First Challenges for the ALICE Tier2 centre at GSI

K. Schwarz
Gesellschaft für Schwerionenforschung, Planckstr. 1, D-64291 Darmstadt, Germany
E-mail: k.schwarz@gsi.de

Abstract. This article describes the work of the GSI Grid Group with the aim to enlarge and operate an ALICE tier2 centre within the global environment of the LHC Computing Grid and to prototype a possible FAIR grid environment.

1. GSI in the context of ALICE and FAIR
GSI (Gesellschaft für Schwerionenforschung) [1] operates a large and in many aspects worldwide unique accelerator facility and employs more than one thousand people. Researchers from Europe and from around the world conduct here experiments extending from nuclear and atomic physics to plasma and materials research, and encompassing biophysics and cancer therapy.

GSI maintains several local experiments, e.g. HADES [2], but participates also in the ALICE [3] experiment at CERN [4]. GSI and the surrounding universities are responsible for building the detector components TRD and TPC of the ALICE detector. At the same time the GSI related groups are developing corresponding software and are doing calibration and data analysis, mainly related to the local detector components. Moreover GSI operates a Tier2 centre within the ALICE computing Grid aiming at about 20% of the global Tier2 capacities.

Centered on GSI, in the years to come an international structure named FAIR (Facility for Antiprotons and Ion Research) [5] will evolve. The international FAIR GmbH was founded on October 4, 2010 and the first beam is expected by 2018. As can be seen in fig. 1 FAIR will practically use GSI as injector. Additionally, two new synchrotrons (SIS100 and SIS300), a Super-FRS, and several storage rings will be built. Among these one facility will be used for antimatter research.

The investment in computing at GSI was up to now dominated by the needs of the ALICE experiment but is shifting more and more towards FAIR requirements.

2. The ALICE tier2 centre at GSI
2.1. Technical setup
The deployment of the ALICE tier2 centre has been finished in time for the startup of the Large Hadron Collider [6] at CERN. The focus of the work moved to operating and upgrading the new infrastructure.

In the core of the ALICE tier2 centre is the central batch farm. The computing capacity has been increased up to 2700 cores.

The GSI batch farm is used by classical batch jobs arriving mainly via LSF [7] submitted by local users and as Grid jobs from ALICE and FAIR. Next to that interactive PROOF [8] sessions are activated by local ALICE users on the farm.
A first evaluation of (Sun) Grid Engine (SGE) as batch system for the future has been promising and first job families are being moved to the new environment. One of the first large scale tests with SGE is currently running by using the ALICE Grid jobs.

By mixing batch and interactive sessions as well as SGE on top of LSF (SGE jobs can be submitted depending on the current load of the target machines) a performance increase of the batch farm up to 10% has been achieved.

As operating system Debian Linux is being used at GSI.

The bandwidth of GSI to the world is currently one 1 Gbit/s which is mainly used by the ALICE tier2 centre. The plan for the near future is to extend this by a dedicated 10 Gb link to the German tier1 centre at GridKa [9] (Karlsruhe Institute of Technology KIT). Corresponding to the ALICE computing model GSI users should get a large extend of their data from the close tier1 centre which will be supported by the new infrastructure.

The available disk space at the GSI tier2 is distributed among an xrootd [10] cluster of 300 TB capacity and Lustre [11] (1 PB). The main xrootd cluster based on currently 42 file servers serves as strategic or permanent storage and is configured with a single point of entry (redirector) for GSI users and from the Grid. It is configured as a Grid Storage element and globally accessible. All files which are copied there are registered in the AliEn [12] File Catalogue. The Lustre file system with it’s POSIX [13] interface is used for local data storage. It is mounted by every worker node in the central batch farm (about 4000 cores) as well as by many interactive machines (in total about 100 cores).

The whole infrastructure, the local cluster as well as the whole ALICE grid infrastructure in which the ALICE tier2 centre at GSI is embedded, is monitored in detail by using MonALISA
The MonALISA system is an ensemble of agent-based subsystems registered as dynamic services, that are able to collaborate and cooperate in performing a wide range of information gathering and processing tasks. MonALISA provides agents to supervise applications, restart or reconfigure them and to notify other services when certain conditions are detected. Moreover it provides a very user-friendly and highly customizable Graphical User Interfaces to visualize complex information.

### 2.2. Cluster Virtualisation

The Scientific Computing department at GSI started a project to develop an Infrastructure as a Service (IaaS) prototype, as well as strategies and corresponding utilities to use these resources efficiently [15]. On 16 physical boxes a cluster of 100 virtual machines has been made available using OpenNebula [16] for cloud building, libvirt [17] as abstraction layer, and KVM as hypervisor. The cluster is being managed by using Chef [18] - via predefined recipes a complete PBS - cluster including server and virtual worker nodes with fitting applications already on board can be created on the fly. One of the first applications was a virtual ALICE Grid site which was running in production mode and computed about 42000 jobs for the ALICE VO.

Based on experiences gained at GSI the corresponding tools have been ported successfully to the environment of Frankfurt Cloud [19]. By using these virtual clusters on demand FLUKA [20] simulations for the GSI Radiation Safety department and nuclear structure calculations have been produced.

### 2.3. D-Grid and PROOF on Demand

Although the D-Grid project [21] officially came to an end the in this context developed ”PROOF on Demand” (PoD) utilities enjoy a high popularity. PoD [22] is a set of software tools developed at GSI via which each user can dynamically create individual PROOF clusters. As default 10 PROOF workers are being created per PoD instance, but this parameter is variable.

At GSI the transition process from a static PROOF cluster with xrootd processes being tightly integrated into the system to a PoD only system has been fully completed.

The development is going on and new releases are being published. Next to more supported plugins (currently PoD environments can be started via gLite, LSF, SGE, PBS, and ssh) also the user community is increasing. Altogether the software package has been downloaded so far 394 times with increasing frequency towards 2011. About 30% of the PoD user community is from Germany. The majority of users is distributed over many countries with Switzerland, France, USA, Russia, and Ukraine holding the largest share.

In the ALICE context PoD starts to be established as standard, and first ATLAS [23] installations have been done in Munich.

### 2.4. ALICE tier2 operations

Throughout the whole year GSI participated in all centrally managed ALICE production and data analysis activities. In the year 2010 the ALICE tier2 centre at GSI computed more than 290000 ALICE jobs. Many of these jobs run for several hours and produce several GigaByte of output. At maximum 800 Grid jobs are allowed to run in parallel. The overall job share of GSI has been 4% of all globally computed ALICE Grid jobs worldwide. The contribution of GSI, including local activities, comprises about 20% of the total ALICE Tier2 capacity.

Due to the first heavy ion runs at LHC end of 2010 the ALICE tier2 centre met it’s first “real data” challenge. The main tasks of a tier2 centre are Monte Carlo simulations and individual data analysis. Additionally, alignment and calibration are being processed for the detector components TPC and TRD. The Pb-Pb runs have been transferred to GSI continuously, and first corresponding papers have been published.
3. Preparation for FAIR tier0

PANDA-Grid [?] has been extended to 18 sites and upgraded to the newest AliEn version. The available resources sum up to 1200 CPUs and 300 TB of disk space. Among these sites 3 universities from the Philippines and 1 university from Thailand are participating. The collaboration with these sites has been strengthened by several working visits of Asian scientists at GSI.

The complete Panda framework has been installed on the Grid and the full chain of simulation, digitization, reconstruction, and data analysis has been exercised. First Fast Simulation productions started.

CBMGrid is up and running. First small productions have been running successfully at GSI using the CBM Grid infrastructure. The largest CBM Grid site is Dubna, supported by the JINR-BMBF grant.

The collaboration between ALICE and the FAIR experiments intensified significantly. Since the common software stack used by all experiments is quite large there have been many areas identified, especially in software development, where synergy effects are being experienced. Successful collaborative work has been shown in the development of the Virtual Monte Carlo (VMC) interface, the development of the experiment frameworks AliRoot and FAIRRoot [?] as well as in the context of Grid computing. Recent success stories are the fact that the cmake based build system has been taken over by the AliRoot/ROOT team, the development of the AliEn - Oracle interface as well as the prototyping of an inter VO Grid interface. A joined FAIRGrid - AliEn developers week has been held at CERN.

3.1. AliEn - Oracle interface

AliEn has originally been developed to run only by using MySQL as backend database. To make the system more flexible is has been necessary to enable the Middleware to run on any other database. As a first specific implementation Oracle has been chosen. AliEn uses the DBI (Database Interface) [?] module from Perl for creating the communication with the database. DBI makes porting an application from one database to another fairly easy as long as vendor specific SQL is not used. To avoid restrictions in functionality and performance it is needed, though, to use specific SQL functionality. The implementation has been chosen, therefore in such a way that most of the code is written in a compatible way for both databases, MySQL and Oracle, in the generic Database interface module of the Alien Middleware. Specific SQL expressions and functions have been implemented in submodules, one for each database in use. So far the performance when using MySQL or Oracle as backend database is comparable. But tuning work is going on. Currently CBMGrid is being ported to Oracle.

4. summary and outview

GSI IT is engaged in LHC computing since 2001. As part of the Worldwide LHC Computing Grid an ALICE Tier2 centre has been set up and is operated at GSI. GSI is developing software and operating procedures for LHC and is actively taking part in LHC boards and committees, e.g. the Technical Advisory Board and the Overview Board of the corresponding Tier1 centre at GridKa [9]. All these combined experiences in LHC computing are being used for the setup of the Tier0/1 centre at GSI and surrounding universities for the upcoming FAIR project. When comparing data rates and duty cycles FAIR computing will be in the same order of magnitude as LHC computing in terms of computing power and storage requirements. FAIRRoot started as a new implementation following the design of AliRoot and is now heavily used within the FAIR community. Many ideas from FAIRRoot go back to ALICE and CERN. FAIRGrid, as implemented for the PANDA [24] experiment within the PandaGrid project, is in production since 2003 and now heavily used on many sites over the Globe. PandaGrid can be considered as one of the first production ready components of the Panda experiment. The work in CBMGrid
started recently but it is fast evolving while profiting from the experiences gained in ALICE and PANDA. By using Grid and Cloud technology resources can be added to the FAIR computing infrastructure from various sources.

FAIR is not some vague idea in the far future. There are people, buildings, and work done. FAIR IT is used right now.

References

[1] GSI website, http://www.gsi.de
[2] experiment portal, http://www-hades.gsi.de/
[3] The ALICE experiment at the CERN LHC, The ALICE Collaboration et. al, 2008, JINST 3 S08002
[4] CERN website, http://public.web.cern.ch/public/
[5] project website, http://www.gsi.de/portrait/fair.html
[6] web page of the LHC, http://lhc.web.cern.ch/lhc/
[7] web page of Platform LSF, http://www.platform.com/workload-management/high-performance-computing/add-on-products/platform-lsf-family
[8] web page of the PROOF project, http://root.cern.ch/drupal/content/proof
[9] web page of GridKa, http://www.gridka.de/
[10] web page of the xrootd project, http://xrootd.slac.stanford.edu/
[11] web page of the Lustre file system, http://wiki.lustre.org
[12] web page of the AliEn project, http://alien2.cern.ch/
[13] wiki page about POSIX, http://en.wikipedia.org/wiki/POSIX
[14] web page of the MonALISA project, http://monalisa.caltech.edu/monalisa.htm
[15] Cluster Virtualisation at GSI, GSI Annual Report 2010.
[16] web page of the OpenNebula project, http://opennebula.org/
[17] web page of the libvirt project, http://libvirt.org/
[18] wiki page of the Chef project, http://wiki.opscode.com/display/chef/Home
[19] Radiation Simulation and Nuclear Structure Calculation at Frankfurt Cloud, GSI Annual Report 2010.
[20] The official FLUKA site, http://www.fluka.org/fluka.php
[21] web site of the D-Grid project, http://www.dgrid.de/
[22] PROOF on Demand, http://pod.gsi.de/, PROOF on Demand, GSI Annual Report 2010.
[23] web site of the ATLAS experiment, http://atlas.web.cern.ch/Atlas/Collaboration/
[24] PANDA experiment portal, http://www.gsi.de/panda
[25] CBM experiment, http://www.gsi.de/forschung/fair_experiments/CBM