CASTOR development news

Sébastien Ponce
CERN IT DSS

CASTOR Face to face meeting - 28-30 Nov 2012

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

Overall program

- Quick reminder on 2.1.12 serie
- Overview of what is new in 2.1.13
  - Stager DB, tape side
  - Nameserver security
  - Monitoring
  - Miscelaneous
- (almost) done for next version
  - Xroot for tape
  - Traffic shaping
  - Diskserver monitoring (also known as tape scheduling)

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

DSS 2.1.12 (1)

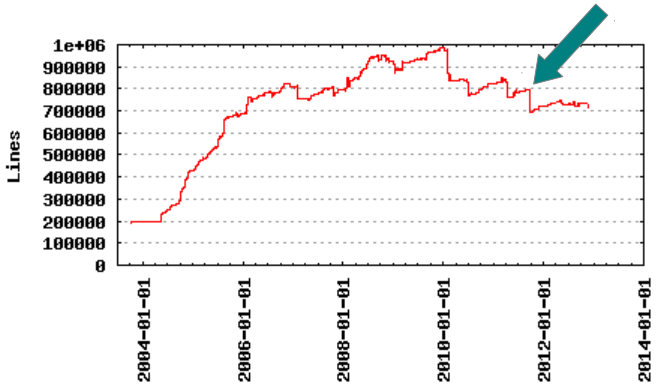
CERN IT Department

- Stager schema changed for migrations
  - tapepools have nb drives, minAmountData, minNbfiles, maxFileAge
  - migrationRouting table replaces policies
    - Maps fileclass/copyNb to TapePool
  - Migration decision taken at open time
  - Multiple concurrent migrations issues solved
- New admin tools
  - print/insert/deletesvcclass/tapepool/diskpool/...
  - More readable, proper man pages

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

2.1.12 (2)

- Repack fully rewritten (See Daniele's talk)
  - Simple admin tool
    - no daemon, no DB
    - any stager is a repack instance
  - Much more efficient
    - 20 tapes of 200K files start in 2h
- Id2type has been dropped
  - allows 250 files handling per second now
- Major code cleanup
  - 75K lines of code dropped
    - 8% of total code base



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

2.1.13 overview

- Stager schema changed for recalls
- Secure namespace
- Monitoring
- Miscellaneous
  - Logging improvements
  - Bulk interfaces in stager for tapegateway
  - New xroot and gridFTP versions
  - Dynamic configuration
  - LRUPin GC policy
  - SLC6 support

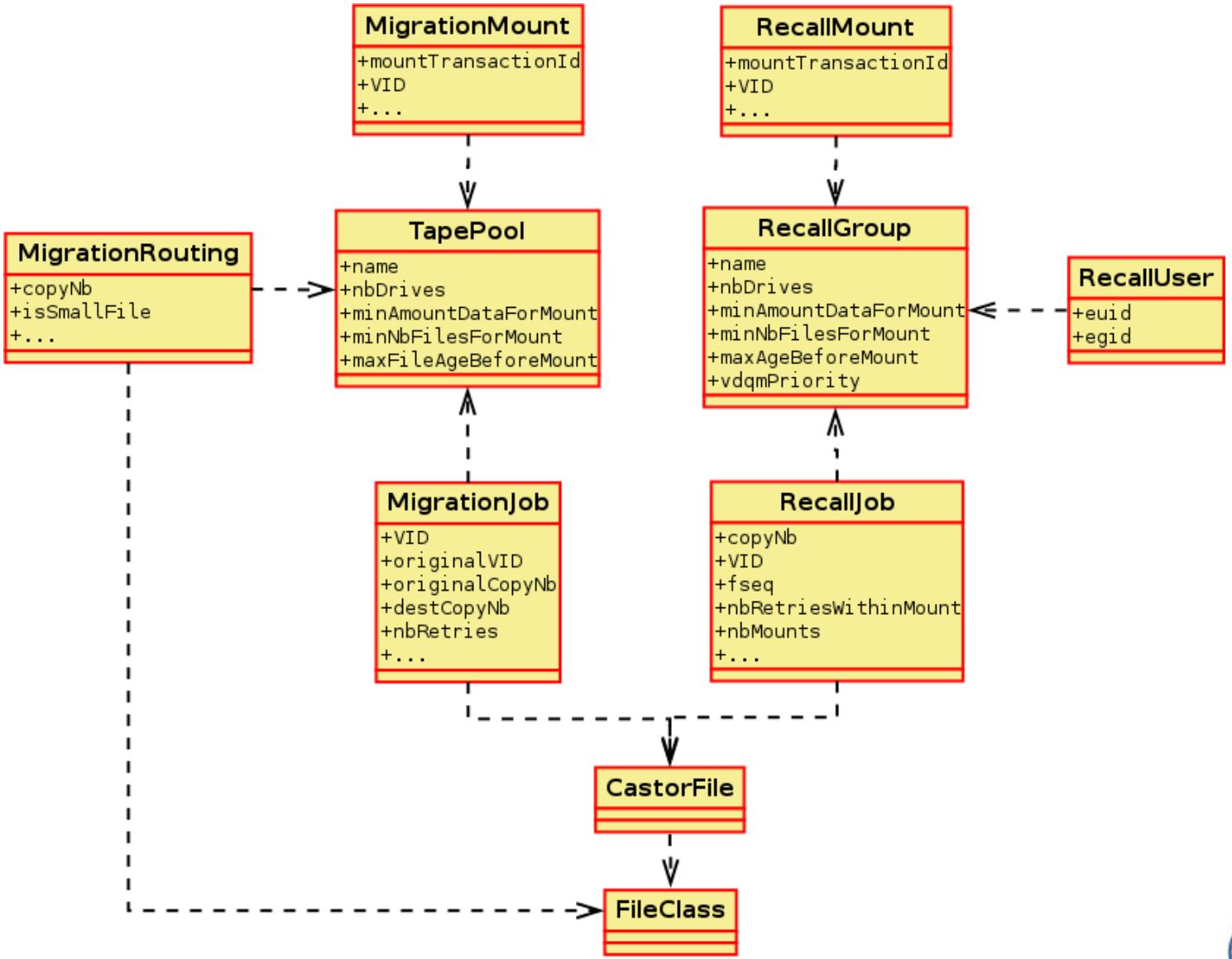
CERN IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it

New stager DB schema for tape

- Main ideas
  - Split recall and migrations
  - Have clear entities for mounts and jobs
    - Recall/MigrationJob/Mount
  - Have a way to limit nb drives used
    - By tapepool for migrations
    - By recallGroup for recall
  - Simplify complex policies
    - Replaced by simple DB job
    - Decision to mount taken on couple of numbers
      - MinAmountData, minNbFiles, maxAge
    - Simplified migration routing
      - Based on fileClass and copyNb only

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

New stager DB schema for tape



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

New tools

CERN IT Department

```
[root@c2cmssrv401 ~]# printmigrationroute
      FILECLASS COPYNB ISSMALLFILE        TAPEPOOL LASTEDITOR           LASTEDITION
-------------------------------------------------------------------------------------
        c3_copy      1             -        cms_user        root 12-Mar-2012 11:08:35
            cms      1             - cmsfamily_new1        root 12-Mar-2012 11:05:10
cms_production       1             -     cms_prod_08        root 12-Mar-2012 11:05:10
        cms_raw      1             -      cms_raw_08        root 12-Mar-2012 11:05:10
   cms_streamer      1             -   cms_stream_08        root 12-Mar-2012 11:05:10
       cms_temp      1             -    cms_testdata        root 12-Mar-2012 11:05:11
       cms_test      1             -      cms_csa_07        root 12-Mar-2012 11:05:11
       cms_user      1             -        cms_user        root 12-Mar-2012 11:08:55
```

```
[root@c2cmssrv401 ~]# printrecallgroup
     NAME NBDRIVES MINAMOUNTDATA MINNBFILES MAXFILEAGE VDQMPRIORITY          ID LASTEDITOR          LASTEDITION
-----------------------------------------------------------------------------------------------------------------
  default       20         10GiB         10        4h             0 23575325766    gcancio 12-Nov-2012 08:02:55
      vip       40        100GiB       1000      10mn           100 23575331556       root 08-Oct-2012 16:30:57
immediate        5           1B          1        5s          1000 23983267548       root 01-Nov-2012 18:27:07
```

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

Nameserver security

- Nameserver can be started in secure mode
  - Requires kerberos authentication
  - Runs on separate port
- Clients are “backward-compatible”
  - They try to go secure
  - If it fails, they try unsecure mode
- A white list exists for old clients
  - Used for some data taking
- “local” nameservers of the stagers
  - Should now be made 100% local
    - i.e. should only accept connections from localhost

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

Nameserver security

- Nameserver can be started in secure mode
  - Requires kerberos authentication
  - Runs on separate port
- Clients are “backward-compatible”
  - They try to go secure
  - If it fails, they try unsecure mode
- A white list exists for old clients
  - Used for some data taking
- “local” nameservers of the stagers
  - Should now be made 100% local
    - i.e. should only accept connections from localhost

More details coming on Friday by Andrea

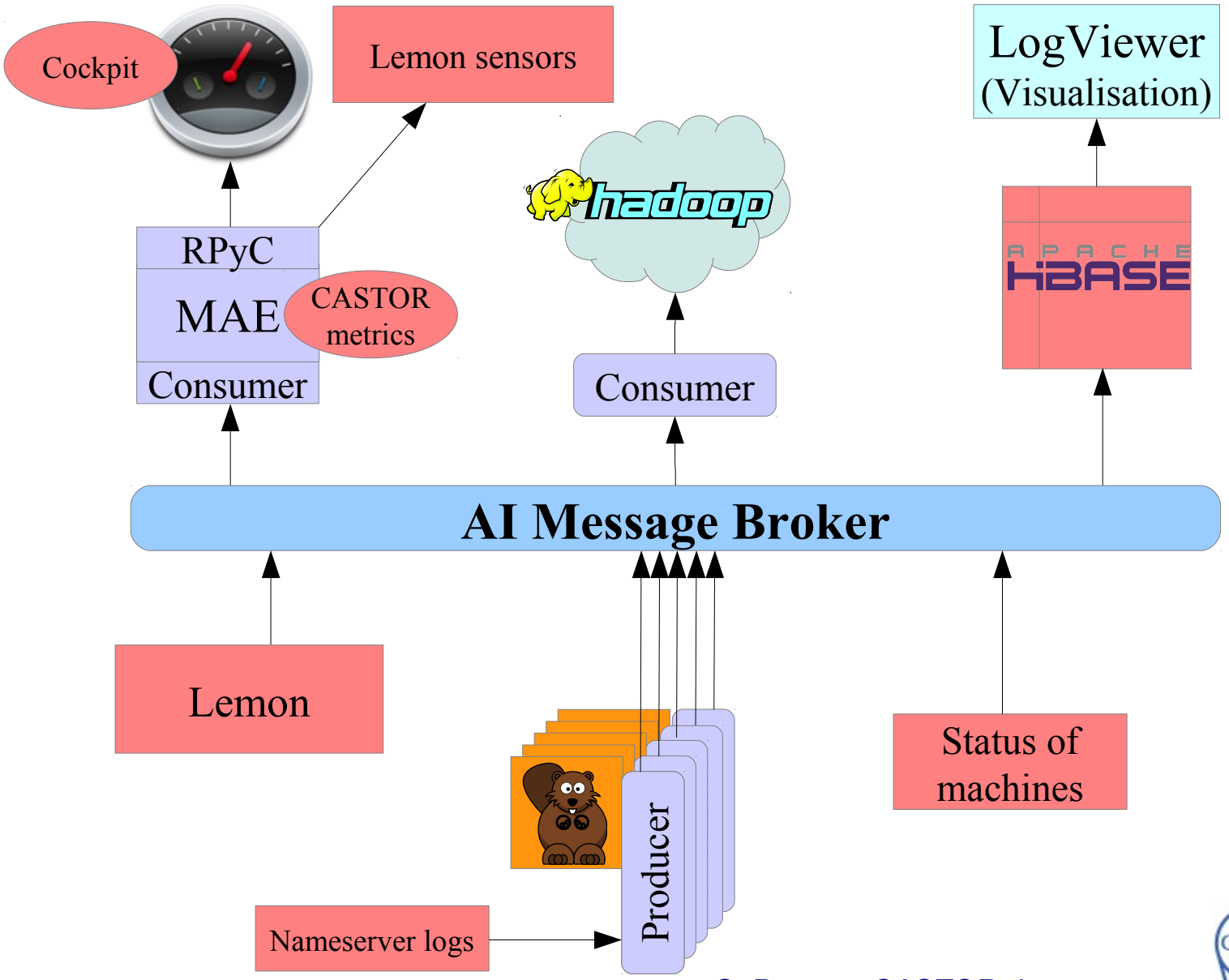
CERN IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it

New monitoring

- Main ideas
  - Work on the flow of log
    - Rather than greping/selecting afterward
  - Use a bus approach and have several producers/consumers of logs
    - And provide common producers/consumers
      - producers from CASTOR log files
      - consumers for “cockpit” plots
  - Have easy to define metrics
    - Handle by a metric analysis engine
    - Computed online on the flow of logs
  - Provide up to date user interface
    - The cockpit

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

New monitoring



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

How to display a new metric

- First define the metric

```
<metric>
 name: ClientVrsDistribution
 unit: Hz
 category: General
 window: 60
 conditions: MSG=="Reply sent to client" and DAEMON=="rhd"
 groupbykeys: ClientVersion
 data: CounterHz(COUNT)
 handle_unordered: time_threshold
 nbins: 1
</metric>
```

- Copy it to /etc/mae/metrics
- Enjoy it immediately in the cockpit

More details coming on Friday by Benjamin

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

Miscellaneous (1)

- Logging from DB
  - The PL/SQL code can now easily log
    - e.g. a DB job is logging migration/recall queues
- Bulk interfaces in stager for tapegateway
  - Makes the DB updates more efficient
    - The whole chain is now bulk
- New xroot and gridFTP versions
  - Xroot 3.2
  - GridFTP from EMI
  - Prerequisite for SLC6 support

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

Miscellaneous (2)

- Dynamic configuration
  - castor.conf is read every 5mn by default
  - so restart of daemons are not needed anymore
    - But some daemons still cache parameters
- New LRUPin GC policy
  - Least Recently Used
  - But taking setFileGcWeight into account
    - up to 1 month
- SLC6 support
  - Code is ready and tested
  - RPMs are provided for 2.1.13-6
  - Issues with default SLC6 rsyslog
    - Work around should come in next release
    - Rsyslog may be fixed (developers contacted)

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

Coming in next version

- Xroot in Tape transfers
  - Migrations and recalls done via xroot
  - Work done by Victor Kotlyar during summer
- Traffic shaping
  - Allow tape streams to have precedence
  - Work done by Eric
    - See details in his presentation
- New DiskServer monitoring
  - a.k.a. drop rmmaster/node and shared mem
  - a.k.a. Tape scheduling
    - Allow to avoid collisions of tape streams on diskservers

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

DiskServer Monitoring

- In practice
  - rmmasterd integrated into transfermanagerd
  - rmnoded integrated into diskmanagerd
  - No more shared memory
    - Single source of information : the stager DB
  - rmAdminNode/rmGetNode/movediskserver gone
    - print/modify/deltediskserver replace them
  - Status fo all streams of all filesystems available for tape scheduling within 1s
    - To be used by recalls' and migrations' scheduling

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

That's it for now

Any questions ?

Any thing you want details on ?

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