Introduction to OGSA-DAI

Principles and Architectures for Structured Data Integration:
OGSA-DAI as an example
ISSGC06 (Ischia, Italy), 17 July 2006

Outline

• OGSA-DAI in a Nutshell
• Design Principles
• Overview
• What’s in the box? – Core Features
• Extensibility Points
OGSA-DAI In One Slide

• An extensible framework for data access and integration.
• Expose heterogeneous data resources to a grid through web services.
• Interact with data resources:
  – Queries and updates.
  – Data transformation / compression
  – Data delivery.
• Customise for your project using
  – Additional Activities
  – Client Toolkit APIs
  – Data Resource handlers
• A base for higher-level services
  – federation, mining, visualisation,…

OGSA-DAI Design Principles – I

• Efficient client-server communication
  – Minimise where possible
  – One request specifies multiple operations
• No unnecessary data movement
  – Move computation to the data
  – Utilise third-party delivery
  – Apply transforms (e.g., compression)
• Build on existing standards
  – Fill-in gaps where necessary
  – DAIS specifications from DAIS WG at GGF
OGSA-DAI Design Principles – II

• Do not hide underlying data model
  – Users must know where to target queries
  – Data virtualisation is hard

• Extensible architecture
  – Modular and customisable
  – e.g., to accommodate stronger security

• Extensible activity framework
  – Cannot anticipate all desired functionality
  – Activity = unit of functionality
  – Allow users to plug-in their own

OGSA-DAI Services

• OGSA-DAI uses data services to represent and provide access to a number of data resources
OGSA-DAI platforms

- OGSA-DAI 2.2 is written in Java and supports three platforms:
  - Globus Toolkit 4.0.2 (Axis 1.2RC2)
  - Axis 1.2.1
  - OMII container (Axis 1.2RC3)

OGSA-DAI Components

A Data Service accesses a number of data service resources …

- Files Resource
- Exist Resource
- MySQL Resource
Each data service resource supports a set of activities …

Activities

- Activities are operations
  - Typically data-related
  - Specific to a data resource type (e.g. JDBC or File System)
  - Data can flow from one activity to another
The OGSA-DAI Framework

Application

Client Toolkit

OGSA-DAI service

Engine

readFile

ToCSV

GridFTP

JDBC

XMLDB

File

SQL Server

eXist

SWISS PROT

Data Service Accessors

Activities

OGSA-DAI service

Engine

SQLBag

Multiple DSR

Multiple Data Service Resources

OGSA-DAI service

Engine

SQLBag

Multiple DSR

SQL

JDBC

SQL

JDBC

SQL

JDBC

SQL

JDBC
Core features of OGSA-DAI – I

- A framework for building applications
  - Supports data access, insert and update
    - Relational; XML; Files
  - Supports data delivery
    - SOAP over HTTP
    - GridFTP; FTP
    - Inter-service
    - E-mail
  - Supports data transformation
    - XSLT
    - ZIP + GZIP Compression
    - Projection; Random Samples
    - Handling BLOBs
  - Supports security
    - X.509 certificate based security

Core features of OGSA-DAI – II

- A framework for building data clients
  - Client toolkit library for application developers

- A framework for developing functionality
  - Extend existing activities, or implement your own
  - Mix and match activities to provide functionality you need

- Highly-extensible
  - Customise our out-of-the-box product
  - Provide your own services, client-side support and data-related functionality

- Comprehensive documentation and tutorials
Basic Service Interactions

Client

Data Sink

Data Service

Activity

Activity

Activity

<?xml?>

<perform>

…

</perform>

<?xml/>

<response>

…

</response>

<?xml/>

<response>

…

</response>

Perform Documents

- Perform documents
  - Encapsulate a serialisation of multiple interactions with a service into a single interaction
  - Abstract each interaction into an “activity”
  - Data can flow from one activity to another
  - No control constructs present
    - no conditionals, loops or variables
Activities

- An Activity dictates an action to be performed
  - Query a data resource
  - Transform data
  - Deliver results
- Subset of activities available to a Data Resource
  - Specified in configuration files
- Data can flow between activities

Extensibility Points

- Extensibility points in OGSA-DAI can be used to implement very powerful additions
- Allow applications to be tailored for specific tasks
- More about this later…
Future Features

- Multi-Resource Perform Documents
- New Activity Framework
  - New interfaces
  - Simplify activity development for implementers
  - Add hooks for monitoring, authorisation, …
  - More out-of-the-box activities
- New Pipe Model
  - Improved concurrent processing
  - New streaming model
- Extended support for BLOB and binary data extraction
- Persistency support through a configuration database
- Dynamic Service Configuration and Management

Conclusions

- OGSA-DAI provides an extensible, data service based framework which makes it easier to implement data integration
- Can be used “stand-alone” (with Apache Axis) or integrates with Globus Toolkit and OMII frameworks
- Multiple operations (activities) are encapsulated in a single interaction with a service
- The architecture is designed for extensibility
The End

• Questions?