

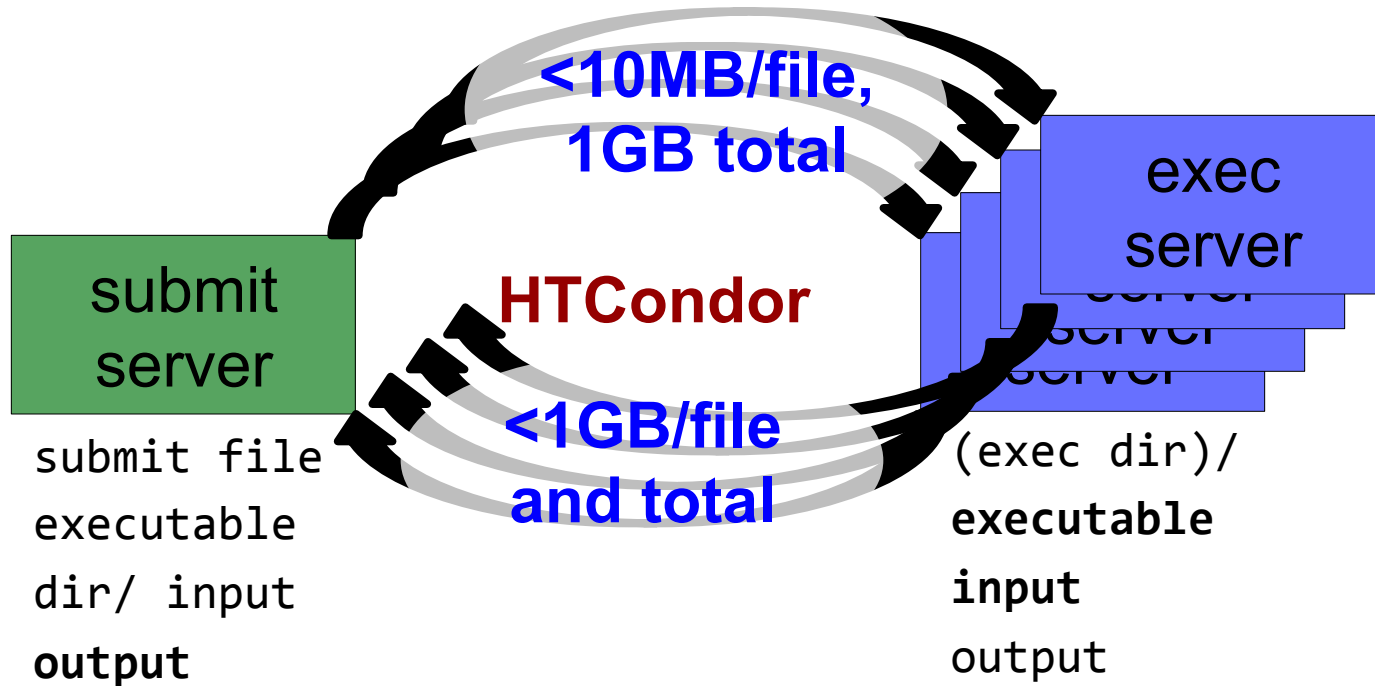
# Large Output and Shared File Systems

Thursday PM, Lecture 2

Lauren Michael

CHTC, UW-Madison

# Per-job transfer limits



# What's Different for Output?

---

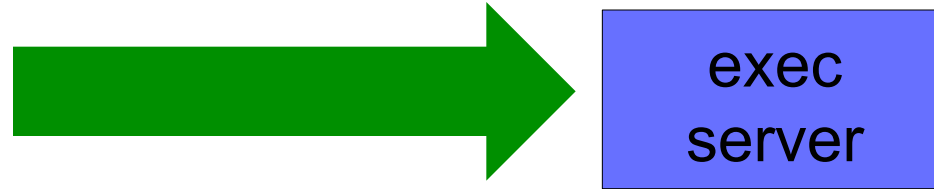
- always unique (right?)
- caching won't help
- files not associated with your local username
  - security barriers outside of local context
- security issues with world-writability
  - (versus okay world-readability for input)

# Output for HTC and OSG



file size	method of delivery
words	<del>within executable or arguments?</del>
tiny – <u>1GB</u>	HTCondor file transfer (up to 1 GB total per-job)
<b>1GB+</b>	<b>shared file system (local execute servers)</b>

# 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)
<b>10 GB – TBs, unique or shared</b>	<b>shared file system (local copy, local execute servers)</b>

# (Local) Shared Filesystems

---

- data stored on file servers, but network-mounted to local submit and execute servers
- use local user accounts for file permissions
  - Jobs run as YOU!
  - readable (input) and writable (output, most of the time)
- *MOST* perform better with fewer large files (versus many small files of typical HTC)

# Shared FS Technologies

---

- *via network mount*
  - NFS
  - AFS
  - Lustre
  - Gluster (may use NFS mount)
  - Isilon (may use NSF mount)
- *distributed files systems (data on many exec servers)*
  - HDFS (Hadoop)
  - CEPH

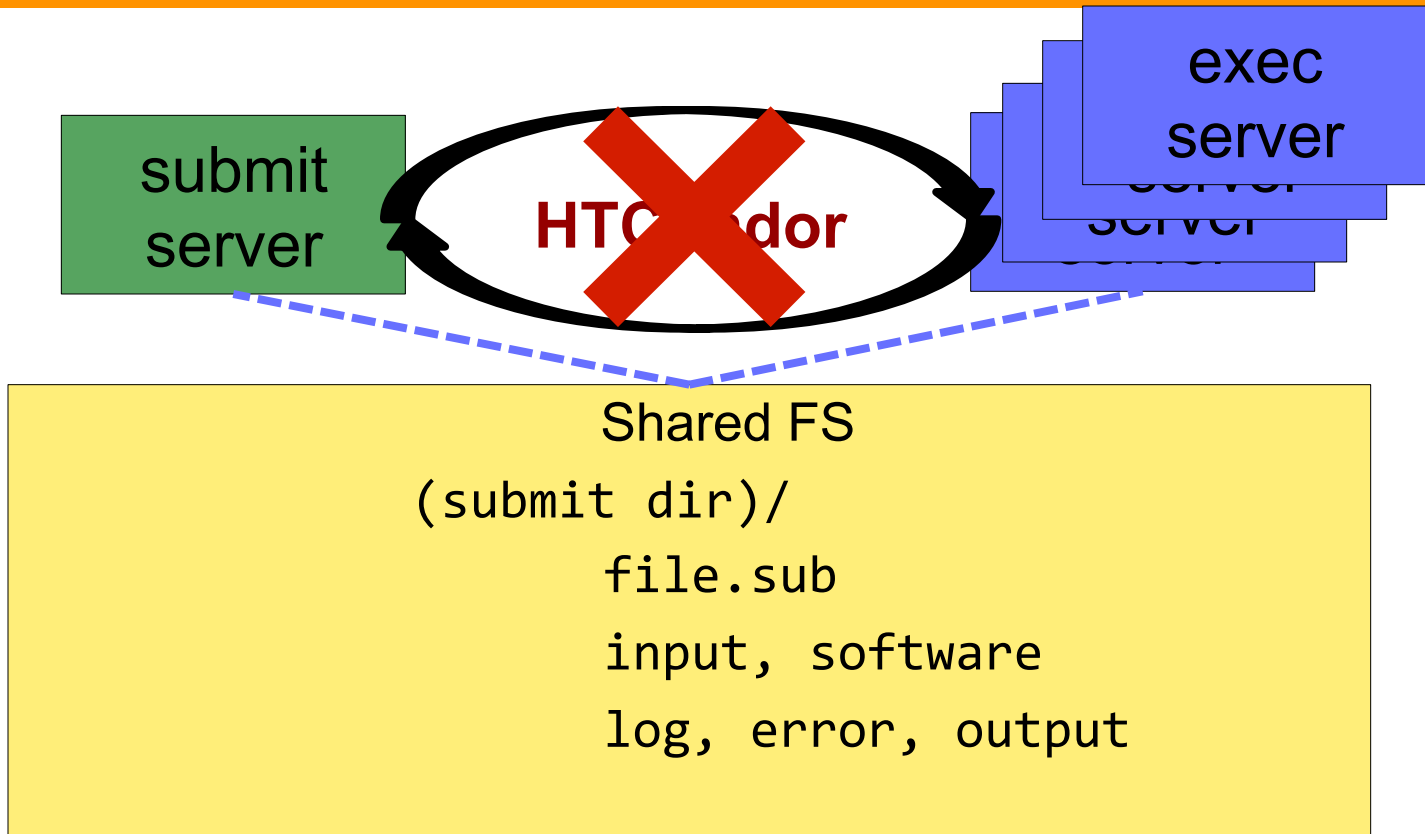
# Shared FS Configurations

---

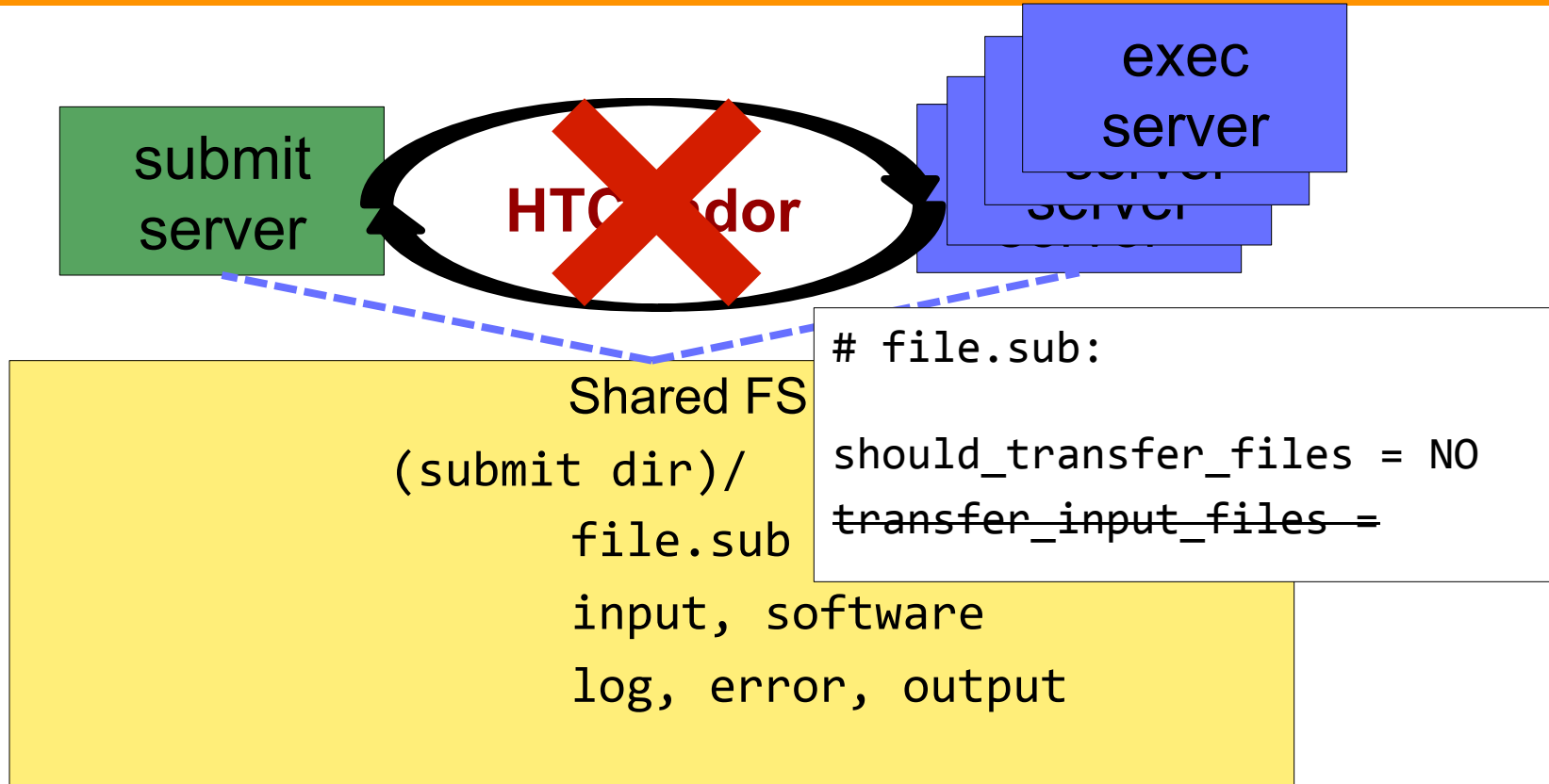
1. Submit directories *WITHIN* the shared filesystem
  - most campus clusters
  - limits HTC capabilities!!
2. Shared filesystem separate from local submission directories
  - supplement local HTC systems
  - treated more as a repository for VERY large data (>GBs)
3. Read-only (input-only) shared filesystem
  - Treated as a repository for VERY large input, only



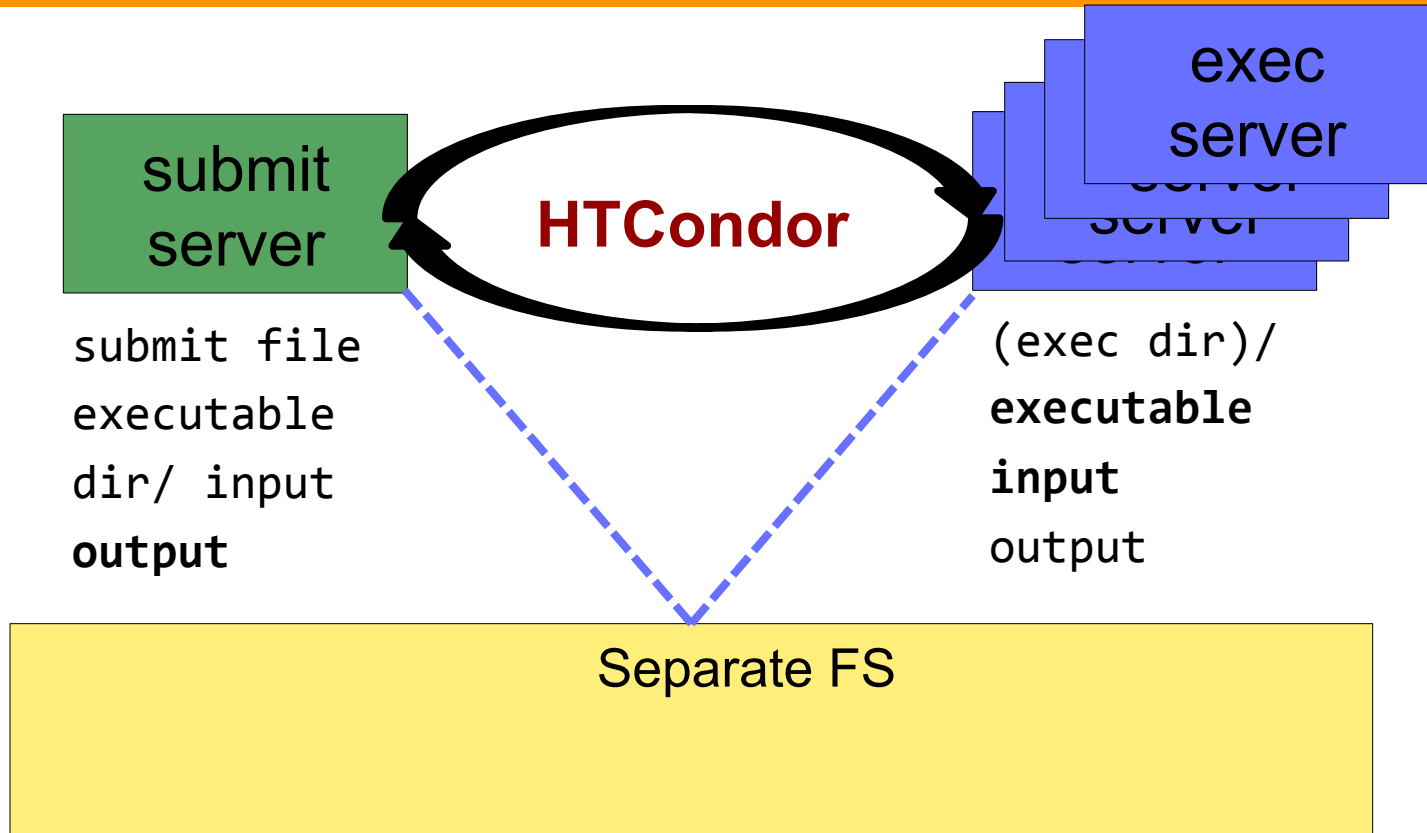
# Submit dir within shared FS



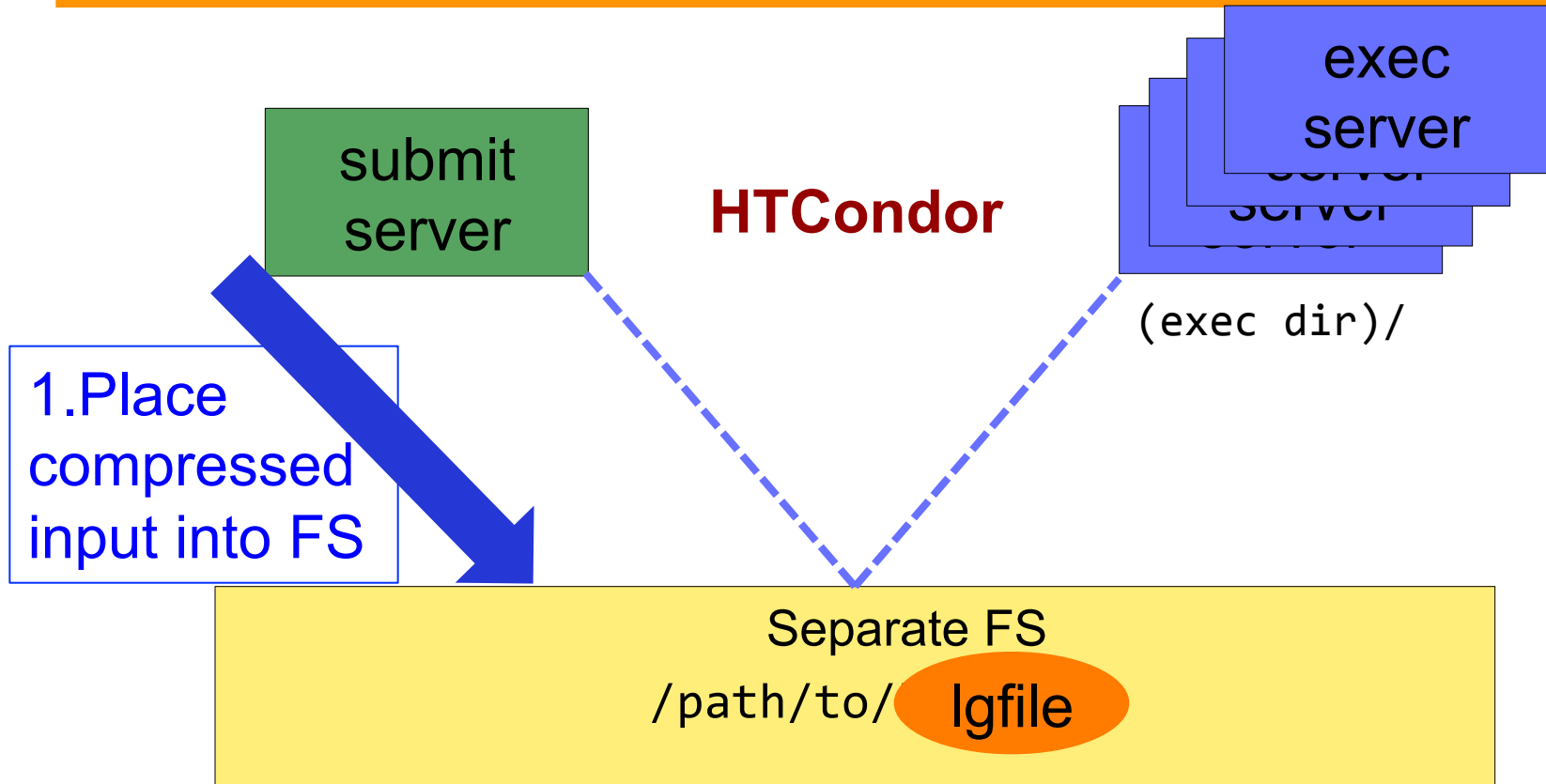
# Submit dir within shared FS



# Separate shared FS

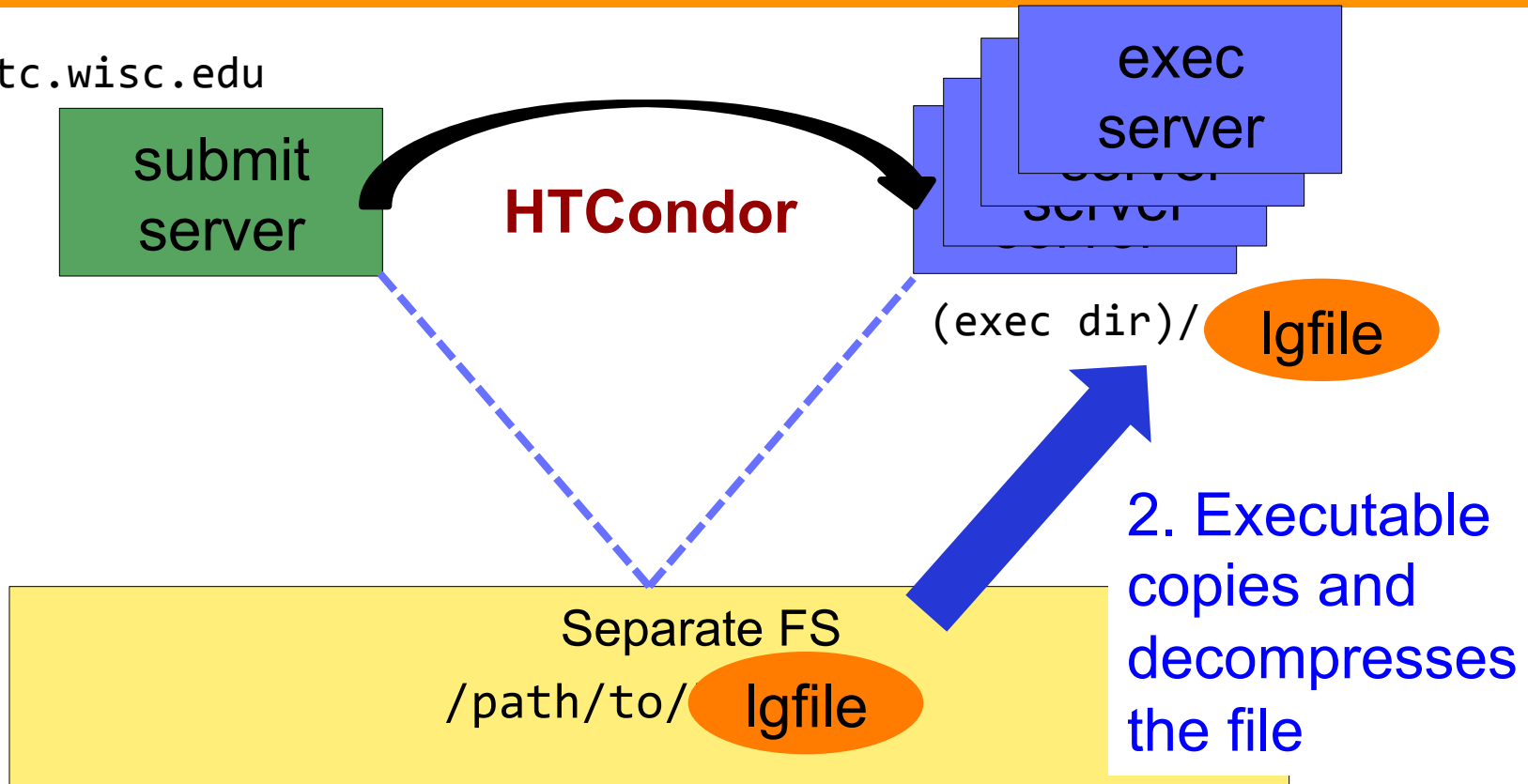


# Separate shared FS - Input



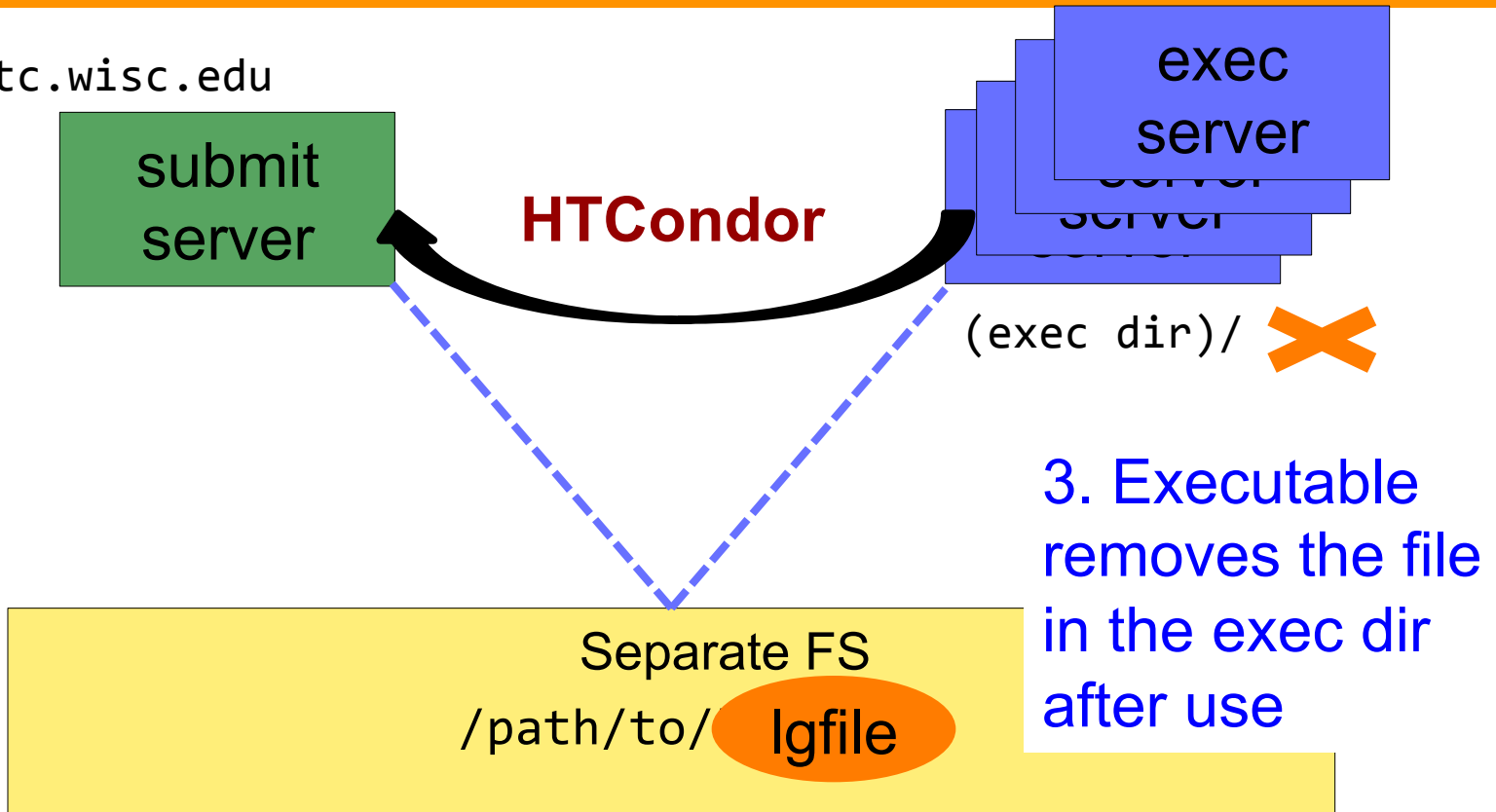
# Separate shared FS - Input

learn.chtc.wisc.edu



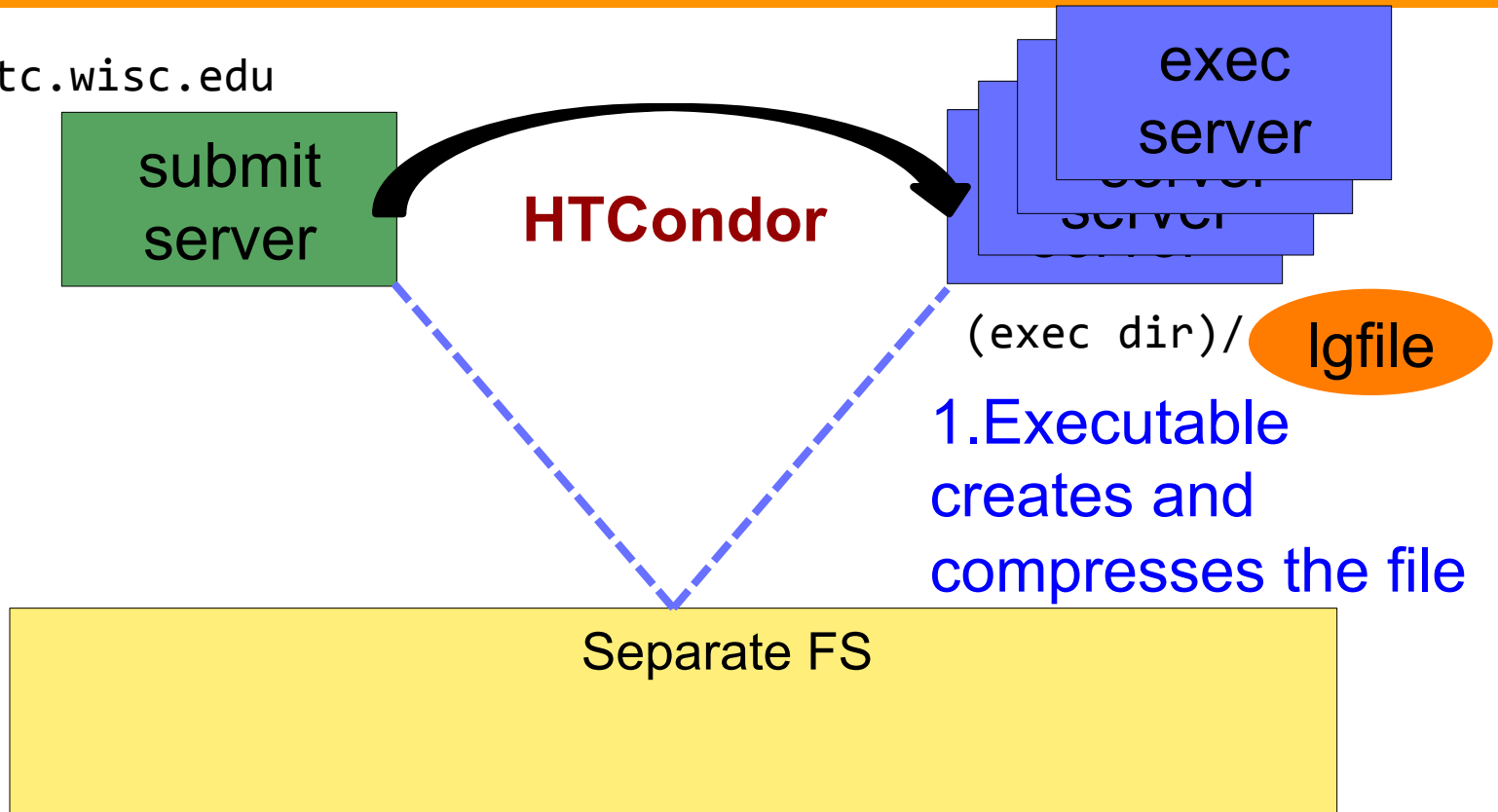
# Separate shared FS - Output

learn.chtc.wisc.edu



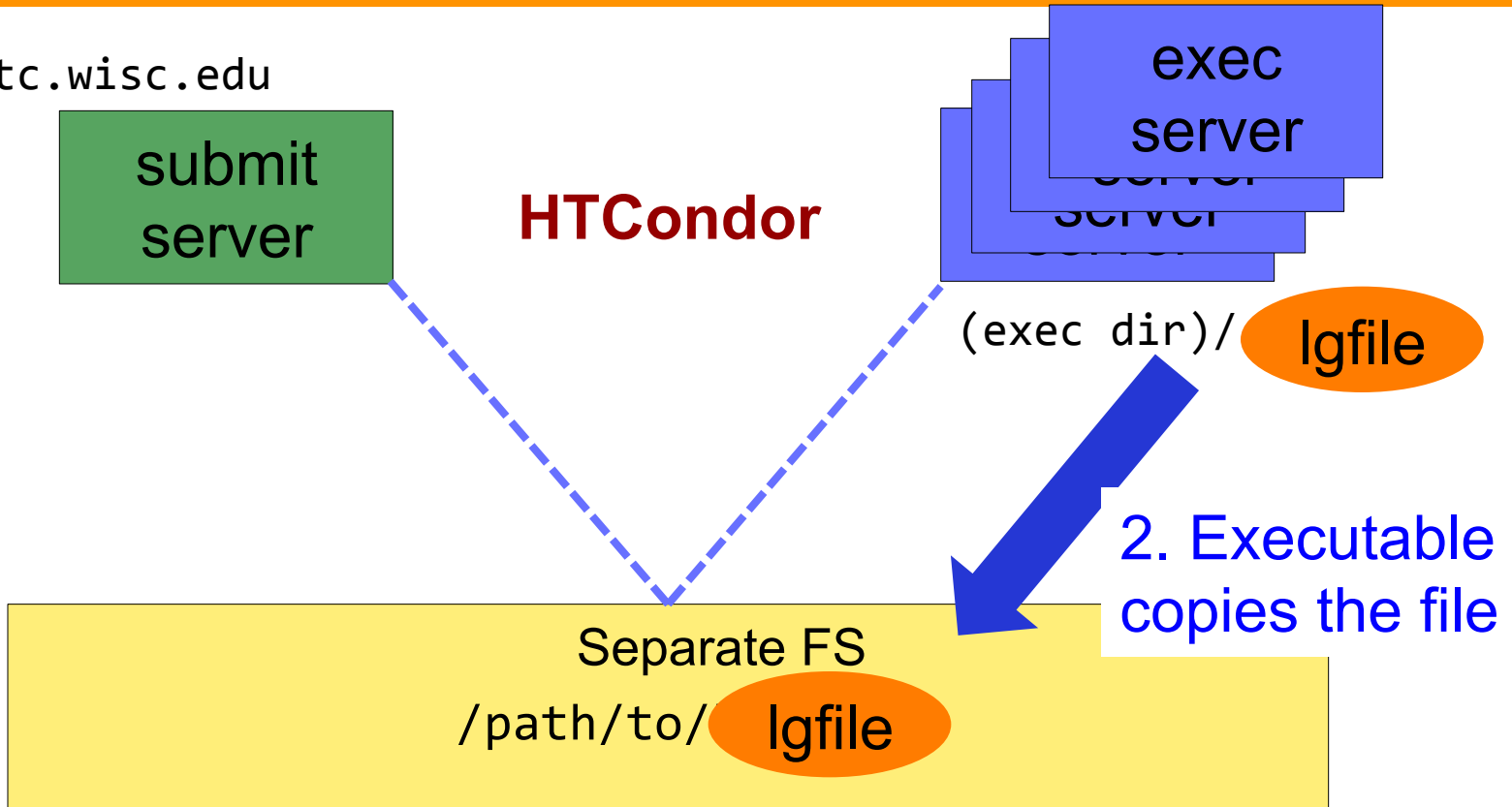
# Separate shared FS - Output

learn.chtc.wisc.edu



# Separate shared FS - Output

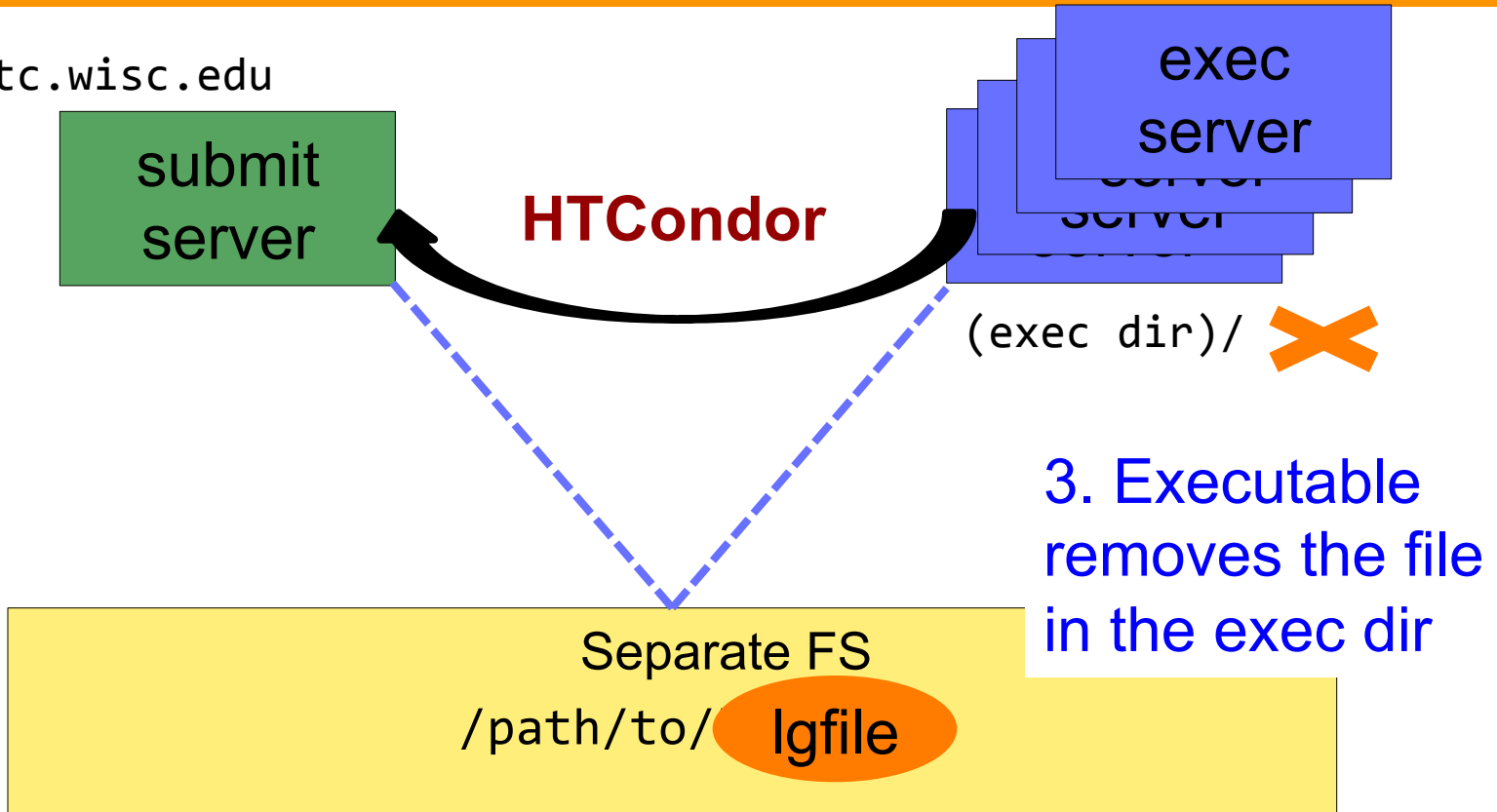
learn.chtc.wisc.edu





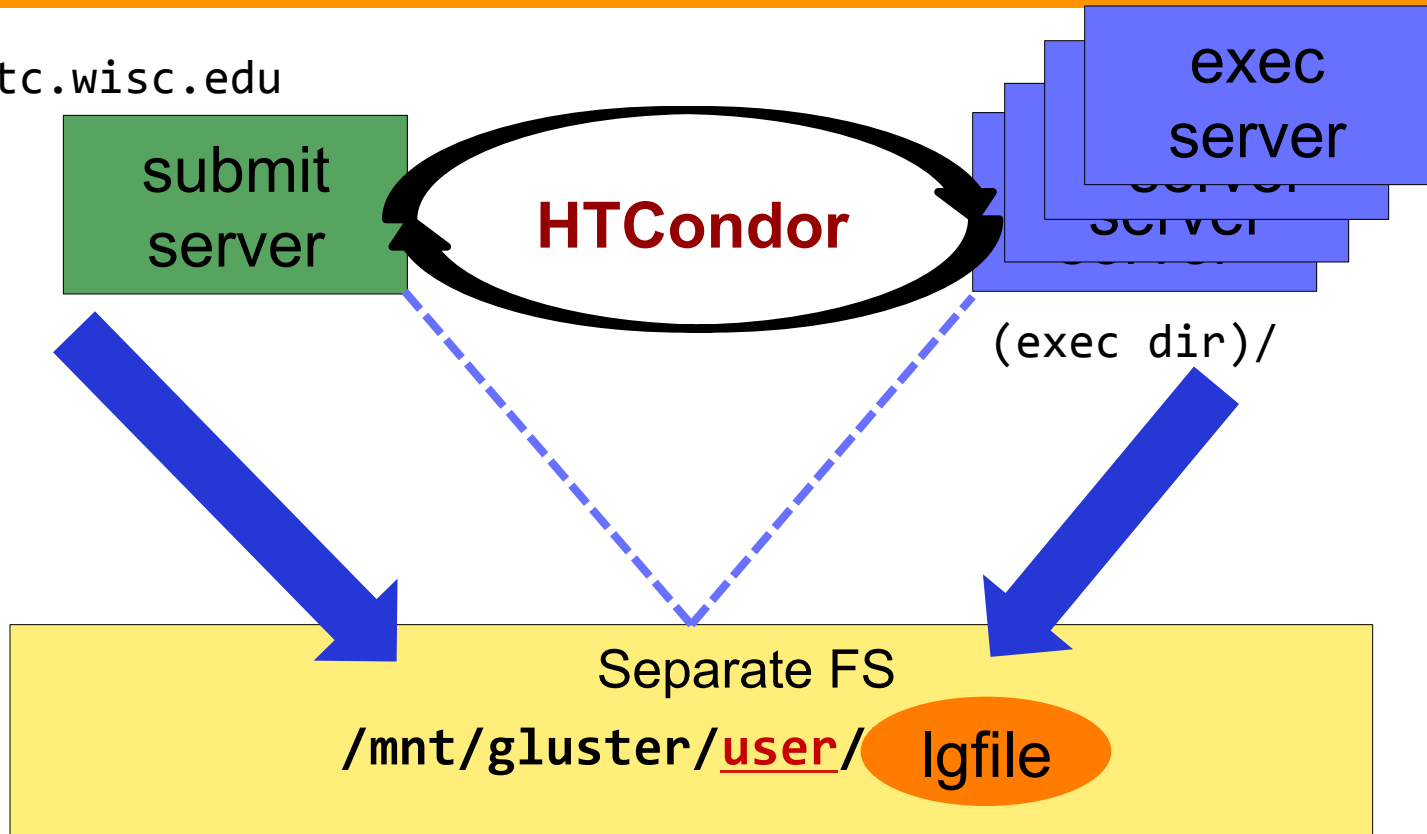
# Separate shared FS - Output

learn.chtc.wisc.edu



# At UW-Madison (Ex. 4.1-4.2)

learn.chtc.wisc.edu

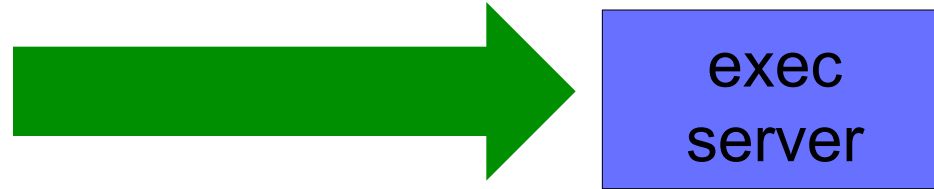


# Shared FS Configurations

---

1. Submit directories *WITHIN* the shared filesystem
  - most campus clusters
  - limits HTC capabilities!!
2. Shared filesystem separate from local submission directories
  - supplement local HTC systems
  - treated more as a repository for VERY large data (>GBs)
3. Read-only (input-only) shared filesystem
  - Treated as a repository for VERY large input, only

# 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)
<b>10 GB – TBs, unique or shared</b>	<b>shared file system (local copy, local execute servers)</b>

# Output for HTC and OSG



file size	method of delivery
words	<del>within executable or arguments?</del>
tiny – <u>1GB</u>	HTCondor file transfer (up to 1 GB total per-job)
<b>1GB+</b>	<b>shared file system (local execute servers)</b>

# Review

Option	Input or Output?	File size limits	Placing files	In-job file movement	Accessibility?
HTCondor file transfer	Both	10 MB/file (in), 1 GB/file (out); 1 GB/tot (either)	via HTCondor submit node	via HTCondor submit file	anywhere HTCondor jobs can run
Web proxy	Shared input only	1 GB/file	specific to VO	HTTP download	anywhere, by anyone
StashCache	Shared and unique input	10 GB/file (will increase!)	via OSG Connect submit server	via stashcp command (and module)	OSG-wide (90% of sites), by anyone
Shared filesystem	Input, likely output	TBs (may vary)	via mount location (may vary)	use directly, or copy into/out of execute dir	local cluster, only by YOU (usually)

# Exercises

---

- 4.1 Shared Filesystem for Large Input
- 4.2 Shared Filesystem for Large Output

# Questions?

---

- Feel free to contact me:
  - [lmichael@wisc.edu](mailto:lmichael@wisc.edu)
- Next: Exercises 4.1-4.2
- Later: Wrap-up