iRODS usage at CC-IN2P3

Jean-Yves Nief

Talk overview

- What is CC-IN2P3 ?
- Who is using iRODS ?
- iRODS administration:
  - Hardware setup.
- iRODS interaction with other services:
  - Mass Storage System, backup system, Fedora Commons etc...
  - iRODS clients usage.
- Architecture examples with collaborating sites.
- Rules examples.
- SRB to iRODS migration.
- To-do list and prospects.

CC-IN2P3 activities

- Federate computing needs of the french scientific community in:
  - Nuclear and particle physics.
  - Astrophysics and astroparticles.
- Computing services to international collaborations:
  - CERN (LHC), Fermilab, SLAC, ….
- Opened now to biology, Arts & Humanities.

DSM cea IRFU

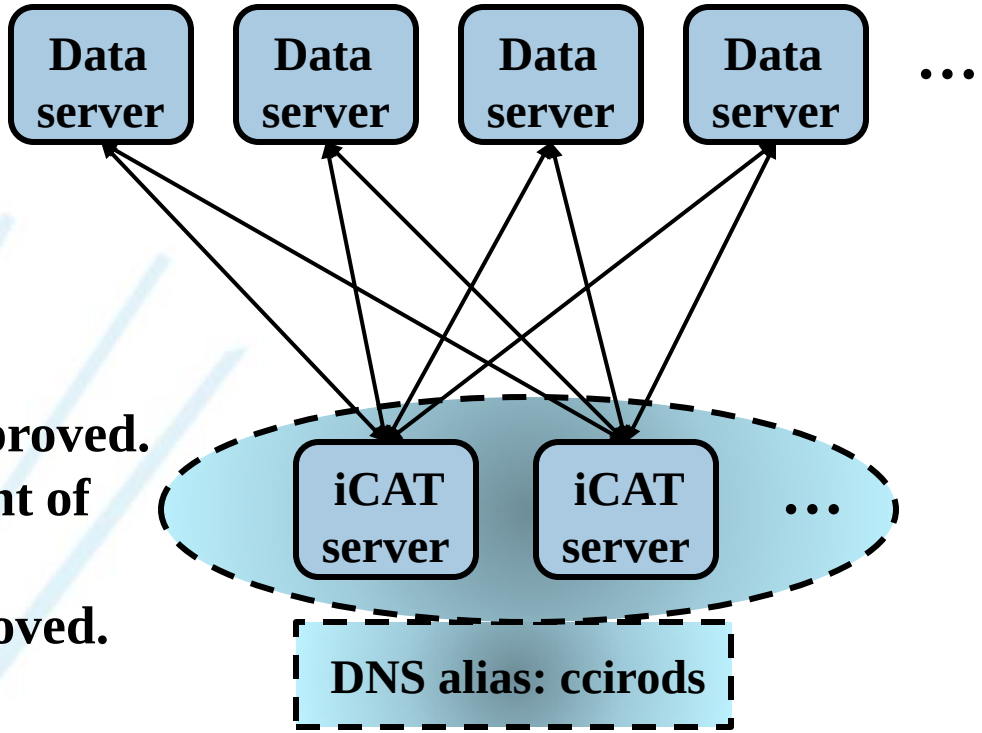
CCIN2P3

CNRS IN2P3
Institut National de Physique Nucléaire et de Physique des Particules

iRODS setup @ CC-IN2P3

- In production since early 2008.
- 14 servers:
  - 2 iCAT servers (metacatalog): Linux SL4, Linux SL5
  - 12 data servers (520 TB): Sun Thor x454 with Solaris 10, DELL v510 with Linux SL5.
- Metacatalog on a dedicated Oracle 11g cluster.
- Monitoring and restart of the services fully automated (*crontab* + *Nagios*).
- Automatic weekly reindexing of the iCAT databases.
- Accounting: daily report on our web site.

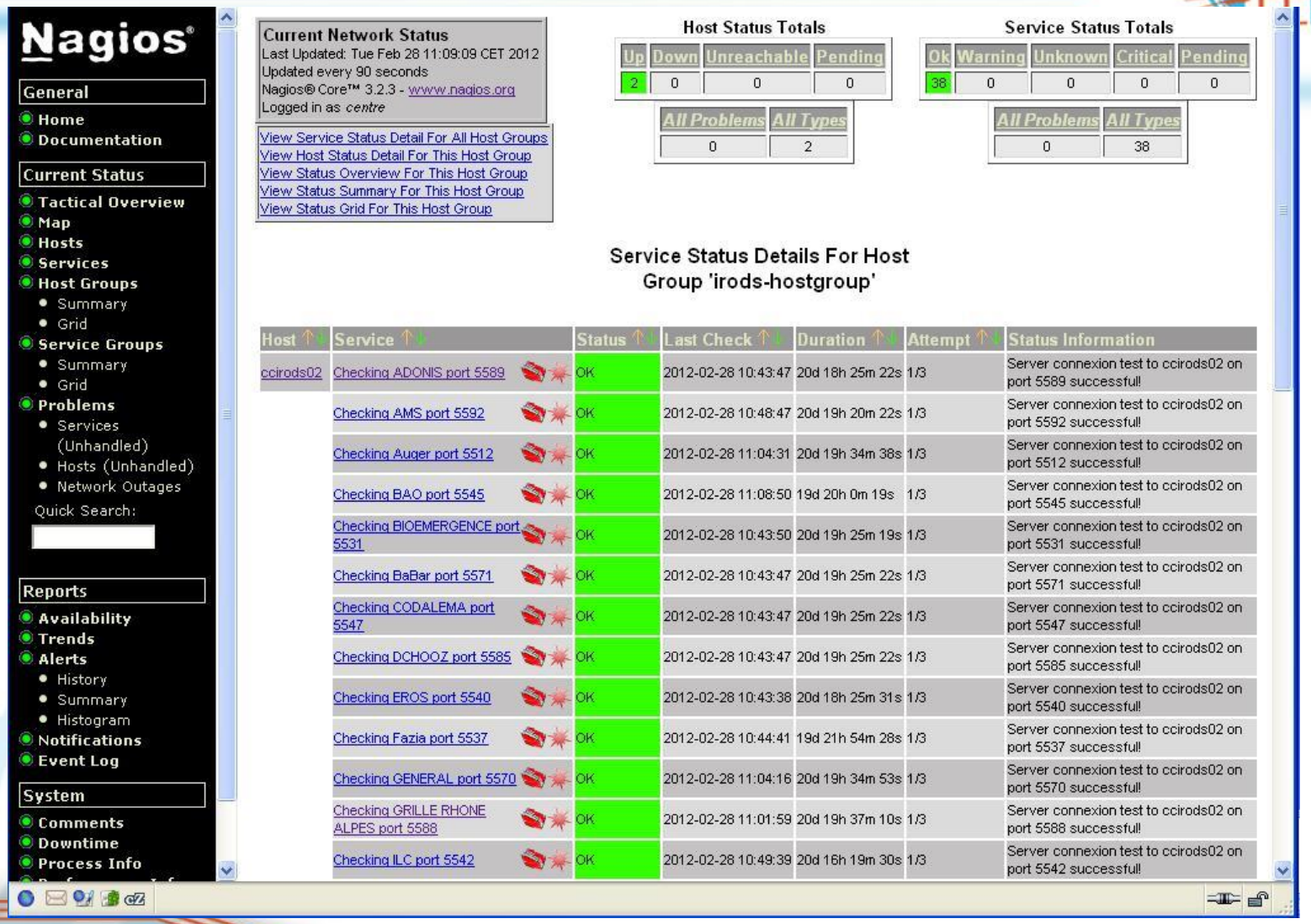
iRODS setup @ CC-IN2P3



DNS alias:

- **load balanced.**
- **redundancy improved.**
- **avoid single point of failure.**
- **scalability improved.**

iRODS monitoring: Nagios



iRODS interaction with other services

- **Mass storage system**: HPSS.
  - Using compound resources.
  - Interfaced using the universal MSS driver (RFIO protocol used).
  - Staging requests ordered by tapes using Treqs.
- **Backup system**: TSM.
  - Used for projects who do not have the possibility to replicate precious data on other sites.
- **Fedora Commons**:
  - Storage backend based on iRODS using FUSE.
  - Rules to register iRODS files into Fedora.
- **External databases**:
  - Rules using RDA (see Yonny's talk).

iRODS clients

- Clients: from laptop to batch farms.
  - Authentication: password or X509 certificates.
- iCommands: most popular.
  - From any platform: Windows, Mac OSX, Linux (RH, CentOS, Debian…), Solaris 10.
- Java APIs: interaction with iRODS within workflows.
- C APIs: direct access to files (open, read, write) to do « random access ». Drivers for some viewer such as OsiriX (biomedical apps).
- FUSE for legacy web sites and Fedora Commons.
- Windows explorer and iDrop.

Who is using iRODS ?

- **High energy and nuclear physics**:
  - BaBar: data management of the entire data set between SLAC and CC-IN2P3: total foreseen 2PBs.
  - dChooz: neutrino experiment (France, USA, Japan etc…): 400 TBs.
- **Astroparticle and astrophysics**:
  - AMS: cosmic ray experiment on the International Space Station (280 TBs).
  - TREND, BAOradio: radioastronomy (170 TBs).
- **Biology and biomedical applications**: phylogenetics, neuroscience, cardiology (50 TBs).
- **Arts and Humanities**: Adonis (46 TBs).

Who is using iRODS ?

iRods disk space usage & files per experiment at IN2P3 Computing Center

Area maintained by Thomas Kachelhoffer

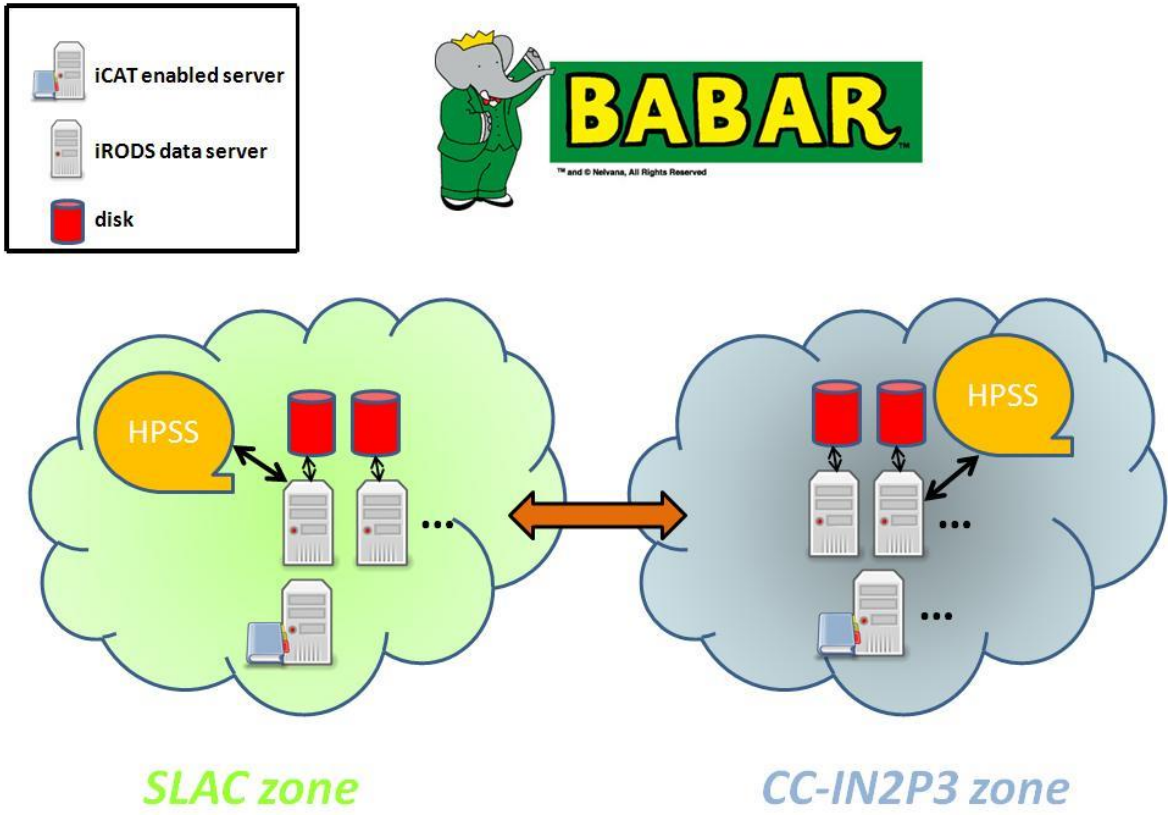
Description:

1 630 TB are used at this time. These values were collected the ***2012-02-27*** at ***11:27:01***. By clicking on the instance name below, you will find the values corresponding to the selected instance and their evolutions.

List of iRods instances:

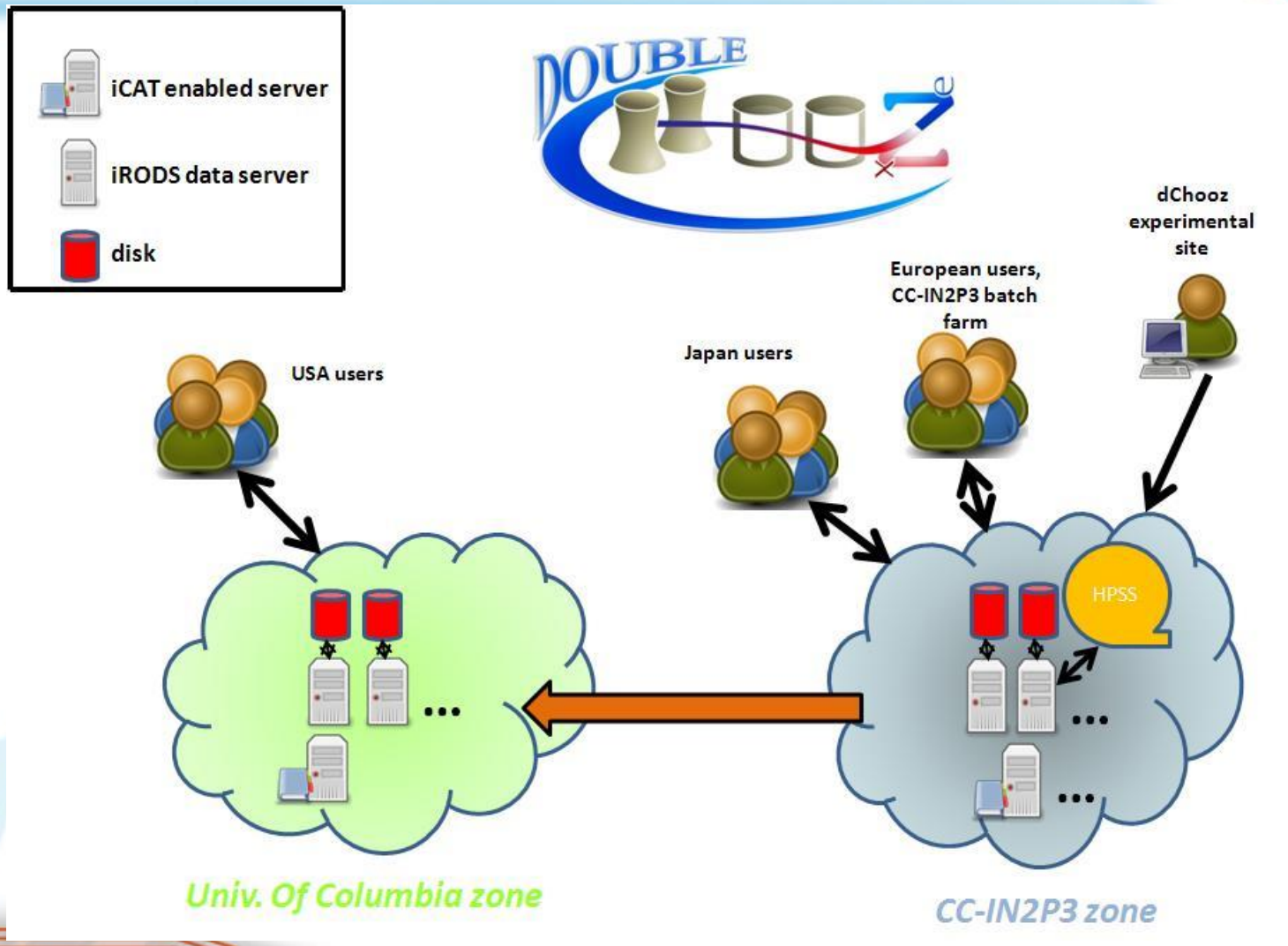
adonis	46 624 GB	3 816 847 files
ams	280 561 GB	115 913 files
babar	713 022 GB	60 507 files
bao	142 277 GB	1 621 935 files
bioemergence	12 618 GB	3 749 525 files
codalema	1 925 GB	512 243 files
dchooz	400 477 GB	529 425 files
fazia	2 744 GB	7 015 files
general	4 354 GB	843 984 files
imxgam	544 GB	19 636 files
ipm	553 GB	210 files
test	54 GB	13 347 files
tidra	19 162 GB	7 384 479 files
tidra-neuro	17 144 GB	2 890 211 files
trend	27 357 GB	532 899 files
	1 669 415 GB	22 098 176 files

Architecture example: BaBar

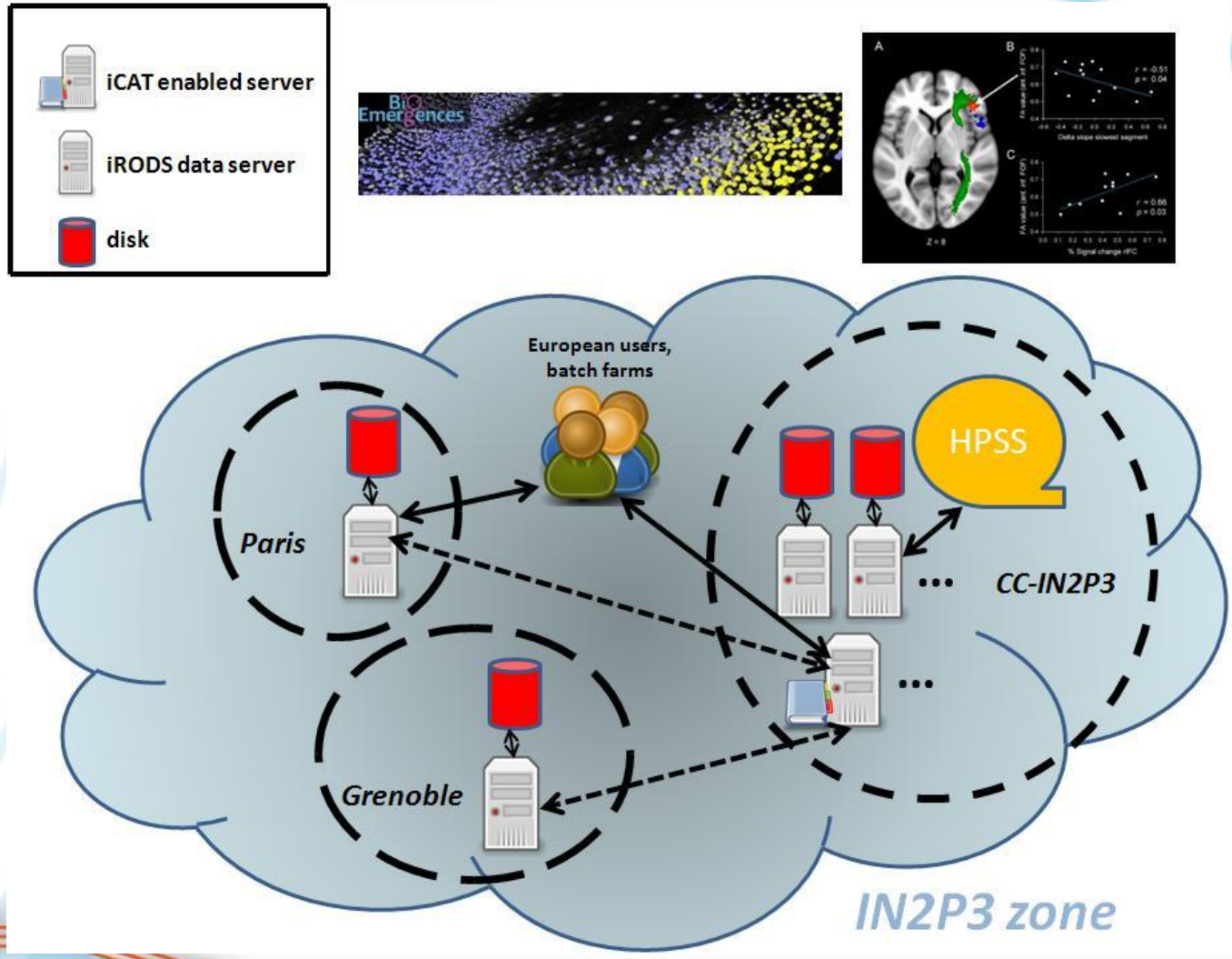


- archival in Lyon of the entire BaBar data set (total of 2 PBs).
- automatic transfer from tape to tape: 3 TBs/day (no limitation).
- automatic recovery of faulty transfers.
- ability for a SLAC admin to recover files directly from the CC-IN2P3 zone if data lost at SLAC.

Architecture example: dChooz



Architecture example: embryogenesis and neuroscience



Rules examples (I)

- Delayed replication to the MSS:
  - Data on disk cache replication into MSS asynchronously (1h later) using a delayExec rule.
  - Recovery mechanism: retries until success, delay between each retries is doubled at each round.
- ACL management:
  - Rules needed for fine granularity access rights management.
  - Eg:
    - 3 groups of users (**admins**, **experts**, **users**).
    - ACLs on /<zone-name>/*/rawdata => **admins** : *r/w*, **experts** + **users** : *r*
    - ACLs on all others subcollections => **admins** + **experts** : *r/w*, **users** : *r*
- External database interface: using RDA to build rules for the DTM tool (see Yonny's talk).

Rules examples (II)

- Fedora Commons:
  - Tar balls content stored in iRODS are automatically registered into Fedora Commons.
  1. Automatic untar of the files + checksum on the iRODS side: *msiTarFileExtract.*
  2. Automatic registration in Fedora-commons (delayed rule): *msiExecCmd* of a java application.
- Automatic metadata extraction from DICOM files (neuroscience…):
  - A given predefined list of metadata is extracted from the files using DCMTK (thanks to Yonny), then user metadata are created for each file.

SRB to iRODS migration

- SRB still used: 3.7 PBs so far.
- Migration to iRODS already made for BioEmergence (embryogenesis) in 2010:
  - Data workflow was using Jargon: transparent.
  - Migration from Scommands to icommands was needed.
  - 2 hours of downtime to complete the migration (scripts were needed).
- Needs to migrate all the other projects by the end of 2012, beginning of 2013:
  - SRB is deeply embedded in data management workflows and projects can't live without SRB.

  ➔ Main issue: migration should be as « transparent » as possible in order to keep up with the data activity.

To-do list

- Complete SRB to iRODS migration.
- Connection control:
  - Connections can come from anywhere especially batch farms on the data grid.
  - Servers can be overwhelmed (network, disk activity for hundreds of connection in //).
  - Causes clients to exit with an error ➔ not good.
  - Improved version of CCMS (connection control) is needed.
- Conversion to rules of the scripts used to manage cache space on compound resources.
- Dealing with filename with accentuated characters for iCommands on Windows.
- Provide a light weight transfer tool for every single users (ship files between CC-IN2P3 to distant site).
- Centralized administration through a GUI (15 instances of iRODS running so far).

Prospects

- 1.6 PBs in iRODS as of Feb 2012 (should be **5 PBs** at the end of this year).
- Future projects:
  - Biomedical field: research in cardiology, MS (anonymization) with data from > 10 hospitals.
  - Private companies (data encryption needed ?).
  - Astrophysics.
- Grid: iRODS officially promoted by the French NGI.

Acknowledgement

- Thanks to:
  - Pascal Calvat.
  - Yonny Cardenas.
  - Rachid Lemrani.
  - Thomas Kachelhoffer.
  - Pierre-Yves Jallud.