

Database Access and Integration Working Group Meeting

Date: Monday 22 June 2002

Location: GGF5, Edinburgh

Present: 149 people

Chairs: Norman Paton, Dave Pearson, Leanne Guy

Norman chaired the session, beginning with an overview of the DAIS charter and mandate. Norman emphasised that the goal of the DAIS working group was to develop recommendations leading to standards for database access and services in a Grid environment, to encourage the adoption of these standards and to converge on a consensus as to what should be developed and implemented to achieve this goal. The group is currently at the stage of linking requirements to functionalities and forging working relationships amongst the various groups working on related topics. Norman hinted that the DAIS WG may evolve into one or more research groups in the future.

Norman then discussed what was in the scope of the DAIS WG: what functionalities are suitable for standardisation and what are not. Transparency to the materialisation of data, whether it be data in databases or file systems is a important topic. This implies an obvious overlap with the Data Replication WG and it was agreed that strong interaction between the two was essential. It was also agreed in the Data Replication WG that strong interaction between the two was necessary.

Presentation from five groups working in areas related to database access and integration were the given.

Dave Pearson, Oracle corporation:

Grid Database Access and Integration Requirements and Functionalities

Dave presented the paper on requirements and functionalities for grid database access and integration. How will the DAIS services support discovery, referencing, interrogation of data in heterogeneous databases? There are few standards that anyone conforms to presently; once data is discovered, what form will it be in? Performance, availability and scalability must be built into DAIS services. He focused on the priorities of the WG and the scope, stressing that overlap with other WGs must be avoided.

Inderpal Narang, IBM:

Data Access and Management Services on the Grid.

Inderpal presented a paper on data access, stating that data in any format should be considered and not just data in databases, i.e. file systems or anything that can expose data to a data virtualisation service. The challenge is to provide transparent access to any data sources within the grid. Inderpal stated that IBM are in an exploratory phase and will make no product commitment at this stage. Inderpal presented transparencies for data that are beyond the capabilities of current data management technologies and that

should be pursued. He defined data virtualisation services including core services to be used by clients as well as auxiliary services to be used by core services to improve and enhance performance and access. A proposal for Grid data virtualisation services to be developed in the near term was presented. Such a service aims to mask the underlying distributed, heterogeneous and autonomous nature of the data sources

Simon Laws IBM:

Grid Data Services - Relational Database Management Systems

Simon presented their paper on the development of relational database services for Grids. Usage scenarios to identify use cases and drive requirements were discussed focusing what database functions are or might be required, such as querying, bulk loading, management, etc. There is a lot of overlap with the work of the spitfire project.

Rob Baxter:

XML database specification

Rob presented the draft specifications for XML database services within the OGSA framework. This work is being done in the context of the UK eScience programme and focuses on OGSA based data access and integration services. They are focusing on XML databases and also data in files. This is similar to the work presented by Simon Laws (IBM) and Gavin McCance (Glasgow) on Spitfire however Rob and colleagues are considering XML databases. They have defined specifications for grid services for accessing XML databases and grid data service factories. The Grid XML data service provides a factory to create a grid data service upon request and a service for a user session to a database. They are currently focusing on formalising the architecture, incorporating the IBM work and intend to bring put a prototype at the end of the year for Grid based access XML and relational databases. They are looking to GGF to get input to assist with the formalisation of the architecture

**Gavin McCance Glasgow, Grid data management, EU DataGrid Project
Project Spitfire: Towards web service Databases.**

Gavin presented an update and status report on the latest stable Spitfire release. The Spitfire service grid-enables a wide range of relational database systems, providing a uniform service interface, data and security model, and network protocol. The recent work on progressing towards a web service based architecture, the security model and the client APIs was presented. Spitfire has currently been trailed using MySQL and PostgreSQL databases. The web services alpha release should be available soon.

Open Discussion and questions

Question: Who has prototype implementations and when will they be released?

Simon: not until September

Gavin: Spitfire version 1.1.0 is in production, the web services alpha will be available soon.

Rob Baxter pointed out that an open source licence must be agreed upon soon, hopefully soon after GGF5

Question: How should data from 1000000 files be streamed efficiently across the Grid?

Gavin: Spitfire is optimised for metadata access and has not considered this scenario

Malcolm: Currently defining functionality and APIs, we have ideas about how to send that sort of data efficiently, but this functionality would not be in an early deliverable.

Question: What services exist for the management of data distributed across multiple data stores?

Inderpal: commented that the federated management of data, the integrity of data across multiple data sources and tracking and management of data was important.

The meeting was closed due to lack of time and further discussions took place the day after.