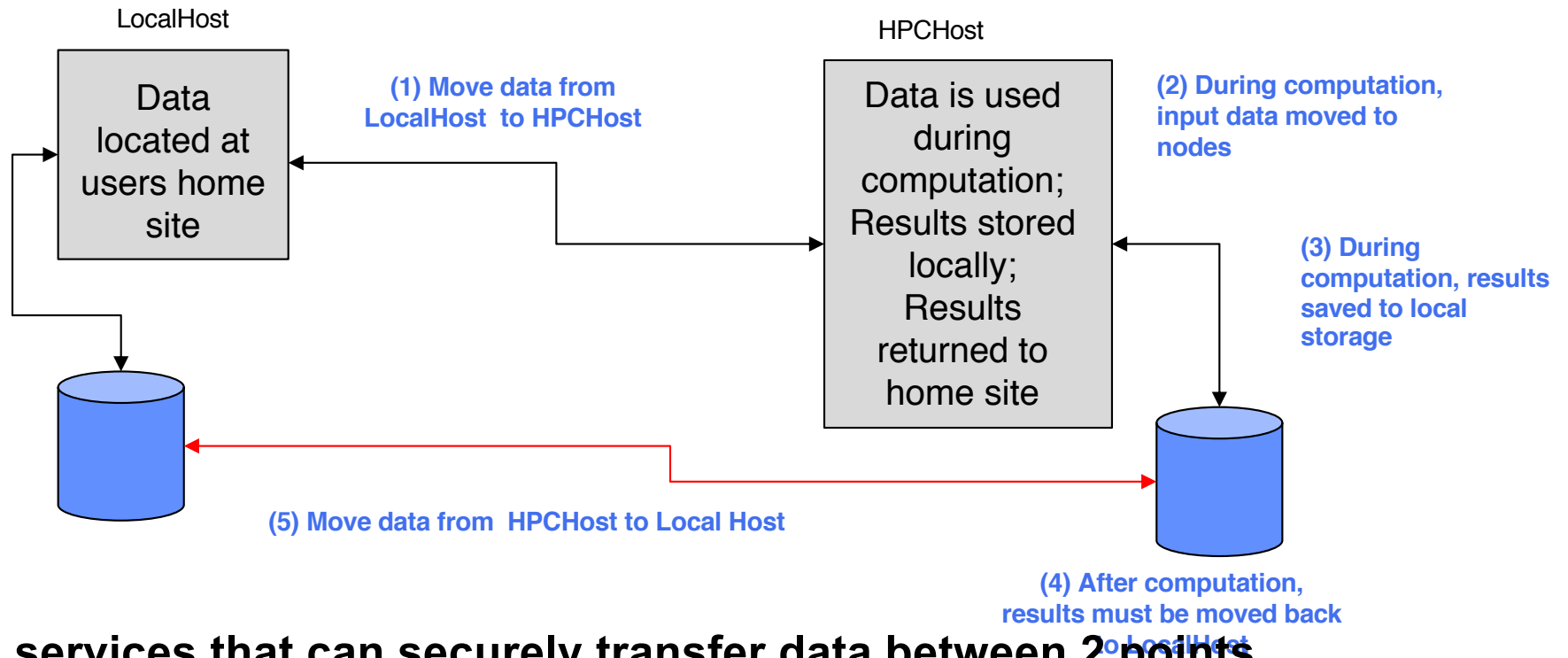


TSCC Bootcamp: Introduction to Accessing and Running Jobs on the TSCC System

Moving Data

Moving Data To/From TSCC



- **Need services that can securely transfer data between 2 points.**
 - Hosted on compute nodes, storage devices,
- **Most common commands are SCP, SFTP, wget**
 - Apps run from command line, from scripts, or other apps

Factors Affecting TSCC Data Transfer

- **Network bandwidth:**
 - CPU network: 10GbE (QDR InfiniBand optional)
 - Home-to-work: TCP 10 MB/sec, Cox optical ~ 400MB/sec
 - Time to xfer 1 TB with Cox optical fiber to SDSC:
$$\text{Txfer} \sim (1000 \text{ GB}) \cdot (8 \text{ Bits/Byte}) / (400 \text{ MBits/sec}) \sim 20,000 \text{ secs} \sim 5.5 \text{ hours}$$
- **IO hardware on node, storage, servers**
- **Data size and type**
- **Other activity on node, network, servers**

Ways to Move Data

- **Direct Transfers:**
 - **Secure copy protocol (SCP):** secure transfer of computer files between a local host and a remote host or between two remote hosts.
 - **SSH File Transfer Protocol (SFTP):** SSH + remote file system management functionality, from within a shell
 - **GNU wget:** used for retrieving files using HTTP, HTTPS, FTP and FTPS. Best for web page scraping.
 - **GUI Apps:** Fetch, Filezilla
 - **Many other applications** (IBM aspera,)
- **3rd party file transfers: move data between endpoints (services); independent of requester.**
 - Globus IO:
- **TSCC Data Mover Node:** tscs-dm1.sdsc.edu
 - need to login to move data from external location

SCP Copy from Laptop to HPC System

```
[mthomas@gidget:~/working/bigdatafiles] scp 10MB.dat.tar.gz mthomas@comet.sdsc.edu:bigdatafiles/10MB.dat.tar.gz
10MB.dat.tar.gz                                     100% 29KB 444.1KB/s 00:00

[mthomas@gidget:~/working/bigdatafiles] ssh -l mthomas comet.sdsc.edu "ls -al bigdatafiles"
total 29
drwxr-xr-x  2 mthomas use300    3 Jan 24 04:37 .
drwxr-x--- 23 mthomas use300   45 Jan 24 04:35 ..
-rw-r--r--  1 mthomas use300 29262 Jan 24 04:37 10MB.dat.tar.gz
[mthomas@gidget:~/working/bigdatafiles]
```

```
[mthomas@comet-ln2:~] cd bigdatafiles/
[mthomas@comet-ln2:~/bigdatafiles] ll
total 20
drwxr-xr-x  2 mthomas use300  2 Jan 24 04:35 .
drwxr-x--- 23 mthomas use300 45 Jan 24 04:35 ..
[mthomas@comet-ln2:~/bigdatafiles] ls -al
total 29
drwxr-xr-x  2 mthomas use300    3 Jan 24 04:37 .
drwxr-x--- 23 mthomas use300   45 Jan 24 04:35 ..
-rw-r--r--  1 mthomas use300 29262 Jan 24 04:37 10MB.dat.tar.gz
[mthomas@comet-ln2:~/bigdatafiles]
```

Syntax:

`scp [options] username1@source_host:directory1/filename1
username2@destination_host:directory2/filename2`

Globus Data Transfer

- 3rd Party file transfer
- Service provider hosts "endpoints"
- Client installs local software or visits website.
- Client submits request to transfer files from A→B
 - Globus gets to work moving/monitoring transfer of data
- Client goes and gets coffee.
- Faster than SCP/SFTP



<https://www.globus.org/data-transfer>