iRODS GridFTP File Driver

Documentation

SPONSORED BY THE

Federal Ministry
of Education
and Research

Content

Author: Christian Loeschen, TU Dresden
Dresden, 07/14/2010

1 Introduction

This is the documentation of the GridFTP file driver, developed for the iRODS Data-management System. Using this file driver it is possible to transfer files from iRODS to GridFTP servers and back. It is possible to GSI proxy certificates for every iRODS user, or the iRODS server can be run using a single certificate.

It was developed within the German DGrid Integrationsprojekt 2 (DGI2) founded by the German Federal Ministry of Education and Research (BMBF).

The iRODS GridFTP file driver depends on the iRODS GridFTP module. This must be installed on an iRODS server in order to use the iRODS GridFTP file driver. It is available at the project homepage.

2 Installation

1. Install the iRODS GridFTP module first. Therefore refer to the iRODS GridFTP module documentation, please.
2. Copy file `gridftpFileDriver.h` into directory: `server/drivers/include/`
3. Copy file `gridftpFileDriver.c` into directory:`server/drivers/src/`
4. Apply changes to file `server/drivers/include/fileDriverTable.h`:

   - add:

   ```
   #include "gridftpFileDriver.h"
   ```

   - add to `fileDriver_t FileDriverTable[]`:

   ```
   {GRIDFTP_FILE_TYPE, intNoSupport, intNoSupport, intNoSupport,
   intNoSupport, intNoSupport, intNoSupport, intNoSupport,
   intNoSupport, longNoSupport, intNoSupport, intNoSupport,
   intNoSupport, intNoSupport, intNoSupport, intNoSupport,
   intNoSupport, intNoSupport, intNoSupport, longNoSupport,
   intNoSupport, gridFtpStageToCache, gridFtpSyncToArch},
   ```

5. Apply changes to `lib/core/include/objInfo.h`:

   - add to `typedef enum fileDriverType_t`:

   ```
   GRIDFTP_FILE_TYPE,
   ```

6. Apply changes to `lib/core/include/rcGlobal.h`:

   - add to `rescTypeDef_t RescTypeDef[]`:

   ```
   {"gridftp", FILE_CAT, GRIDFTP_FILE_TYPE, DO_CHK_PATH_PERM},
   ```

7. Apply changes to `server/Makefile`:

   - add to `SVR_DRIVERS_OBJS`:

   ```
   $(svrDriversObjDir)/gridftpFileDriver.o
   ```

8. Apply changes to `modules/gridftp/microservices/src/gridftp_put.c`:

   - replace on line 233 `#else` with `#endif`

- remove on line 364 `#endif`

9. Apply changes to `modules/gridftp/microservices/src/gridftp_get.c`:
   - replace on line 13 `#else` to `#endif`
   - remove on line 15 `#endif`
   - replace on line 254 `#else` with `#endif`
   - remove on line 404 `#endif`
10. Run `irodssetup`.

3 Usage

3.1 Create a GridFTP resource

The following steps describe, how a new GridFTP resource is added to iRODS. All steps are done within the `iadmin` tool. Doing `help <command>` within the `iadmin` tool, a comprehensive explanation of the single commands is displayed. For better understanding, all the changes done are reflected using appropriate commands.

So, first let's add the new GridFTP resource type:

```
~> iadmin
iadmin>at resc_type "gridftp"
iadmin>lt resc_type
unix file system
hpss file system
windows file system
s3
MSS universal driver
gridftp
```

We have to create a local cache resource for staging files:

```
iadmin>mkresc helenaCache 'unix file system' cache localhost
    /tmp/helenaCache
iadmin>lr helenaCache
resc_id: 10020
resc_name: helenaCache
zone_name: tempZone
resc_type_name: unix file system
resc_class_name: cache
resc_def_path: /tmp/helenaCache
free_space:
free_space_ts:  : 1970-01-01.01:00:00
resc_info:
r_comment:
resc_status:
create_ts: 01278670303 : 2010-07-09.12:11:43
modify_ts: 01278670303 : 2010-07-09.12:11:43
```

And we create our GridFTP resource:

```
iadmin>mkresc helena 'gridftp' compound helena.zih.tu-dresden.de
    gsiftp://helena.zih.tu-dresden.de/pnfs/zih.tu-dresden.de/data/ghep
iadmin>lr helena
resc_id: 10022
resc_name: helena
zone_name: tempZone
resc_type_name: gridftp
resc_class_name: compound
resc_net: helena.zih.tu-dresden.de
resc_def_path: gsiftp://helena.zih.tu-dresden.de/pnfs/zih.tu-
dresden.de/data/ghep
free_space:
free_space_ts:  : 1970-01-01.01:00:00
resc_info:
r_comment:
resc_status:
create_ts: 01278670372 : 2010-07-09.12:12:52
modify_ts: 01278670372 : 2010-07-09.12:12:52
```

The local cache and our GridFTP resource have to be combined within a resource group, so the cache resource is known for staging files for the GridFTP resource:

```
iadmin>mkgroup helenaGroup
iadmin>atrg helenaGroup helena
iadmin>atrg helenaGroup helenaCache
iadmin>lrg helenaGroup
resc_group_name: helenaGroup
resc_id: 10020
resc_name: helenaCache
create_ts: 01278670388 : 2010-07-09.12:13:08
modify_ts: 01278670388

resc_group_name: helenaGroup
resc_id: 10022
resc_name: helena
create_ts: 01278670385 : 2010-07-09.12:13:05
modify_ts: 01278670385 : 2010-07-09.12:13:05

iadmin>q
```

That's it.

3.2 Working with the GridFTP resource

The first example shows how a file is put directly to the GridFTP resource using the `iput` command. The file is stored on the cache resource belonging to the GridFTP resource first, then transferred to the GridFTP resource. To clean the copy on the cache resource after that, `itrim` is issued:

```
~> ils
/tempZone/home/rods:
  user-proxy
~> iput -R helena /bin/sh foo.bar
~> ils -L
/tempZone/home/rods:
   rods   0 helenaCache   606864 2010-07-09.12:14 & foo.bar
        /tmp/helenaCache/home/rods/foo.bar               helenaGroup
```

```
   rods   1 helena        606864 2010-07-09.12:14 & foo.bar
        gsiftp://helena.zih.tu-dresden.de/pnfs/zih.tu-
dresden.de/data/ghep/home/rods/foo.bar                    helenaGroup
   rods   0 demoResc        3677 2010-07-09.11:19 & user-proxy
        /home/christian/iRODS/Vault/home/rods/user-proxy
~> ll /tmp/helenaCache/home/rods/
drwx------ 5 christian users 4,0K  9. Jul 12:14 .
drwx------ 3 christian users 4,0K  7. Mai 10:14 ..
-rwx------ 1 christian users 593K  9. Jul 12:14 foo.bar
~> itrim -N 1 -S helenaCache foo.bar
~> ll /tmp/helenaCache/home/rods/
drwx------ 5 christian users 4,0K  9. Jul 12:15 .
drwx------ 3 christian users 4,0K  7. Mai 10:14 ..

~> ils -L
/tempZone/home/rods:
   rods   1 helena        606864 2010-07-09.12:14 & foo.bar
        gsiftp://helena.zih.tu-dresden.de/pnfs/zih.tu-
dresden.de/data/ghep/home/rods/foo.bar                    helenaGroup
   rods   0 demoResc        3677 2010-07-09.11:19 & user-proxy
        /home/christian/iRODS/Vault/home/rods/user-proxy
```

The second example explains how to replicate an already existing iRODS file to the GridFTP resource. The cache resource is cleaned up again after that:

```
~> iput /bin/sh bar.foo
~> ils
/tempZone/home/rods:
  bar.foo
  foo.bar
  test.file
  user-proxy
~> irepl -R helena bar.foo
~> ils -L
/tempZone/home/rods:
   rods   1 helena        606864 2010-07-09.12:14 & foo.bar
        gsiftp://helena.zih.tu-dresden.de/pnfs/zih.tu-
dresden.de/data/ghep/home/rods/foo.bar                    helenaGroup
   rods   0 demoResc      606864 2010-07-09.11:19 & bar.foo
        /home/christian/iRODS/Vault/home/rods/bar.foo
   rods   1 helenaCache   606864 2010-07-09.13:52 & bar.foo
        /tmp/helenaCache/home/rods/bar.foo                helenaGroup
   rods   2 helena606864 2010-07-09.13:52 & bar.foo
        gsiftp://helena.zih.tu-dresden.de/pnfs/zih.tu-
dresden.de/data/ghep/home/rods/bar.foo                    helenaGroup
   rods   0 demoResc        3677 2010-07-09.11:19 & user-proxy
        /home/christian/iRODS/Vault/home/rods/user-proxy
~> itrim -N 1 -S helenaCache bar.foo
```

3.3 Removing iRODS user home directory from GridFTP URL

For a more common usage of GridFTP resources it is possible to remove the iRODS user home directory from the used GridFTP URL. So it is possible to replace a URL with a maybe inconvenient iRODS user home directory inside like

```
gsiftp://helena.zih.tu-dresden.de/
                        pnfs/zih.tu-dresden.de/data/ghep/home/rods/foo.bar
```

with something more common like

```
gsiftp://helena.zih.tu-dresden.de/
                                    pnfs/zih.tu-dresden.de/data/ghep/foo.bar
```

As you can see, the iRODS user home directory (`/home/rods`) is removed from the GridFTP URL. To enable this URL editing, define `GRIDFTP_DRIVER_CHANGE_URL` in `gridftpFileDriver.h` file. Note that this change won't be saved within the iRODS file catalogue, it won't be displayed using `ils -L for example`.

3.4 What will not work

With this module, it is regrettably currently only possible to put iRODS files to a GridFTP resource and get them back to iRODS, once they are put there. Any other function will not work:

- no directory creation/removal/listing
- no file removal/renaming
- no visibility of already existing files on the GridFTP server, only they who are put there via the iRODS GridFTP file driver are visible and gettable

But you can use the GridFTP module to do actions like this. Please note, that if a file stored on a GridFTP resource using the GridFTP file driver is deleted somehow, it is still listed on iRODS file catalogue. You have to use `irm -U <file>` to remove it from the file catalogue.

4 Further Information

4.1 License

This software is released under BSD license.

4.2 Contact

You can contact the author and developer via email: *christian.loeschen@tu-dresden.de*

4.3 Links

IRODS www.irods.org/

Globus Toolkit www.globus.org/toolkit/

DGI2 homepage http://dgi.d-grid.de/index.php?id=456&L=1

Project homepage

http://tu-dresden.de/die_tu_dresden/zentrale_einrichtungen/zih/forschung/grid_computing/iRODS

Appendix A: Files

```
gridftpFileDriver.h          GridFTP file driver header file
gridftpFileDriver.c          GridFTP file driver source file
```

Appendix B: API Functions

```
/*
 *    Gets a file from a GridFTP file server.
 *
 *    @param rsComm             iRODS server communication handle
 *    @param cacheFileType      unused
 *    @param mode               unused
 *    @param flags              unused
 *    @param filename           destination GridFTP file URL
 *    @param cacheFilename      source file name in cache resource
 *    @param dataSize           unused
 *    @param condInput          unused
 *    @return                   0 on success, 1 on failure
 */
int gridFtpStageToCache(
        rsComm_t *rsComm,
        fileDriverType_t cacheFileType,
        int mode,
        int flags,
        char *filename,
        char *cacheFilename,
        rodsLong_t dataSize,
        keyValPair_t *condInput);


/*
 *    Puts a file to a GridFTP file server.
 *
 *    @param rsComm             iRODS server communication handle
 *    @param cacheFileType      unused
 *    @param mode               unused
 *    @param flags              unused
 *    @param filename           source GridFTP file URL
 *    @param cacheFilename      destination file name in cache resource
 *    @param dataSize           unused
 *    @param condInput          unused
 *    @return                   0 on success, 1 on failure
 */
int gridFtpSyncToArch(
        rsComm_t *rsComm,
        fileDriverType_t cacheFileType,
        int mode,
        int flags,
        char *filename,
        char *cacheFilename,
        rodsLong_t dataSize,
        keyValPair_t *condInput);


/*
 *    Removes the iRODS user home directory from the GridFTP URL.
 *
 *    @param url                GridFTP URL to change
 *    @param rsComm             iRODS server communication handle containing
 *                              user information
 */
void changeFilename(char *url, rsComm_t *rsComm);
```