

iRODS: Setup and Use of a National Data Management System in the French NGI.

Wednesday, 21 May 2014 15:10 (20 minutes)

Scientific communities acquire more and more data and the need for tracking and managing of data in a simple framework is essential for providing scientific results in a timely manner. In France, several laboratories have gained experience with iRODS (Rule-Oriented Data Management System) which acts as a data grid by providing a transparent access to data which can be spread over different physical locations on heterogeneous storage technologies.

Five of these laboratories (CCIN2P3 in Villeurbanne, IPHC in Strasbourg, LPSC in Grenoble, MCIA and CBiB in Bordeaux) have joined their efforts by providing a single iRODS infrastructure in the framework of the French NGI.

This system is intended for hosting users from any scientific domain in need of storage or data management, and with little experience or dedicated personnel in the domain of data management. In this contribution, an overview of the iRODS features is given, followed by the description of the national iRODS instance.

The hosting of end users and the early adopter experience are also presented, as well as the development of the project towards building a national pool of expertise for the use and the administration of the iRODS product.

Wider impact and conclusions

The installation of the national iRODS infrastructure in France is fairly new. However, through the share of expertise the partners have build a powerful synergy. The fact that the partners come from different backgrounds is also an asset which allows them to better understand the needs of the user and to accommodate them.

Since the start of the project, we have been receiving positive feed-back from a number of communities and users. Following the implementation of our infrastructure, ten users have signed up to use the service. We will give special attention to the feed-back of these early adopters in order to further develop the service.

We have also been contacted by communities that are in the process of installing local iRODS systems and eager to share expertise regarding data management and user support. In this spirit, we also plan to develop a national pool of experts in the use and the administration of iRODS systems.

URL(s) for further info

<http://www.france-grilles.fr/Pour-les-chercheurs-ou-ingenieurs?lang=en>

<https://forge.in2p3.fr/projects/france-grilles-documentation/wiki>

<https://www.irods.org/>

<http://cc.in2p3.fr/>

<http://www.iphc.cnrs.fr/>

<http://lpsc.in2p3.fr/>

<http://www.mcia.univ-bordeaux.fr/>

<http://www.cbib.u-bordeaux2.fr/>

Description of work

The CCIN2P3 in Villeurbanne, the IPHC in Strasbourg, the LPSC in Grenoble and the MCIA and CBiB in Bordeaux have joined their efforts to provide a reliable data management system, based on iRODS, to be used by any French Scientific community. This project is operated in the framework of the French NGI which coordinates the project.

iRODS presents itself as a grid-like storage though iRODS also integrates a “rule engine” which implements data management policies. iRODS is highly configurable making iRODS appropriate for a diverse set of use

cases.

Technically, the components of our iRODS system are hosted by the current partners located in different regions in France: the CCIN2P3 hosts the iRODS database and rule engine, the other partners provide storage space for a total of 80 TB. We have also connected iRODS with the grid computing sites supporting the France-Grilles VO by deploying iRODS clients on them.

The infrastructure is working well and early adopters have just signed up. They come from various scientific fields like astrophysics, biology or human and social sciences.

Our team has also worked on the user hosting and the resource allocation process. In addition to write a comprehensive documentation, we have defined a procedure to welcome users. This includes training, support and a specific AUP describing the life cycle of the data. We also develop and distribute tools (like RPMs or VM)

to easily get the user started. In addition, we plan to review all the user projects each year.

This will help us to better understand the usage of our service and to globally adjust the resource allocation if needed.

The sharing of expertise between the partners has already build a powerful synergy. Our team has set up a working infrastructure, released user documentation, started a cycle of tutorials and early adopters have signed up. With their feed-back, we are working on consolidating the service, for example by organising a shared user support.

Primary authors: BISCARAT, Catherine (CNRS); PANSANEL, Jerome (CNRS)

Co-authors: HIROUX, Benoît (Université de Bordeaux); BENABEN, David (INRA); Dr NIEF, Jean-Yves (CNRS); GAY, Pierre (CNRS); Dr CARDENAS, Yonni (CNRS)

Presenter: PANSANEL, Jerome (CNRS)

Session Classification: New data management solutions for EGI

Track Classification: Requirements and solutions for data management and computing (Track Leaders: B. Konya, H. Heller, S. Tarkoma)