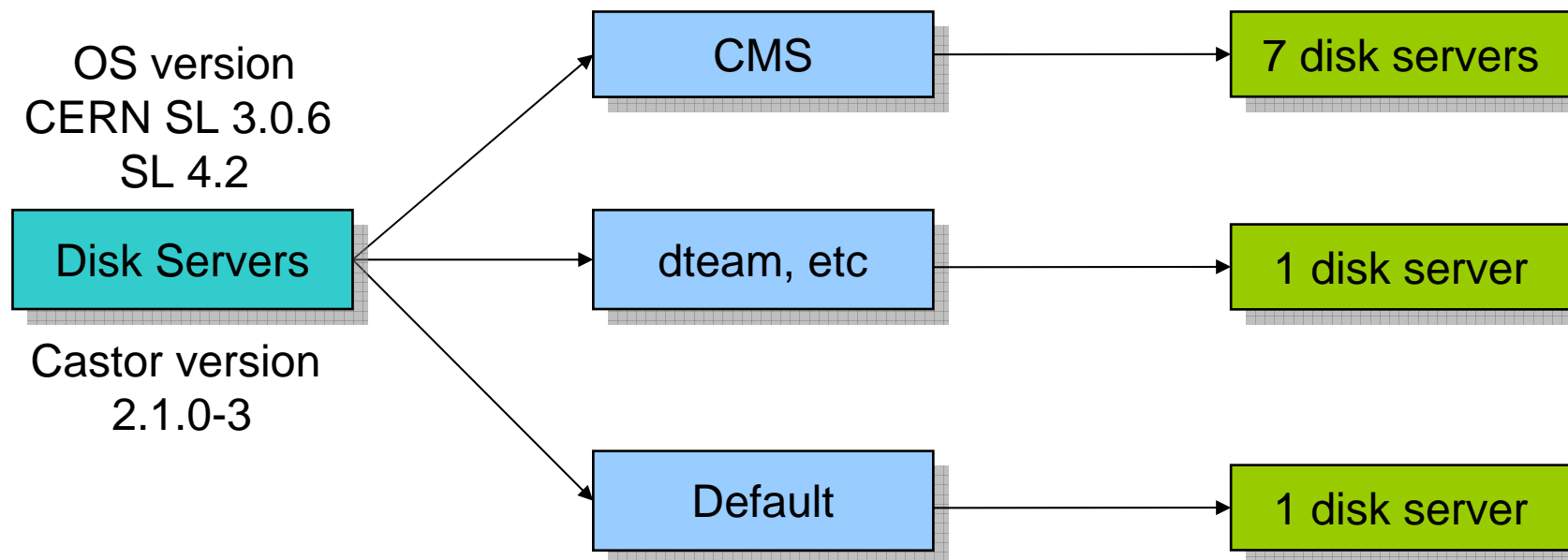
Currently supports only:

- CMS for CSA06
  - dteam, ops,alice,lhcb
  - default, for testing
- Test platform

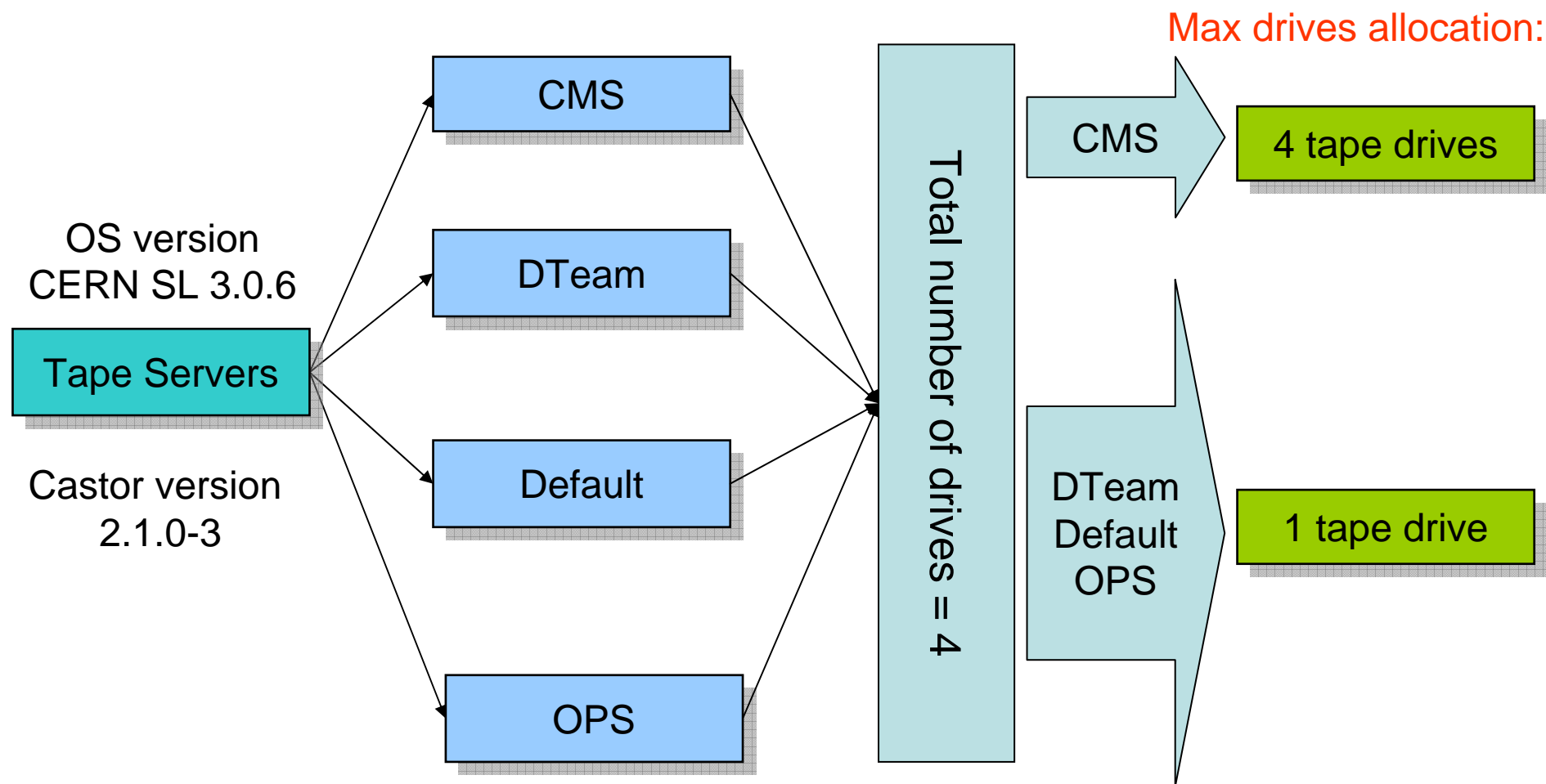
# Current pre-production Castor infrastructure (1/4):



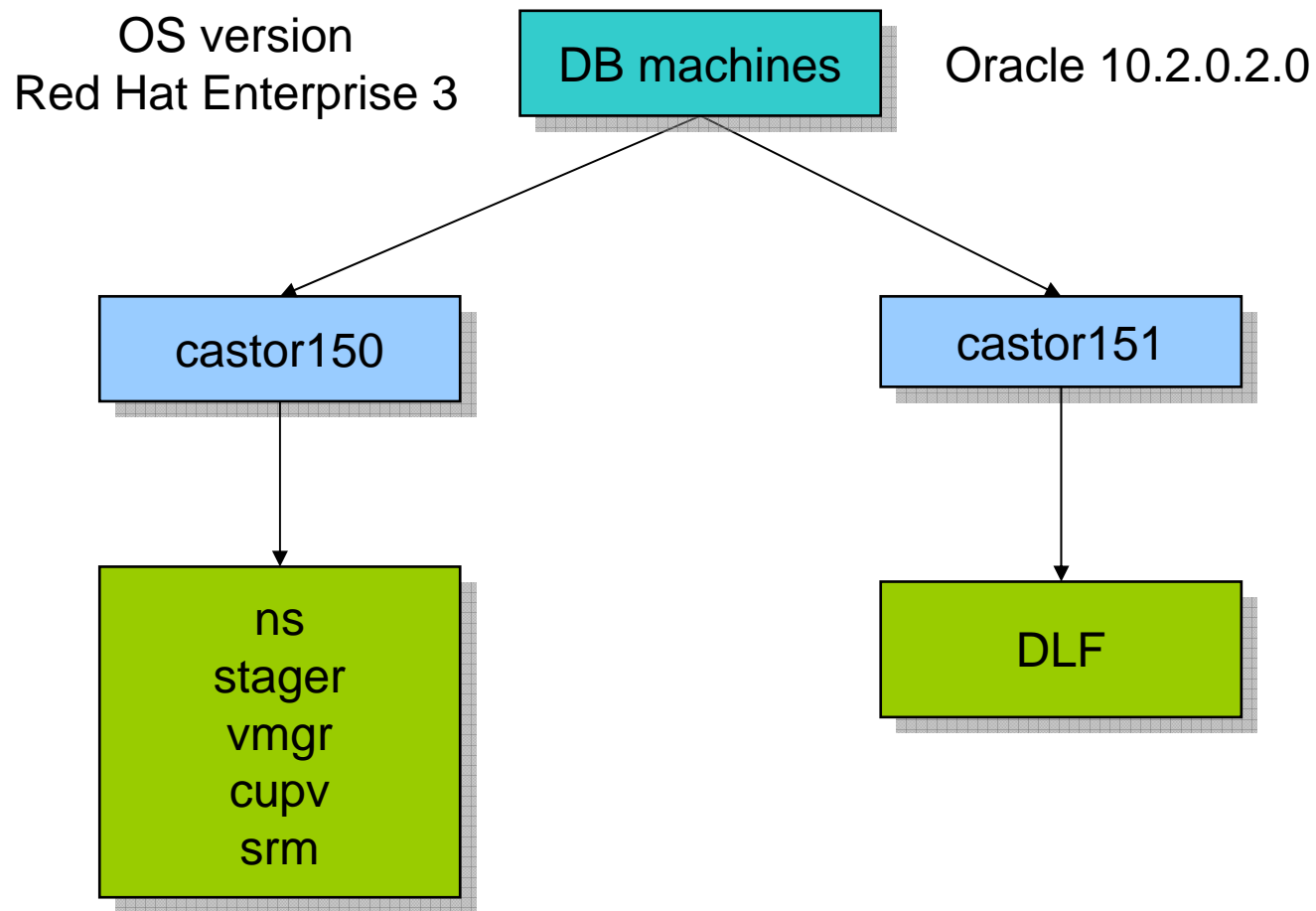
## Current pre-production Castor infrastructure (2/4):



# Current pre-production Castor infrastructure (3/4):



# Current pre-production Castor infrastructure (4/4):



## Pre-production platform:

- Pre-production - fully functional platform where number of disk servers will be increase soon.

	c/s	d/s	t/s	DB	SRM
in used:	4	10	4	2	1
Not in use yet:	0	0	6	0	2





# Current test Castor infrastructure (1/2):

OS version  
CERN SL 3.0.6

Castor version  
2.1.1-[4|9]

## Central Servers

Tcastor100:  
NameServer  
Stager

Tcastor102:  
DLF  
LSF

rfiodCupvdaemon  
Msgdaemon  
Nsdaemon  
Vdqmserv  
Vmgrdaemon  
Rfiod  
Rhserver  
Stager  
Rtcpclientd  
MigHunter

dlfserver  
Rmmaster  
expertd

## DB machines

OS version  
Red Hat Enterprise 3

Oracle 10.2.0.2.0

lcg0610

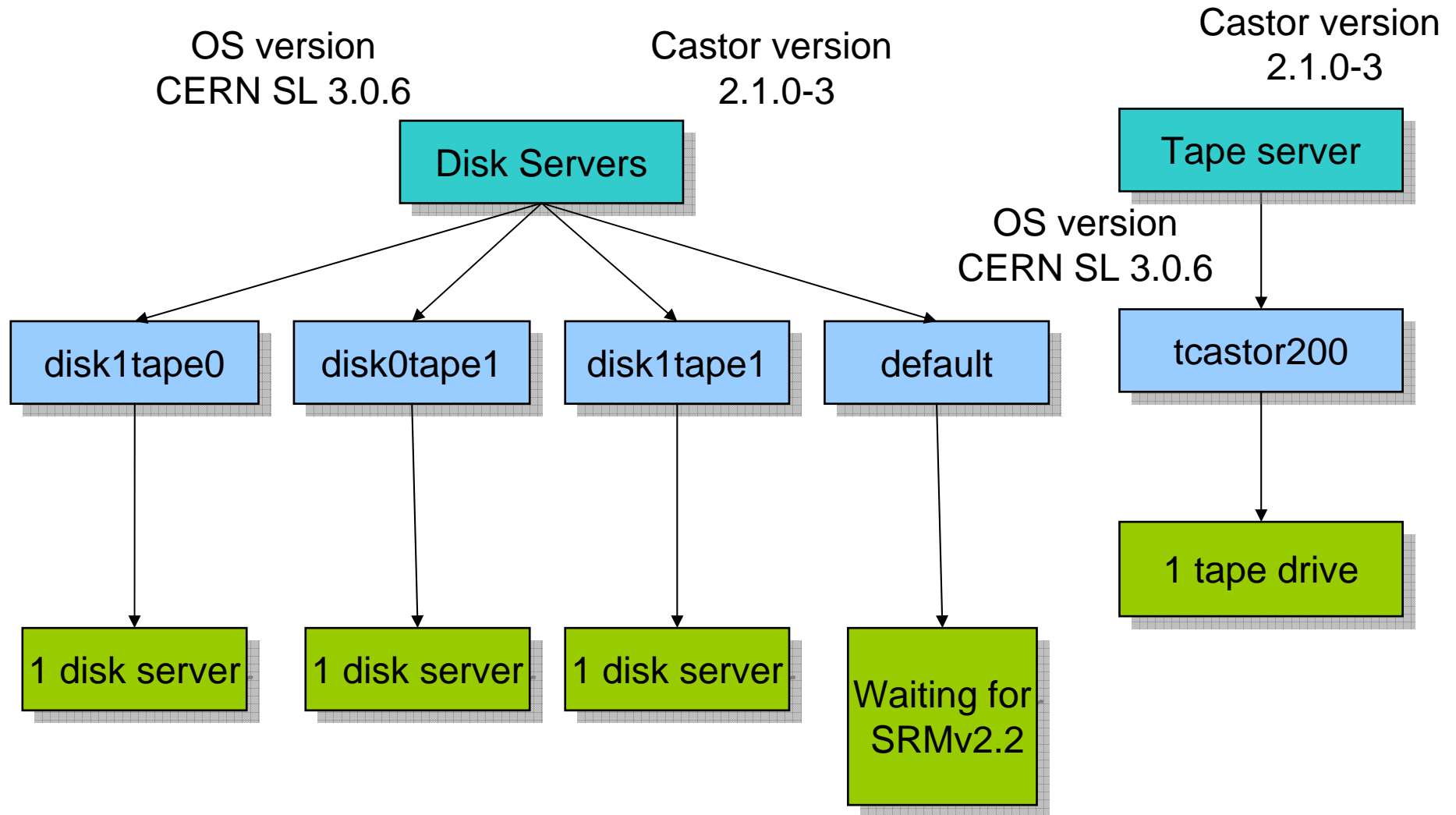
lcg0611

ns  
stager  
vmgr  
cupv  
srm

DLF

Bonny Strong  
RAL

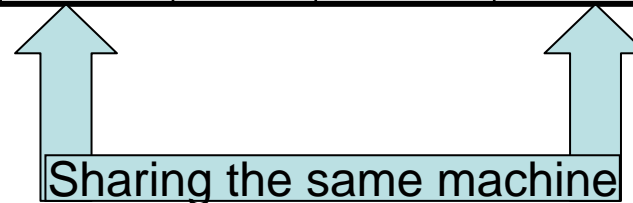
## Current test Castor infrastructure (2/2):



## Test platform:

- Test – still under construction and configuration.

	c/s	d/s	t/s	DB	SRM
in used:	2	3	1	2	1
Not in used yet:	0	1	0	0	0



## **2. Tape Systems: Status and Issues**

Bonny Strong  
RAL

---

- 5 STK T10000 tape drives dedicated to castor
- 10 tape servers, 5 in use.
  - 4GB memory per server (enough?)

- Doing non-disk IO ~120MB/sec
- Disk IO limited to disk speeds
  - Tape IO depends on disk activity
    - No disk activity can get ~100MB/sec
    - With disk activity ~15-20MB/sec

- No re-pack (yet). Affects our resourcing
- /var/tmp/RTCOPY vanishes.
  - We do not run the script to remove old files, so why vanishing?
- When VDQM restarts drives left in UNKN state with date of 1 Jan
- tpusage does not work

- Problems in deployment left us behind schedule for mid-Oct startup.
- Had to enlist extra help:
  - Olof, visit to RAL to help trouble-shoot
  - Shaun and Jens, who delayed other commitments
  - RAL managers, assisted coordination
- Coordination meetings with CMS
  - 3 times weekly
  - reviewed problems and progress
- We made it - just in time!



- Pre-production platform frozen after start of CSA06; no changes made except adding diskservers
- Were hoping to upgrade to 2.1.1-4, but too many problems on test instance (see later)

# CMS Service Classes

- CMS originally planned 3 service classes
- Due to delays in deploying diskserver, decided to run with only 1 service class
- CMS wanted disk1tape0, promising to manage disks themselves to not fill
- We chose not to run without GC due to known LSF meltdown problems
- Set GC at 90% and kept tape copy
- Found filesystems on same diskserver filled unevenly, so one FS may hit 90% while another only at 50%
  - Believed fixed in later release

- Achieved overall throughput of 250 MB/sec to disk
- Now have about 100TB data on tape
- End result is that CMS are very happy with castor performance
  - Found it big improvement over dcache

# Most Troublesome Problems for CMS deployment

- Diskservers in different network domain from central servers
- Tape looping
- Problems with castor\_gridftp
  - Some of which were our fault, some of which could have been solved with better documentation
- Problems with database upgrades and hotfixes
  - Now much better addressed in 2.1.1-x

- Rfio issues
  - ‘Address in use’ with rfio under load
    - 64 parallel stream xfering 1 GBfiles
    - Only at RAL?
- Tape looping
  - Problem solved by Olof – stager entries for non existent disk server
- ‘Internal Error’
  - rfcg to any service class except ‘default’
- GC intermittent
  - Can not trigger

1. Problems with database upgrades
2. Castor\_gridftp
3. Error messages don't have enough info
4. “Normal” errors fill logs – especially from scheduler

5. Need more admin tools
  - Would like to not access database directly
6. Limited by only one diskpool per diskserver, especially for testing
7. Uneven filling of filesystems on diskservers
  - Is this an SL4 issues?
8. Clearing out bad files/data