

Kepler - Bug #1548

consolidating data access user interfaces

04/30/2004 11:52 AM - Matt Jones

Status:	In Progress	Start date:	04/30/2004
Priority:	Immediate	Due date:	
Assignee:	Jing Tao	% Done:	0%
Category:	data access	Estimated time:	0.00 hour
Target version:	3.X.Y	Spent time:	0.00 hour
Bugzilla-Id:	1548		
Description <p>Currently Kepler contains several distinct methods for binding data sources to a workflow. These include the EML200DataSource actor, the JDBC data source actor(s), the incipient EcoGrid access interfaces, the GridFTP actor, and probably others. Each of these exposes the data in a different way, and is therefore multiply representing data in a confusing way. We need to consolidate these approaches to find a single UI that can encapsualte all of the data access approaches.</p> <p>This proposal is to use and adapt the user interface described in kepler/docs/dev/screenshots and related design documents to data access in EcoGrid, GridFTP, JDBC, and other sources. This would allow a user to view data uniformly in the workflow, regardless of which data access protocol is used to get the data. This would also allow the user to specify subsetting constraints (WHERE clause) uniformly, and to choose which attributes from the joined relations are exposed to the workflow. Finally, it would allow us to use richer metadata descriptions of underspecified data sources (like those found at the other end of JDBC connections) so that the user (and ultimately the SEEK SMS system) can reason about these data sources effectively.</p>			
Related issues:			
Is duplicate of Kepler - Bug #1887: Consolidate grid actors			Resolved 01/20/2005

History

#1 - 01/20/2005 10:16 AM - Chad Berkley

uniform put and get will be looked at by jing and efrat.

#2 - 01/20/2005 10:44 AM - Jing Tao

The get function should be extened to accept sql query. So user can get partial data object.

#3 - 01/21/2005 12:38 PM - Chad Berkley

- Bug 1887 has been marked as a duplicate of this bug. ***

#4 - 04/04/2005 02:57 PM - Jing Tao

Here is the summary after the meeting in March 11:

Federate Metadata across different communities: Create a unified metadata object, called DataProxy. The DataProxy can get the metadata (probably\uffffusing a\uffffDataSystem class\uffffas described below) and parse\uffffit using different metadata formats interpreter, such as EML, Darwin Core, ADN, FGDC, etc... After parsing the\uffffmetadata, the DataProxy object will have the info to download\uffffthe\uffffdata object\uffffas described by the metadata specification\uffffand pass the info to proper DataSystem class to download the data. The API will include the following functionality:

```
InputStream getFullMetadata(String id, String endPoints);
DataSystem parseMetadata(InpuStream metadata);
void downloadData(DataSystem object);
\uffff
```

The DataSystem class: a generic class to handle get data object (including metadata object) from different data sources (data system).

API requirements:

InputStream getData(String identifier, String endPoints);

InputStream getData(other signatures).

Extending classes: EcoGridDataSystem, MetacatDataSystem (for metacats that don't implement the ecogrid interface), JDBCDataSystem, etc...

Certificate authority: Create a single centralized certificate authority to provide a shared infrastructure to access and maintain different sites' certificate authorities, e.g., the GEON portal, the seek different sites' CAs. - follow up with Karan for more

information about the Grid Account Management Architecture (GAMA) used in the GEON portal authentication.

.

A unified web service access to the datasources:

- In order to support other clients than the Kepler interface to access the various datasources.

- A web service access to datasources with no additional requirements (such as registering). Communities can benefit from accessing each other datasources directly.

The GEON and SEEK datasources access architectures are very similar - a follow up meeting is required on consolidating datasources access through a unified web service with Kai, Ashraf, Karan, Sandeep, Efrat, Chaitan from GEON and folks from SEEK.

.

Unified query for data sources in Kepler either by adding more datasources querying classes (besides EMLDataSource and DCDataSource), or

once there is a unified web service access, using a generic web service actor to query all the data sources.

#5 - 11/02/2005 12:48 PM - Matt Jones

put/get should be implemented for 1.0.0, but the fully unified data access interface will not be, so changing milestone.

#6 - 03/27/2013 02:17 PM - Redmine Admin

Original Bugzilla ID was 1548