Large Input in DHTC

Thursday PM, Lecture 1

Derek Weitzel

OSG

Transfers
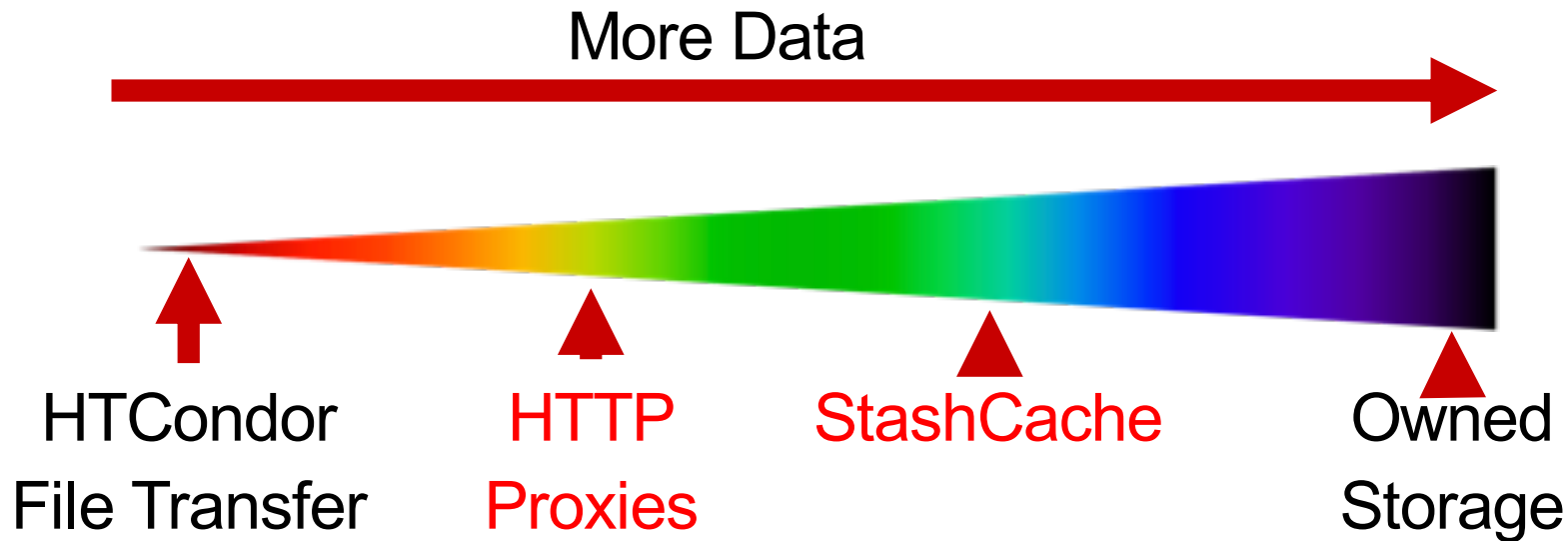
Hardware transfer limits

<10MB/file, 1GB total

submit server

HTCondor

exec server

<1GB/file and total

submit file
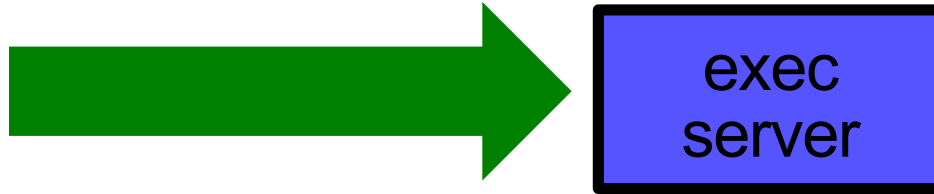executable
dir/ input
output

(exec dir)/
executable
input
output
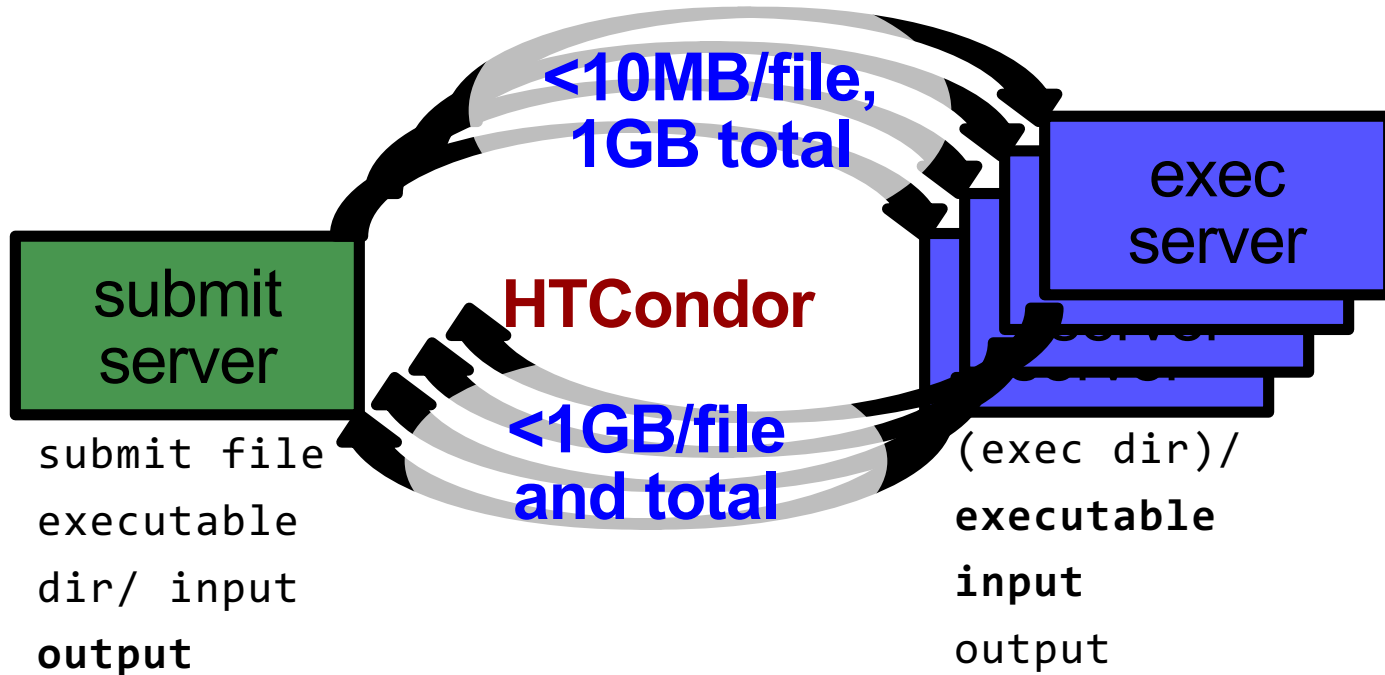
Reducing data needs

- ***An HTC best practice!***

- split large input for better throughput *and* less per-job data

- eliminate unnecessary data

- compress and combine files

Large input in HTC and OSG

file size	method of delivery
words	within executable or arguments?
tiny – 10MB per file	HTCondor file transfer (up to 1GB total per-job)
10MB – 1GB, shared	download from web proxy (network-accessible server)
1GB - 10GB, unique or shared	StashCache (regional replication)
10 GB - TBs	shared file system (local copy, local execute servers)

Using a Web Proxy

- Place the file onto a local, proxy-configured web server
- Have HTCondor download via HTTP address
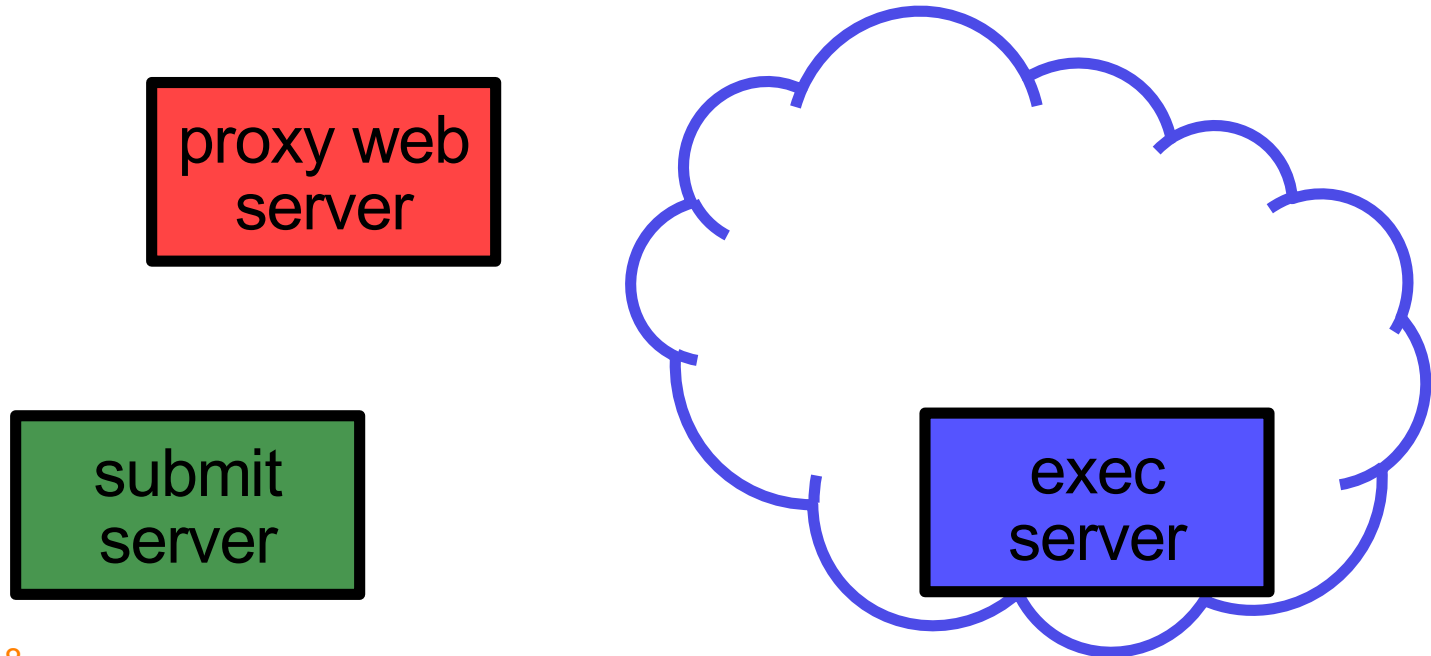
proxy web server

submit server

exec server

Using a Web Proxy

- Place the file onto a proxy-configured web server

- Have HTCondor download via HTTP address
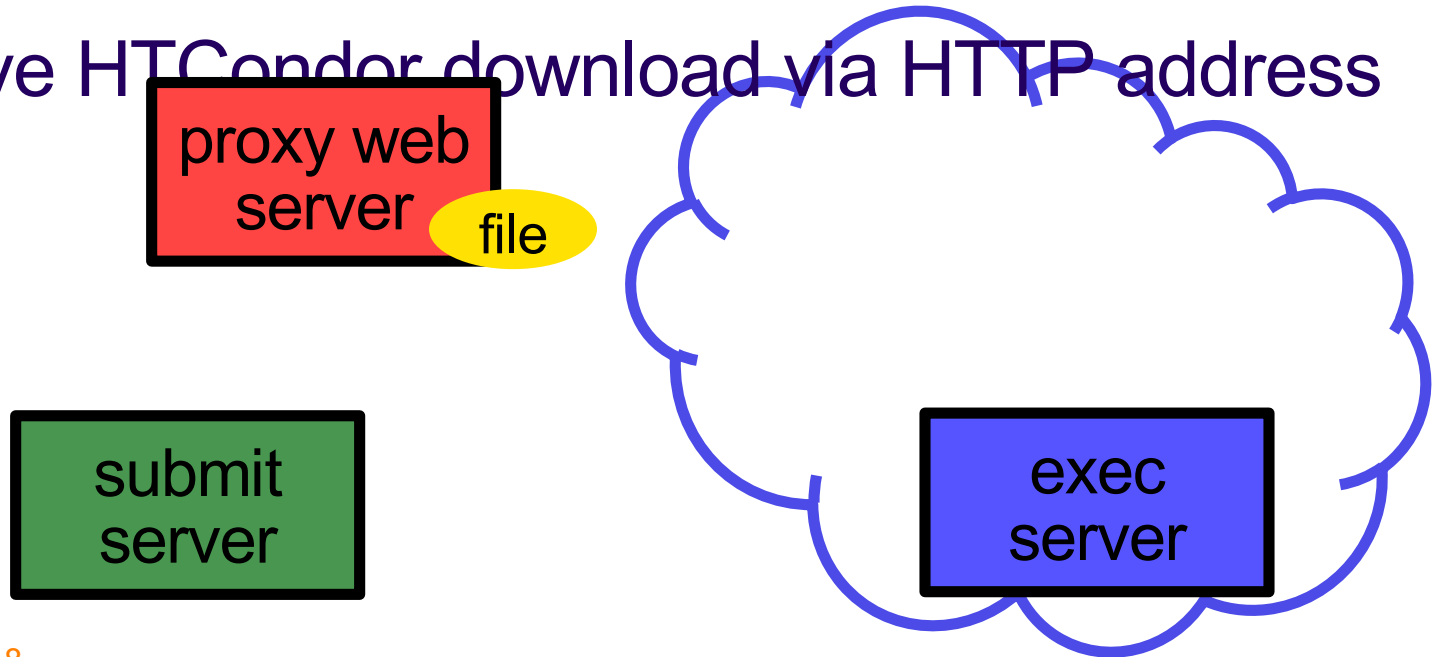
proxy web server

file

submit server

exec server

Using a Web Proxy

- Place the file onto a proxy-configured web server

- Have HTCondor download via HTTP address

Using a Web Proxy

- Place the file onto a proxy-configured web server

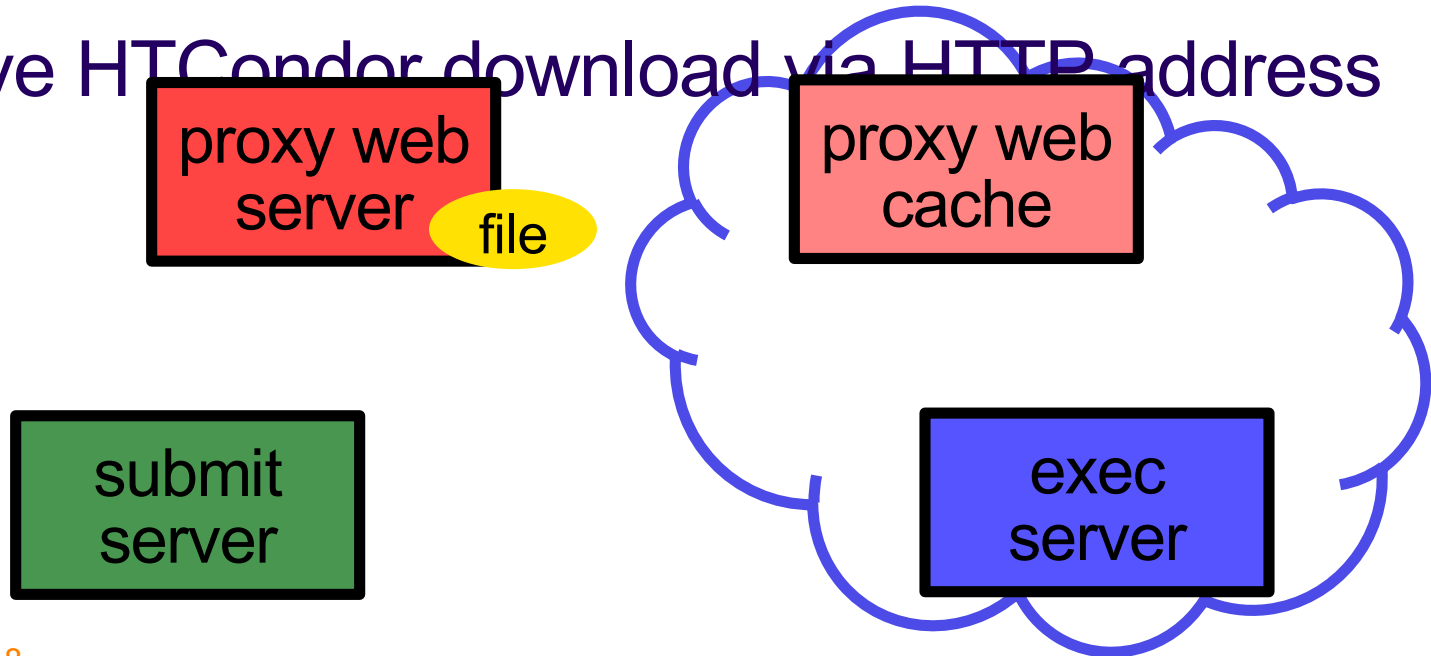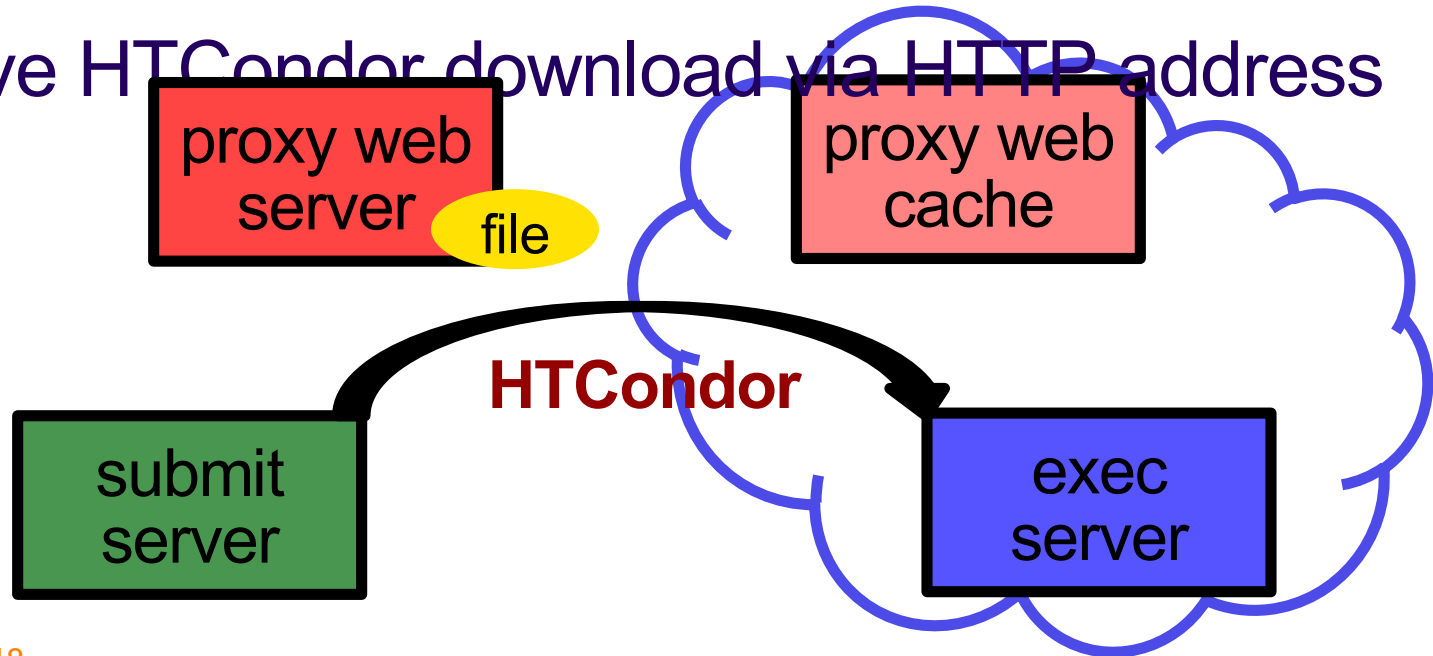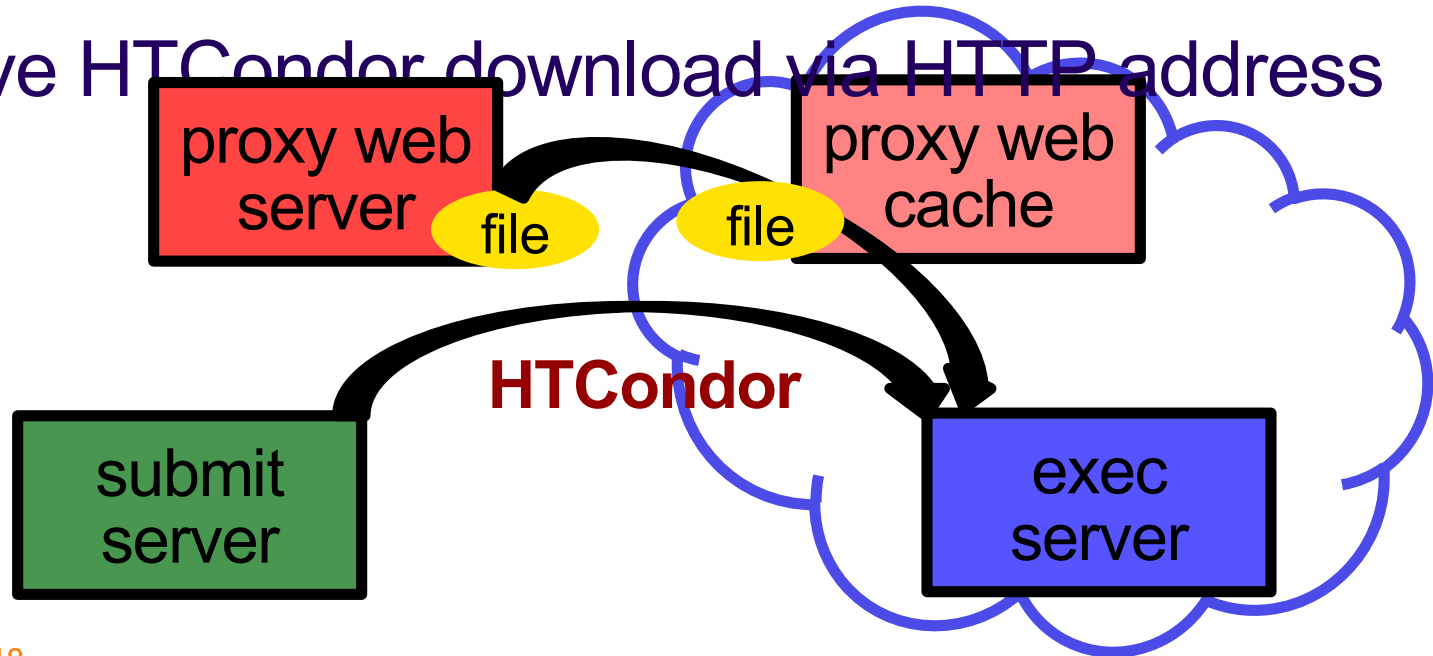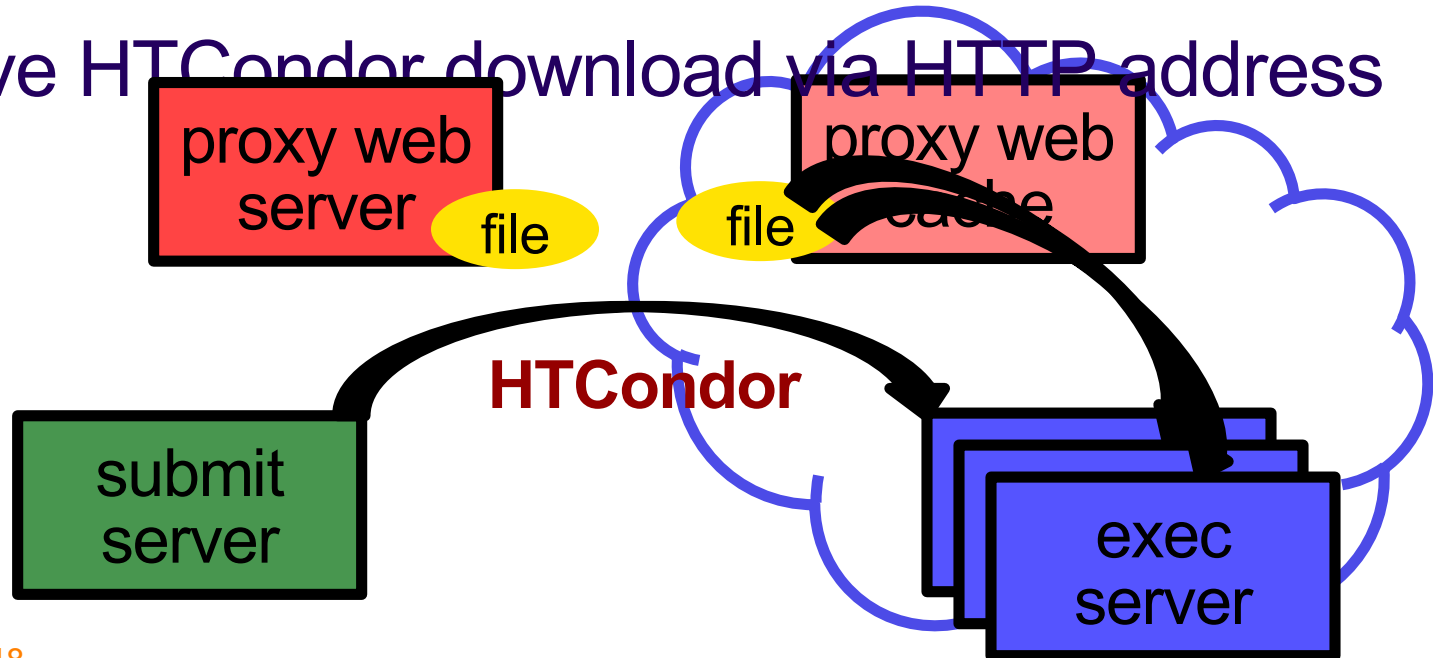- Have HTCondor download via HTTP address

Using a Web Proxy

- Place the file onto a proxy-configured web server

- Have HTCondor download via HTTP address

Using a Web Proxy

- Place the file onto a proxy-configured web server

- Have HTCondor download via HTTP address

proxy web server

proxy web cache

file

file

HTCondor

submit server

exec server

Downloading Proxy Files

- HTCondor submit file:
  ```
  transfer_input_files =
  http://host.univ.edu/path/to/shared.tar.gz
  ```

- Anywhere (in-executable, or test download)
  ```
  wget http://host.univ.edu/path/to/shared.tar.gz
  ```
 - in-executable: make sure to delete after un-tar or at the end of the job!!! (HTCondor thinks it's 'new')

Web Proxy Considerations

- Managed per-VO

- Memory limited, **max file size: 1 GB**

- Local caching at OSG sites

  - good for _shared_ input files, only

  - perfect for software and common input

  - need to rename changed files!!!

- Files are downloadable by **ANYONE** who has the specific HTTP address

  - Will work on 100% of OSG sites, though not all sites will have a local cache

At OSG Connect (Ex. 3.1)

- place files in $HOME/stash/public

- address: `http://stash.osgconnect.net/~user/shared.tar.gz`
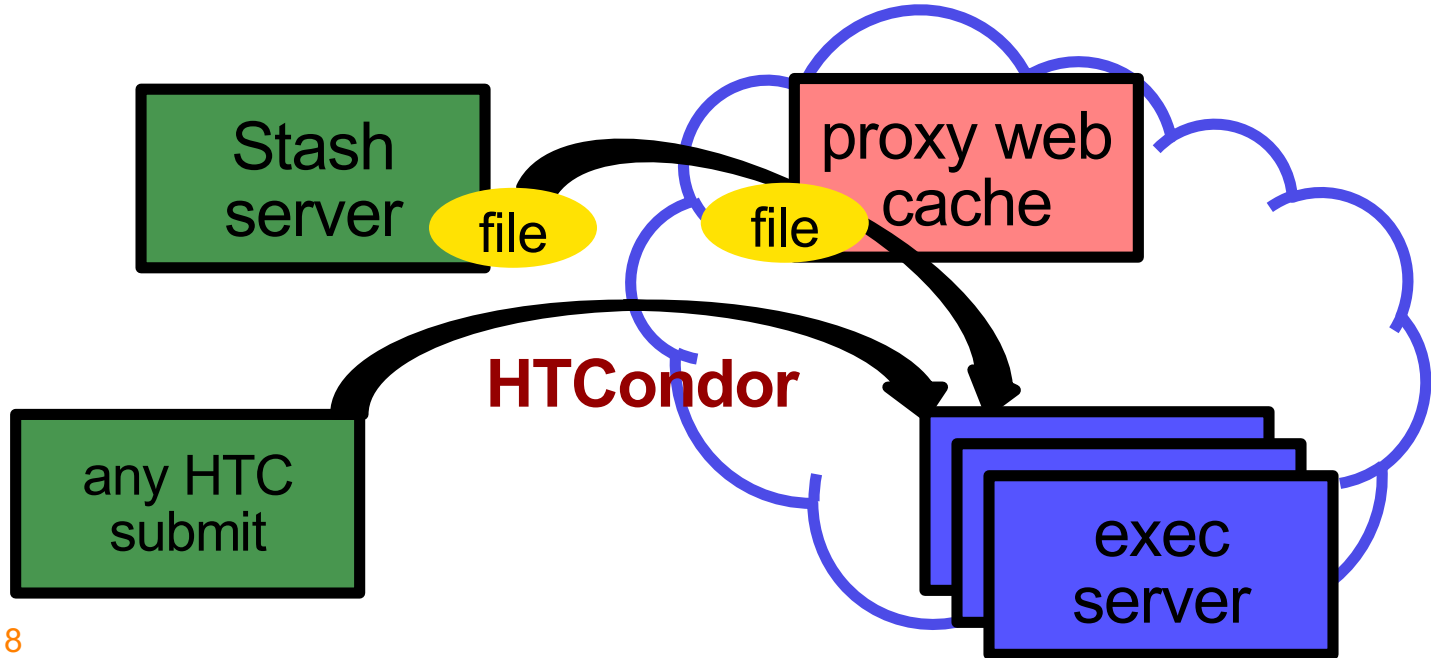
Large input in HTC and OSG

exec server

file size	method of delivery
words	within executable or arguments?
tiny – 10MB per file	HTCondor file transfer (up to 1GB total per-job)
10MB – 1GB, shared	download from web proxy (network-accessible server)
1GB - 10GB, unique or shared	StashCache (regional replication)
10 GB - TBs	shared file system (local copy, local execute servers)

Using StashCache for Input

- regionally-cached repository managed by OSG Connect

Stash origin: ⭐

OSG Caches: 🟠

Placing Files in StashCache

- place files in /home/**user**/public on osgconnect.net

osgconnect.net

local server

/home/**username**/public

"Stash" origin

file

any OSG submit

regional cache

exec server
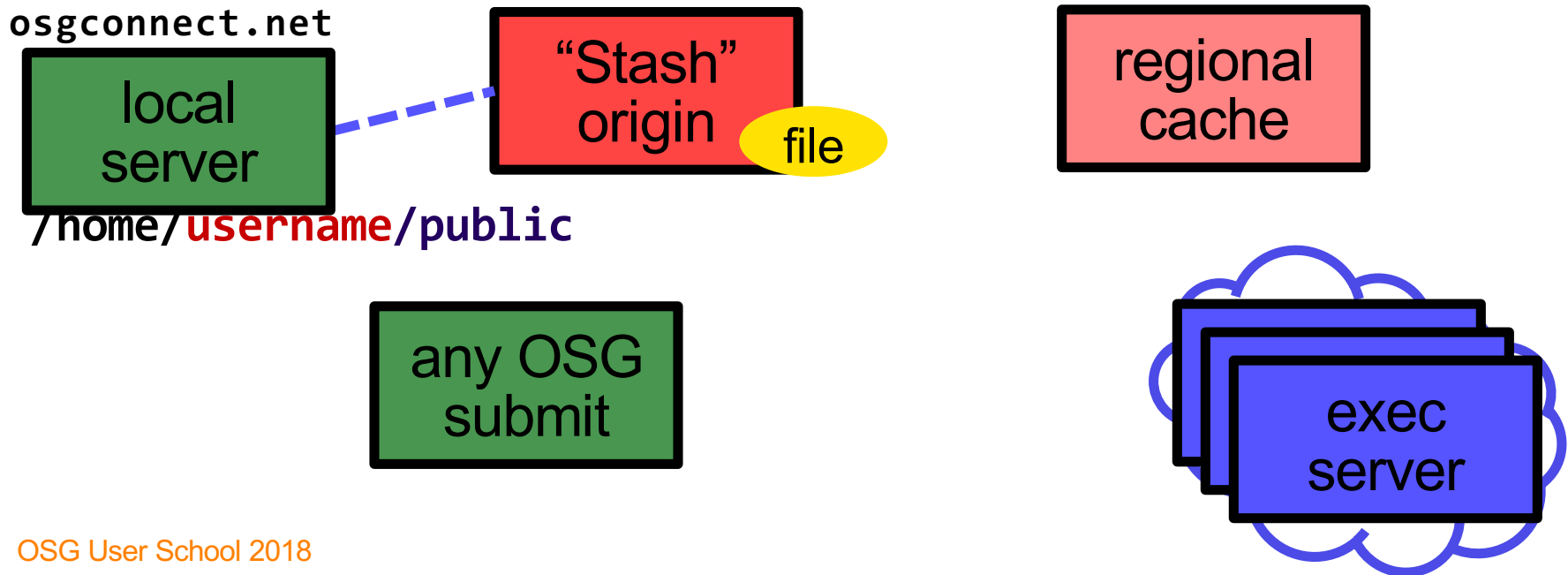
Obtaining Files in StashCache

- Use HTCondor transfer for other files

`osgconnect.net`

local server

"Stash" origin

file

file

regional cache

`/home/username/public`

any OSG submit

HTCondor

exec server
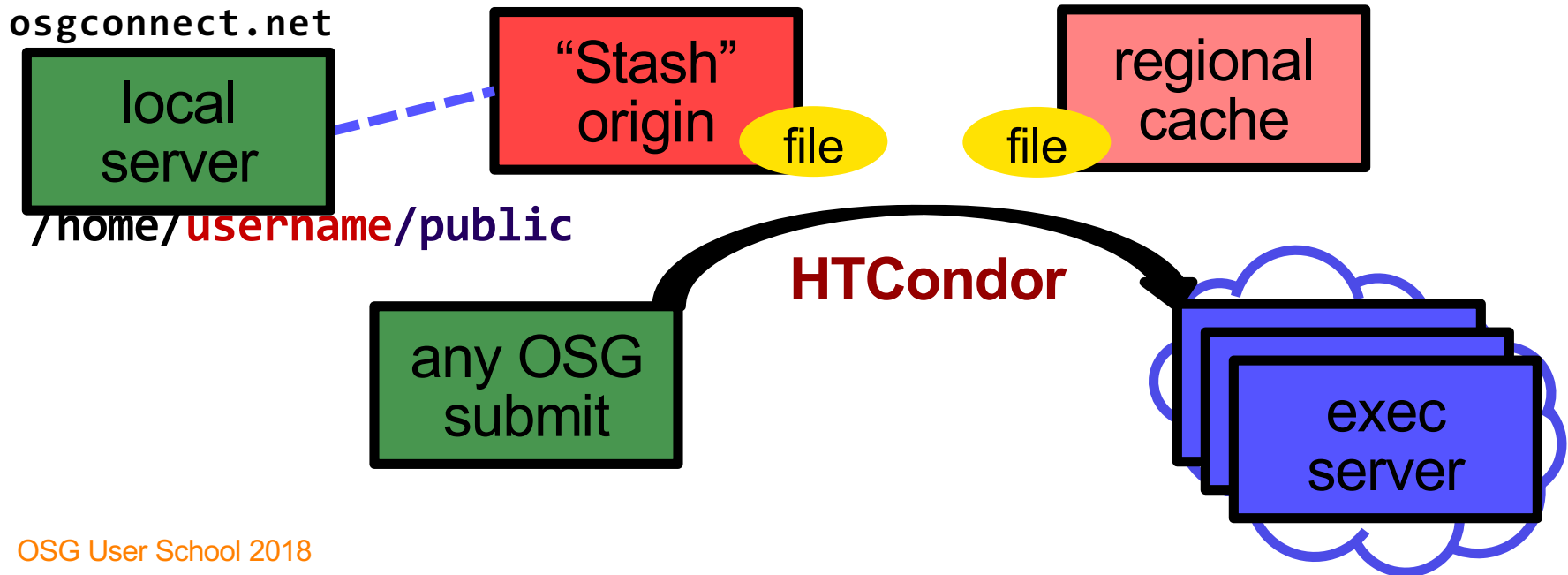
Obtaining Files in StashCache

- Download using `stashcp` command (available as an OASIS software module)
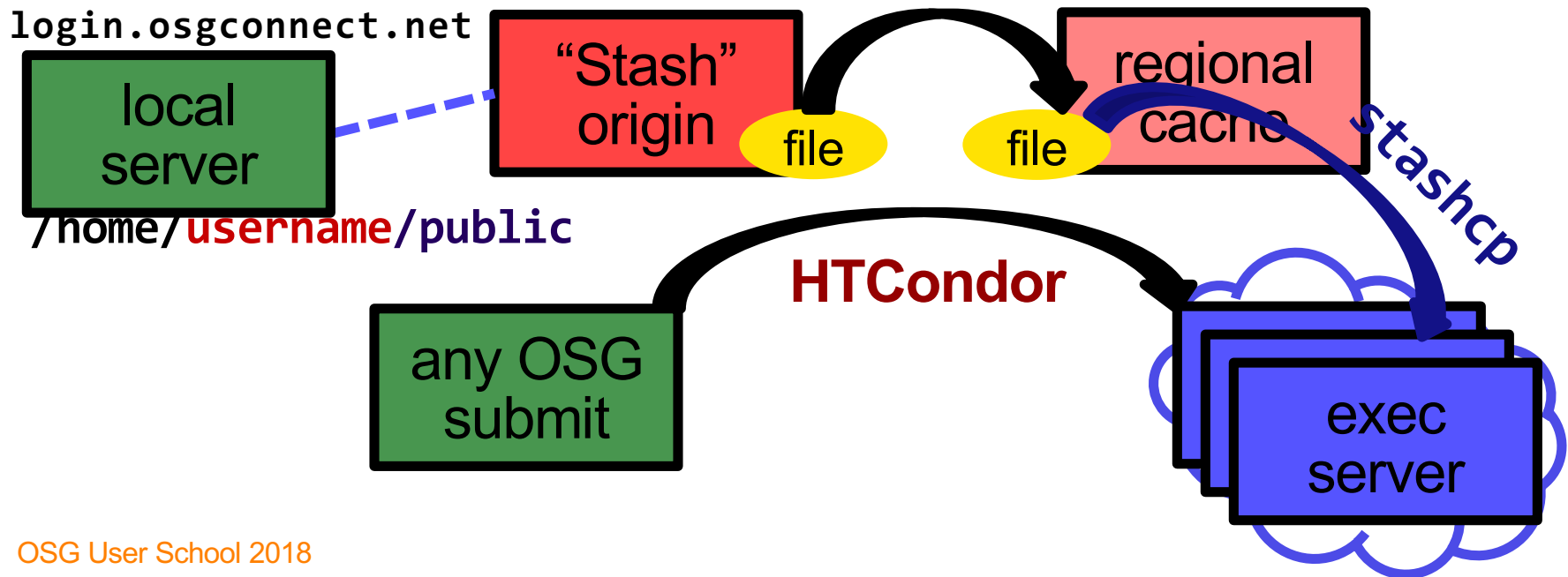
In the Submit File

- Require StashCashe sites in the submit file

`+WantsStashCache`

- Require sites with OASIS modules (for `stashcp`)

`Requirements = <other &&> (HAS_MODULES =?= true)`

In the Job Executable

```bash
#!/bin/bash
# setup:
module load stashcp
stashcp /user/username/public/file.tar.gz ./


<untar, then remove the file>
<job commands>


<remove all files from StashCache>
# END
```
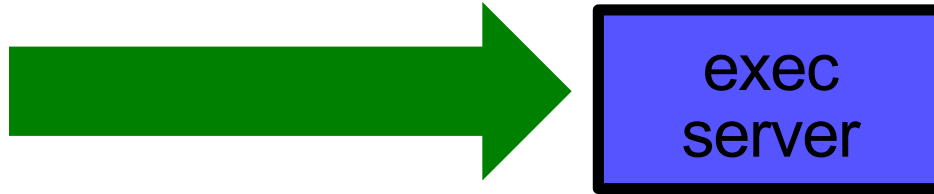
StashCache Considerations

- Available at ~90% of OSG sites
- Regional caches on *very fast* networks
  - **Max file size: 10 GB**
  - *shared* OR *unique* data
- Can copy multiple files totaling >10GB
- Just like HTTP proxy, change name when update files

Large input in HTC and OSG

file size	method of delivery
words	within executable or arguments?
tiny – 10MB per file	HTCondor file transfer (up to 1GB total per-job)
10MB – 1GB, shared	download from web proxy (network-accessible server)
1GB - 10GB, unique or shared	StashCache (regional replication)
10 GB - TBs	shared file system (local copy, local execute servers)

Other Options?

- Some distributed projects with LARGE, shared datasets may have project-specific repositories that exist only on certain sites
  - (e.g. CMS, Atlas, LIGO?, FIFE?, others?)
  - Jobs will require specific sites with local copies and use project-specific access methods
- OASIS?
  - Best for lots of small files per job (e.g. software)
  - StashCache and Proxies better for fewer larger files per job

Cleaning Up Old Data

- For StashCache *AND* web proxies:

  make sure to delete data when you no longer need it in the origin!!!

- StashCache and VO-managed web proxy servers do NOT have unlimited space!
  - Some may regularly clean old data for you. Check with local support.

Other Considerations

- Only use these options if you MUST!!
  - Each comes with limitations on site accessibility and/or job performance, and extra data management concerns

file size	method of delivery
words	within executable or arguments?
tiny – 10MB per file	HTCondor file transfer (up to 1GB total per-job)
10MB – 1GB, shared	download from web proxy (network-accessible server)
1GB - 10GB, unique or shared	StashCache (regional replication)
10 GB - TBs	shared file system (local copy, local execute servers)

Exercises

- 3.1 Using a web proxy for shared input
  - place the blast database on the web proxy

- 3.2 StashCache for shared input
  - place the blast database in StashCache

- 3.3 StashCache for unique input
  - convert movie files

Questions?

- Feel free to contact me:
  - dweitzel@cse.unl.edu

- Next: Exercises 3.1-3.3
- Later: Large *output* and shared filesystems