



RELIABLE FILE TRANSFER

Contact: Madduri@mcs.anl.gov

Web: http://dev.globus.org/wiki/Reliable_File_Transfer

The Globus Toolkit's Reliable File Transfer Service is a Web Services Resource Framework (WSRF) compliant web service that provides "job scheduler"-like functionality for data movement.

While globus-url-copy and GridFTP in general are a very powerful set of tools, there are characteristics which may not always be optimal. First, the GridFTP protocol is not a web service protocol (it does not employ SOAP, WSDL, etc). Second, GridFTP requires that the client maintain an open socket connection to the server throughout the transfer. For long transfers this may not be convenient, such as if running from your laptop. While globus-url-copy uses the robustness features of GridFTP to recover from remote failures (network outages, server failures, etc), a failure of the client or the client's host means that recovery is not possible since the information needed for recovery is held in the client's memory. What is needed to address these issues is a service interface

based on web services protocols that persists the transfer state in reliable storage. We provide such a service and call it the Reliable File Transfer (RFT) service.

With RFT, the user simply provides a list of source and destination URLs (including directories or file globs) and then the service writes your job description into a database and then moves the files on your behalf. Once the service has taken your job request, interactions with it are similar to any job scheduler. Service methods are provided for querying the transfer status, or you may use standard WSRF tools (also provided in the Globus Toolkit) to subscribe for notifications of state change events.

We provide the service implementation which is installed in a web services container (like all web services) and a very simple client. There are Java classes available for custom development, but due to lack of time and resources, work is still needed to make this easier.

New features coming in RFT in GT 4.2:

- Command-line client in C: A new Command-line client in C is provided to get around the JVM startup time of the current Java command line client.
- Transfer Time prediction: New Resource properties that would provide transfer time predictions based on performance markers from the gridftp servers and based on historic data of the transfers.
- Implement priorities in transfers
- Provide Data Scheduler functionality in RFT so data transfers can be "scheduled" and managed.
- Providing support for multiple transfer protocols. Right now RFT supports third-party gridftp transfers only.
- Ability to modify a transfer request after the request is submitted. This will also provide the ability to add new transfers to an existing RFT resource.