GLAS-PPE/2002-05

UNIVERSITY
of
GLASGOW

Department of Physics & Astronomy
Experimental Particle Physics Group
Kelvin Building, University of Glasgow
Glasgow, G12 8QQ, Scotland
Telephone: ++44 (0)141 339 8855 Fax: +44 (0)141 330 5881

GRIDPP – Project Elements

A.T. Doyle
Department of Physics and Astronomy
University of Glasgow, Scotland
Email: a.doyle@physics.gla.ac.uk

On behalf of the GridPP Collaboration

Abstract

Key Results: Contributions to EU DataGrid and US-based (SAM) middleware. Rapid deployment of a testbed within the UK using EU DataGrid tools. Adaptation and interfaces to a range of particle physics applications.

GRIDPP – PROJECT ELEMENTS

Prof. A. T. Doyle (University of Glasgow), on behalf of the GridPP Collaboration.

Key words to describe the work: EU DataGrid, SAM, Particle Physics, HEP.

Key Results: Contributions to EU DataGrid and US-based (SAM) middleware. Rapid deployment of a testbed within the UK using EU DataGrid tools. Adaptation and interfaces to a range of particle physics applications.

Motivation (problems addressed): LHC Computing Challenge and the requirements of existing US-based particle physics experiments to harness distributed CPU and storage on an unprecedented scale.

UK Physicists are currently preparing for the turn on of the Large Hadron Collider (LHC) at CERN in 2007. UK Physicists are also participating in a number of US-based experiments that are already producing data. GridPP [1] is a 3-year £17m project which started on 1/9/01 and is funded by PPARC. There are three main developments within GridPP: Grid-enabled applications; Grid software (*middleware*); and, provision of computing infrastructure in the UK and CERN. The GridPP Collaboration aims to develop and deploy a large-scale science Grid in the UK for use by the worldwide particle physics community. Code is open source and (predominantly) used on linux platforms.
The elements of the project are shown in Figure 1. "CERN" {£5.7m} represents the LHC Computing Grid (LCG) which started on 11/3/02 and aims to establish the Grid as a 24×7 service on a world-wide distributed computing fabric where the LHC experiment applications can run. The high-level planning for the LCG is now complete.

"DataGrid" {£3.8m} started on 1/1/01 and refers to the EU development of Grid services, appropriate for particle physics, biomedical and earth observation applications. Initial technology and requirements surveys have been completed and 16 services have now been implemented, based on the Globus-2 toolkit and Condor-G. The project is characterised by rapid development, testing and deployment of middleware within a European-wide testbed. The performance of 140,000 lines of code has been evaluated within a testbed consisting of hundreds of servers distributed around Europe.
"Applications" {£2.0m} are the existing experiments' software which require interfaces to the Grid middleware. Typically O(1million) lines of code are required to analyse data from a collider experiment. Each of these experiments typically distributes Monte Carlo simulation jobs and has an existing (relational) database structure with a data hierarchy to enable efficient analysis. The process of (reverse-)engineering the code requires careful evaluation of Grid and existing experiment code dependencies prior to implementation.

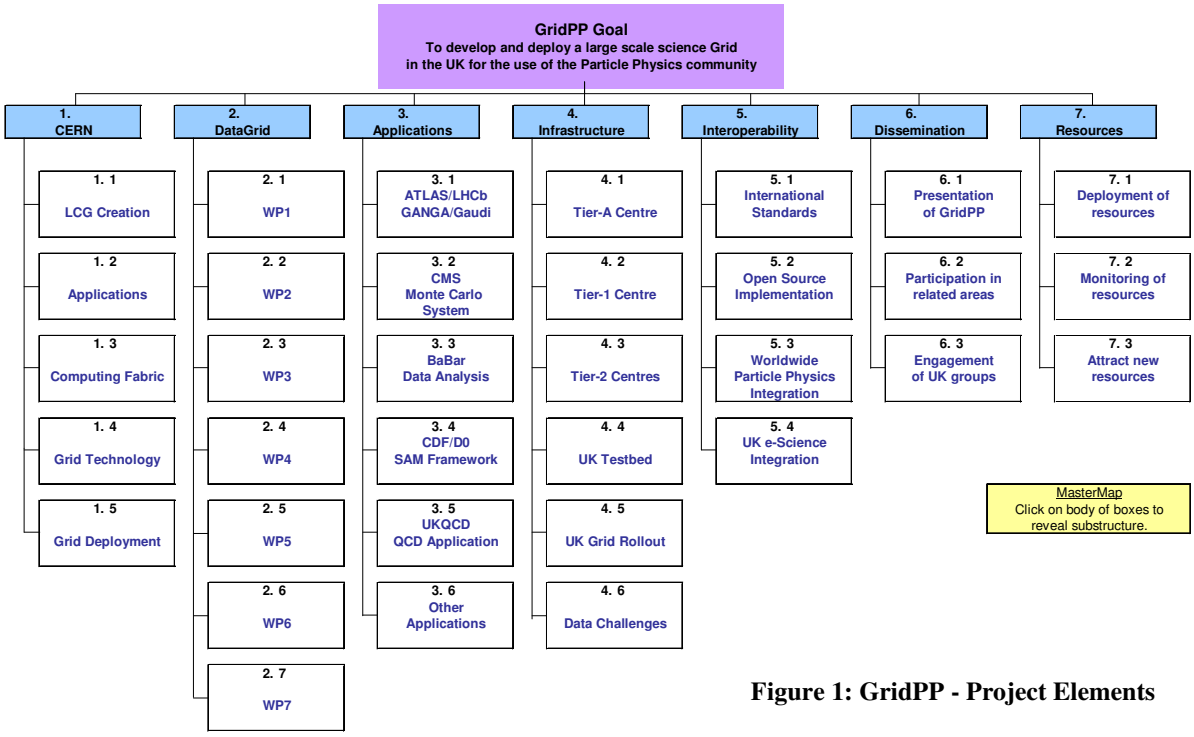

Figure 1: GridPP - Project Elements

"Infrastructure" {£3.7m} refers to the UK-based hardware and associated system management expertise necessary to deploy a Grid. The plan is to have O(2000) CPUs and O(1 PByte) of data accessible via the Grid within three years. A hardware assessment group determines the Tier-1/A annual purchase and balance of CPU and storage, appropriate for the LHC, BaBar and other experiments. Tier-2 resources are distributed around the UK and are being integrated via a UK testbed, which aims to roll out the Grid to the Institutes. In parallel, the data challenges of the experiments place immediate production requirements on this infrastructure.

"Interoperability" refers to (typically non-technical) issues such as the adoption of (emerging) international standards, adherence to open source implementations, integration with non-UK developments within particle physics and with the core e-science programme within the UK. "Dissemination" aims to inform external groups and Institutes within GridPP of relevant developments. "Resources" enables the allocated funding to be appropriately monitored. "Operations" {£1.8m} enables travel/participation in this world-wide activity by O(100) physicists and computer scientists, as well as project management. The project management tool is accessible via the web.

The presentation will concentrate on highlights from the first year of operations, emphasising UK deployment and developments within DataGrid. A number of issues arise and the aim will be to emphasise those of general applicability.

WP1: Workload Management builds upon Condor and provides a job submission system recognised throughout the testbed: UK effort is aimed at deploying and testing the system.

WP2: Data Management encompasses file replication and management of metadata using GDMP. In the UK work is focussed on secure access to replicated distributed data using Spitfire to access various databases, and investigations of how to optimise access to replicated files using the Optor simulation.

WP3: A Relational Grid Monitoring Architecture (R-GMA) has been developed in the UK providing an information service which allows application monitoring via servlets. These receive an http(s) request and provide an XML response on the current status of CPU, disk or round trip times between sites.

WP4: Configuration management uses the LCFG package, first developed in the UK, and adapted to meet DataGrid requirements, to manage and automatically install consistent software on each of the compute and storage nodes.

WP5: Existing storage systems require an interface to Grid file access methods such as GridFTP, GridRFIO, SlashGrid and future OGSA developments. The Storage Element aims to provide a generic interface with minimal access overheads.

WP6: A testbed is a necessity for any middleware development project. In the UK, the testbed team of four people provides the information and tools necessary to make the component parts work together and reacts to bugs reported via Bugzilla. The process of integration and testing of the middleware itself generates development of useful components. A spin-off is the development of GridSite which accepts Globus certificates and enables GridPP members to update pages on a central web server via a browser.

WP7: Network monitoring is fundamental to the development of a Grid. Existing tools such as Pinger and iPerf over TCP/IP or UDPmon provide the necessary input to the (existing MDS or future R-GMA) information service which can in turn be used to manage workload and optimise future data transfer/replication.

Security work is associated to this work package. Authentication is provided via 11 National Certificate Authorities (CAs). The current UKHEPGRID CA has issued 200 personal certificates using relatively primitive methods, but the plan is to migrate to the UK e-Science CA in the near future. The authorisation grid-mapfile is constructed from seven Virtual Organisations (4 LHC experiments, Biomedical, Earth Observation and Guidelines) using LDAP-based management tools from Globus.

WP8: In addition to providing the necessary interfaces to the HEP applications themselves, this work package provides requirements and evaluates whether those requirements are met in an iterative process.

A number of key elements are operational in the UK and have been tested and reported by the BaBar, UKDMC and LISA experiments. These tools also provide a framework for the UK QCDGRID project which aims to develop/manage XML-tagged data.

In addition to DataGrid development, GridPP is contributing to the development of the SAM system developed by the D0 experiment and recently adopted by the CDF experiment at Fermilab. The SAM system is Grid-like in many respects and has been recently interfaced to Condor in the UK.

The current status of the testbed can be found on the GridPP web site, see Figure 2.

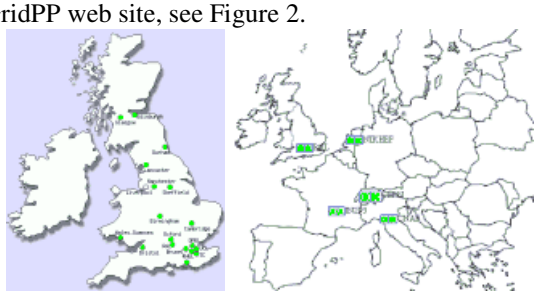

Figure 2: Maps providing the status of the testbed.

A number of GridPP posters are currently available. Three live demonstrations: "Exploiting the Grid to Simulate and Design the LHCb Experiment", "Building the Grid for BaBar" and "Dynamic Grid Optimisation" are currently available and new/updated demonstrations are under development.

References

1. The UK Grid for Particle Physics: http://www.gridpp.ac.uk/ and associated links.