

The gCube Metadata Framework: integrated environment for managing Metadata Objects and relationships on top of Grid-enabled Storage Systems

Wednesday, February 13, 2008 11:00 AM (20 minutes)

The framework allows to: i) store, update, validate, manipulate, and retrieve metadata through the Metadata Catalog; ii) arbitrarily transform metadata through the Metadata Broker; iii) index metadata through the XML Indexer and discover them through XQuery and XPath expressions; iv) manage annotations through the Annotation Management stack.

The granularity of each operation varies from a single metadata entity, to bulk, passed by-reference, entities that allow managing entire collections. The outputs of the operations can be a static or dynamic, continually updated products of their inputs.

Each component is a well defined Web Service. The framework itself has been factored to support inclusion of new services at any (even run-) time. Moreover, apart the service that manages the upload and the relationships among the metadata items and the objects they describe, the rest of the services can be omitted, if the provided functionalities are not desired.

3. Impact

The gCube Metadata Framework has been successfully adopted within the DILIGENT infrastructure, where, hundreds of thousands of metadata objects, potentially outsourced onto the gLite Data Management System through the gCube Data Management API, have been stored and manipulated along the project's lifetime. It proved to be scalable and efficient, well capable of serving the needs of heterogeneous communities.

The XML Indexer has been exploited in the query workflows by gCube's native Search Engine, satisfying complex queries in acceptable response times, thanks to the transparent partitioning mechanism implemented. Alongside this, several transformation programs have been employed by the Metadata Broker, in order to generate new metadata collections, in different formats, towards facing interoperability and presentation challenges. On top of these, semi-structured annotations, over diverse types of content, add new potential to the exploitation of annotations in Information Retrieval.

URL for further information:

<http://www.gcube-system.org/>

4. Conclusions / Future plans

The design and the implementation of the framework will evolve to simplify the plug-ability of different storage systems as backend support. Eventually storing metadata directly on the gLite Data Management System, overcoming the gCube Data Management layer, will become feasible.

Furthermore new services dedicated to the management of specialized object-to-object relationships will be analysed and integrated in the framework to serve application specific needs.

Provide a set of generic keywords that define your contribution (e.g. Data Management, Workflows, High Energy Physics)

Metadata Management, gCube, DILIGENT, D4Science, XML

1. Short overview

A metadata object is any kind of data about other data. Any system aiming at managing content has to deal with them. Typically, systems are targeted on a limited set of metadata formats and they built their own

semantics for such formats.

The gCube Metadata Framework provides an efficient and generic API, exploitable by domain-specific services, that does not care about the format or semantics of the metadata. It rather focuses on management along with efficient storage and retrieval facilities.

Primary author: SIMI, Manuele (CNR-ISTI)

Co-authors: Dr KAKALETRIS, George (University of Athens); Dr PAGANO, Pasquale (CNR-ISTI)

Presenter: Dr PAGANO, Pasquale (CNR-ISTI)

Session Classification: Data Management

Track Classification: Existing or Prospective Grid Services