Introduction to iRODS

Jean-Yves Nief

dapnia
cea
saclay

CENTRE NATIONAL
DE LA RECHERCHE
SCIENTIFIQUE

Storage virtualization

- Scientific collaborations spread world-wide:
  - Data can also be spread among different sites.
- Using heterogeneous:
  - storage technologies (disk/tape, file systems/hierarchical storage system/databases/(home grown) information systems/…).
  - operating systems (Linux: Red Hat/Ubuntu/CentOS/Suse…, Unix: BSD/Solaris 10/AIX/Mac OSX, Windows).
- Virtual organization needed:
  - Authentication and access rights to the data.
- Storage virtualization:
  - To be independent from technology and hardware evolution.
  - To be independent of local organisation of the files (servers, mount point etc…).

➔ Logical view of the data independent of the physical location.

Requirements/features (I)

- Logical organization of the data decoupled from the physical organization:

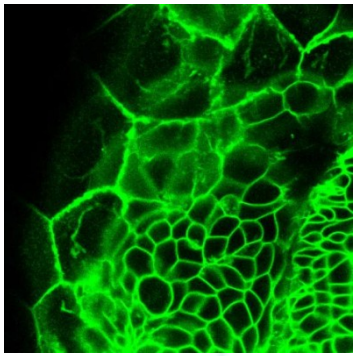
Nom Donnée	Date Creation	Propriétaire	Num. Replica	Num. Version	Taille	Type Donnée
nul01.evt.gz	2006-10-24-14.15.30	ant_write@ccin2p3	0	0	27808157	generic
nul02.evt.gz	2006-10-24-14.15.30	ant_write@ccin2p3	0	0	27778433	generic
nul03.evt.gz	2006-10-24-14.15.28	ant_write@ccin2p3	0	0	27209758	generic

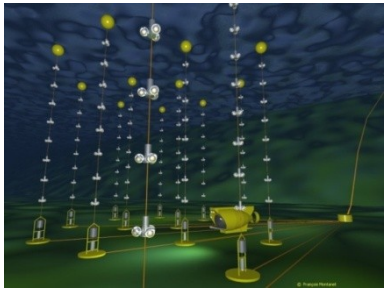
- Various client tools: GUI, Web, APIs (PHP, C, Java etc…), shell commands (icd, imkdir, iput, iget….).
- Authentication: password, certificate X509.
- Organization of the users' space by:
  - Type (sysadmin, domainadmin, simple user…).
  - Zones, domains, groups.
- ACL on the objects and directories.
- Tickets: temporary rights on a file.

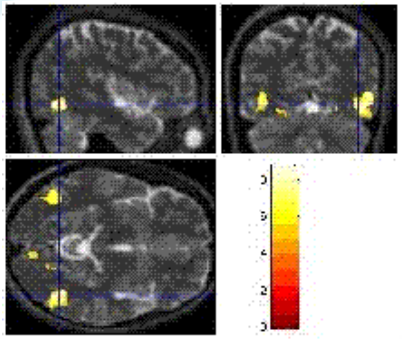
Requirements/features (II)

- Replica and versions handling.
- Access to the data from their attributes instead of their name and physical location.
- File search by metadata associated to them.
- Files annotations.
- Auditing: record of all the actions on the system.
- Storage resources hierarchy:
  - Logical resources: set of physical resources.
  - Compound resources: cache resource (temporary eg: disks) + 1 archival resource (ex: tapes).
- Ability to interface the system with any kind of information systems, storage system (like HPSS).

SRB (iRODS predecessor)







....

- used in HEP, astroparticles, biology, biomedical projects since 2003.
- 3.7 PBs of data referenced and handled by SRB in 2012.
- will be phased out in 2013.

Beyond storage virtualization

- Storage virtualization not enough.
- For client applications relying on these middlewares:
  - No safeguard.
  - No guarantee of a strict application of the data preservation policy.
- Real need for a data distribution project to define a coherent and homogeneous policy for:
  - data management.
  - storage resource management.
- Crucial for massive archival projects (digital libraries …).
- No grid tool had these features until 2006.

Virtualization of the data management policy

- **Typical pitfalls:**
  - No respect of given pre-established rules.
  - Several data management applications may exist at the same moment.
  - Several versions of the same application can be used within a project at the same.

  ➔ *potential inconsistency.*
- **Remove various constraints for various sites from the client applications.**
- **Solution:**
  - Data management policy virtualization.
  - Policy expressed in terms of rules.

A few examples of rules

- **Customized access rights to the system:**
  - Disallow file removal from a particular directory even by the owner.
- **Security and integrity check of the data:**
  - Automatic checksum launched in the background.
  - On the fly anonymization of the files even if it has not been made by the client.
- **Metadata registration:**
  - Automated metadata registration associated to objects (inside or outside the iRODS database).
- **Small files aggregation before migration to MSS.**
- **Customized transfer parameters:**
  - Number of streams, stream size, TCP window as a function of the client or server IP.
- **… up to your needs …**

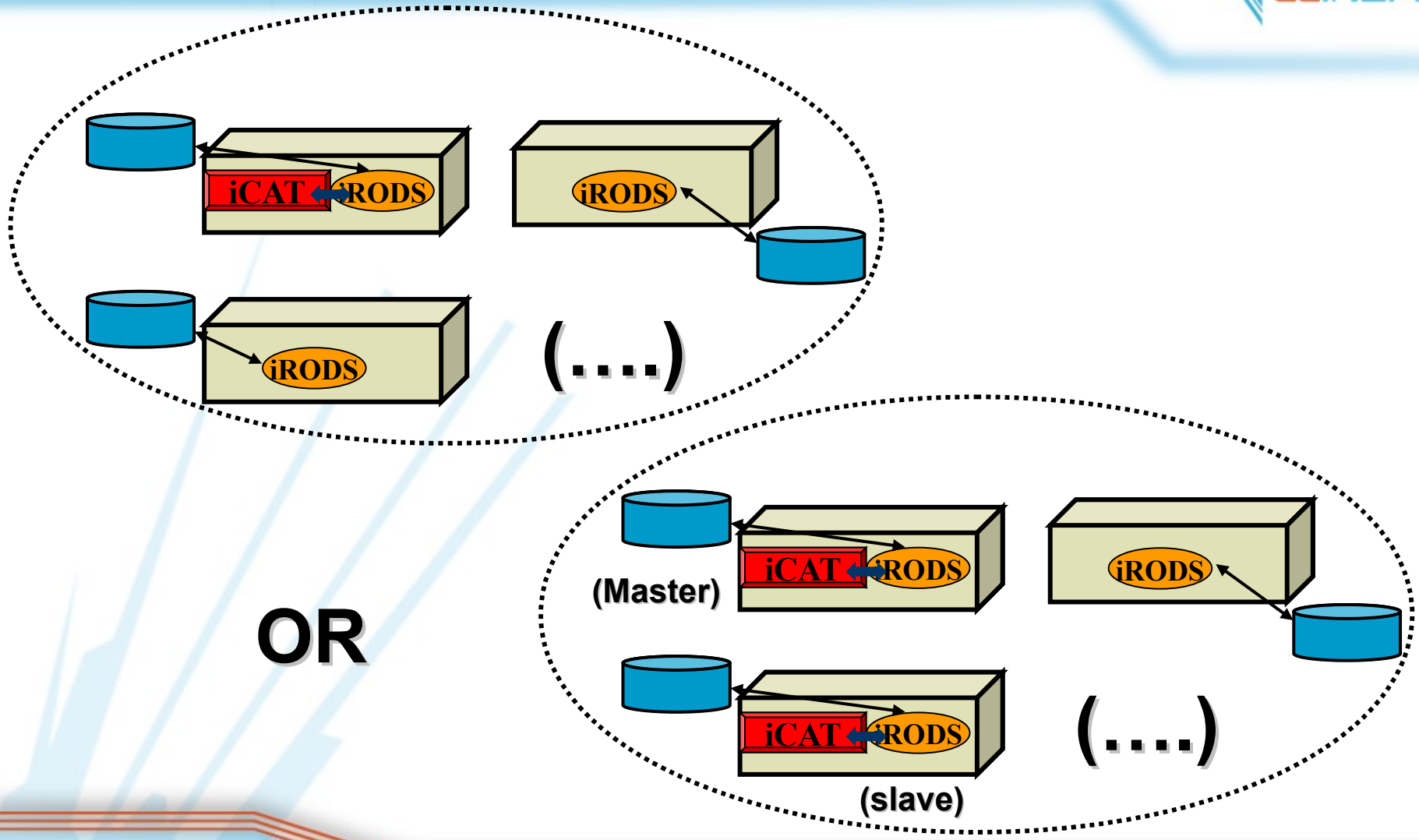
iRODS

- iRule Oriented Data Systems.
- Project begun in January 2006, led by DICE team (USA).
- First version official in December 2006 (v 0.5).
- Open source.
- Financed by: NSF, NARA (National Archives and Records Administration).
- CC-IN2P3 (France), e-science (UK), ARCS (Australia): collaborators.

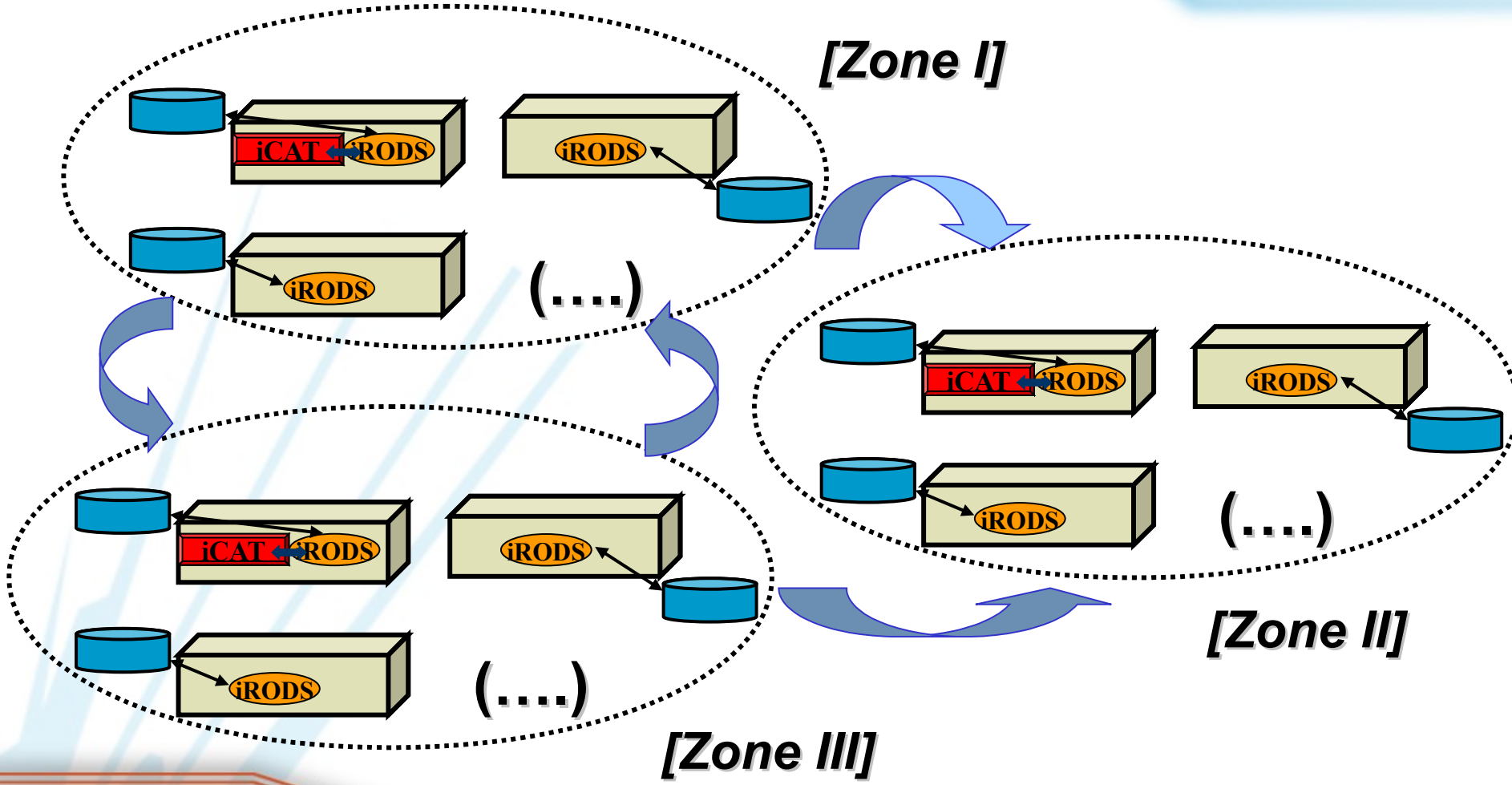
iRODS

- Based on the same ideas used in SRB.
- iCAT for iRODS ⇔ MCAT for SRB.
- But goes much further:
  - Data management based on rules build on the server side.
  - System can be fully customized without modifying any single line of the iRODS code.
  - Write your own services by adding your own modules.
  - Virtualization of the data management policy.
  - Logical name space for the rules:
    - Clustering in sets of rules.
    - Chaining the rules in a complex workflow (with a C like language).
    - Versioning handling.

iRODS architecture (1 federation)



iRODS architecture (2 federations)



Clients and APIs

- **APIs and some of the clients:**
  - C library calls …… Application level
  - .NET …….. Windows client API
  - Unix / Windows commands …. Scripting languages
  - Java I/O class library (JARGON) Web services, sites
  - Php, python …….. Web sites
  - SAGA …… Grid API
  - Web browser (Java-python) ….. Web interface
  - Windows browser …… Windows interface
  - WebDAV …….. iPhone interface
  - Fedora digital library middleware Digital library middleware
  - Dspace digital library …….. Digital library services
  - Kepler workflow …… Grid workflow
  - Fuse user-level file system ……. Unix file system
  - iDROP …… Drag and drop GUI

iRODS: the rules (I)

- A rule (prefix *ac*) contains:
  1. Name.
  2. Condition.
  3. Function(s) call: other rule(s) or micro-services.
  4. Recovery in case of error.

- A micro-service (prefixe *msi*):
  - Does a given task, can rely on internal functionnalities of iRODS.
  - 256 micro-services available (extensible at will).

iRODS: the rules (II)

- this rule will be executed automatically after a new file has been put into iRODS.
- It replicates the newly created file on two other resources and add a comment metadata on the first replica.

```
acPostProcForPut {
    on ($objPath like "/tempZone/home/cines-test/archivage/*.tar") {
        acAdonis1($objPath,"test");  # untar the file and register all the created files in iRODS
    }
    msiSysReplDataObj("demoResc4","null"); # replication on an other disk space called “demoResc4”
    msiSysReplDataObj("diskcache“,"null");  # replication on an other disk space called “diskcache”
    msiSysMetaModify("comment++++numRepl=0",“salut");  # modify the file metadata called comment for replica number 0
}
```

Example of iRODS users

- Irods @ USA/Canada
  - Canadian Virtual Observatory
  - NASA
  - iPLANT
  - NARA (US national archives)
  - Private companies (DataDirect Network ….)
- Irods @ France
  - BNF (national archives)
  - Cines
  - Ciment
  - Observatoire de Strasbourg (in production for the International VO services)
- Australia, Japan…

iRODS dev: to do list

- LDAP authentication.
- Shibboleth support.
- Msi structured object (allows request/transfer to any htpp, ftp etc… servers).
- Extension of NetCDF support.
- Support iCAT servers on Windows (with SQL server).
- Users request (from the chat, user group meeting).

iRODS @ CC-IN2P3

- iRODS: 1.7 PB with a growth of 5 TBs/day
  http://cctools.in2p3.fr/mrtguser/compta_irods.php
- Arts and Humanities (Adonis): 40 TB interfaced with Fedora Commons.
- HEP (BaBar): SLAC/LYON 3TB/day.
- Astroparticle (AMS, Double chooz): 2 TB/day.
- Biology / biomedical apps (Phylogenetics, embryogenesis, neuroscience, cardiology): 30 TB.

- Volume estimated for Dec 2012: 5 PBs.

iRODS @ CC-IN2P3 (II)

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

Area maintained by Thomas Kachelhoffer

Description:

1 786 TB are used at this time. These values were collected the *2012-05-09* at *11:27:02*. 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	49 805 GB	3 907 220 files
ams	349 407 GB	157 385 files
babar	713 063 GB	60 508 files
bao	142 816 GB	1 625 404 files
bioemergence	12 738 GB	3 756 235 files
codalema	1 952 GB	517 656 files
dchooz	478 146 GB	638 287 files
fazia	3 352 GB	8 470 files
general	7 145 GB	894 740 files
imxgam	547 GB	19 742 files
ipm	555 GB	215 files
test	45 GB	10 806 files
tidra	21 754 GB	8 034 479 files
tidra-neuro	17 360 GB	2 908 742 files
trend	30 620 GB	609 010 files
	1 829 305 GB	23 148 899 files

http://cctools.in2p3.fr/mrtguser/compta_irods.php

Some information

- For test accounts and questions:
  - contact: nief@cc.in2p3.fr
- Documentation:
  https://www.irods.org/index.php/Main_Page
- Discussion list: irod-chat@googlegroups.com