Journal of Physics: Conference Series **119** (2008) 062006 doi:10.1088/1742-6596/119/6/062006

# ALICE - ARC integration

**C Anderlik[1], A R Gregersen[1], J Kleist[2], A Peters[3], P Saiz[3]**

[1] NDGF - Nordic Data Grid Facility, Kastruplundgade 22(1), DK-2770 Kastrup
[2] NDGF and Aalborg University, Department of Computer Science, Selma Lagerlöfsvej 300, DK 9220 Aalborg SØ
[3] CERN

E-mail: csaba@ndgf.org, arg@ndgf.org, kleist@ndgf.org, andreas.peters@cern.ch, pablo.saiz@cern.ch

**Abstract.** AliEn or Alice Environment is the Grid middleware developed and used within the ALICE collaboration for storing and processing data in a distributed manner. ARC (Advanced Resource Connector) is the Grid middleware deployed across the Nordic countries and gluing together the resources within the Nordic Data Grid Facility (NDGF). In this paper we will present our approach to integrate AliEn and ARC, in the sense that ALICE data management and job processing can be carried out on the NDGF infrastructure, using the client tools available in AliEn. The inter-operation has two aspects, one is the data management part and the second the job management aspect. The first aspect was solved by using dCache across NDGF to handle data. Therefore, we will concentrate on the second part. Solving it, was somewhat cumbersome, mainly due to the different computing models employed by AliEn and ARC. AliEN uses an Agent based pull model while ARC handles jobs through the more "traditional" push model. The solution comes as a module implementing the functionalities necessary to achieve AliEn job submission and management to ARC enabled sites.

## 1. Introduction

ALICE[1] is one of four experiments at the LHC[2], it is also one of the experiments to which the Nordic countries contributes. It has not been possible to situate a central Tier 1 data center in one of the Nordic countries, as each host country already has its computational resources in its respective countries.

The Nordic approach has been to utilize the existing heterogeneous computational and storage resources, and tie them together so it seems to come from one Tier 1 data center, the Nordic Data Grid Facility (NDGF), an organization described in the following section.

This article describes how NDGF expects to implement a solution that will make the computational resources for the ALICE experiment seem as one homogeneous resource for the AliEn (ALICE environment), the grid middleware used by ALICE.

In the next section we describe what NDGF is, followed by an overview of the Advanced Resource Connector, a grid middleware used in the Nordics in section 3. Section 4 gives an overview of ALICE and AliEn followed by a section on the AliEn/ARC interfacing in section 5. Section 6 deals with the data management challenges followed by a conclusion in section 7.

## 2. Nordic Data Grid Facility (NDGF)

The Nordic countries are each quite small (0.5-10 mio people), but together they represent a population of 25 mio people and rank as one the richest population groups in the world. In order to participate in huge international research projects it is more beneficial to pool resources with a common administrative and management entry point.

A trend during the last decades is the data explosion – today scientific experiments produce vast amounts of data and it is becoming increasingly difficult to store, access and process these data to produce scientific discoveries. So just like scientific computer centers today address the need for assistance in optimizing algorithms for the researchers there is also a need for optimizing access to databases and a need for storing these and accessing them, as they might be stored at several places.

The Nordic Data Grid Facility (NDGF) as a meta center connecting several existing scientific computing facilities tries to address exactly these needs - to enable access to many and distributed resources, to enable a pan Nordic storage system, and to enable easy access to all this through a single point interface.

The NDGF has been established with the primary goal of establishing a joint, Nordic production grid facility, leveraging the scientific computing resources and grid facilities already existing in the four participating countries. As such, the purpose of NDGF is to foster operational collaboration and joint policy, not to establish or own new hardware resources.

In addition, NDGF has a number of other goals:

- To act as a single point of contact for e-Science in the Nordic countries.
- To maximize visibility and impact of Nordic grid efforts outside the Nordic region, and to represent the community in international efforts.
- To host major e-science communities and projects, providing resources for coordination and project management.
- To provide middleware development efforts for specific requirements or projects
- To support and guide the development efforts of the NorduGrid community.

The main customers of NDGF are Nordic science communities with a need for collaborative work. The current example of this is the LHC[2] community – or, rather, the ALICE[1], ATLAS[3], and to some extent CMS[4] communities. However, NDGF is also working with other science communities with similar needs, in particular environmental science and bio-informatics.

## 3. Advanced Resource Connector

NorduGrid is a Grid Research and Development collaboration aiming at development, maintenance and support of the free Grid middleware, known as the **Advanced Resource Connector (ARC)** [5, 6].

ARC provides a reliable implementation of the fundamental grid services, such as information services, resource discovery and monitoring, job submission and management, brokering and data management and resource management. The middleware builds upon standard Open Source solutions like the OpenLDAP[7], OpenSSL[8] and Globus Toolkit libraries[9].

ARC consists of several services. First and foremost, sites running computations runs the Grid Manager, an ARC GridFTP server and the Local Information Service. The Grid Manager is a service running on a resource which takes care of jobs and a cache area to which input and output data is staged. Job submission and pre- and post-job data staging are made through the GridFTP server. Information services provides information about the capabilities of sites and detailed information about jobs by populating the information database stored in the Globus-modified OpenLDAP back-ends.

Indexing services are used in several places. Sites register themselves to a setup of several Globus GIIS back-end which allows building a hierarchical mesh of Grid-connected sites. This

allows clients to locate resources by querying the index. Data indexing gives the ability to replicate data a several locations to improve availably, as well as it give more flexibility in locating data through meta-data information. ARC middleware can use a variety of data indexing services, such as the Globus RC, RLS and the LCG File Catalogue (LFC).

ARC provides a light-weight command line interface to submit, monitor and manage jobs, move data and obtain resource information. This interface has a built-in broker, which is able to select the best matching resource for a job. Another special client is the Web-based Grid Monitor.

Storage Elements (SE) offer Grid-enabled secure access to disk-based storage capacity. ARC provides two solutions: the Conventional SE is based upon the ARC GridFTP server and supports advanced Grid-identity based authorization via GridSite GACL. The Smart SE implements a Web-service based solution featuring automatic reliable replication, increased data integrity and flexible access control. ARC can also interact with Storage Resource Manager (SRM)[10] based storage, such as the service that is used for storing data in NDGF.

## 4. ALICE and AliEn

**ALICE** - A Large Ion Collider Experiment at LHC [1], is one of the four LHC (Large Hadron Collider) experiments, currently being built at CERN, Geneva. Next year when the experiment starts running, it will collect data at a rate of up to 2PB per year and probably run for 20 years while generating more than $10^9$ data files per year in more than 50 locations worldwide. The data will be analyzed by thousands of scientists from hundreds of geographically distributed institutes, implying a highly distributed data flow, and therefore it is a very good candidate to benefit from the advances in the area of Grid computing.

The ALICE Grid analysis system is based on AliEn and ROOT and the jobs are controlled by an intelligent workload management system. The analysis starts with a meta-data selection in the AliEn file catalogue, followed by a computation phase. Analysis jobs are sent to the sites where the data is located, thus minimizing the network traffic. Both batch and interactive jobs are fully supported. The latter are "spawned" on remote computing elements and report the results back to the user's workstation.

**AliEn** [11, 12] - Alice Environment is a Grid framework built from a large number of Open Source modules and provides the following functionalities: shell like client access, virtual file catalog including meta data management, monitoring system, software management service.

In contrast to the push model traditionally implemented in other Grid systems, the AliEn Workload Management is based on pull architecture which is also applied to Data Management. The job description (JDL)[13] in the form of the Condor ClassAd is kept in a Task Queue while waiting for the Computing Elements to advertise their status and capabilities and to request jobs. While the jobs are waiting in the Task Queue, the Job Optimizers will inspect the JDLs, optimize and order the requests. The system can trigger a file replication to make a job eligible to run on a specific site in order to balance the overall load or enforce specific policies.

The system has been deployed for ALICE users at the end of 2001 for distributed production of Monte Carlo data, detector simulation and reconstruction at over 50 sites located on four continents. Up to now, more than 600,000 ALICE production and analysis jobs have been run under AliEn control worldwide during Physics and Data Challenge exercises.

## 5. AliEn/ARC interfacing

In order to integrate the ALiEn model with the NDGF facility we employed the VO-box approach, where there is a separate machine running all the AliEn/ALICE related services: CE, ClusterMonitor, PackMan and MonaLisa[14] and relays communication to the actual computing resources. To comply with the single entry point view of NDGF the final goal is to have a single

ALICE VO-box for the whole NDGF infrastructure, just as shown in figure 1. However, since this is somewhat cumbersome, we plan to achieve this in two stages.

In stage one, which is the current setup, we have installed an ALICE VO-box at each member site in NDGF, with jobs being executed through the appropriate backend interfaces. Let us take a closer look at the AliEn job model [11]: the AliEn Computing Elements monitor local resources and, once they have some shares available, they advertise themselves to the central service (CPU Server) by presenting their own ClassAds. The CPU Server will carry out the ClassAd matching against the descriptions of all tasks in the queue taking into account overall priorities and policies and, if any matching task is found, it will be given to the CE for execution. This will trigger the submission of an Agent job to the local resource manager system (LRMS).

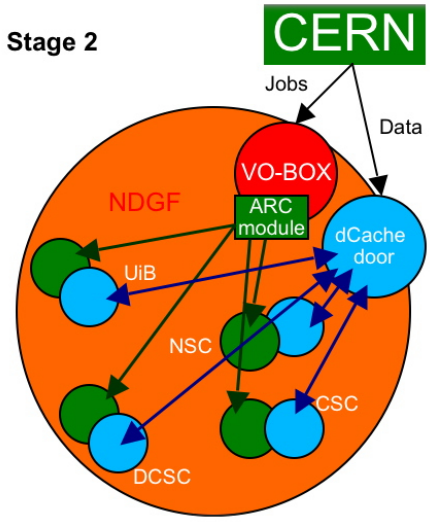


**Figure 1.** Single ALICE VO-box for NDGF T1, submitting JAs to ARC servers (green circles) at each participating node, data management is done using a single dCache door for NDGF; each local storage element will run as a dCache pool (blue circles) for the central dCache server.

The translation of the source JDL, and all the necessary environment settings are done through a Perl module tailored to the LRMS. NDGF being a distributed facility uses ARC as resource manager, therefore we implemented the corresponding Perl module to handle JDL to xRSL [15] translation as well job control using the ARC client interface (ng- commands).2. Once the JobAgent starts and picks up a task from the central AliEn task queue, the task will be monitored within the AliEN system, and the output of the task execution will be registered in the AliEN file catalog. This can then be accessed through the AliEn client shell. The ARC module is now a part of the AliEn distribution and can be installed optionally through the AliEn installer script.

The ARC interface is currently employed at a number of NDGF sites, and tested to see any performance and latency issues. We are also investigating what are the issues which need to be solved, if possible, before the second stage, single VO-box scenario can be achieved. The fact that a major part of the information about the different sites is stored in a static manner in the AliEN Central information system, is a big limitation in case of the single VO-box scenario and set of heterogeneous site characteristics. For example, the paths to the temporary and log directories differ from site to site.

Another important issue for the management of the distributed Tier-1 is accounting. The current solution employed by NDGF is the accounting system used in the Swedish National Grid project called SGAS (SweGrid Accounting System) [16], a JAVA based Grid

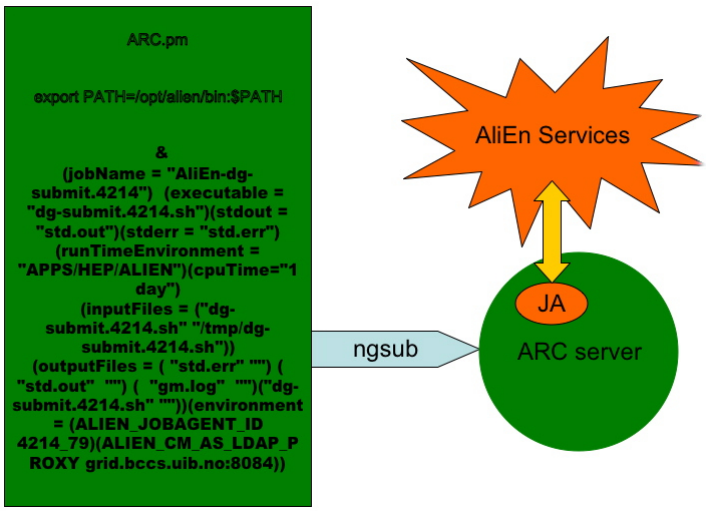


**Figure 2.** ARC module: sets up env, translates JDL to xRSL job description for the AliEn JobAgent (JA), JA xRSL submitted using ARC. The JA will then communicate with the AliEn services outside ARC. Agent monitoring is done through ARC as well.

banking/accounting system well integrated with ARC. There is ongoing work done by NDGF to integrate MonALISA[14], the monitoring system used by AliEn with SGAS for NDGF accounting. However this is out of the scope of this paper.

## 6. Data Management

The Nordic LCG Tier 1 is in contrast to other Tier 1 centers distributed over most of Scandinavia. A distributed setup was chosen for both political and technical reasons, this provides a number of unique challenges[17]. dCache is well known and respected as a powerful distributed storage resource manager, and was chosen for implementing the storage aspects of the Nordic Tier 1.

Basically dCache aggregates a multitude of disk and tape pools distributed over the Nordic region so they seem as one central piece of storage. This storage can be interacted with over a number of protocols via interfaces known as doors in dCache parlance. One of these doors is the xrootd door.

An NDGF ALICE VO-box is configured with dCache as its storage element via the xrootd door. The rest of the VO-boxes in the Nordic region are configured with this VO-box as their close storage element. Transfers are always done between the pool and the ALICE site. The central VO-box exists only to make the Nordic storage seem homogeneous and as such is not becoming a bottleneck with increasing amounts of storage and transfers.

Using this approach we have succeeded in having the distributed Nordic ALICE storage appear as a single unified storage element.

## 7. Conclusions

In this paper we present the current status of the integration work of the AliEn software components with the Nordic Data Grid Facility. We also address some of the challenges and their possible solution, encountered on the way to achieve the single access point view of NDGF for the ALICE experiment.

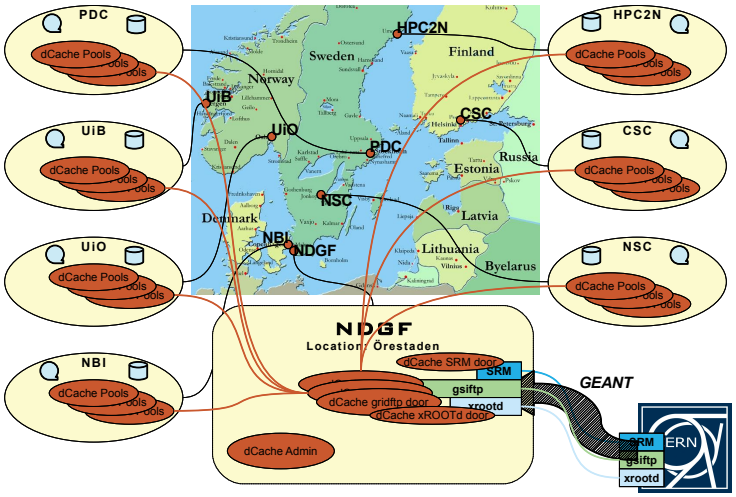


**Figure 3.** NDGF T1 storage setup, a central dCache installation masquerades the multitude of storage pools in the Nordic countries as one single storage element. A central ALICE VO-box is connected to the dCache installation via its xrootd interface, the rest of the ALICE VO-boxes share storage with the central VO-box making the Nordic storage seem homogeneous from within AliEn.

We show, that with some effort, it is possible to create a meta Tier 1 center with distributed ressources. Such a center has the advantage of being robuste, as ressources are distributed throughout the Nordic countries, and being possible to integrate from exisisting ressources.

## References

[1] ALICE Technical Proposal for A Large Ion Collider Experiment at the CERN LHC Tech. rep.
[2] Large Hadron Collider URL `http://lhc.web.cern.ch/lhc/`
[3] 1994 Technical Proposal for a General Purpose pp Experiment at the Large Hadron Collider at CERN Tech. rep.
[4] The Compact Muon Solenoid Experiment URL `http://cms.cern.ch/`
[5] M Ellert *et al* 2007 *Future Generation Computer Systems* **23** 219–240
[6] G Behrmann *et al* 2007 *ATLAS DDM integration in ARC* Proceedings of CHEP 2007
[7] OpenLDAP URL `http://www.openldap.org`
[8] OpenSSL URL `http://www.openssl.org`
[9] The Globus alliance URL `http://www.globus.org`
[10] The Storage Resource Manager Interface Specification URL `http://sdm.lbl.gov/srm-wg/doc/SRM.v2.2.html`
[11] P Buncic *et al* 2003 *The AliEn system, status and perspectives* Proceedings of CHEP 2003
[12] L Betev *et al* 2004 *The ALICE physics data challenge and the ALICE distributed analysis prototype*
[13] Condor Classified Advertisements URL `http://www.cs.wisc.edu/condor/classad`
[14] Monitoring Agents using a Large Integrated Services Architecture URL `http://monalisa.caltech.edu/`
[15] Extended Resource Specification Language URL `http://www.nordugrid.org/documents/xrsl.pdf`
[16] SweGrid Accounting System URL `http://www.swegrid.se`
[17] G Behrmann *et al* 2007 *A Distributed Storage System with dCache* Proceedings of CHEP 2007