

Lawrence Berkeley National Laboratory

Lawrence Berkeley National Laboratory

Title

Summary of Parallel Session I: Grid testbeds and applications

Permalink

<https://escholarship.org/uc/item/025035vk>

Author

Olson, Douglas L.

Publication Date

2002-10-10

Summary of Parallel Session I: Grid testbeds and applications *

D. L. Olson^a

^aLawrence Berkeley National Laboratory,
Berkeley, CA, 94720, USA

This paper is a summary of talks presented at ACAT 2002 in parallel session I on grid testbeds and applications. There were 12 presentations on this topic that show a lot of enthusiasm and hard work by many people in bringing physics applications onto the grid. There are encouraging success stories and also a clear view that the middleware has a way to go until it is as robust, reliable and complete as we would like it to be.

1. Introduction

The 12 presentations included in this summary are mentioned briefly in the next section with an expression of the main point of the talk, from this author's point of view. They cover a rather broad range of activities in Europe and the US and together provide an encouraging look at what it takes to build and run applications on a grid. A few observations and conclusions are given at the end.

2. Presentations, in no particular order

2.1. The Nordugrid Project

Aleksandr Konstantinov described "The Nordugrid Project: Using Globus toolkit for building Grid infrastructure", [1] a short term (18 month) project started in spring 2001, with partners in Denmark, Norway, Sweden and Finland, and has support from NorduNet2. This project shows a successful example of building a grid for the ATLAS experiment using the Globus toolkit [10] and other ingredients from the EU-DataGrid (EDG) project [2]. Globus provides the defacto standard for low level grid services. Scheduling, staging, accounting and user interface were developed from EDG components. A key design consideration used for the developments was to avoid hav-

ing any single point of failure.

2.2. GENIUS: a web portal to the grid

Roberto Barbera described the Grid Enabled web eNvironment for site Independent User job Submission (GENIUS) [3], a portal to the grid being developed as part of EU-DataGrid for the LHC experiments as well as other applications. A web portal provides access to grid services from anywhere, a common user interface to different grid dialects, security, and should be easily understood. The GENIUS portal provides access to services for security, data management, jobs, information, monitoring, interactive work and virtual organization (VO) management.

2.3. Interfacing Interactive Analysis Tools with the Grid

Doug Olson describe the work started in PPDG [16] on interfacing interactive analysis tools with the grid. A workshop [4] held in June in Berkeley was well attended and a set of use cases related to interactive data analysis were discussed. There was also a review of several existing technologies and tools, such as JAS, ROOT, Clarens, MCAT, and bitmap indices. A document [5] describing these use cases and requirements on grid services was prepared and includes identification of important issues to work on. These issues include needs for: standard data definitions, "grid object" definitions, debugging tools, common metadata catalog schema, match-making services and user interface portals.

*Authorship of this paper is supported by the U.S. Department of Energy Office of Science, the Office of High Energy and Nuclear Physics and the Office of Advanced Scientific Computing Research under contract No. DE-AC03-76SF00098.

2.4. Development of an Interdisciplinary Fragment of the Russian Grid Segment

Alexei Joutchkov described the start of a grid segment being developed in Russia for an interdisciplinary set of applications including IT, physics, biomedical, knowledge networks as well as portal development to support a broad range of science applications. The participating sites to date, in and around Moscow, include SINP MSU, IOC, EIMB, ITEP, IITP and RSI. There have been successful tests of running the BLAST application with globus. The next steps for this grid are to develop the requirements for the facilities to be used on this grid (hardware, software, middleware), metadata formats for the Knowledge Network, and requirements for the portal (user interface).

2.5. Data Challenges in ATLAS

Alexandre Vaniachine described the data challenges (DC) in ATLAS [6] and the integration with the developing grid middleware. The current DC1 is a scale of 10M reconstructed events and future DC's are at geometrically larger scales leading up to the start of the LHC in 2007/8. Reproducibility of results is important for grid computing and ATLAS has procedures to validate grid nodes to ensure they produce correct results. This is aided by use of the virtual data catalog (VDC) of the GriPhyN project[7] which has capabilities to encapsulate the prescription to generate data.

2.6. Remote Job Submission over GRID with IMPALA/BOSS

This described an example of interfacing the IMPALA/BOSS simulation system of the CMS experiment[8] to run jobs on a grid using the EU-DataGrid resource broker. The Dolly package was developed to provide this interface where the grid replaces the role of a traditional batch system. Two of the 4 stages of simulation for the Spring 2002 CMS production have been tested (CMKIN, CMSIN) and the additional two stages are underway.

2.7. Using Grid prototype infrastructure for QCD background study on Alliance Resources

Vladimir Litvin described a simulations production for CMS to generate QCD background processes in order to estimate their contribution to the Higgs decay to 2 photon channel. Condor[9] and Globus[10] were used for the grid middleware. Job control, via Condor Master, was run at Caltech using a Condor pool at Wisconsin and the NCSA[11] linux cluster, with data archived in the NCSA Unitree system. There were about $3 \cdot 10^9$ primary events generated and $1.6 \cdot 10^6$ events carried through full detector simulation and reconstruction. It was determined that the QCD processes account for about 15% of the total background for this decay channel. This "proof of concept" encourages continued development of the grid computing model to get real work done for physics experiments, though there is considerable work still to do.

2.8. ALiEn - ALICE environment on the GRID

Pablo Saiz described the ALICE[12] environment on the GRID, ALiEn.[13] Using a collection of open source software and components from Condor (Class Ads) and Globus (security) a rather complete environment for distributed computing for ALICE has been constructed, deployed and used for production. The available services include authentication, distributed file catalog, a task queue, configuration and information for computing and storage elements, a metadata catalog and a web portal. Nearly 6000 events, taking 100K CPU hours were generated using 13 cluster across 9 sites in Europe and the US during the last data challenge. As the middleware from the grid projects matures and stabilizes it will be integrated/interfaced with ALiEn in a way that leads to minimal changes to the user interface for physicist users.

2.9. The SAM-GRID Project, Architecture and Plans

Gabrielle Garzoglio described the SAM-GRID project where grid middleware is being integrated with the SAM data handling system[14]

for D0[15] (and now CDF). This work is part of the PPDG[16] (US) and GridPP[17] (UK) grid projects. This integration results in a Job Management Infrastructure and the Information and Monitoring (JIM) system additions to SAM. The Job Management infrastructure uses Condor matchmaking for the resource broker, Condor-G[9] for job scheduling and Globus GRAM[10] as interface to local batch systems. The information and monitoring services use Globus MDS to publish information from sensors on the compute nodes and file replica information from SAM. This is currently running at several sites in the US and the UK, and will be widely deployed for use by the physicists in a year.

2.10. Experience with OO Database for CMS Events Distributed Between Two Sites

Olga Kodolova described the experience of distributing data for the CMS experiment[8] with a single Objectivity/DB federation spread across two sites. From spring 2000 to spring 2002 nearly 1M events were generated at SINP MSU and ITEP and stored in Objectivity/DB. Using AMS in Objectivity it was possible to have a single federation.

2.11. Distributing Applications in Distributed Computing Environment

Natalia Ratnikova described a tool called DAR (Distribution after Release) developed at Fermilab for CMS for packaging and distributing releases of the CMS software. It will package up a local installation including the runtime environment, which can then be installed anywhere and reproduce the same environment. This is a very useful feature in a grid environment where the location and environment of nodes where jobs will run is not well known ahead of time.

2.12. Secure Automated Request Processing Software for DataGrid Certification Authorities

Lev Shamardin described the secure automated request processing for the Russian certificate authority, part of the X509 PKI needed for grid computing. The trust one places in the grid authentication certificate relies on the trust one has

in the certification authority (CA) that issued the certificate, so secure operation of the CA is very important. Using secure email (S/MIME) as the transport protocol end users can send requests to their registration authority (RA). After the RA verifies the information it is transferred once a day, using removable media, to an offline CA that signs the certificates. The generated certificates are then carried back to the online host where certificates, CRL, etc. are transferred back to the requestors using secure email and web.

3. Observations and Conclusions

It is clear from the level of activity and enthusiasm presented in talks here the the work to bring applications to the grid is expanding geographically and to more disciplines, both of which are very healthy signs for the grid. However, it is also clear that we are still in the early days of grid deployment where concerns about functionality and stability of the middleware is foremost in the minds of the application developers who are making use of it. The success stories of actually getting application work done on the grid give confidence to those who are just embarking on the effort. Interoperability of grid services is essential for the ultimate success of the grid model for the world-wide distributed applications like physics experiments. Perhaps a good summary statement of the current status is that while the management of grid projects think INTER-operability the people in the trenches building the grids think inter-OPERABILITY.

REFERENCES

1. <http://www.nordugrid.org/>
2. <http://www.eu-datagrid.org/>
3. <https://genius.ct.infn.it/>
4. <http://www.ppdg.net/mtgs/18jun02-lbl/agenda.htm>
5. http://www.ppdg.net/pa/ppdg-pa/idat/papers/analysis_use-cases-grid-reqs.pdf
6. <http://atlas.web.cern.ch/Atlas/Welcome.html>
7. <http://www.griphyn.org/>
8. <http://cmsinfo.cern.ch/Welcome.html>

9. <http://www.cs.wisc.edu/condor>
10. <http://www.globus.org/>
11. <http://www.ncsa.edu/>
12. <http://alice.web.cern.ch/Alice/>
13. <http://alien.cern.ch/>
14. <http://d0db.fnal.gov/sam/>
15. <http://www-d0.fnal.gov/>
16. <http://www.ppdg.net/>
17. <http://www.gridpp.ac.uk/>