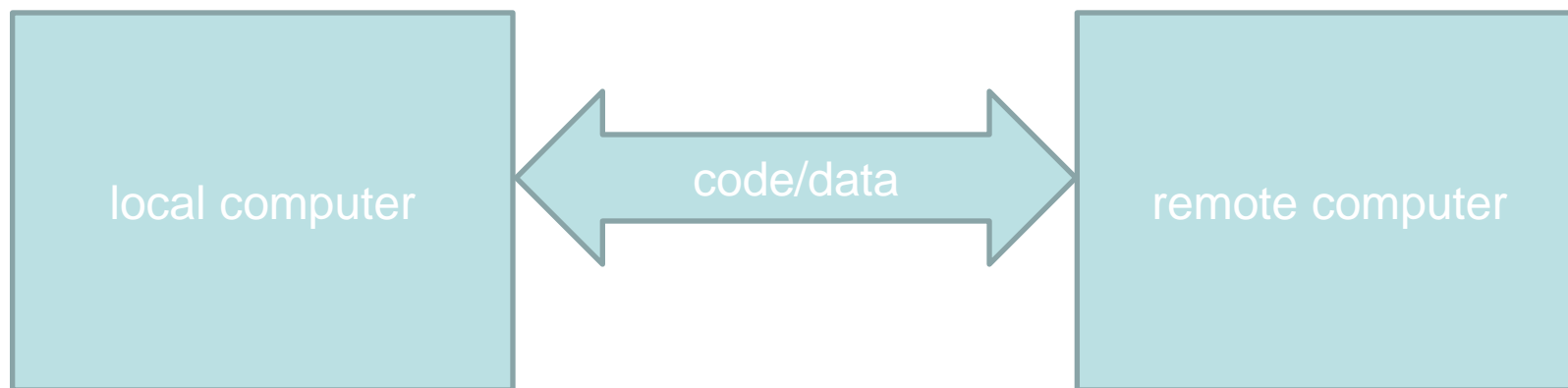# Data Transfer and Storage

Linda Woodard

Consultant

Cornell CAC

Workshop: Data Analysis on Ranger, January 19, 2012

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 Ranger

File system	Total Size	User Quota	Shortcut	Backup Policy	Purpose
$HOME cwd at login	100TB	6GB	cdh	Nightly	store source code; build executables
$WORK run compute jobs here	236TB	200GB	cdw	None	store large files
$SCRATCH global Lustre file system	845TB	None	cds	Purged after 10 days	store temporary files

Data storage options on Lonestar

File system	Total Size	User Quota	Shortcut	Backup Policy	Purpose
$HOME cwd at login	1.5TB	1GB	cdh	Nightly	store source code; build executables
$WORK run compute jobs here	440TB	250GB	cdw	None	store large files
$SCRATCH global Lustre file system	880TB	None	cds	Purged after 10 days	store temporary files

Long term data storage options

- Ranch
  - Mass storage server called Ranch (ranch.tacc.utexas.edu)

  - Uses Sun's Storage Archive Manager File system to move files in and out of a tape archival system

  - 57 TB of online storage; 10 PB of offline tape storage

  - Tar files before moving to Ranch; compression not necessary

  - Files should be "staged" (retrieved from tape) before attempting to access them
    ```
    ssh $ARCHIVER stage "file list"
    rcp $ARCHIVER:"file list"
    ```

Data Transfer Software

- Easy secure transfer for small files (< 2 GB)

    SCP (secure copy protocol)

    SFTP (secure FTP) like SCP, but with browsing capability

    File Manager from XSEDE User Portal

- Third party GridFTP transfer among XSEDE machines

    File Manager from XSEDE User Portal—use striping with large files

    globus-url-copy—not interactive, so better for scripts

    uberFTP—more control

- General Third party GridFTP transfer

    Globus Online

Data Transfer for Small Files--Linux

- SCP

  remote -> local computer
  [local] $ scp userName@ranger.tacc.utexas.edu:big localBig   (Ranger)
  [local] $ scp userName@linuxlogin1.cac.cornell.edu:big localBig (CAC)

  local -> remote computer
  [local] $ scp localBig userName@ranger.tacc.utexas.edu:big2

- SFTP

  [local] $ sftp ranger.tacc.utexas.edu or sftp linuxlogin1.cac.cornell.edu

  remote -> local computer
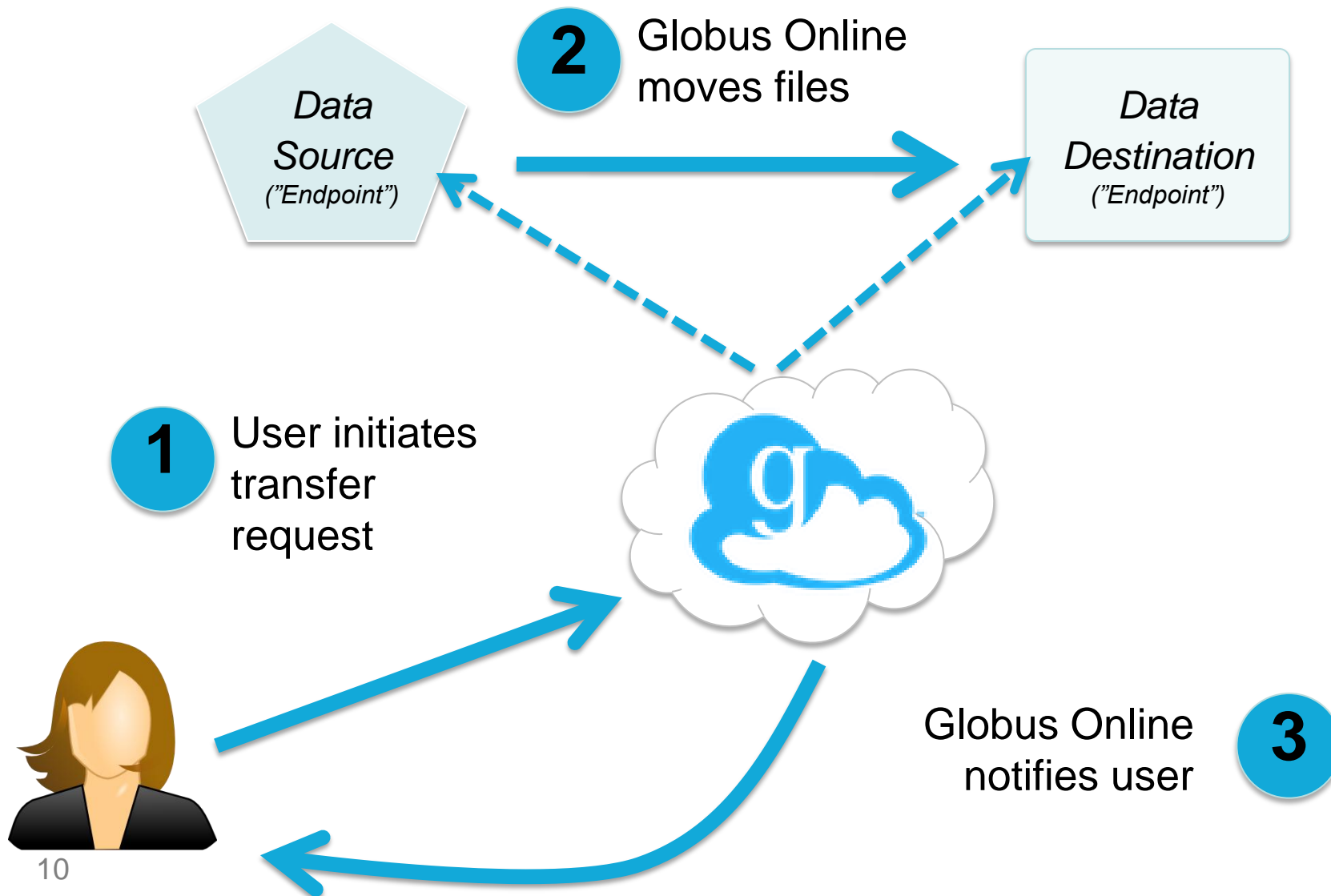 get big
  local -> remote computer
  put big

Data Transfer for Small Files--Windows

- PSCP (can be downloaded from http://tartarus.org/~simon/putty-snapshots/x86)

  remote -> local computer
  > pscp userName@ranger.tacc.utexas.edu:big localBig   (Ranger)
  > pscp userName@linuxlogin1.cac.cornell.edu:big localBig (CAC)

  local -> remote computer
  > pscp localBig userName@ranger.tacc.utexas.edu:big

- PSFTP (can be downloaded from http://tartarus.org/~simon/putty-snapshots/x86)

  [local] $ psftp userName@ranger.tacc.utexas.edu or psftp userName@linuxlogin1.cac.cornell.edu

  remote -> local computer
  get big
  local -> remote computer
  put big

Data Transfer using the XSEDE User Portal File Manager
https://portal.xsede.org/

- Small Files
  - Select Resources->File Manager from the XSEDE User Portal
  - A Java Applet will run and give you two screens
  - Open the resources on these screens 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

- Large files
  - Before opening the resources, right click and edit the resources
  - Click Stripe Transfers
  - Choose the file(s) to transfer from the source resource
  - Drag and drop the file(s) to the destination resource

2 Globus Online moves files

Data Source *("Endpoint")*

Data Destination *("Endpoint")*

1 User initiates transfer request

Globus Online notifies user **3**

10

Globus

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

Globus

- Install Globus Connect https://www.globusonline.org/xfer/StartTransfer?gcdialog=true

Globus

- Transfer file   https://www.globusonline.org/xfer/StartTransfer

Globus

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

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 from the Command Line

- Enable globus account for ssh

  add SSH public key

  https://www.globusonline.org/account/ManageIdentities

- ssh to cli.globusonline.org

  ssh cli.globusonline.org

- Commands include:

  transfer

  scp

- Documentation:

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

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

References

- TACC User Guides
  - https://www.xsede.org/web/guest/tacc-ranger
  - https://www.xsede.org/web/guest/tacc-lonestar
  - https://www.xsede.org/tacc-ranch

- Globus
  - www.globusonline.org
  - support@globusonline.org