

Usage of grid for TREND data processing

Introduction

These are the conclusions of the meeting held at IHEP computing centre on April 12th 2012. The goal of the meeting was to discuss the possible ways to exploit the grid infrastructure supporting the France-Asia virtual organization for the data processing needs of the TREND experiment.

Agenda

<http://indico.in2p3.fr/conferenceDisplay.py?confId=6826>

Attendees

- TREND: Olivier Martineau, Thomas Saugrin, Fabio Hernandez (secretary)
- CC-IN2P3: Ghita Rahal, Yonny Cardenas

Conclusions

1. Since April 11th, and thanks to the intervention of Xiaofei YAN (IHEP) and Yonny CARDENAS (CC-IN2P3), a computing element operated by IHEP is configured to support the members of the France-Asia VO. There is no storage element because at this stage this does not seem to be necessary.
2. Thomas presented the current status of the TREND data processing chain (see slides). Three stages are implemented, including the phases of data acquisition, data transfer and data reduction. The offline data processing is performed at CC-IN2P3 using the local cluster. The reason of this choice is the proximity of the computing power to the data repository, which is managed by the iRods instance at CC-IN2P3.
3. Yonny presented an example of how to submit jobs to the grid sites supporting the France-Asia VO with detailed information relevant for the TREND use-case.
4. The agreed actions are:
 - Olivier and Thomas (and the future post-doc) will request a user certificate to the appropriate certification authorities (CNRS or IHEP). With this certificate, they will be able to register as members of the France-Asia VO.
 - The first goal is to prepare the tools for submitting jobs via the grid interface for execution at CC-IN2P3 only. This work will certainly require some modification of the current tools used by TREND for executing their workflow at CC-IN2P3. The jobs submitted via the grid interface will execute with a different identity (Unix id and group id) than the one used when the jobs are submitted locally to the batch farm at CC-IN2P3. In addition, the grid jobs will very likely lack the appropriate permissions to access the storage spaces currently used by TREND (AFS, SPS, etc.). This is considered the first step to prepare the TREND software to be site-agnostic and therefore to be executed in other sites than CC-IN2P3.
 - The second goal, once the first one is reached, is to modify the first stage of the offline data processing (see Thomas' presentation) in order not to depend on Matlab. According to Thomas and Olivier, this could be rewritten in Python or C so that it can be executed in another site where Matlab licenses may not be available. This however is a significant work that may take time.

- Yonny insisted that each individual must use his/her own iRods identifier when using iRods. The access writes to the multiple iRods directories belonging to TREND need to be modified to reflect the right owner and authorized individuals.
- 5. In addition, the possibility to deploy an instance of iRods at IHEP could be foreseen in the future.