



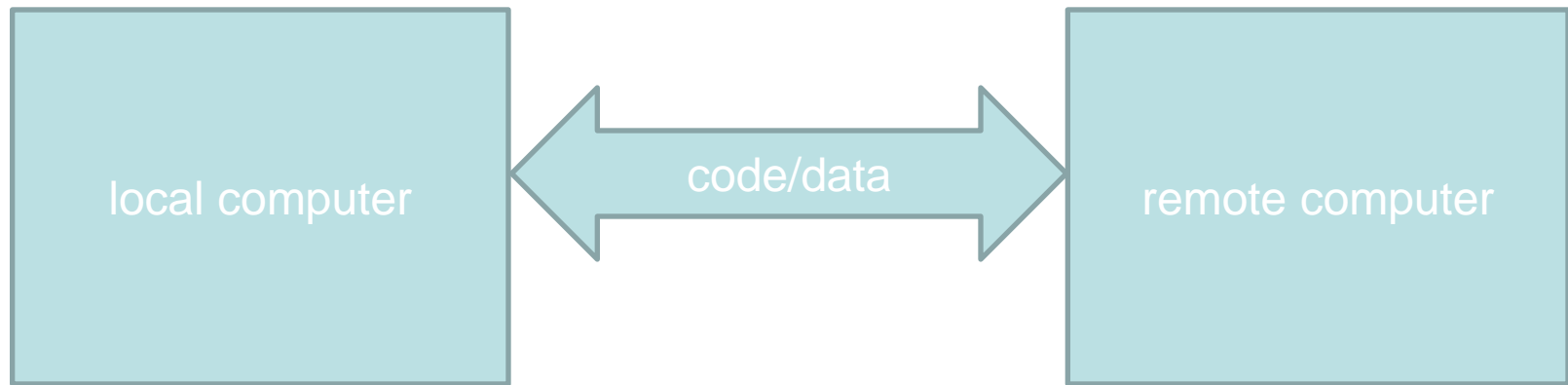
Data Transfer

Linda Woodard
Consultant
Cornell CAC

Workshop: Parallel Computing on Stampede: October 30, 2013



Data Transfer-how do I move my data from here to there?



- Needs to be a secure transfer
- Speed becomes important as the amount of data increases



Data storage options on Stampede

File system	Total Size	User Quota	Short cut	Backup Policy	Purpose
\$HOME cwd at login	524TB	5GB	cdh	nightly	store source code; build executables
\$WORK	1.1PB	400GB	cdw	none	store large files
\$SCRATCH	7.5PB	none	cds	purged after 30 days	store temporary files
/tmp on each compute node	250GB	none		Purged after job completes	store files during job processing



Archival Storage on Stampede

Ranch (<http://www.tacc.utexas.edu/user-services/user-guides/ranch-user-guide>)

- Mass storage server called Ranch (ranch.tacc.utexas.edu) with 50 TB of online storage; 60 PB of offline tape storage; not backed up
- Uses Sun's Storage Archive Manager File system to move files in and out of a tape archival system
- Tar files before moving to Ranch; works best with large files (< 10 GB)
- Running jobs cannot access Ranch directly
- Files on tape need to be “staged” before attempting to access them



Data Transfer Software

- Easy secure transfer for small files (~15 MB/s)
 - SCP (secure copy protocol)
 - SFTP (secure FTP) like SCP, but with browsing capability
 - rsync--only copies parts of files or directories that differ between machines
- Transfers using GridFTP protocol
 - GUI Interface
 - XUP (XSEDE User Portal); requires Java
 - Globus Online
 - Command Line Interface (~125 MB/s)
 - Globus Online CLI
 - Globus-url-copy



Data Transfer for Small Files--Linux

- SCP—requires password for every transfer

local -> remote computer

```
[local] $ scp localBig userName@stampede.tacc.utexas.edu:/path/to/project/directory
```

remote -> local computer

```
[local] $ scp userName@stampede.tacc.utexas.edu:big localBig
```

- SFTP—requires password for initial connection

```
[local] $ sftp stampede.tacc.utexas.edu
```

local -> remote computer

```
put big
```

remote -> local computer

```
get big
```



Data Transfer with RSYNC—Linux Only

- Copies only those parts of a file that have changed, making it significantly faster and more efficient than other ssh transfers

```
rsync source.c userName@stampede.tacc.utexas.edu:/path/to/project/directory
```

- Directory changes can also be copied recursively with rsync

```
rsync -avtr ./Source userName@stampede.tacc.utexas.edu:/path/to/project/NewSource
```

– Options

- a archive mode preserves symbolic links, devices, attributes, permissions, ownerships, etc
- t keeps modification times
- v verbose increases the information displayed during transfer
- r transfers the files recursively



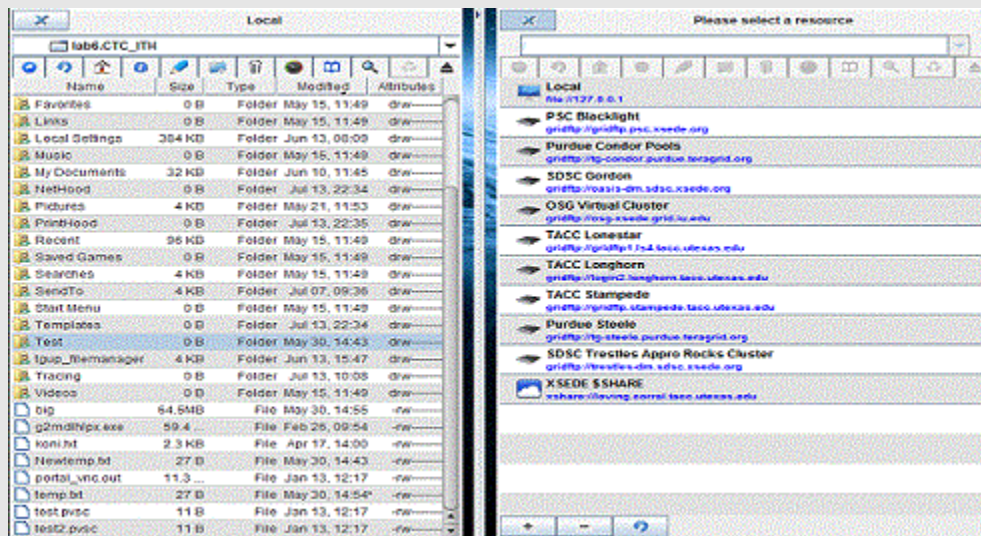
Data Transfer for Small Files--Windows

- There are a number of SCP and SFTP clients for windows
 - **Putty for both SCP and SFTP**
(<http://www.chiark.greenend.org.uk/~sgtatham/putty/download.html>)
 - **FileZilla for SFTP**
(<https://filezilla-project.org/>)
- The syntax of the commands is the same for Windows and Linux



Data Transfer using the XSEDE User Portal File Manager

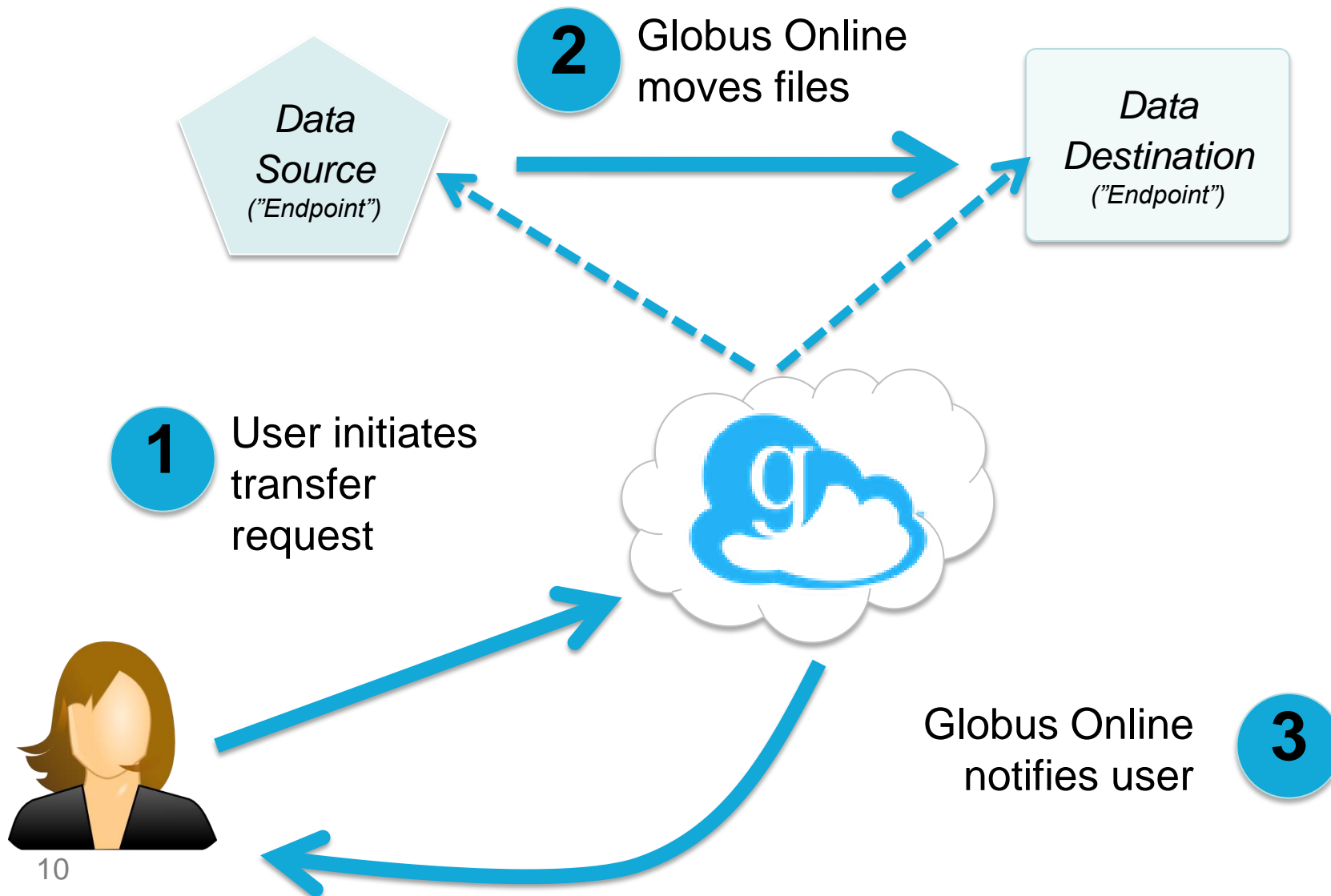
- Select Resources->File Manager from the XSEDE User Portal
- A Java applet will run and give you two screens



- Open the resources you want to transfer between
- Choose the file(s) to transfer from the source resource
- Drag and drop the file(s) to the destination resource



How It Works





Globus

- Get a Globus account <https://www.globusonline.org/SignUp>

Sign Up... Already a member

Full Name

Email

Username
Your username can only contain lower case letters and must begin with one. It may contain numbers.

Password
Better passwords are longer, use mixed case letters with punctuation and numbers.

Show Password

I have read and agree to the [Globus Online Terms of Service and Privacy Policy](#).

Please email me updates about Globus Online



Globus

- Install Globus Connect https://www.globusonline.org/globus_connect/
Available for Linux, Windows, Mac OS X
- Use Globus Online <https://www.globusonline.org/dashboard/Main>

Transfer Summary

Requested Today

0 active transfers.

0 transfers completed successfully.

0 inactive transfers.

0 transfers failed.

Requested This Week

0 active transfers.

0 transfers completed successfully.

0 inactive transfers.

0 transfers failed.

Lifetime

0 active transfers.

31 transfers completed successfully.

0 inactive transfers.

11 transfers failed.



File Transfer

Use your browser to move data securely and reliably.

Start Transfer

View Transfers

Manage Endpoints



My Profile

View and change your account settings, including contact information and security credentials



Globus Connect

Use Globus Connect to transfer files between your computer and any Globus Online endpoint.



Globus

- Transfer files <https://www.globusonline.org/xfer/StartTransfer>

Transfer Files

Get Globus Connect
Turn your computer into an endpoint.

The screenshot displays the Globus online file transfer interface. It features two endpoint panes. The left pane is for endpoint 'cac#home' and shows a list of files with their sizes. The right pane is for endpoint 'xsede#stampede' and shows a list of folders and files. At the bottom, there is a 'Label This Transfer' field containing the text 'Test Transfer'.

Endpoint	Path	Item Name	Size	Type
cac#home	/~/	bookmarks.html	147.82 kB	File
		figfile.bt	928 b	File
		gzip.exe	89.5 kB	File
		hello	28.05 kB	File
		hello.c	928 b	File
		localryn.bt	192.52 kB	File
		localtemp.bt	27 b	File
		mykey.private	57 b	File
		mysit	13.95 kB	File
		pscp.exe	304 kB	File
		psftp.exe	320 kB	File
		pythonExample.py	2.54 kB	File
		rsync.bt	192.52 kB	File
		temp.bt	27 b	File
		test.csv	150 b	File
		test.py	1.3 kB	File
		test.sh	188 b	File
		test2.sh	188 b	File
test3.sh	223 b	File		
test4.sh	282 b	File		
xsede#stampede	/~/	envi		Folder
		hybrid		Folder
		intel		Folder
		mic		Folder
		openmp		Folder
		pimpi		Folder
		test1		Folder
		bigfile.bt	928 b	File
		file2	35 b	File
		newtest	1.56 GB	File
rsync.bt	192.52 kB	File		
temp.bt	27 b	File		



Globus

- Options

▼ less options

Label This Transfer

This will be displayed in your transfer activity.

Transfer Settings

- only transfer new or changed files ?
- delete files on destination that do not exist on source ?
- preserve source file modification times ?
- verify file integrity after transfer ?
- encrypt transfer ?

- “only transfer new or changed files” operates like rsync
- “encrypt transfer” will slow transfer, but can be important for certain types of data



Globus

- Check on file transfers <https://www.globusonline.org/xfer/ViewTransfers>

Transfer Activity

Cancel View 25 Records

◀◀ 1 of 1 ▶▶

Status	Label	Task Progress	Completion Time	Request Time
<input type="checkbox"/> ✓	Task Id:4b8864b2-38a4-11e1-81e6-1231381bd061	1 / 1	01/06/2012 08:26 PM	01/06/2012 08:24 PM
<input type="checkbox"/> ✓	RangerText	1 / 1	01/06/2012 08:04 PM	01/06/2012 08:02 PM

Cancel View 25 Records

◀◀ 1 of 1 ▶▶

Email Notification:

Task ID : c30dc1b2-389a-11e1-81e6-1231381bd061
Task Type : TRANSFER
Status : SUCCEEDED
Request Time : 2012-01-06 20:02:40Z
Deadline : 2012-01-07 20:02:39Z
Completion Time : 2012-01-06 20:04:14Z
Total Tasks : 1
Tasks Successful : 1
Tasks Canceled : 0
Tasks Failed : 0
Command : API 0.10 GO
Label : RangerText
Files : 1
Files Skipped : 0
Directories : 0
Bytes Transferred: 104857600
Bytes Checksummed: 0
Mbits/sec : 8.924



Globus Online CLI

- Create a Globus Online Account
No need to download Globus Client Software
- Enable globus account for ssh
add SSH public key <https://www.globusonline.org/account/ManageIdentities>
- ssh to cli.globusonline.org
ssh username@cli.globusonline.org
- Transfer files using globus scp
scp -D xsede#stampede:file.txt cac#home:newfile.txt
use the -D option to run the transfer in the background

<https://www.globusonline.org/usingcli/>

<https://www.globusonline.org/beyondbasics/>



Globus-url-copy

- Transfer between sites with GridFTP servers or via a 3rd party
- Preferred method for transferring files between XSEDE sites (including to and from Ranch)
- Necessary steps for transferring files on XSEDE
 - module load globus
 - Grid-proxy-info (check for a valid proxy)
 - myproxy-logon (if you don't have a valid proxy)
- Syntax for transferring files—can be incorporated in a script
globus-url-copy gsiftp://sourceURL gsiftp://destinationURL
- XSEDE GridFTP server name without the “:2811”
<https://www.xsede.org/web/guest/data-transfers#table12>



Tips

- Small files will transfer faster with scp or sftp than those using GridFTP protocol
- Globus Online and Globus Online CLI have the same transfer rates
- If you are transferring a large number of files, tar them; aim for a tar ball < 10GB
- When updating files, use rsync or the similar option in Globus Online
- Data transfer is resource intensive
 - limit simultaneous transfers to 3 or less
 - only one globus-url-copy should be active at a time
 - avoid using the recursive (-r) flag with large transfers
- Beware of cross platform issues with filenames
 - avoid spaces in the names
 - Linux is case sensitive and Windows is not



References

- TACC User Guides
 - <https://www.xsede.org/web/guest/tacc-stampede>
 - <https://www.xsede.org/tacc-ranch>

- Globus
 - <http://www.globusonline.org>
 - support@globusonline.org