iRODS at CC-IN2P3: managing petabytes of data

Jean-Yves Nief
Pascal Calvat
Yonny Cardenas
Quentin Le Boulc'h
Rachid Lemrani

CCIN2P3

cnrs

IN2P3

What is CC-IN2P3 ?

- **IN2P3**:
  - one of the 10 institutes of CNRS.
  - 19 labs dedicated to research in high energy, nuclear physics, astroparticles.
- **CC-IN2P3**:
  - computing resources provider for experiments supported by IN2P3 (own projects and international collaborations).
  - resources opened both to french and foreign scientists.

CC-IN2P3: some facts and figures

- CC-IN2P3 provides:
  - Storage and computing resources:
    - Local, grid and cloud access to the resources.
  - Database services.
  - Hosting web sites, mail services.

- 2100 local active users (even more with grid users):
  - including 600 foreign users.

- ~ 140 active groups (lab, experiment, project).

- ~ 20000 cores batch system.

- ~ 40 PBs of data stored on disk and tapes.

Storage at CC-IN2P3: disk

Hardware

Direct Attached Storage servers (DAS):

- Dell servers (R720xd + MD1200)
- **~ 240** servers
- Capacity: **12 PBs**

Disk attached via SAS:

- Dell servers (R620 + MD3260)
- Capacity: **1.7 PBs**

Storage Area Network disk arrays (SAN):

- IBM V7000 and DCS3700, Pillar Axiom.
- Capacity: **240 TBs**

Software

Parallel File System: GPFS (**1.9 PBs**)

File servers: xrootd, dCache (**10.6 PBs**)

- Used for High Energy Physics (LHC etc…)

Mass Storage System: HPSS (**600 TBs**)

- Used as a disk cache in front of the tapes.

Middlewares: SRM (none), iRODS (**840 TBs**)

Databases: mySQL, PostGres, Oracle **(57 TBs)**

Storage at CC-IN2P3: tapes

Hardware

4 Oracle/STK SL8500 librairies:

- **40,000** slots (T10K and LTO4)
- Max capacity: **320 PBs** (with T10KD tapes)
- **106** tape drives

1 IBM TS3500 library:

- **3500** slots (LTO6)

Software

Mass Storage System: HPSS

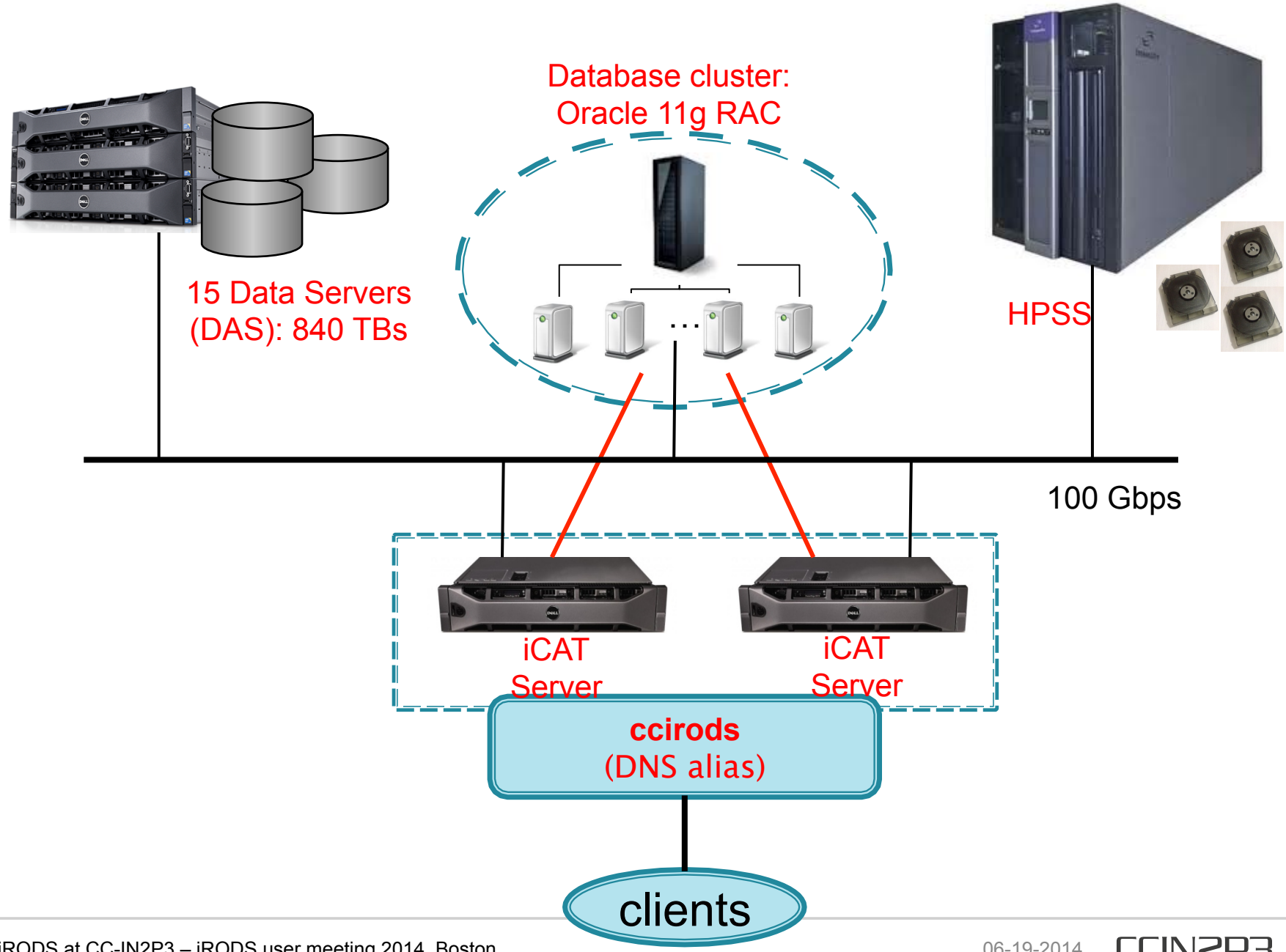
- **24 PBs**
- Max traffic (from HPSS): **100 TBs / day**
- Interfaced with our disk services

Backup service: TSM (**1 PB**)

SRB – iRODS at CC-IN2P3: a little bit of history

- *2002*: first SRB installation.
- *2003*: put in production for CMS (CERN) and BaBar (SLAC).
- *2004*:
  - CMS: data challenges.
  - BaBar: adopted for data import from SLAC to CC-IN2P3.
- *2005*: new groups using SRB: biology, astrophysics…
- *2006*: first iRODS installation, beginning contribution to the software.
- *2008*: first groups in production on iRODS.
- *2010*: 2 PBytes in SRB.
- *2009 until now*:
  - SRB phased out (2013) and migration to iRODS.
  - Evergrowing number of groups using our iRODS services.

Server side architecture



Features used on the server side

- iRODS interfaced with:
  - HPSS.
  - Fedora Commons (fuse).
  - Web servers (fuse).

- Rules:
  - iRODS disk cache management (purging older files when quota reached).
  - Automatic replications to HPSS or other sites.
  - Automatic metadata extraction and ingestion into iRODS (biomedical field).
  - Customized ACLs.
  - External database feeding within workflows.

iRODS users' profile @ CC-IN2P3

- Researchers of various disciplines:
  - Data sharing, management and distribution.
  - Data processing.
  - Data archival.

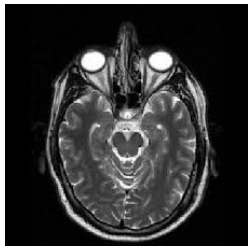
  - Physics:
    - High Energy Physics
    - Nuclear Physics
    - Astroparticle
    - Astrophysics
    - Fluid mechanics
    - Nanotechnology

  - Biology:
    - Genetics, phylogenetics
    - Ecology

  - Biomedical:
    - Neuroscience
    - Medical imagery
    - Pharmacology (in silico)

  - Arts and Humanities:
    - Archeology
    - Digital preservation
    - Economic studies

  - Computer science



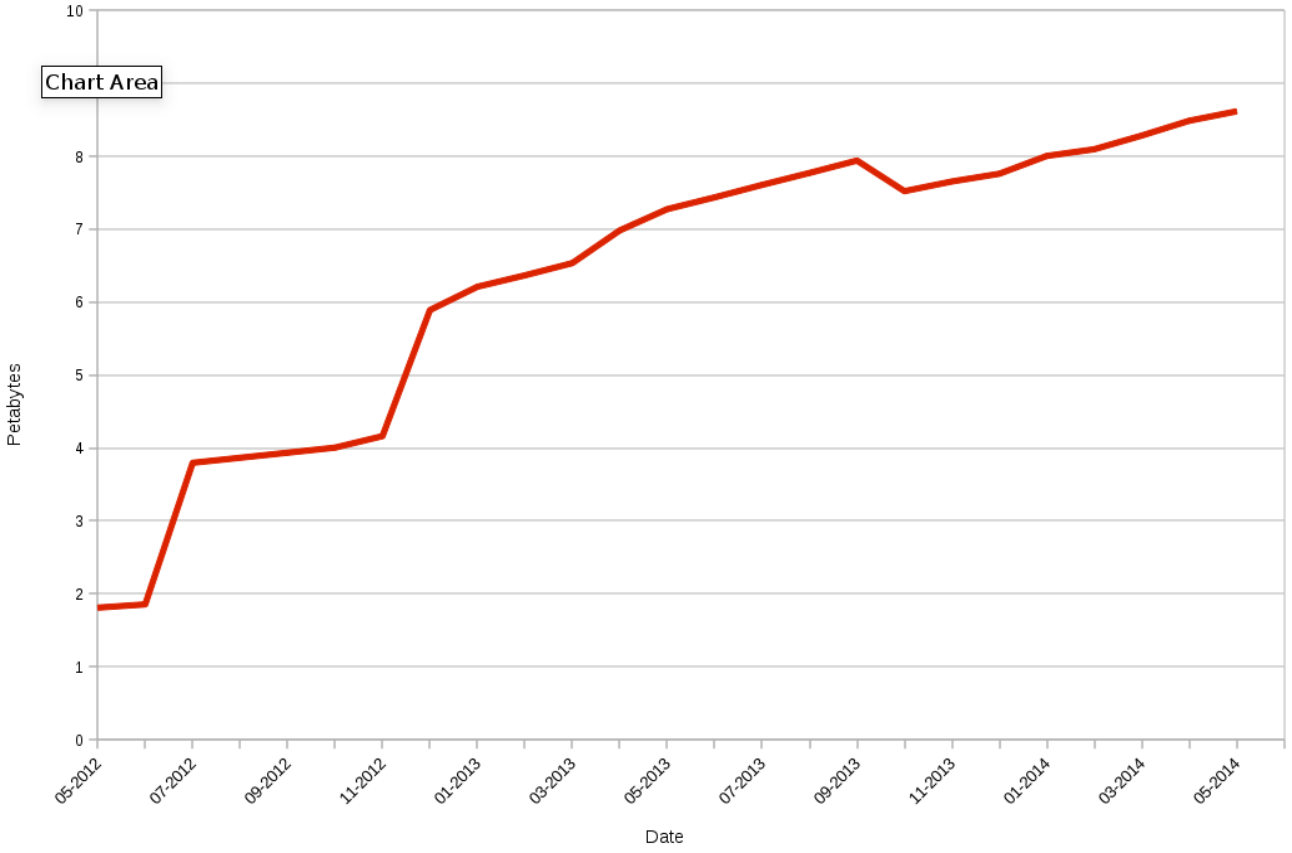
iRODS @ CC-IN2P3: some of the users

iRODS in a few numbers

- 23 zones.
- 39 groups.
- 469 users:
  - Maximum of 800k connections per day.
  - Maximum of 6.4m connections per month.
- 80 millions of files.
- 8560 TBs of data as of today:
  - Up to +30 TBs growing rate per day.

iRODS in a few numbers

iRODS storage evolution (PBs): 2012-2014

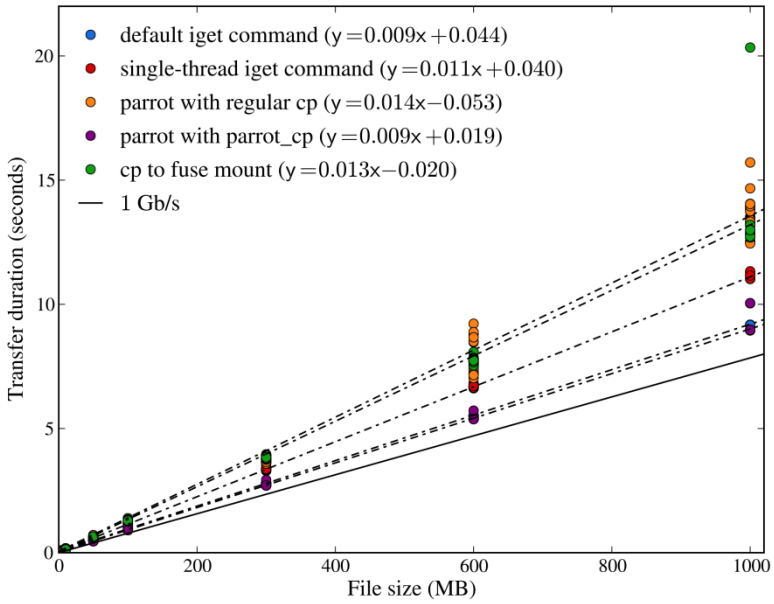


On the client side

- Clients' OS:
  - Linux (Ubuntu, Debian, Suse, Scientific Linux, CentOS…)
  - Mac OSX
  - Windows

- Using:
  - icommands
  - C or Java APIs
  - Fuse
  - Parrot
  - CC-IN2P3 provides the PHP web browser:
    - Also testing iRestServer from myirods

Performance tests: parrot, fuse, icommands



Upload files of different size to iRODS using:

- fuse
- iput
- Parrot

➔ Parrot performances closed to icommands

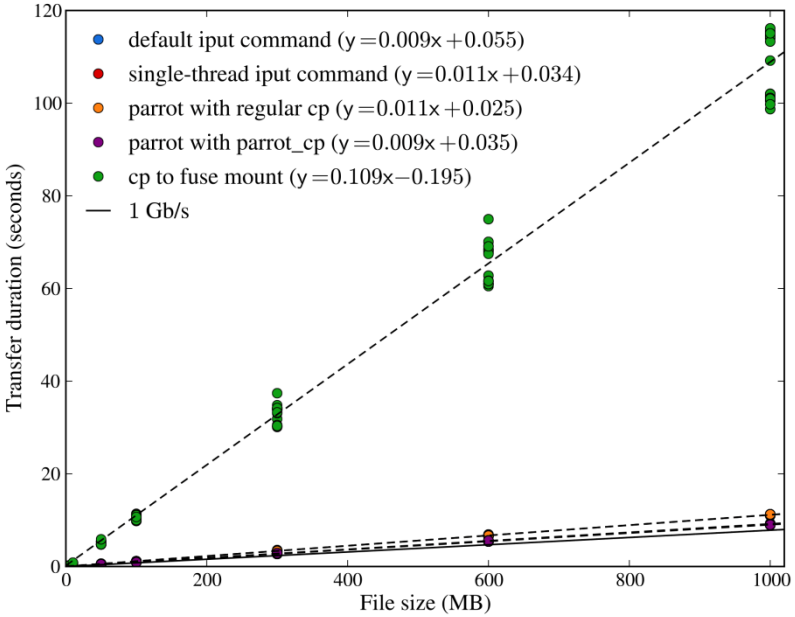
➔ Fuse performances differences between uploads and downloads.

(Credit: Quentin Le Boulc'h)

Download files of different size from iRODS using:

- fuse
- iget
- Parrot

➔ Parrot performances closed to icommands



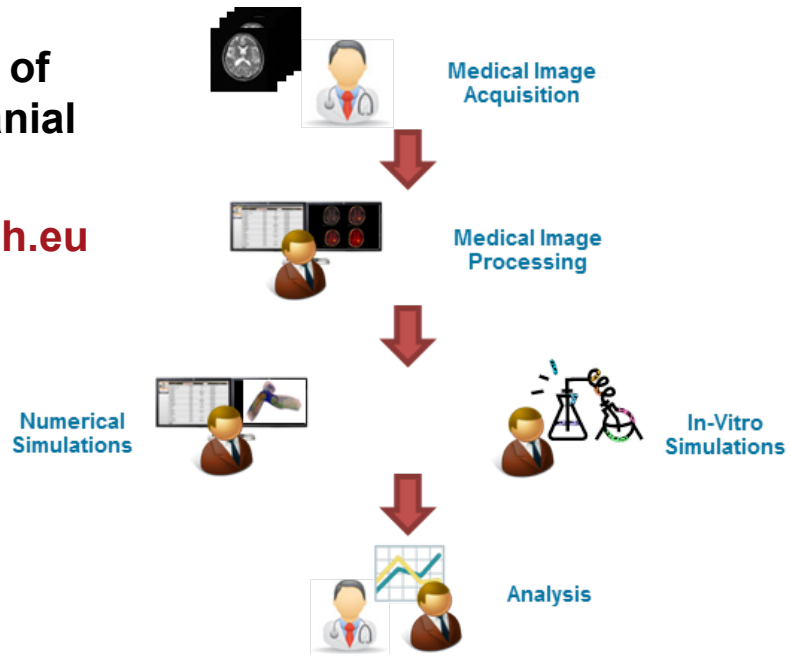
Biomedical example

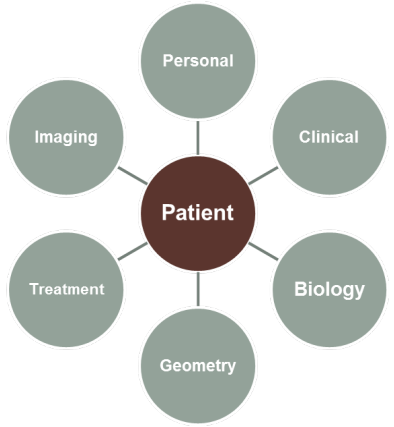
thrombus

A quantitative model of thrombosis in intracranial aneurysms

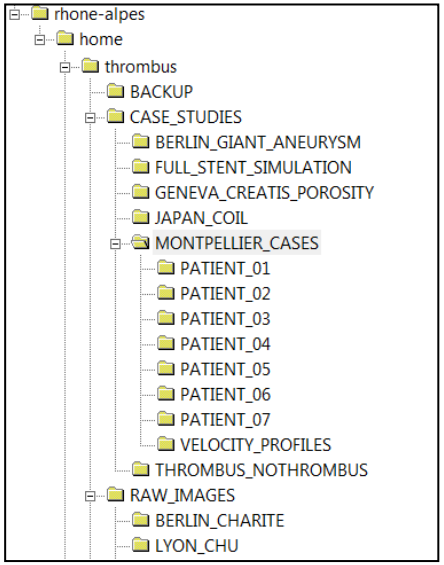
http://www.throbus-vph.eu

Virtual simulation of the thrombosis. Partners to correlate any type of data in case simultaneous multidisciplinary analysis is required.

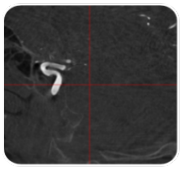




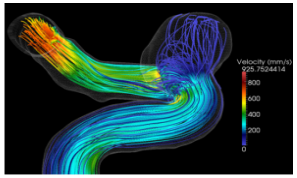
Multiple Patient Data



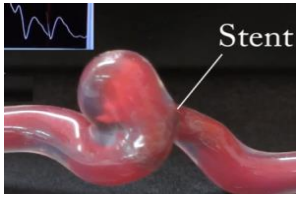
Data flow

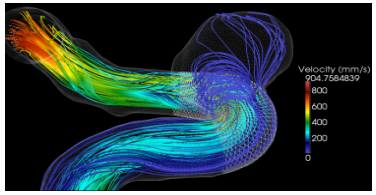


Medical Image



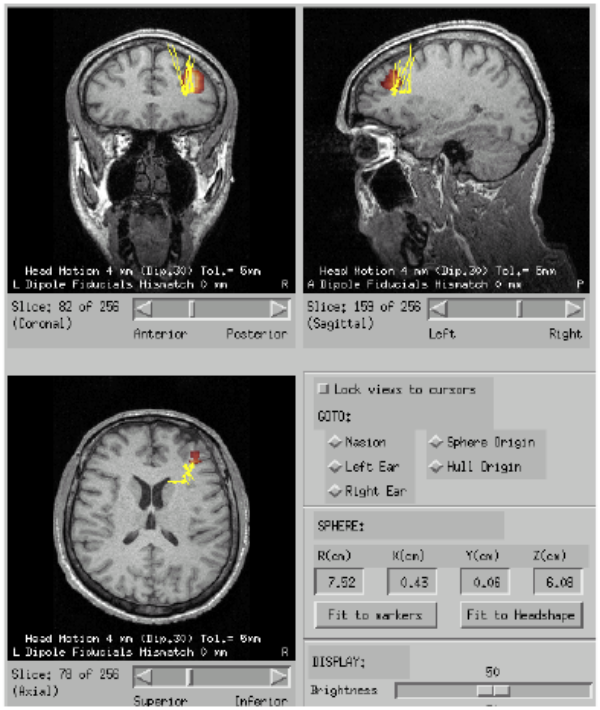
Numerical Simulation



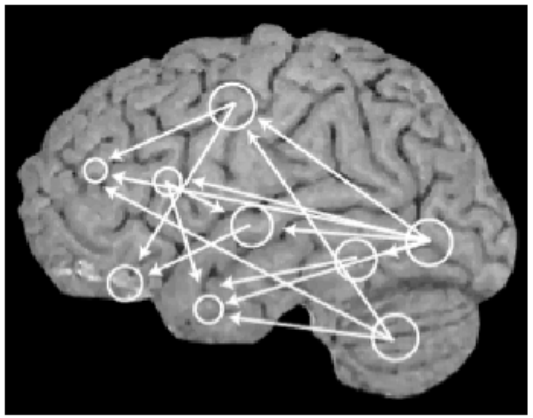


Biomedical example: neuroscience

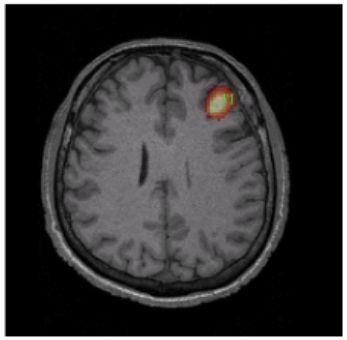
cermep
imagerie du vivant



Epilepsy treatment

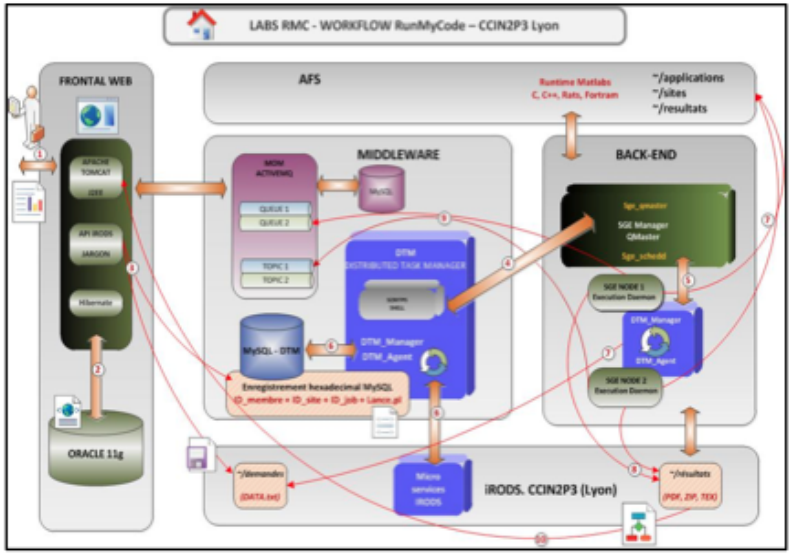


causality (RSVP)



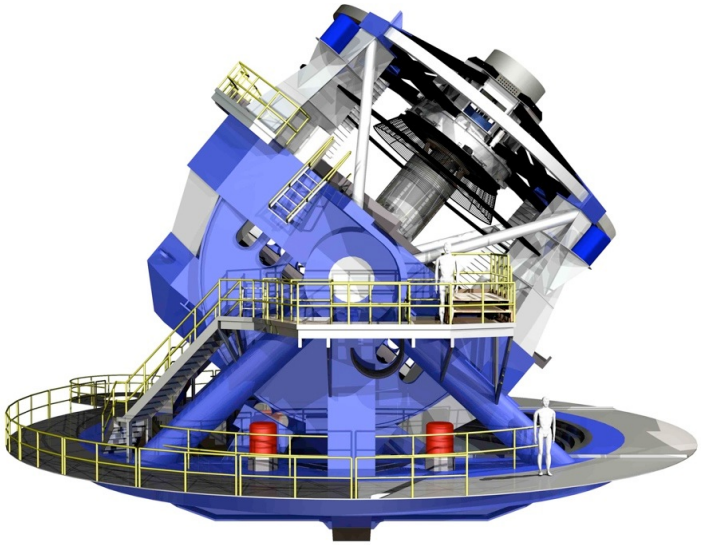
Arts and Humanities example

runmycode



Astrophysics example: LSST

- CC-IN2P3 :
  - half of Data Release Production
  - will host all the processed data

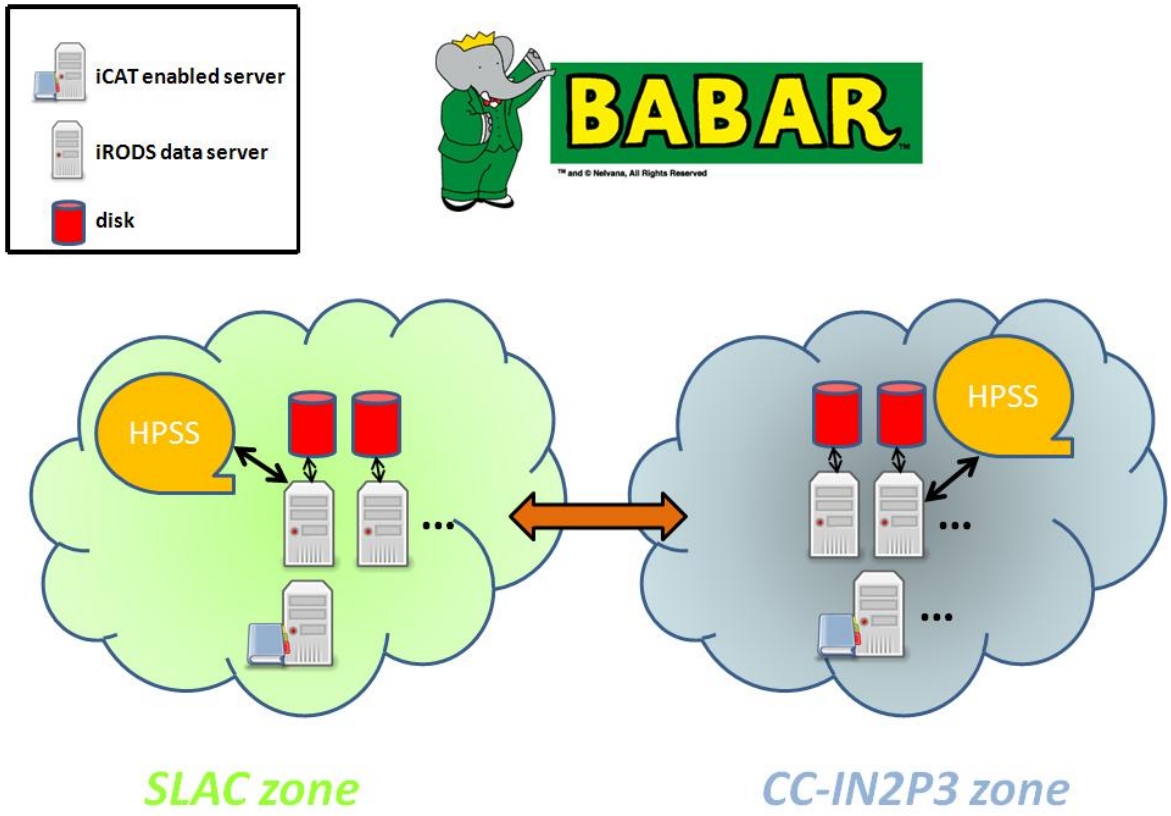


- iRODS :
  - Data Management :
    Raw images & Processed data
  - Data Transfers CC-IN2P3 ⇔ NCSA
  - Archival of 1 PB of data for the camera studies produced at SLAC ?

- Data Challenge 2013 :
  - SDSS Data processed by CC-IN2P3 and NCSA
  - Results shared using iRODS : **~100 TB**
  - Disks interfaced with Tapes (HPSS)

- **~100s of PBs** expected in 2030 ?

High Energy Physics 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.

Prospects

- iRODS:
  - key application for IN2P3 data management.
  - new big projects joining: LSST, Euclid.
  - user community still growing.

- Our concerns:
  - scalibility: database connections pooling needed.
  - iRODS v4.x :
    - OS portability on various systems.
    - « build in place » installation.
    - Oracle support.

- Our needs:
  - improvement in the Connection Control mechanism: interested to participate.
  - rule naming and priorities on rules (can have tens of thousands of rules to be executed).
  - SSL for uploads and downloads.
  - REST APIs.

Acknowledgement

At CC-IN2P3:

- Pascal Calvat (*user support: biology/biomedical apps, client developments*)
- Yonny Cardenas (*user support: biology/biomedical apps, client developments, rules*)
- Rachid Lemrani (*user support: astroparticle/astrophysics*)
- Quentin Le Boulc'h (*user support: astroparticle/astrophysics*)
- Thomas Kachelhoffer (*MRTG monitoring*)

At Huma-num:

- Pierre-Yves Jallud (*user support: Arts and Humanities*)

At SLAC:

- Wilko Kroeger (*iRODS administrator*)