

PIC
port d'informació
científica

dCache upgrade plans at PIC

Gerard.Bernabeu@pic.es

- Current dCache deployment
- Target
- 1.9.10 experience
 - New config files
- PIC Tier1 dCache instance upgrade plan

Tape dCache (Tier1/2 dCache)

- 1.9.5-23 (~4.5PB in SATA disks)
- Enstore Tape backend (~4PB LTO3-5 in 2 robots)
- Currently supporting 10 experiments (ATLAS, CMS and LHCb from WLCG)
- and several protocols
 - HTTP with home made browsing and user/password authentication
 - [gsi]dCap, GridFTP
 - Xrootd – actually not used by any experiment
 - a few tests done by CMS failed. Not able to open a file.
- 3 dCache servers: dccore, dcip, srm (2*2Core@2.3, 8GB RAM)
- 1 pnfs (1*4core@3.5, 48GB RAM, 80GB FusionIO, 6 HDD RAID6)
- + pnfs RO (postgresql9 Streaming Replication)
- 4 doors (1 CPU L3406+4GB+2x1GE aggregated)

Disk dCache (Tier3)

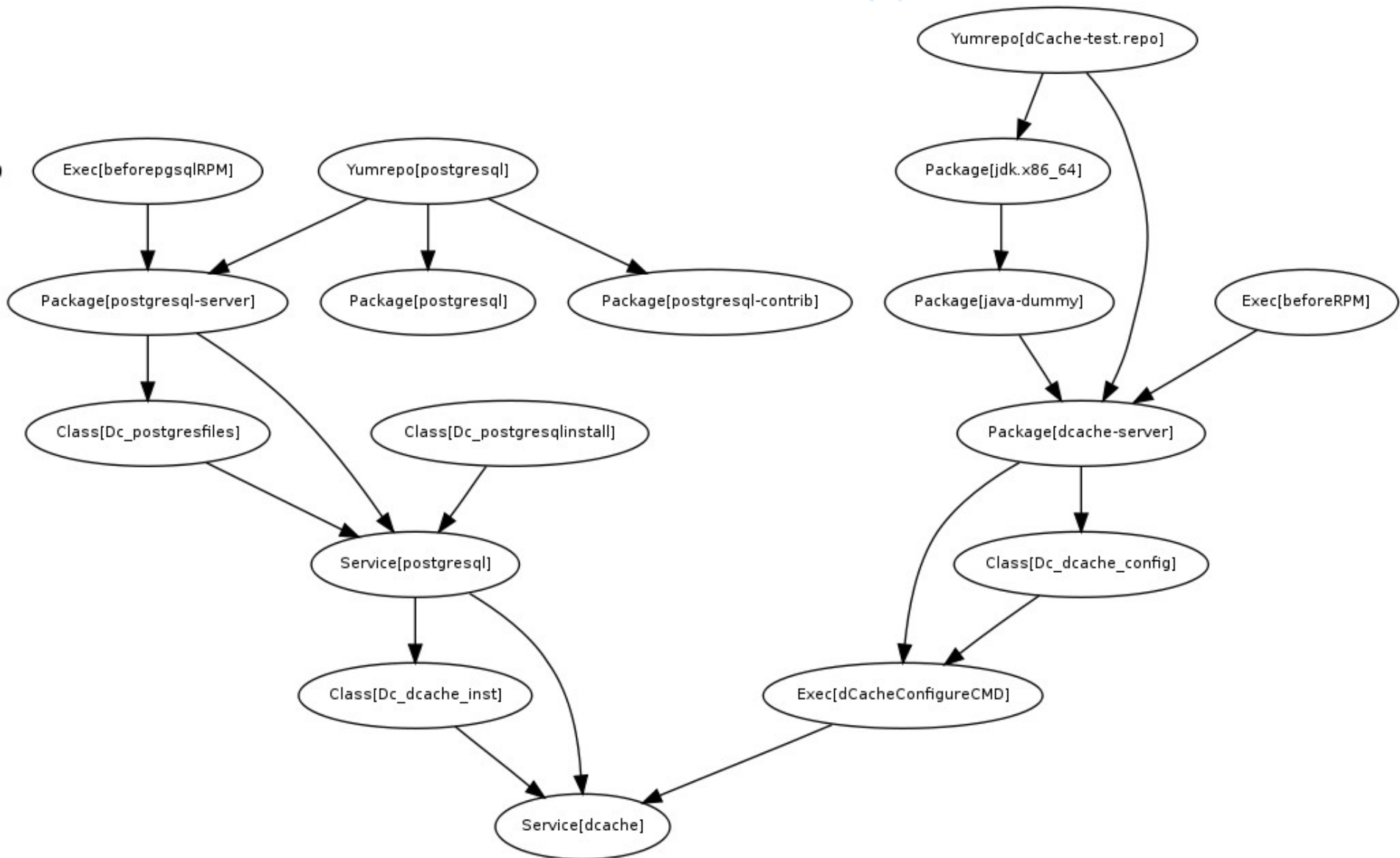
- 1.9.10-5 (60TB)
 - deployment delayed because of xrootd bugs. Solved in -5
 - users worried about bug RT#6010
- No Tape backend
- 1 dCache server & Chimera (1*4core@3.5, 48GB RAM, 6 HDD RAID6)
- 2 doors (1 CPU low power+4GB+2x1GE aggregated)
- Protocols in use: gridFTP, WebDAV/HTTP, [gsi]dCap, Xrootd

Test dCache (Tape dCache)

Virtual Machines, for functional tests before upgrading Tier1/2 dCache.

All servers are managed by Puppet, including PostgreSQL and dCache install&config.

Puppetized dCache



- Move Tier3 to disk-only instance
- Move Tier2 to disk-only instance
- Upgrade both dCache instances to 1.9.12 (Golden release)
 - Move the Tape dCache to Chimera
 - Start using NFSv4.1
 - Reduce the number of active dCache servers from 4 (pnfs, srm, poolmanager, info) to 3 (pnfs, srm, catch-all)

1.9.10 experience

- NFSv4 is focused on reads, works but special kernel is required for SLC5.
- WebDav/HTTP is more efficient than dCap in LAN bulk data transfers (wget vs dccp).
 - Using ROOT client doesn't work. HTTP reimplemented in 1.9.11.
- Nice HTTP browsing capabilities
 - lack of simple user/pass authentication (we'll use a proxy to workarround).
 - Not customizable (ie: add file size), this will come with 1.9.12.
- Several bugs found in Xrootd, solved in 1.9.10-5
- We still need to get rid of some dependencies with the old info provider (gone after 1.9.5), like the user-view space monitoring

1.9.10 experience

Fri Mar 11 16:35:03 CET 2011

atlas;;Free:540782;Total:2133993;Used:835906;Reserved:757305;Cached:248457

ATLASDATADISK;Free:86005;Total:650000;Used:563995;TokenID:11643

ATLASMCDISK;Free:295124;Total:300000;Used:4876;TokenID:11746

ATLASDATATAPE;Free:62994;Total:64000;Used:1006;TokenID:11747

ATLASMCTAPE;Free:19914;Total:20000;Used:86;TokenID:11748

ATLASSCRATCHDISK;Free:101997;Total:110000;Used:8003;TokenID:17684300

ATLASHOTDISK;Free:3889;Total:5000;Used:1111;TokenID:28255280

ATLASdefaultToken;Free:0;Total:0;Used:0;TokenID:30730542

IFAEdefaultToken;Free:0;Total:0;Used:0;TokenID:30730551

ATLASGROUPDISK;Free:7216;Total:20000;Used:12784;TokenID:5269825

ATLASPRODDISK;Free:2500;Total:2500;Used:0;TokenID:14939115

T2ATLASLOCALGROUPDISK;Free:4566;Total:8000;Used:3434;TokenID:5278301

T2ATLASPRODDISK;Free:23208;Total:25000;Used:1792;TokenID:3517583

T2ATLASGROUPDISK;Free:17379;Total:50000;Used:32621;TokenID:32671737

T2ATLASDATADISK;Free:97737;Total:273530;Used:175793;TokenID:2813561

T2ATLASHOTDISK;Free:4158;Total:5000;Used:842;TokenID:28072951

T2ATLASSCRATCHDISK;Free:8124;Total:28000;Used:19876;TokenID:17930260

ATLAST2spaceAccounting;Free:272752;ToAssign:128878;InTokens:389530;Total:518408;TokenID:-

ATLAST1spaceAccounting;Free:776878;ToAssign:421585;InTokens:1194000;Total:1615585;TokenID:-

withoutToken;Free:540782;Total:-;Used:9687;TokenID:-

atlas;Tape-Recall;Free:1049630;Total:2133993;Used:0;Reserved:0;Cached:1084363

cms;;Free:792050;Total:938420;Used:42413;Reserved:103957;Cached:658379

CMSunmerged;Free:46123;Total:50000;Used:3877;TokenID:48064436

CMSbackfill_1;Free:7798;Total:25000;Used:17202;TokenID:48064482

CMSbackfill_test01;Free:25000;Total:25000;Used:0;TokenID:48064484

CMStemp;Free:25000;Total:25000;Used:0;TokenID:48676186

withoutToken;Free:792050;Total:-;Used:21334;TokenID:-

cms;Tape-Recall;Free:237628;Total:938420;Used:0;Reserved:0;Cached:700792

lhcb;;Free:25122;Total:294675;Used:160656;Reserved:108896;Cached:39232

LHCb_USER;Free:3367;Total:25000;Used:21633;TokenID:1613981

LHCb_RAW;Free:0;Total:0;Used:0;TokenID:30518868

LHCb_MC_M-DST;Free:892;Total:25000;Used:24108;TokenID:3652126

LHCb_RAW;Free:2206;Total:2206;Used:0;TokenID:42138

LHCb_M-DST;Free:15478;Total:35000;Used:19522;TokenID:42140

LHCb_DST;Free:78871;Total:140000;Used:61129;TokenID:42141

LHCb_FAILOVER;Free:4799;Total:5000;Used:201;TokenID:42142

LHCb_RDST;Free:2206;Total:2206;Used:0;TokenID:45329

LHCb_MC_DST;Free:1081;Total:25000;Used:23919;TokenID:5254380

withoutToken;Free:25122;Total:-;Used:10144;TokenID:-

lhcb;Tape-Recall;Free:94787;Total:294675;Used:0;Reserved:0;Cached:199888

<http://demon.pic.es/SpaceMonitorOutput.html>

With the new config files it is now very easy to customize services in domains (java procs).

- After spending a couple hours with the new config files I wouldn't go back.
- Very useful to go through `/opt/d-cache/share/defaults` (all possible properties are there!)
- When referring to variables/properties, use `${name}`, it is not bash (don't miss the `{}`)
 - In 1.9.10 the system doesn't complain if a property is not properly defined/recognized
- Would be good to have a recommended daemon-service layout based on cell relationships

New config files

```
cat ./opt/d-cache/etc/layouts/dccore.conf.disk
#dccore
[dCacheDomain]
[dCacheDomain/poolmanager]
[dCacheDomain/dummy-prestager]
[dCacheDomain/broadcast]
[dCacheDomain/loginbroker]
[dCacheDomain/srm-loginbroker]
[dCacheDomain/topo]
[dCacheDomain/gplazma]
[dCacheDomain/gsi-pam]
[dCacheDomain/pinmanager]

[infoDomain]
[infoDomain/billing]
[infoDomain/httpd]
[infoDomain/webadmin]
[infoDomain/admin]
[infoDomain/info]
[infoDomain/statistics]

#dcns
[nfsDomain]
#Not possible to run both 4.1 and 3 in same server
[nfsDomain/nfsv3]
[nfsDomain/nfsv41]

[namespaceDomain]
[namespaceDomain/pnfsmanager]
[namespaceDomain/cleaner]
[namespaceDomain/acl]
[namespaceDomain/dir]
```

```
#srm
[srm-${host.name}Domain]
[srm-${host.name}Domain/srm]
[srm-${host.name}Domain/spacemanager]
[srm-${host.name}Domain/transfermanagers]

#Pools
#[${host.name}Domain]
#[${host.name}Domain/pool]
#name=${host.name}_1
#path=/dcpool/vpool1
#waitForFiles=${path}/data
##La config de movers i cues, etc es pot fer estàtica
segons el tipus de node en el fitxer setup dins del pool

#Door servers
#[dcap-${host.name}Domain]
#[dcap-${host.name}Domain/dcap]

#[xrootd-${host.name}Domain]
#[xrootd-${host.name}Domain/xrootd]

#[gridftp-${host.name}Domain]
#[gridftp-${host.name}Domain/gridftp]

#[webdav-${host.name}Domain]
#[webdav-${host.name}Domain/webdav]

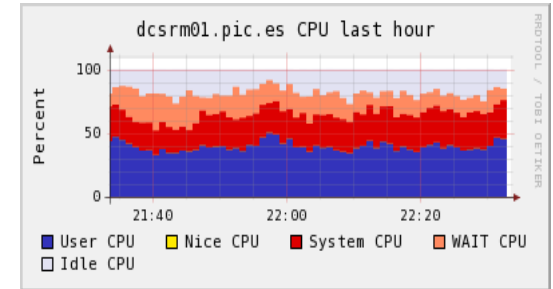
#[gsidcap-${host.name}Domain]
#[gsidcap-${host.name}Domain/gsidcap]

#[ftp-${host.name}Domain]
#[ftp-${host.name}Domain/ftp]
```

PIC Tier1 dCache instance upgrade plan

Disk dCache Plan

- 1) As soon as it is available upgrade to 1.9.12



Tape dCache plan

- 2) Upgrade SRM hardware (8GB RAM and 2x dual core AMD 2218 is not enough anymore)
- 3) Get rid of old infoProvider dependencies (SRM reservations monitoring page)
- 4) Upgrade to 1.9.12 on the Tape dCache. On the upgrade also
 - 1) Join in a single server PoolManager and InfoServices servers
 - 2) Move all postgresSQL DBs to 9.0, migrating from PITR to streaming replication
- 5) Move to Chimera when tested and working with Enstore (Xmas 2011?).