PROTOCOLS, UNIFORM RESOURCE LOCATORS (URL) AND EXTENSIONS SUPPORTED IN ARC

Description

A.Konstantinov*

*aleks@fys.uio.no

File locations in ARC can be specified both as local file names, and as Internet standard *Uniform Resource Locators (URL)*. There are also some additional URL *options* that can be used.

The following transfer protocols and metadata servers are supported:

`ftp`	ordinary *File Transfer Protocol (FTP)*
`gsiftp`	GridFTP, the Globus Toolkit® 2 -enhanced FTP protocol with security, encryption, etc. developed by The Globus Alliance [1]
`http`	ordinary *Hyper-Text Transfer Protocol (HTTP)* with PUT and GET methods using multiple streams
`https`	HTTP with SSL v3
`httpg`	HTTP with Globus Toolkit® 2 GSI
`ldap`	ordinary *Lightweight Data Access Protocol (LDAP)* [2]
`rc`	Globus Toolkit® 2 *Replica Catalog (RC)* [3]
`rls`	Globus Toolkit® 2 *Replica Location Service (RLS)* [4]
`fireman`	Fireman indexing service of EGEE gLite [5]
`lfc`	LFC catalog and indexing service of EGEE gLite [5]
`se`	ARC Smart Storage Element service [6]
`srm`	Storage Resource Manager (SRM) service [7]
`file`	local to the host file name with a full path

An URL can be used in a standard form, i.e.

```
<protocol>://host[:port]/<file>
```

Or, to enhance the performance, it can have additional options:

```
<protocol>://host[:port][;option[;option[...]]]/<file>
```

For a metadata service URL, construction is the following:

```
rc://rc://[location[|location[...]]@]<host>[:port]/<DN>/<lfn>
rls://[url[|url[...]]@]<host>[:port]/<lfn>
fireman://[url[|url[...]]@]<host>[:port]/<service_path>?<lfn>
lfc://[url[|url[...]]@]<host>[:port]/<lfn>
```

For the Smart Storage Element service, the syntax is

```
se://host[:port][;options]/path[?file_id]
```

For the SRM service, the syntax is

```
srm://<host>[:port][;options]/[service_path?SFN=]<file_id>
```

Versions 1.1 and 2.2 of the SRM protocol are supported. The default *service_path* is srm/managerv2 when the server supports v2.2, srm/managerv1 otherwise.

The URL components are:

`location`	`<location_name_in_RC>[;option[;option[...]]]`
`host[:port]`	IP address of a server
`DN`	Distinguished Name (as in LDAP) of an RC collection
`lfn`	Logical File Name
`url`	URL of the file as registered in RLS/Fireman
`service_path`	End-point path of the Web service
`file`	local to the host file name with a full path

The following options are supported for location URLs:

`threads=<number>`	specifies number of parallel streams to be used by GridFTP or HTTP(s,g); default value is 1, maximal value is 10
`cache=yes|no|renew`	indicates whether the GM should cache the file; default for input files is `yes`. `renew` forces a download of the file, even if the cached copy is still valid. (NB: cached files can not be modified by jobs)
`readonly=yes|no`	for transfers to `file://` destinations, specifies whether the file should be read-only (unmodifiable) or not; default is `yes`
`secure=yes|no`	indicates whether the GridFTP data channel should be encrypted; default is `no`
`blocksize=<number>`	specifies size of chunks/blocks/buffers used in GridFTP or HTTP(s,g) transactions; default is protocol dependent
`checksum=cksum|md5`	specifies the algorithm for checksum to be computed (ev. provided to the indexing server). This is overridden by any metadata options specified (see below).
`exec=yes|no`	means the file should be treated as executable
`preserve=yes|no`	specify if file must be uploaded to this destination even if job processing failed (default is `no`)
`pattern=<pattern>`	defines file matching pattern; currently works for file listing requests sent to an `se://` endpoint
`guid=yes|no`	make software use GUIDs instead of LFNs while communicating to indexing services; meaningful for `rls://` only
`overwrite=yes|no`	make software try to overwrite existing file(s), i.e. before writing to destination, tools will try to remove any information/content associated with specified URL
`protocol=gsi|gssapi`	to distinguish between two kinds of `httpg`. `gssapi` stands for implemention using only GSSAPI functions to wrap data and `gsi` uses additional headers as implmented in Globus IO
`spacetoken=<pattern>`	specify the space token to be used for uploads to SRM storage elements supporting SRM version 2.2 or higher

Local files are referred to by specifying either a location relative to the job submission working directory, or by an absolute path (the one that starts with "/"), preceded with a `file://` prefix.

Metadata service URLs also support metadata options which can be used for register additional metadata attributes or query the service using metadata attributes. These options are specified at the end of the LFN and consist of name and value pairs separated by colons. The following attributes are supported:

`guid`	GUID of the file in the metadata service
`checksumtype`	Type of checksum. Supported values are cksum (default), md5 and ad (adler32 checksum)
`checksumvalue`	The checksum of the file

Currently these metadata options are only supported for lfc:// URLs.

Examples of URLs are:

```
http://grid.domain.org/dir/script.sh
gsiftp://grid.domain.org:2811;threads=10;secure=yes/dir/input_12378.dat
ldap://grid.domain.org:389/lc=collection1,rc=Nordugrid,dc=nordugrid,dc=org
rc://grid.domain.org/lc=collection1,rc=Nordugrid,dc=nordugrid,dc=org/zebra/f1.zebra
rls://gsiftp://se.domain.org/datapath/file25.dat@grid.domain.org:61238/myfile02.dat¹
fireman://fireman_host:8443/glite-data-catalog-interface/FiremanCatalog?data.root
file:///home/auser/griddir/steer.cra
lfc://srm://srm.domain.org/griddir@lfc.domain.org//user/file1:guid=\
    bc68cdd0-bf94-41ce-ab5a-06a1512764dc:checksumtype=ad:checksumvalue=12345678²
lfc://;cache=no@lfc.domain.org/:guid=bc68cdd0-bf94-41ce-ab5a-06a1512764d³
```

---

[1] This is a destination URL. The file will be copied to the GridFTP server at se.domain.org with the path datapath/file25.dat and registered in the RLS indexing service at grid.domain.org with the LFN myfile02.dat.

[2] This is a destination URL. The file will be copied to srm.domain.org at the path griddir/file1 and registered to the LFC service at lfc.domain.org with the LFN /user/file1. The given GUID and checksum attributes will be registered.

[3] This is a source URL. The file is registered in the LFC service at lfc.domain.org with the given GUID and can be copied or queried by this URL. Note that as URL options are part of the location (physical) URL, in meta service URLs the options must be part of the location URL, even if the location URL is empty.

References

[1] I. Foster and C. Kesselman, "Globus: A Metacomputing Infrastructure Toolkit," *International Journal of Supercomputer Applications*, vol. 11, no. 2, pp. 115–128, 1997, available at: http://www.globus.org.

[2] M. Smith and T. A. Howes, *LDAP : Programming Directory-Enabled Applications with Lightweigt Directory Access Protocol*. Macmillan, 1997.

[3] H. Stockinger *et al.*, "File and Object Replication in Data Grids," *Cluster Computing*, vol. 5, no. 3, pp. 305–314, July 2002.

[4] A. L. Chervenak *et al.*, "Performance and Scalability of a Replica Location Service," in *Proceedings of the 13th IEEE International Symposium on High Performance Distributed Computing (HPDC'04)*. IEEE Computer Society Press, 2004, pp. 182–191.

[5] "gLite, Lightweight Middleware for Grid Computing," Web site. [Online]. Available: http://glite.web.cern.ch/glite/

[6] A. Konstantinov, *The NorduGrid Smart Storage Element*, The NorduGrid Collaboration, NORDUGRID-TECH-10. [Online]. Available: http://www.nordugrid.org/documents/SE.pdf

[7] A. Sim, A. Shoshani and others, "The Storage Resource Manager Interface (SRM) Specification v2.2," May 2008, GFD-R-P.129. [Online]. Available: http://www.ggf.org/documents/GFD.129.pdf