|epcc|

GridDataServiceFactory Specification for XML Databases

Project Title: OGSA-DAI GridServe

Document Title: GridDataServiceFactory Specification for XML Databases

Document Identifier: EPCC-GDS-WP3-GDSF v1.1

Document filename: GDSF-spec-1.1.doc

Distribution Classification: DRAFT: Unrestricted

Authorship: Amy Krause, Rob Baxter

Document History:

Personnel	Date	Summary	Version
Amy Krause	20/06/2002	First draft	0.1
Amy Krause	24/06/2002	Updated with comments	0.2
Amy Krause	02/07/2002	Minor revisions	0.3
Rob Baxter	02/07/2002	WP-approved	1.0
Amy Krause	04/07/2002	Revised for submission to DAIS-WG	1.1

Contents

1   Introduction

This document defines the OGSA [2] specification for a prototype Grid data service factory (GDSF). In our terminology, a Grid data service factory is a Grid service designed to create Grid data services (GDS) providing access to a data source.

This specification is written from the perspective of access to an XML database, by which we mean a collection of data stored as full XML Schema documents and which are accessed by XML languages such as XPath or XML Query. We refer to the Grid data services designed to access XML databases as Grid XML Data Services (GXDS).

In this document, the OGSA PortTypes which GDSFs must support are defined, as are messages and types associated with these. This document does *not* define or specify additional Grid services related to GDSFs. In particular, the following services will be covered by their own specification documents:

- GridDataService [3]
- GridDataTransportService [4]
- GridDataRegistryService [5]

This work builds on the early specification of the Grid service interfaces [1] and early work carried out by the UK e-Science Data Base Task Force [6],[7]. It also builds upon prototyping experience carried out at EPCC as part of the OGSA-DAI project.

NB this version of this document describes a prototype GDSF specification. This specification is currently mutable and should not be viewed as representing a complete or finalised definition of a Grid data service.

1.1   A note on terminology

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in RFC 2119 [8].

2   XML Database Definition

The framework described here is intended to be generic and applicable to a wide range of XML databases. The requirements for a database to use the framework are therefore relatively loose.

The service data elements of a GXDS contain the supported query, update, load and edit languages as well as a query template schema [3]. These will also be stored as service data elements within the GDFS, but are also very likely to be the kind of information registered with a discovery service such as a Grid Data Registry Service.

3   PortTypes

This section identifies the PortTypes which a Grid Data Services Factory (GDSF) MUST support. Note that a GDSF does not define its own PortType; its distinction from other types of Grid service factory arises in the service data elements defined by the GDSF.

The service type of a GDSF must be as follows:

- There MUST be a *portTypeList* element which contains references to the G*ridServicePortType* and the *FactoryPortType* of this GDSF.
- There MAY be a *compatibilityStatement* element containing describing compatibility assertions.

A GDSF MUST also make available certain service data. Supported query and update languages of the GDS to be created will be stored as service data elements within the GDSF. This is described in section 4.

3.1 GridService PortType

As for all Grid services, a GDSF MUST implement the GridService PortType [1].

The *FindServiceData* operation will provide access to all information about the GDS this GDFS creates, e.g. the query template of the database, the query, update, load notations supported, the transport services available. This information is represented as a logical collection of service data elements (SDE), as defined in the Grid Service Specification [1].

3.2 Factory PortType

A GDSF MUST support the Factory PortType as it will be used to created instances of GridDataServices. The factory port type of a GDSF is identical to the Factory PortType in the Grid Service specification [1].

3.2.1 Factory::CreateService

Input
- *EarliestTerminationTime* (optional): The earliest termination time of the Grid service instance that is acceptable to the client.
- *LatestTerminationTime* (optional): The latest termination time of the Grid service instance that is acceptable to the client. This value must be ≥ *EarliestTerminationTime*.
- *ServiceParameters*: An XML document describing the GDS that this GDSF is being asked to construct, which conforms to an XML schema. The contents depend on the type of the GDSF: if the GDS is created to connect to an existing data source, no service parameters need to be specified; if the data source is created by the GDSF, a document must be provided which specifies the data source.

Output
- *Reference:* A Grid Service Reference to the newly created Grid Data Service instance. (Note: The Grid Service Handle is carried as part of the reference.)
- *ServiceTimestamp:* The time at which the Grid Data Service was created.
- *CurrentTerminationTime:* The Grid Data Service's currently planned termination time.
- *Maximum Extension:* The maximum extension that the Grid Data Service will currently allow a client to request of its termination time.

4 Grid Data Service Factory Service Data Elements

Type: QName
NameSpace: http://www.w3.org/2001/XMLSchema
Number: 1 or more

Name:	GridServiceFactoryCreationTypeCreated
Description:	A qname to a serviceType created by this Factory, e.g. GridXMLDataService.

Type:	QName
NameSpace:	http://www.w3.org/2001/XMLSchema
Number:	0 or more
Name:	GridServiceFactoryCreationInputTypes
Description:	A qname of an XML type supported by this Factory for the ServiceParameters argument of the CreateService operation.

Type:	TransportServiceType
NameSpace:	http://schemas.nesc.ac.uk/gridServices/GDTS
Number:	1 or more
Name:	GridDataServiceTransportServices
Description:	A list of GridDataTransportService types that the constructed GDS can use.

Type:	QueryTemplateType
NameSpace:	http://schemas.nesc.ac.uk/gridServices/GDS
Number:	Exactly 1
Name:	GridDataServiceQueryTemplate
Description:	An XML document which is a template for the database document structure.

Type:	QueryNotationListType
NameSpace:	http://schemas.nesc.ac.uk/gridServices/GDS
Number:	1 or more
Name:	GridDataServiceQueryNotationTypes
Description:	A list of query notation types the constructed database service supports. Query notations are URLs pointing to a query language specification (e.g. for XPath http://www.w3.org/TR/1999/REC-xpath-19991116).

Type:	UpdateNotationListType
NameSpace:	http://schemas.nesc.ac.uk/gridServices/GDS
Number:	0 or more
Name:	GridDataServiceUpdateNotationTypes
Description:	A list of update notation types the constructed GridDataService supports. Update notations are URLs pointing to an update language specification.

Type:	LoadNotationListType
NameSpace:	http://schemas.nesc.ac.uk/gridServices/GDS
Number:	0 or more
Name:	GridDataServiceLoadNotationTypes
Description:	A list of load notation types the constructed GridDataService supports.

NB: This list of SDEs is not expected to be final at this stage.

5   References

[1]   Tuecke, S., Czajkowski, K., Foster, I., Frey, J., Graham, S., Kesselman, C. and Nick, J., *Grid Service Specification*, June 2002. http://www.globus.org/research/papers.html

[2]   Foster,I., Kesselman,C., Nick,J. and Tuecke,S., The Physiology of the Grid: an Open Grid Services Architecture for Distributed Systems Integration, presented at GGF4, Feb. 2002

[3]   NeSC OGSA-DAI team, *Grid Data Service Specification for XML Databases*, in preparation.

[4]   NeSC OGSA-DAI team, *Grid Data Transport Service Specification for XML Databases*, in preparation.

[5]   NeSC OGSA-DAI team, *Grid Data Registry Service Specification for XML Databases*, in preparation.

[6]   Paton, N.W., Atkinson, M.P., Dialani, V., Pearson, D., Storey, T. and Watson, P., *Database Access and Integration Services on the Grid,* UK DBTF working paper, January 2002.

[7]   Atkinson, M.P., *The Open Grid Services Architecture and a Model for Database Access and Integration Services on the Grid*, contribution to UK DBTF discussions, January 2002.

[8]   RFC 2199, http://www.ietf.org/rfc/rfc2119.txt

A.   XML Schema Definitions

A.1   Service Data Elements

```
<?xml version="1.0" encoding="UTF-8"?>
<xs:schema xmlns:xs=http://www.w3.org/2001/XMLSchema>
  <xs:element name="serviceData">
    <xs:complexType>
      <xs:choice>
        <xs:element name="GridServiceFactoryCreationTypeCreated"
                    type="xs:QName"/>
        <xs:element name="GridServiceFactoryCreationInputTypes"
                    type="xs:QName"/>
        <xs:element name="GridDataServiceTransportServices"
                    type="TransportServiceListType"/>
        <xs:element name="GridDataServiceQueryNotationTypes"
                    type="QueryNotationListType"/>
        <xs:element name="GridDataServiceUpdateNotationTypes"
                    type="UpdateNotationListType"/>
        <xs:element name="GridDataServiceLoadNotationTypes"
                    type="LoadNotationListType"/>
        <xs:element name="GridDataServiceEditNotationTypes"
                    type="EditNotationListType"/>
      </xs:choice>
    </xs:complexType>
  </xs:element>
  <xs:complexType name="QueryNotationListType">
    <xs:sequence>
      <xs:element name="QueryNotation" type="xs:anyURI"
                  maxOccurs="unbounded"/>
    </xs:sequence>
  </xs:complexType>
  <xs:complexType name="UpdateNotationListType">
    <xs:sequence>
      <xs:element name="UpdateNotation" type="xs:anyURI"
                  minOccurs="0" maxOccurs="unbounded"/>
    </xs:sequence>
  </xs:complexType>
  <xs:complexType name="LoadNotationListType">
    <xs:sequence>
      <xs:element name="LoadNotation" type="xs:anyURI"
                  minOccurs="0" maxOccurs="unbounded"/>
    </xs:sequence>
  </xs:complexType>
  <xs:complexType name="EditNotationListType">
    <xs:sequence>
      <xs:element name="EditNotation" type="xs:anyURI"
                  minOccurs="0" maxOccurs="unbounded"/>
    </xs:sequence>
  </xs:complexType>
  <xs:complexType name="TransportServiceListType">
    <xs:sequence>
      <xs:element name="TransportService" type="xs:anyURI"
                  maxOccurs="unbounded"/>
    </xs:sequence>
  </xs:complexType>
</xs:schema>
```