Castor Db Services at CERN Tier 0

Nilo Segura
Oracle support – IT/DES

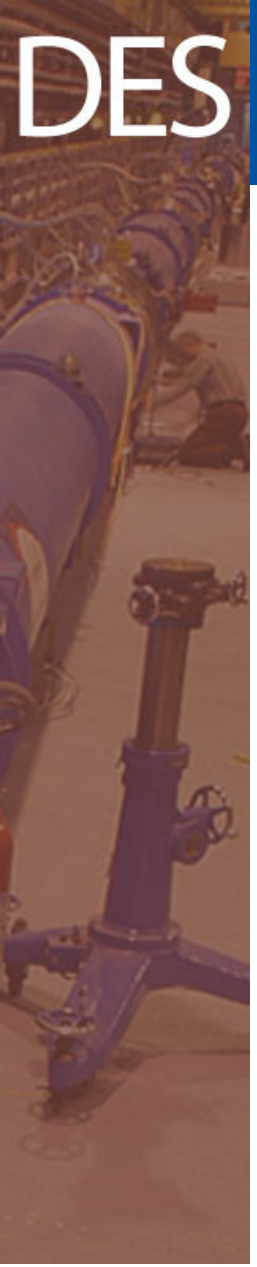
CERN - IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it

Agenda

- Service status
- Software and Hardware status
- What's coming next….
- Pending issues

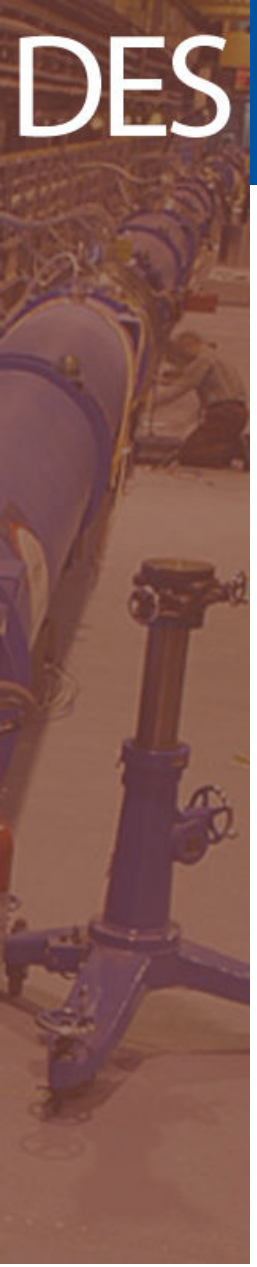
CERN - IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it

Services



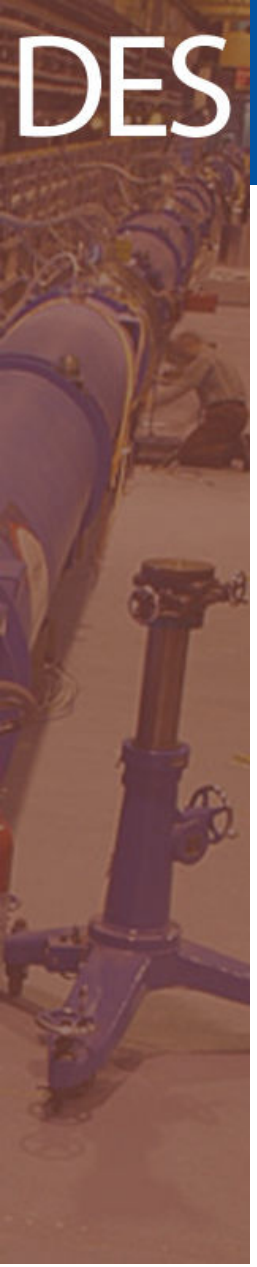
- Each experiment has dedicated Stager, DLF and SRM RAC database
- Plus T3, Repack, ITDC (Stress test) and Name Server
  - Repack server has both DLF and Stager
  - ITDC server has Stager, DLF and SRM.
- Single instances for
  - Tapelog (DLF)
  - Development (2)
  - Preproduction (1)
  - Testing new Oracle releases (1) - Oracle 11gR2
- All Castor DB backups validated automatically on our Automated Recovery Services ☺
  - It would be pretty **inconvenient and unpleasant** to discover that the backup copies are useless ☹ when THE DAY comes

Services



- Performance across all Castor Services is ok.
  - All the tuning done on the application code can be seen now.. ☺
- Atlas SRM:recent heavy row lock contention being investigated ☹
  - Internal SRM version deployed for further debugging..

Hardware

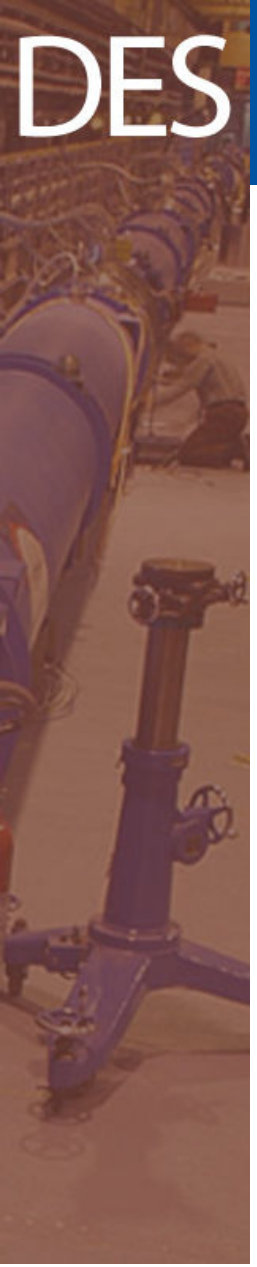


- 2xNodes RAC (almost active-passive)
  - Db jobs on the passive node..
- Completed the migration to new hardware
  - Original boxes run out of warranty 01-Dec-09
  - New boxes with 16/32Gb memory and 8 Cores
    - Intel(R) Xeon(R) CPU E5410 @ 2.33GHz
- Original NAS boxes also replaced for the same reasons

Software

- RDBMS 10.2.0.4 : upgrading to the latest Oracle CPU October patch + recommended one-off list
  - Done : DLF, Development, ITDC and some stagers (Atlas, LHCb, Public)
  - Pending : CMS & Alic stagers,Name Server, SRM, T3, Repack
- Now with large memory (>16Gb) we use HugePages
  - SGA pages are not swapped ☺
  - Greater SGA_MAX_SIZE (larger db_cache_size)
  - More user sessions can live peacefully in the system
    - Name Server and SRM ☺

What's coming next…



- Hardware changes in production services
  - Probably nothing in this area, we have just doubled the systems capacity
  - Change of hardware is however a very quick operation (NFS mount/umount)
- New hardware for the Development services
  - Just installed a couple of temporary replacement servers (thanks to IT/FIO)
  - And waiting for the installation of the new (final) Blades

What's coming…

- Software
  - Performance tests for the new TapeGateway component
  - Security Update (Jan/April/July/Oct 2010)
    - Coupled with our recommended one-offs
  - 10.2.0.5
    - Release date unknown (before Summer 2010 ?)
  - 11gR2
    - Unclear but not before the first patch set is released
    - 2011?
  - Upgrade to Red Hat 5.4 (or later) – coupled with change of hardware (2011?)
  - Netapp OS upgrade to 7.3.2

Pending issues…

- In DLF (only) we see this :

ORA-00600: internal error code, arguments: [kcbgtcr_1], [144165], [533], [646148568], [627], [], [], []~ORA-06512: at "CASTOR_MON.STATSTAPERECALL

```
INSERT INTO TAPERECALL (TIMESTAMP, SUBREQID, TAPEID, TAPEMOUNTSTATE)
(SELECT MES.TIMESTAMP, MES.SUBREQID, MES.TAPEVID, STR.VALUE
FROM CASTOR_DLF.DLF_MESSAGES MES,
 CASTOR_DLF.DLF_STR_PARAM_VALUES STR
WHERE MES.ID = STR.ID AND MES.FACILITY = 22 AND MES.MSG_NO = 57
AND STR.NAME = 'TapeStatus' AND MES.TIMESTAMP >= :B1 AND STR.TIMESTAMP >= :B1
AND MES.TIMESTAMP < :B1 + 5/1440 AND STR.TIMESTAMP < :B1 + 5/1440)
LOG ERRORS INTO ERR_TAPERECALL REJECT LIMIT 100000
```

It does not affect the DLF functionality…

Known Bug (5231155) but we need to request a MLR

Some figures…

Rows…	Castorfile	Tapecopy	Diskcopy
ALICE	1788333	156235	1788836
CMS	8412587	11161	8560961
ATLAS	6975786	1552	7251145
LHCb	1257546	3325	1296762
Public	1550108	68634	1603180

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