PROCEEDINGS OF SCIENCE

Interoperating AliEn and ARC for a distributed Tier1 in the Nordic countries.

Philippe Gros
Lund University, Div. of Experimental High Energy Physics, Box 118, 22100 Lund, Sweden
philippe.gros@hep.lu.se

Anders Rhod Gregersen

NDGF - Nordic DataGrid Facility and Aalborg University, Aalborg, Denmark

Costin Grigoras

CERN - European Organization for Nuclear Research, Geneve, Switzerland

Jonas Lindemann

LUNARC, Lund University, Lund, Sweden

Pablo Saiz

CERN - European Organization for Nuclear Research, Geneve, Switzerland

Andrey Zarochentsev

St Petersburg State University, St Petersburg, Russia

To reach its large computing needs, the ALICE experiment at CERN has developed its own middleware called AliEn, centralized and relying on pilot jobs. One of its strength is the automatic installation of the required packages.
The Nordic countries have offered a distributed Tier-1 centre for the CERN experiments, where the job management should be done with the NorduGrid middleware ARC.
We have developed an interoperation module to allow unifying several computing sites using ARC, and make them look like a single site from the point of view of AliEn. A prototype has been completed and tested out of production. This talk will present implementation details of the system and its performance in tests.

13th International Workshop on Advanced Computing and Analysis Techniques in Physics Research, ACAT2010
February 22-27, 2010
Jaipur, India

PoS(ACAT2010)087

1. Introduction

ALICE [1] is a heavy ion experiment at the Large Hadron Collider (LHC) at CERN. To access the large computing power and storage required for the processing of the data created, the ALICE collaboration has developed a set of grid tools called AliEn (ALIce ENvironment [3]). AliEn allows to access the resources uniformly, and manage them centrally from CERN. The sites are managed using a front-end called a VO-box, which uses the local batch system to manage the resources, without any specific configuration on the nodes themselves.

The Nordic countries are individually small, but represent together an important contributor to the LHC. The decision was taken to pool their resources to create a distributed Tier-1 for the LHC computing grid. An entity called NDGF (Nordic Data Grid Facility [7]) was created to manage the Nordic resources and acts as a single entry point for CERN. The ARC middleware [6], developed by NorduGrid in the Nordic countries was chosen for the management of the resources.

The goal of this work was to develop an interface between AliEn and ARC to create such a distributed site. The data management is already unified using dCache [2]. The interoperation solution presented here addresses only the job management.

In the first part, a short description of the two middleware AliEn and ARC is given. Then we will introduce the principle and the motivation for the interface. In the fourth section, we show the mechanism in the realized interface. Finally, we describe the testing facilities and prototype performance, before concluding.

2. AliEn and ARC

The Grid has become a very important tool for science, but over time different strategies for job and data management are chosen by different communities.

This project involves two pieces of middleware, with a hierarchical relation. AliEn is the master, and the very structure of the system has to be understood to create an interface. ARC is used as a client, and only the client's features are relevant, and not the actual structure of the grid.

AliEn: AliEn is the grid environment developed by the ALICE collaboration at CERN. It is using some centralized services at CERN. The computer clusters are managed by a set of service running on a front-end, called the *VO-box*.

The AliEn job management is done using a pull model and pilot jobs. The computer clusters are managed by a so-called VO-box, running several services to monitor and manage the pilot jobs. These services manage and monitor so-called "Job Agents" (JAs), which run on the nodes and get and run the actual jobs, following a "pull" model.

A strength of AliEn is that it automatically installs the software required by the jobs. The package manager (PackMan) downloads, whenever required, a package (tar ball) and installs (unpacks and runs a configuration script) on a shared file system.

ARC: ARC (Advanced Resource Connector [6]), developed originally by the Nordic community, and now one of the most widely deployed grid middleware, is a highly distributed, multipurpose grid middleware. In ARC, the jobs are directly submitted from the client (user) to the batch system of a cluster, following a "push" model. The packages are installed on the sites by

the system administrators. Then, a RunTime Environment (RTE) script [8] is created to make the package available to jobs.

3. Interoperation

In order to create a distributed Tier1, the AliEn and ARC software have to be interfaced. This interoperation should allow creating a machine which would appear as a front-end to the AliEn grid, but can manage jobs over a pool of computer clusters, using ARC. This task has been initiated (see [2]), but no fully functional interface is used yet.

The very different strategies used by the two middleware are a challenge. The AliEn paradigm is based on a single user (the ALICE collaboration), which receives jobs from the members of the collaboration, and submits them to the computing resources. ARC, on the other hand, provides a tool for many users, from different communities (different Virtual Organizations). An interface between AliEn and ARC would bring advantages for several aspects of resource management:

Accounting and Operation Both accounting of resource usage and operation of the infrastructure are simplified. For the central ALICE management, only one entry exists, so the region is managed through one single point. For the Nordic community, the ALICE site would use the resources in the same way as any other Virtual Organization, and would be managed with the same tools. For that, they would benefit of the expertise already available in the region.

Flexibility From a technical point of view, such a hierarchical system is more appealing. It should improve the scalability of the system. More clearly it greatly improves the flexibility, allowing to easily move and create new resources at a regional level, keeping only the total to the pledged value.

On the other hand, we increase the load on a single gateway machine, but for the Nordic Tier-1, the combined size of all the federated sites should be comparable to a normal Tier-1 site.

The main requirement of such an interface is simplicity. For ALICE, the new VO-box should not create an exception: it has to be managed with the same configuration and management tools as any other VO-box. Therefore the code must be kept as much as possible to some simple modules, non invasive to the normal system. For NDGF, the ARC "sub-sites" should not run any special services. The VO-specific configuration should be done through an ARC RunTime Environment [8].

4. Interfacing AliEn with ARC

We have developed an interface to allow a VO-box to submit and manage jobs using the ARC middleware. ARC can easily be used for the submission of Job Agents on a local cluster. However, to have a single VO-box submitting to remote sites via ARC, the package management has to be adapted.

On an AliEn site, ARC can be used as a Local Resource Manager (LRMS) to manage the JAs. If the VO-box is set up for a single site (local area network and file system shared with the nodes), this works like any other AliEn site. The corresponding module is now included in AliEn and used in production on several site with high efficiency. However, if we want the VO-box to submit the JAs to a remote site, communication with the JAs through firewalls has to be considered.

Fortunately, all vital communication is done from the JA to the VO-box using SOAP. Therefore, only outbound HTTP connectivity is required, which is the most common situation.

In the case of a distributed AliEn site, the PackMan cannot rely on a shared file system to make the packages available to the nodes. The packages are therefore installed on a shared file system on each sub-site, using ARC jobs with special writing privileges.

The installation jobs use an AliEn command to install the package in the exact same way as the PackMan. That way, the package installation is as reliable as on any other AliEn site and the installation script does not require maintenance in itself.

Once the package is installed, an ARC RTE is created. The RTE contains all the relevant environment variables for the package, as created by the PackMan's post-installation script. It also instantly (within the one minute refreshing time of the database server of the ARC grid manager) advertizes the existence of the software on the site. The PackMan on the VO-box can therefore check the available packages on its pool of sub-sites. The RTE can then be required when submitting JAs with package requirements.

If the installation script fails, no RTE is created, and a flag is put on the VO-box to retry later.

4.1 Implementation

The AliEn native code is based on Perl. Besides a few minor modifications on the base code, the interface is contained into two Perl modules. One handles the submission of the Job Agents using ARC, in the same way as other LRMS like PBS. The second handles the package installation. The configuration of the VO-box in the central configuration database allows to use these modules instead of the default ones. It is also used to configure the system (by giving for instance the list of sub-sites).

5. Testing

An interface prototype has been tested in Lund. To avoid interfering with ALICE production, a test bed was created. A more limited version of the interface (not supporting multiple sites) is currently used for production on several Nordic sites. This test allowed to confirm the viability of the interface, and to address some potential issues.

5.1 A Simple Test-bed

To avoid interfering with the production grid at a critical time, a test bed had to be set up to test prototypes of the interface. A complete AliEn grid system was installed in Lund, with a corresponding Virtual Organization (VO). Its purpose was to validate the principle of the interface in a simple environment, not to test it in heavy load situation. The test bed had three components:

AliEn Central Services: The basic AliEn Central Services (CS) were installed on a machine in Lund and were associated to a test VO.

One AliEn VO-box: A single VO-box was created, on the same machine as the CS. There the interface was included (2 specific modules, and some minor modifications on some services). The site was configured accordingly in the CS database.

Two clusters running ARC: Two ARC sites participated in the exercise: LUNARC and Aalborg. These sites are used for grid production. On each site, an AliEn RTE was created. Directories were created for the installation of the packages and their associated RTE. Two grid users were added: one for running JAs, with normal user rights; the other for package installation, having writing privilege on the previous directories. A simple plugin had to be added in LUNARC to bypass a configuration that prevented writing possibility for the RTE directory.

5.2 Results and Observations

A few jobs requiring ROOT [4] packages were submitted to the AliEn CS. The VO-box installed the packages on the two sites. It then submitted JAs. The JAs ran the jobs successfully, producing outputs.

The test was successful as a proof of principle, though the scope of the project did not allow any quantitative measurement. However, it appeared that sites with long queues create minor problems (time out) at the installation stage. This should be improved by setting high priority to the installation jobs. Checking for RTE before retrying the installation greatly reduces the impact of such situations. In any case, even in a full scale system, package installation is not frequent, and this problem should not occur.

Besides, the interface limited to a single sites (and not addressing the question of package management) is being used for production on several Nordic sites (e.g. Aalborg). The efficiency and stability of the interface is comparable and maybe higher than with direct submission to the LRMS [9] (this has been evaluated by system adminstrator's experience, but not directly measured).

6. Conclusion

It is possible to create a distributed Tier-1 for AliEn. A prototype has been successfully tested in a reduced environment. A part of the interface for single sites is already used in production. A fully distributed site needs to be tested with high loads before it is used for ALICE production.

Since the ARC functionalities used in the interface are not very middleware-specific, the same work could probably be applied to other flavors of middleware such as gLite or UNICORE.

References

[1] ALICE Technical Proposal for A Large Ion Collider Experiment at CERN LHC, CERN/LHCC/95-71, 15 December 1995
[2] C Anderlik et al: ALICE - ARC intergration. J. Phys.: Conf. Ser. 119 (2008)
[3] Bagnasco, S. et al.: AliEn: ALICE environment on the GRID. J. Phys.: Conf. Ser. 119 (2008)
[4] Brun, R. Rademakers, F. Nucl. Instr. and Meth. A 389 (1997) 81; http://root.cern.ch/.
[5] Buncic, P. and Peters, A. J. and Saiz, P., and Grosse-Oetringhaus, J.F.: The architecture of the AliEn system, CHEP 2004, Interlaken, Switzerland (2004)
[6] Ellert, M. et al. Advanced Resource Connector middleware for lightweight computational Grids, Future Generation Computer Systems, vol 23, 2007, p. 219-240
[7] Fischer, L., Grønager, M., Kleist, J., Smirnova, O.: A distributed Tier-1. J. Phys.: Conf. Ser. 119 (2008)
[8] ARC Runtime Environment Registry. http://gridrer.csc.fi/
[9] http://pcalimonitor.cern.ch/map.jsp