GRIDPP-2/2.5

METADATA, DATA MANAGEMENT AND DISTRIBUTED DATABASES GROUP

FINAL EVALUATION OF METADATA COMPONENTS

March 2008

The GridPP Metadata, Data Management and Distributed Databases Group was established to understand, improve and catalogue the various methods by which data should be handled in GridPP-2 and GridPP-2.5. The three themes of development work indicated in the group's name were defined, with corresponding deliverables.

This is a final evaluation of the components produced as part of the Metadata activity.

1. Introduction

At the beginning of GridPP2, metadata posts were defined for work on the CDF and ATLAS experiments as well as a 'generic' post to work on issues common to HEP experiments. Over the course of the project, the exact scope of these roles has evolved, as have the people filling the positions. An overview will therefore be given of the activities undertaken by the metadata group, before going on to an evaluation of the final components. Details of the deliverables can be found in the Metadata, Data Management and Distributed Databases Group Logbook [1].

2. Overview of Metadata Activities

2.1 Generic Metadata

The Generic Metadata post was intended to work in the development of Grid technologies within a service-focused architecture for use in metadata based applications for the experiments, delivering fault-tolerant, reliable and manageable software for this purpose and using this technology for the enabling of existing experiments' metadata based products.

Initial survey and planning work was done, including an evaluation of OGSA-DAI, a generic architecture and planning report, and documentation of schema for metadata [2]. Contributions were made to the 'Core Use Cases for Metadata' document [3], which compared use cases from several HEP experiments to gain an understanding of the commonalities between them.

A number of collaborative ventures were then initiated. This included the running of a testbed service at Glasgow, allowing EGEE, LCG/ARDA, ATLAS and LHCb groups to perform testing. A wiki service was also set up at Glasgow, giving a central information point for interested parties. Collaboration was fostered with experiments and with EGEE and LCG/ARDA through the running of an annual workshop to discuss issues related to metadata. Collaboration with EGEE included contribution to the Grid Monitoring Data Exchange (GMDX) standard, including the production of an XML schema for it.

After the planning and setting up stages, it was decided that the most useful software contribution would be a generic monitoring service for a metadata service. After writing a requirements document for such a monitoring service [4], this resulted in the production of MonAMI, which is the main outcome from this post and is evaluated below.

2.2 ATLAS / CDF Metadata

The ATLAS and CDF posts were intended to enable the metadata requirements of the ATLAS and CDF experiments to be met by developing services for data set handling, analysis and job handling requirements on the Grid. Key tasks were to gain a conceptual understanding of the existing ATLAS / CDF metadata structures and the ATLAS / CDF specific use cases that drive them; develop, with reference to the use-cases and interactions with other ATLAS / CDF developers, the metadata necessary to support the navigational use cases; understand the analysis use-cases, particularly with respect to granularity of events to files; and implement fully working and documented solutions, working with the ATLAS / CDF teams to ensure that the developments were fully integrated with the rest of the ATLAS / CDF software, in particular, with the AMI / SAM products.

Due to changes in personnel and a re-targetting of requirements, these posts changed over the course of the project from having one person for CDF metadata and another for ATLAS, working on the ATLAS Metadata Interface (AMI) in particular, to having both posts dedicated to ATLAS – one on AMI and one on event-level metadata (also known as TAG data). Eventually, the AMI postholder was transferred to the ATLAS rolling grant and therefore no longer reported through GridPP. Over the course of the project, however, the following activities were undertaken.

The initial planning stages included documentation of metadata in CDF, particularly the SAM (Sequential Access via Metadata) schema, followed by the documentation and extension of the SAM Dimension Editor. On the ATLAS (AMI) side, documentation was written for the AMI database schema. Together with the Generic Metadata postholder, the 'Core Use Cases for Metadata' document was written.

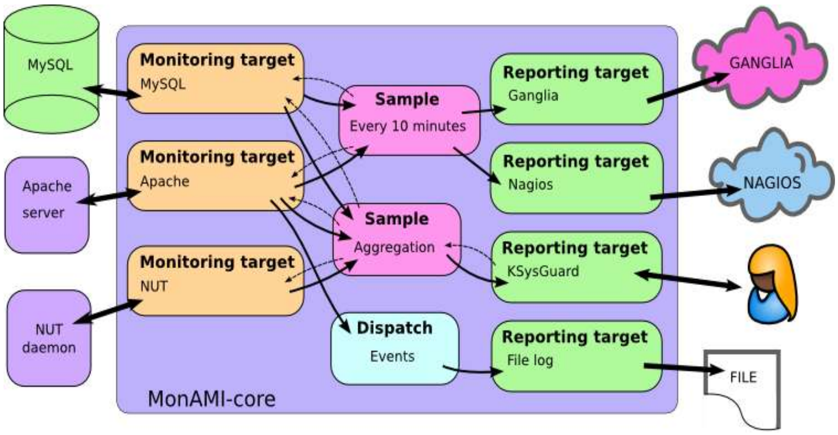
A gLite metadata catalogue interface for AMI was produced, followed by an SQLite backend for AMI, to allow use of AMI without being connected to the main database. The AMI postholder then worked with Gidon Moont at Imperial College, London, to produce Acacia – a generic VOMS proxy delegation solution which could be used by AMI to authenticate and authorise users.

On the TAG side, after familiarisation with the ATLAS event-level metadata scheme and distributed data management system, the Tag Navigator Tool (TNT) was developed to allow interfacing of the two systems. This was then re-worked into a plugin for GANGA, the distributed analysis framework, called GangaTnt. After performance testing and schema development work for the central TAG database, effort was concentrated on the development of ELSSI, the Event-Level Selection Service Interface, a web-based service which gives access to the TAG database and allows analysis using the selected events.

The main components produced through the ATLAS / CDF posts are therefore Acacia, GangaTnt and ELSSI, and these are evaluated below.

3. Evaluation of Generic Metadata Components: MonAMI

MonAMI aims to be a universal sensor framework. It can monitor many different services and send this information to any number of information systems. It does this whilst remaining easy to configure. Its plugin structure means that adding different monitoring targets is easy, as is including support for different monitoring systems. Its architecture is depicted below.



It has a large number of monitoring plugins, including AMGA (the Metadata Catalogue Project), Apache, dCache, Disk Pool Manager, GridFTP and MySQL. It also has a large number of reporting plugins, including those commonly used in the HEP Grid community, such as Ganglia, GridView, MonaLisa, Nagios and SAM (the Site Availability Monitor).

MonAMI is readily available from http://monami.sourceforge.net, where there is excellent documentation and a tutorial for users. It meets the original requirement of software which is fault-tolerant, reliable and manageable. It is unfortunate that, with the postholder leaving towards the end of the project, visibility and uptake of the software by LHC experiments has been lower than hoped, but there has been a steady growth in uptake, as evidenced by the usage statistics on the project page, and it has been used in data management monitoring for ATLAS.

4. Evaluation of ATLAS Metadata Components

4.1: Acacia

Acacia provides a Java implementation of various Attribute Certificate (AC) functionalities. In particular, it provides various client interfaces to the Virtual Organisation Management System (VOMS) system, as well as an AC server that can act as a VOMS Server.

It allows interaction with grid components through a web portal. Current web browsers have no concept of grid proxy certificates or VOMS servers, and the machine on which they are running may have no grid components installed. Acacia allows a short-lived proxy certificate to be generated from the grid certificate stored in a user's browser, and has it uploaded to an experiment's

VOMS server to allow the user to be identified and the grid to be accessed on that user's behalf, with the correct authorisation.

This proxy delegation service has proved crucial to the functioning of the ELSSI Skimming Service (below), allowing ATLAS users to submit jobs to LCG from their TAG selection on the web browser. It also allows VOMS-based authentication to AMI. As more web-based grid applications develop which need VOMS authentication, Acacia will doubtless continue to meet this vital need.

There are various improvements which could be made. The user experience in setting up Acacia for the first time is somewhat convoluted, and requires administrator privileges on the machine on which they are running in order to get the initial setup done. The interface and dialogue boxes could also be made more user-friendly. However, some of these issues are due to technical and legal difficulties with encryption software and beyond the developers' control. Acacia therefore meets its requirements satisfactorily for the present. It is available at http://sourceforge.net/projects/acacia.

4.2: GangaTnt

GangaTnt is a plugin for GANGA, the distributed analysis system used by ATLAS. It was written to allow a GANGA job to run a query on the TAG database, find the required input files by interacting with the Distributed Data Management system, DQ2, and send the user's analysis job to run on the LCG sites where the data is present.

Over the course of the project, it became apparent that most, if not all, queries to the central TAG database should be via the web interface, ELSSI (below). The main role of GangaTnt has therefore become the enabling of analysis using the selected TAG events extracted from the TAG database using ELSSI.

It is part of the regular GANGA releases, in the GangaAtlas package, and meets an important need in allowing physicists to use the event-level metadata system in their analysis. It is documented clearly in https://twiki.cern.ch/twiki/bin/view/Atlas/TagNavigatorToolGangaPlugin.

4.3: ELSSI

ELSSI, the ATLAS Event-Level Selection Service Interface, is the main interface for ATLAS physicists to query the TAG database. It guides the user step-by-step through the essential stages of a query – temporal cuts, data quality, triggers and physics attributes – then gives several possible courses of action. A simple COUNT can be done to check how many events are returned by the query. This can be used to refine it to a sensible number of events for returning, or for getting simple statistics. A selection of attributes can also be returned in this way, and histograms plotted. Finally, when the user is happy with the query, the events which satisfy it can be extracted from the database into a ROOT file for them to download and use in analysis.

Following this, they can choose to launch a 'Skim' job; that is, an analysis job which writes the AOD events corresponding to the TAG events extracted. GANGA and GangaTnt are used to run this analysis on LCG, with the resulting AOD being stored as a dataset in DQ2, the distributed data

management system.
Work on ELSSI will continue outside the GridPP project, as there is still much that can be done to improve it. However, responses from users at tutorial sessions has been very positive, and interest will continue to grow as people begin to use the TAG data in earnest with the coming of real data from the LHC. ELSSI can be accessed by ATLAS users with a valid grid certificate at https://atldbdev01.cern.ch/tagservices and is documented for users in a tutorial at https://twiki.cern.ch/twiki/bin/view/Atlas/EventTagTutorial.

5. Summary

The Metadata, Data Management and Distributed Databases Group have produced several key components as part of the GridPP metadata activity. While the major final products – MonAMI, Acacia, GangaTnt and ELSSI – have been reviewed here, there have also been several important documents and a continuous flow of development work as part of the wider collaboration with AMI, EGEE, ARDA and ATLAS Event-Level Metadata.

References

[1]http://ppewww.physics.gla.ac.uk/~caitrian/QReports/Final/MetadataLogBook_Final.pdf
[2]http://www.gridpp.ac.uk/datamanagement/metadata/Documents/HEPMetadataSchema_v1_01.pdf
[3] http://www.gridpp.ac.uk/datamanagement/metadata/SubGroups/UseCases/CoreUseCases_v10.pdf
[4]http://www.gridpp.ac.uk/datamanagement/metadata/Documents/MonitoringReq.pdf