PRODUCTION SERVICES FOR INFORMATION AND MONITORING IN THE GRID

Rob Byrom[3], Brian Coghlan[7], Andrew W Cooke[1], Roney Cordenonsi[4], Linda Cornwall[5], Abdeslem Djaoui <A.DJAOUI@RL.AC.UK>[5], Laurence Field[9], Steve Fisher[5], Alasdair Gray[1], Steve Hicks[3], Stuart Kenny[7], Jason Leake[8], James Magowan[2], Robin Middleton[5], Werner Nutt[1], David O'Callaghan[7], Manfred Oevers[2], Paul Taylor[2], Norbert Podhorszki[6], John Ryan[7], Antony J. Wilson[3], and Xiaomei Zhu[3].

1: Heriot-Watt, Edinburgh, 2: IBM-UK, 3: PPARC, 4: Queen Mary, University of London, 5: Rutherford Appleton Laboratory, 6: SZTAKI, Hungary, 7: Trinity College Dublin, 8: Objective Engineering Ltd. and 9: CERN

Key words to describe the work: relational, GMA, monitoring, information, EDG, DataGrid, EGEE, database, distributed, Grid

Objectives: To develop a Grid Information and Monitoring system for the European DataGrid, which fully meets the requirements of that project and be the basis for a secure, reliable and scalable web services based system for the follow-on project EGEE. To base it on the relational data model and SQL query language to allow for highly desirable queries such as "joins" to be processed efficiently.

To meet the requirements of the European DataGrid project, we have developed a novel information and monitoring system (R-GMA)[1,2]. R-GMA is a realization of the Grid Monitoring Architecture (GMA), as recommended by GGF[3], that also exploits the power of the relational data model and the SQL query language. For monitoring purposes, all published records carry a timestamp[4]. The biggest challenge during the development of R-GMA was to ensure that it could be scaled to operate in a large Grid reliably.

In the simplest mode of operation, GMA Producers of information publish their existence and the descriptions of the table they produce in a registry. Consumers (wanting information from the Grid) consult the registry to find suitable Producers, and following this "matchmaking" stage connect directly to the Producers either using a request/reply or publish/subscribe interaction pattern. This gives the impression of a virtual distributed database, and allows arbitrary correlations between information from different sources in the Grid. The power (and novelty) of R-GMA results from its "matchmaking/mediation" functionality.

To avoid single points of failure and to improve scalability, the registry needs to be replicated. A peer-to-peer policy for registry replication is adopted within R-GMA. Each registry is responsible for replicating its local data with other registries. Given two registries A and B, Registry A ensures Registry B has a copy of all information it masters and similarly Registry B ensures Registry A has a copy of all information it masters. A checksum is sent between registries to validate replicated data.

Over the last few years R-GMA has been extensively tested on many testbeds within DataGrid. Integration with other middleware and testing by application developers has resulted in a highly reliable implementation. The system is now being used in areas as diverse as resource discovery, job logging and bookkeeping, network monitoring and accounting.

To introduce it into the European DataGrid testbed easily, a special interface was written, which invoked existing LDAP information provider scripts and republished the information into R-GMA. The data was then collated with another Producer and written to a new LDAP database. This allowed R-GMA to co-exist and inter-work with the existing LDAP-based information system without the need for the users to change their code and resulted in widespread use of R-GMA on the application testbed.

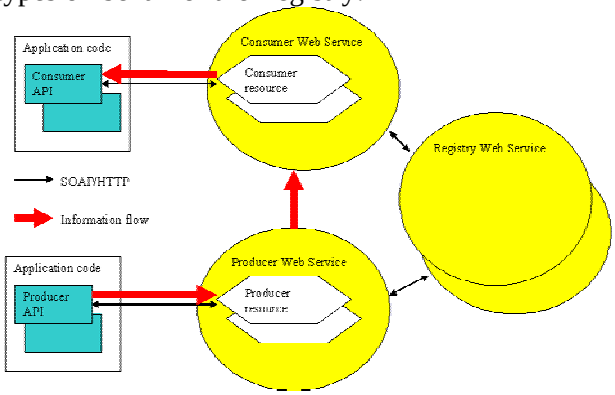
Performance figures taken from CMS particle physics trials of R-GMA for real time job monitoring revealed good scalability to thousands of jobs and many simultaneous users.

Network monitoring information within DataGrid also relies on R-GMA. This is mainly used to calculate the network "cost" function, which is the elapsed time of a file transfer for a file of particular size between two of storage sites. The cost function is available for optimising data replication across DataGrid.

Another use of R-GMA is for generating accounting information for LCG (LHC Computing Grid). Log files are first processed by the LCG individual sites

to generate accounting records which are streamed to the Grid Operations Centre (GOC). R-GMA was chosen because the process requires joins between data derived from files on different machines, for example from the compute farm machine's event log file and the Gatekeeper's log file. The major challenge was ensuring that recovery from network failures between the GOC and the LCG sites was easy.

A web services version of R-GMA is being developed, and the opportunity was seized to improve some aspects of the system. The functionality is exposed to a user (or application) through Producer and Consumer web services WSDL documents. However, for standard users, APIs are provided in many languages including Java and C++. Communication through these APIs with the web services relies on SOAP over HTTP protocols with an XML serialization. The Registry is also a Web service. A Schema web service (not shown in the diagram below) is used to hold the description of the tables, their column names and types on behalf of the Registry.



The thin arrows in the diagram denote invocation of the web services using SOAP over HTTP and the broader ones show the main information flow. In the implementation of the web services, a soft-state registration protocol is used to manage the lifetime of entities and resources that are create, making the system highly fault-tolerant following network or component failure.

The new API supports three categories of Producers: Primary, Secondary and On-demand Producers. Primary Producers are the initial source of the data. A Secondary Producer is used to aggregate streams of data and/or make them persistent. An On-Demand Producer interacts with a user-supplied plug-in that returns data in response to an SQL query. Producers and Consumers support a range of properties in order to cater for the variety of use cases. For example, Producers have a "type" property of either *History* or *Latest* and a "persistency" property of either *Memory* or *Database*. History producers propagate all of the information sent to them, whereas latest producers keep the most recent data. Memory producers are intended to hold data of short-term interest whilst database producers record it in databases so that consumers can recover historic information perhaps months or years old.

The new web services version of R-GMA will support authentication and authorization mechanisms based on the emerging Web services standards. Authorization rules define what actions an individual or certificate holder may carry out. This includes the ability to publish information (via a Producer), to query (via a Consumer) or to discover what Producers exist.

R-GMA is currently being integrated into the first EGEE prototype and should move into production in EGEE, where further testing and hardening will occur. Discussions with previous and current users have introduced a number of new challenging requirements. These will be addressed within EGEE.

Work continues within GGF to define information services for OGSA on the basis of experience with R-GMA. EU support through DataGrid under contract IST-2000-25182 and funding from PPARC through the GRIDPP project are acknowledged.

1. WP3, *Information and Monitoring (WP3) Architecture Report*, DataGrid-01-D1.2-0112-0-3, September 2001 (current version: http://hepunx.rl.ac.uk/edg/wp3/documentation/doc/arch.pdf)

2. Steve Fisher, *Relational Model for Information and Monitoring*, GGF, 2001, (http://www-didc.lbl.gov/GGF-PERF/GMA-WG/papers/GWD-GP-7-1.pdf)

3. Brian Tierney, Ruth Aydt, Dan Gunter, Warren Smith, Valerie Taylor, Rich Wolski and Martin Swany, *A Grid Monitoring Architecture*, GGF, 2001 (revised 2002), (http://www-didc.lbl.gov/GGF-PERF/GMA-WG/papers/GWD-GP-16-2.pdf)

4. Brian Coghlan, Abdeslem Djaoui, Steve Fisher, James Magowan and Manfred Oevers, *Time, Information Services and the Grid*, BNCOD 18, Chilton, UK, July 2001, (http://edms.cern.ch/document/349802)