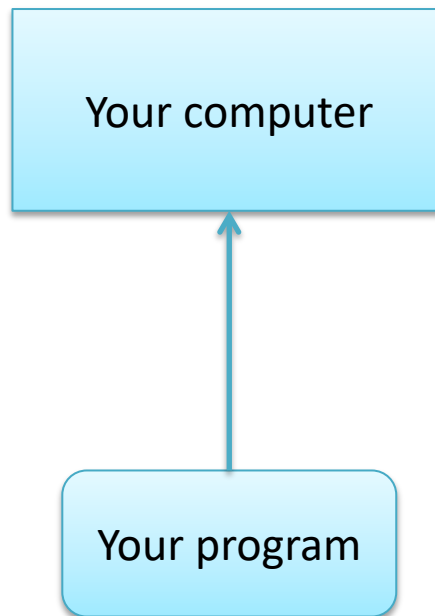


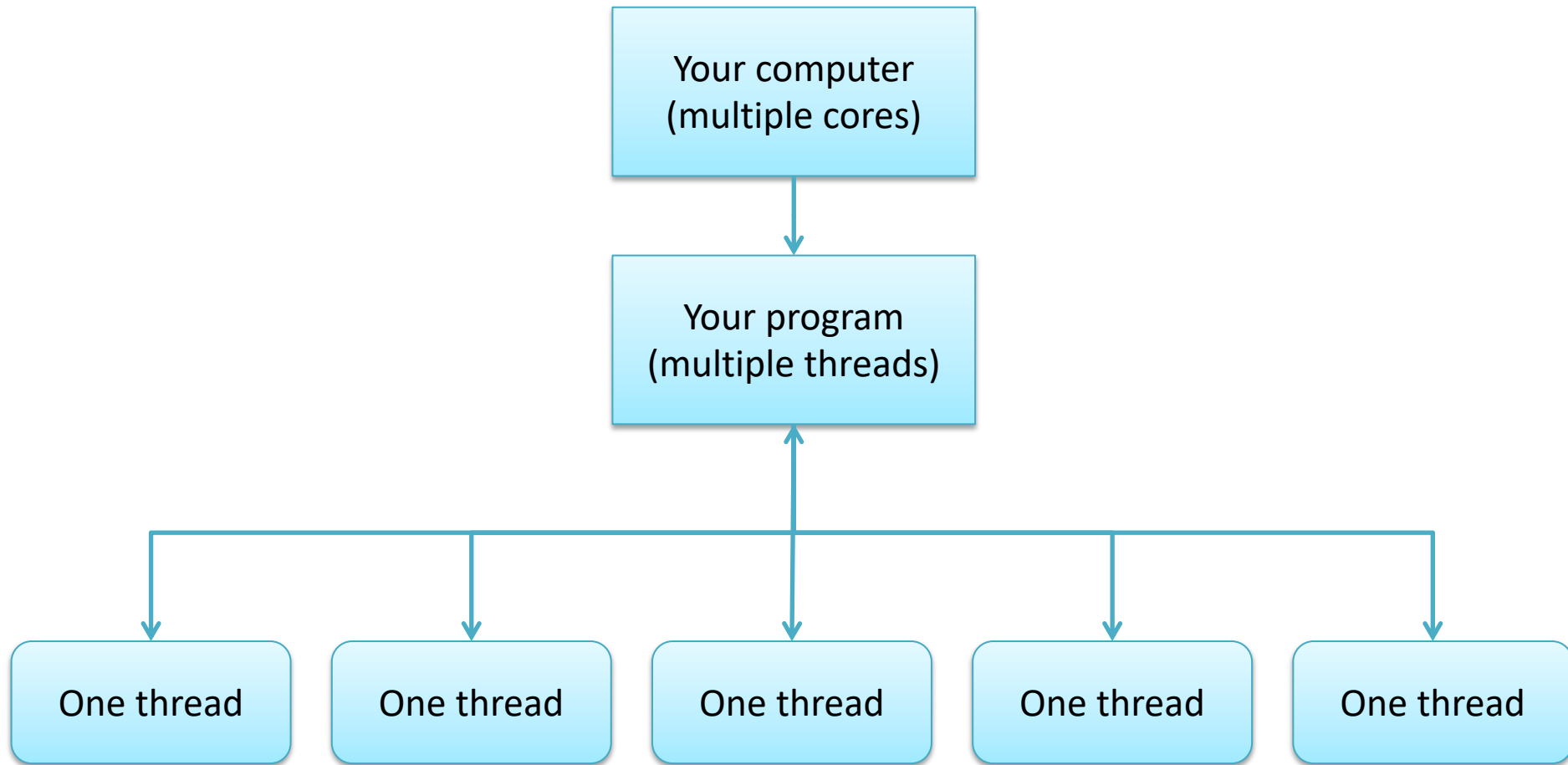
An overview of batch processing

03-Jun-2021

One-on-one



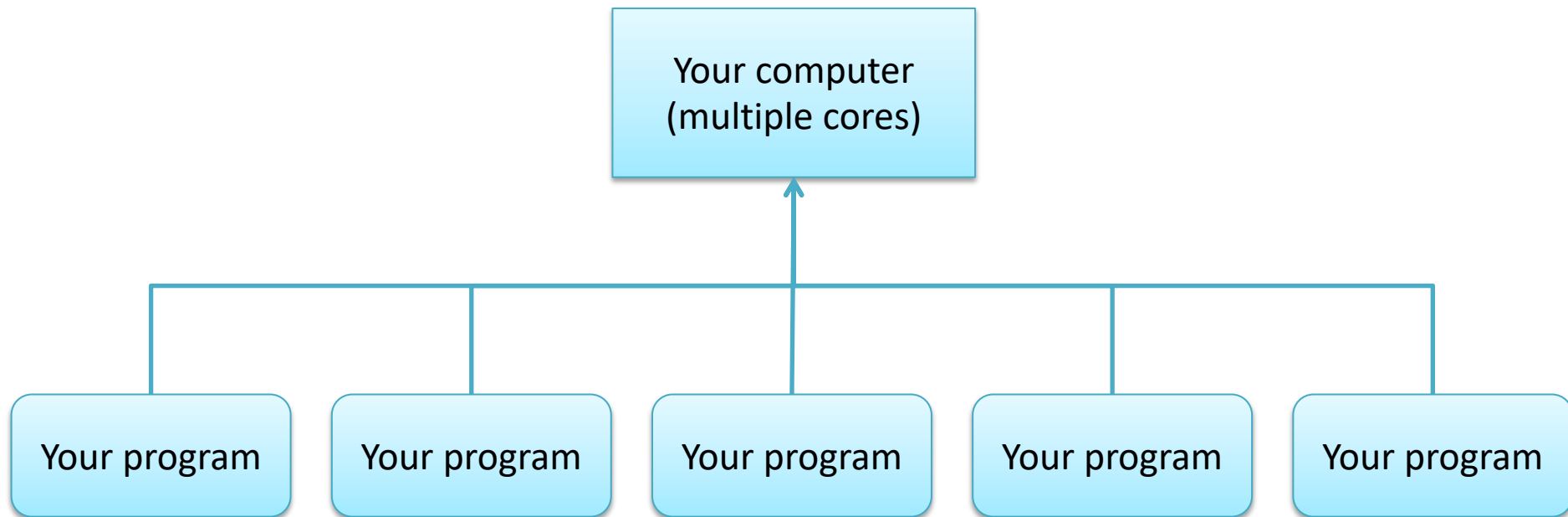
Not to be mentioned in this talk (RDataFrame, PROOF)
because they require thread-safe code



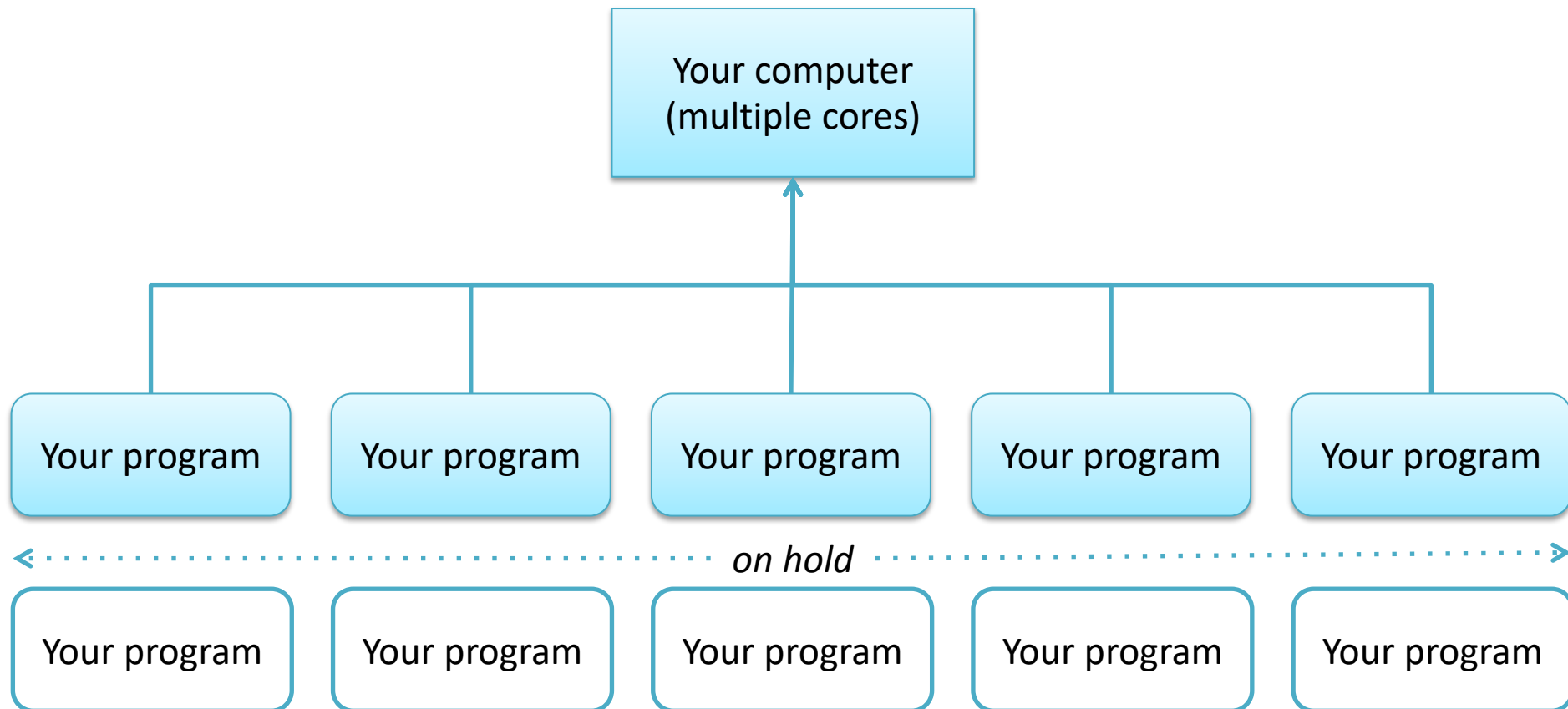
Pro tip: See Part Five of the ROOT tutorial to
explore this option, or my "fast way to analyze"
talk if I give it.

Multiple programs on a single computer (UNIX command “at”)

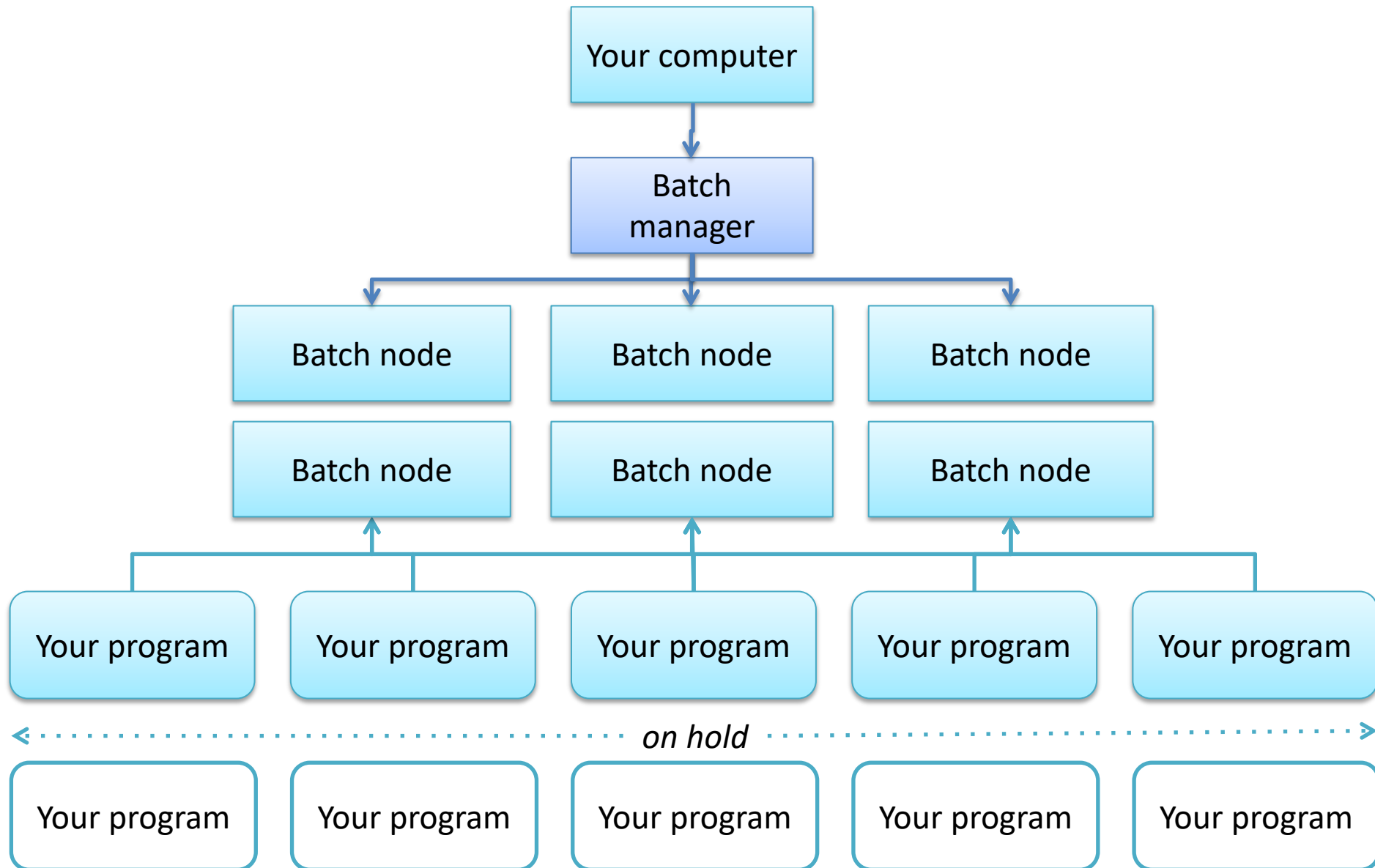
You have to remember not to submit too many jobs or you'll overload the computer



A batch system managing multiple programs on a single computer (UNIX command “batch”)



A batch system managing multiple programs on multiple computers



The standard software for managing batch systems in scientific computing is HTCondor (or just Condor)

Main web page

<http://research.cs.wisc.edu/htcondor/>

Quick start

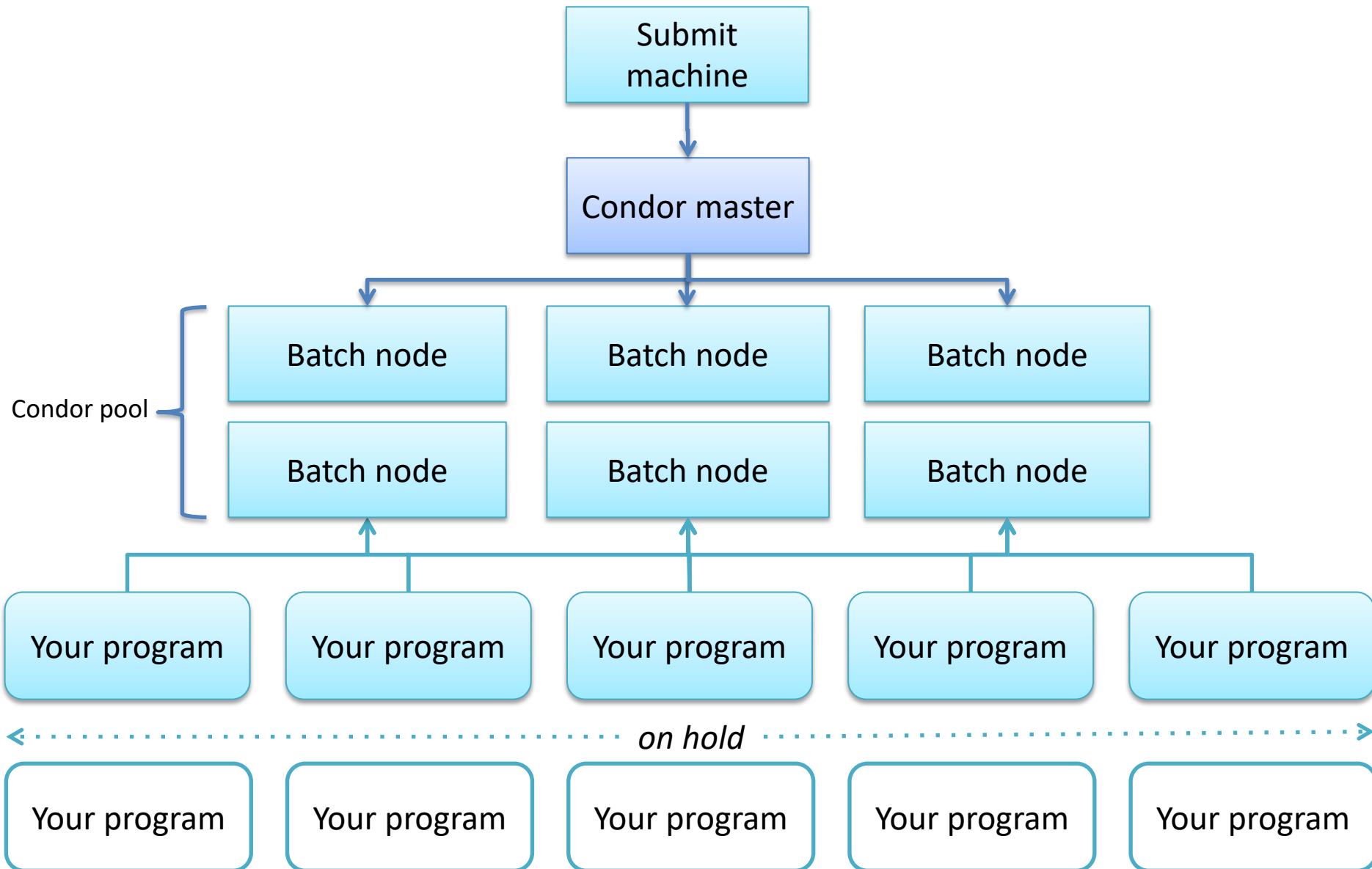
<https://htcondor.readthedocs.io/en/latest/users-manual/quick-start-guide.html>

Full manual

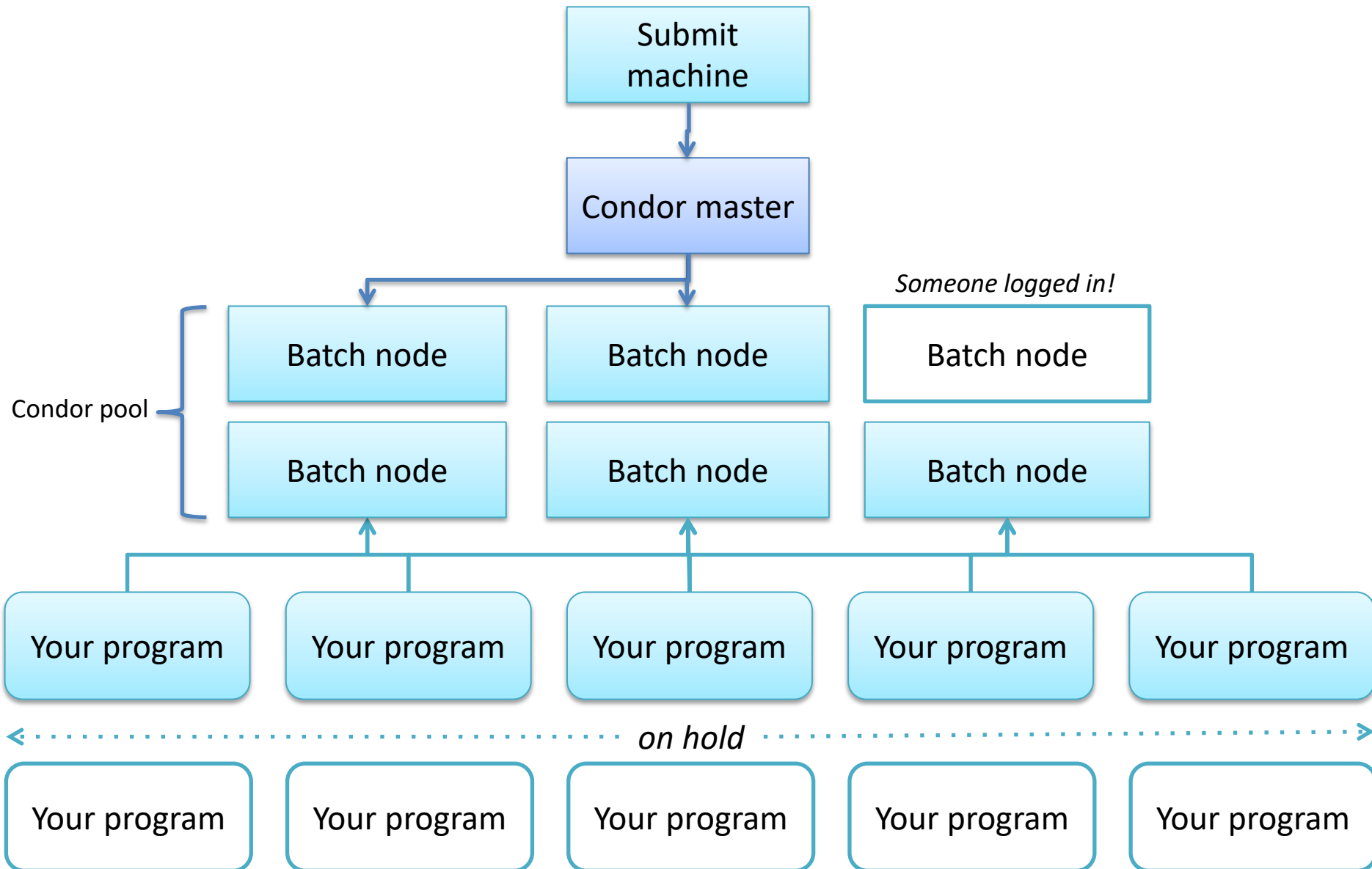
<https://htcondor.readthedocs.io/en/latest/>

- Stick to the “vanilla” universe; the “standard” universe won’t work for ROOT or any other particle-physics software (so you don’t need condor_compile).

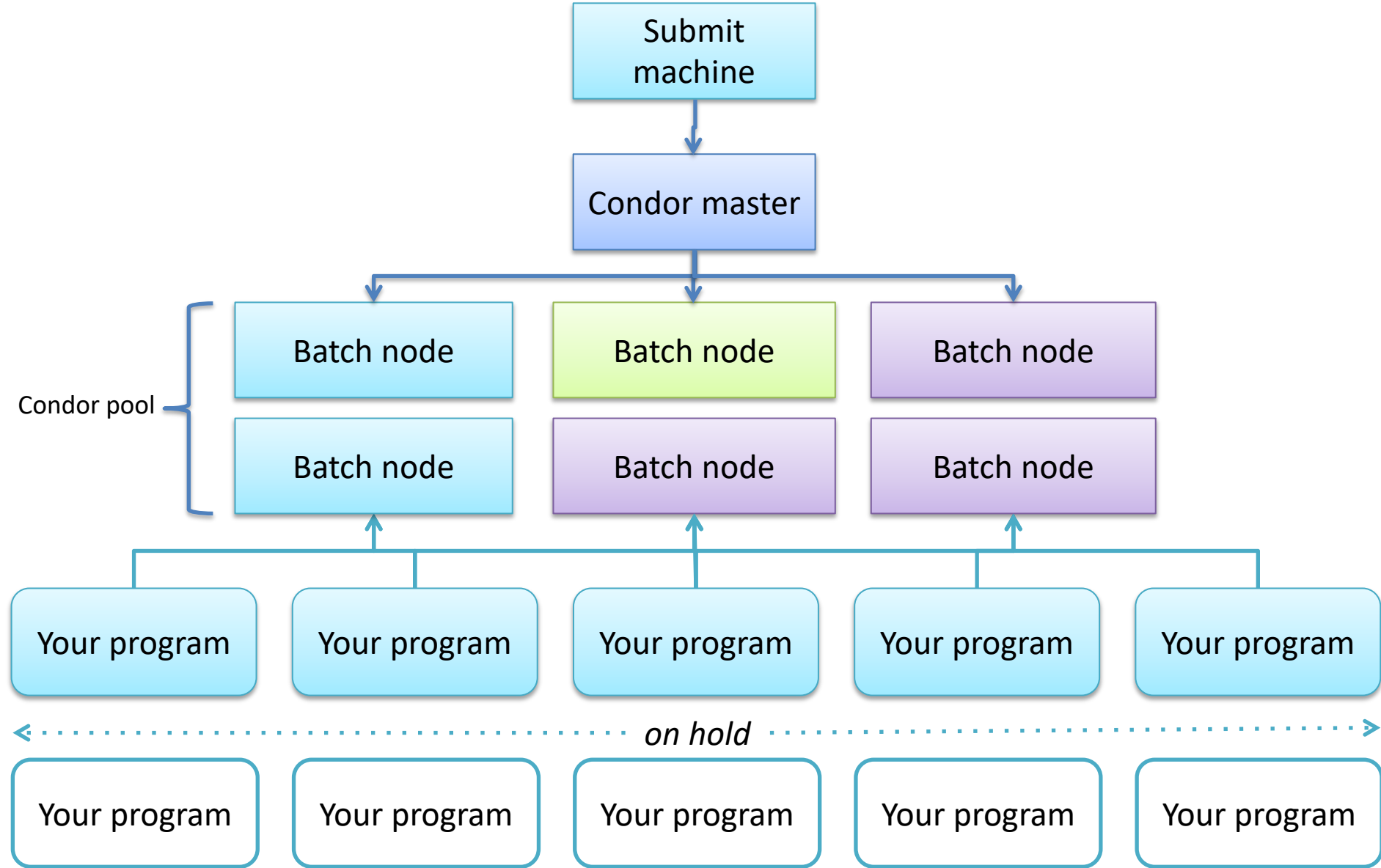
Condor managing multiple programs on multiple computers with multiple queues



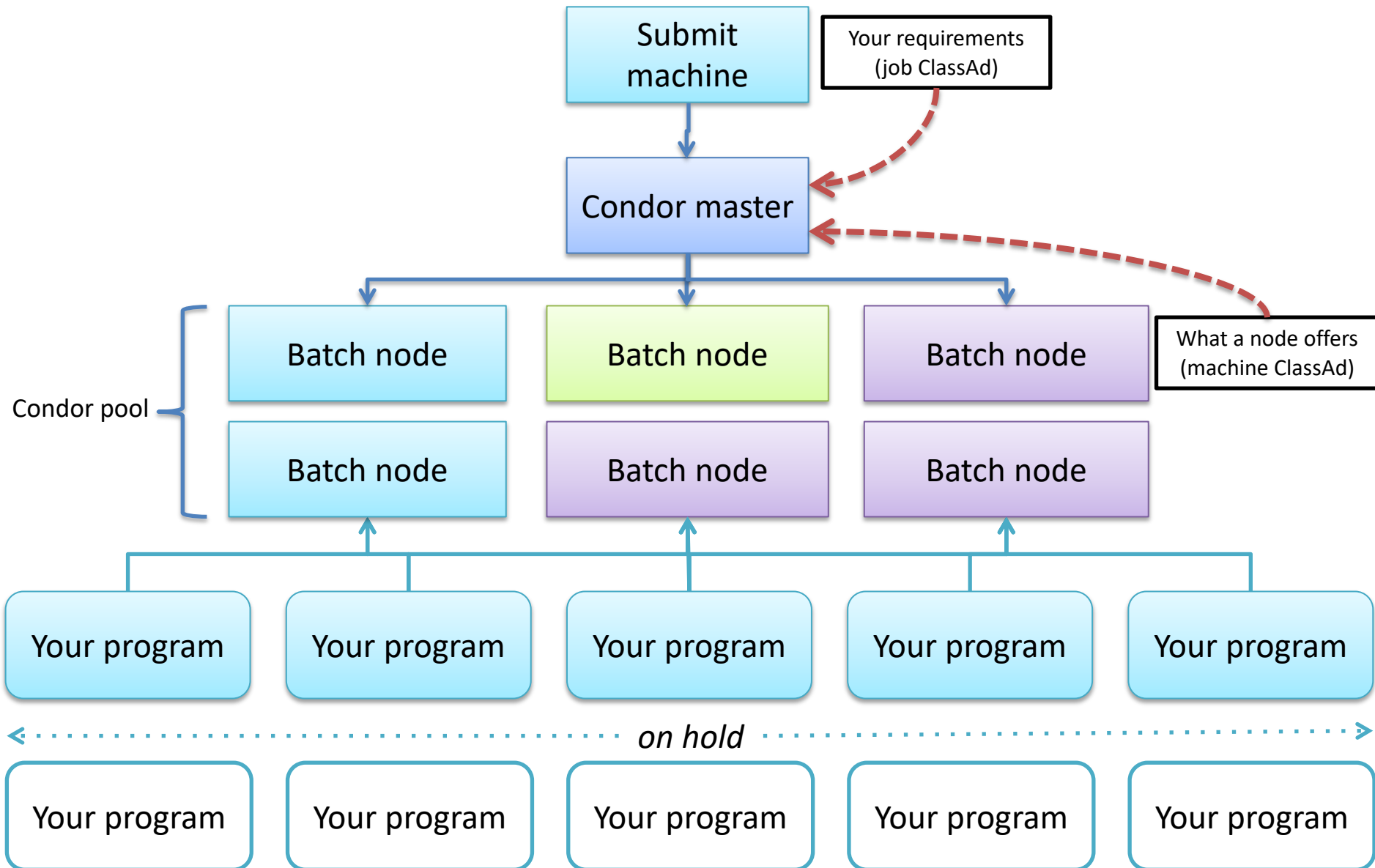
Condor will halt a queue in favor of an interactive program



Condor managing multiple programs on multiple computers with multiple configurations



Condor uses “ClassAds” to match your requirements with what each node offers



Resource Planning

- Condor can't do *everything* for you.
- Think about input files (including programs) and output files and how they'll be accessed.
- Think about disk space. “df -h” and “du -shx *” can help.
- Fun fact: The particle-physics Condor pools **can't** see your home directory!
- Moral: Let condor transfer your files... when possible.

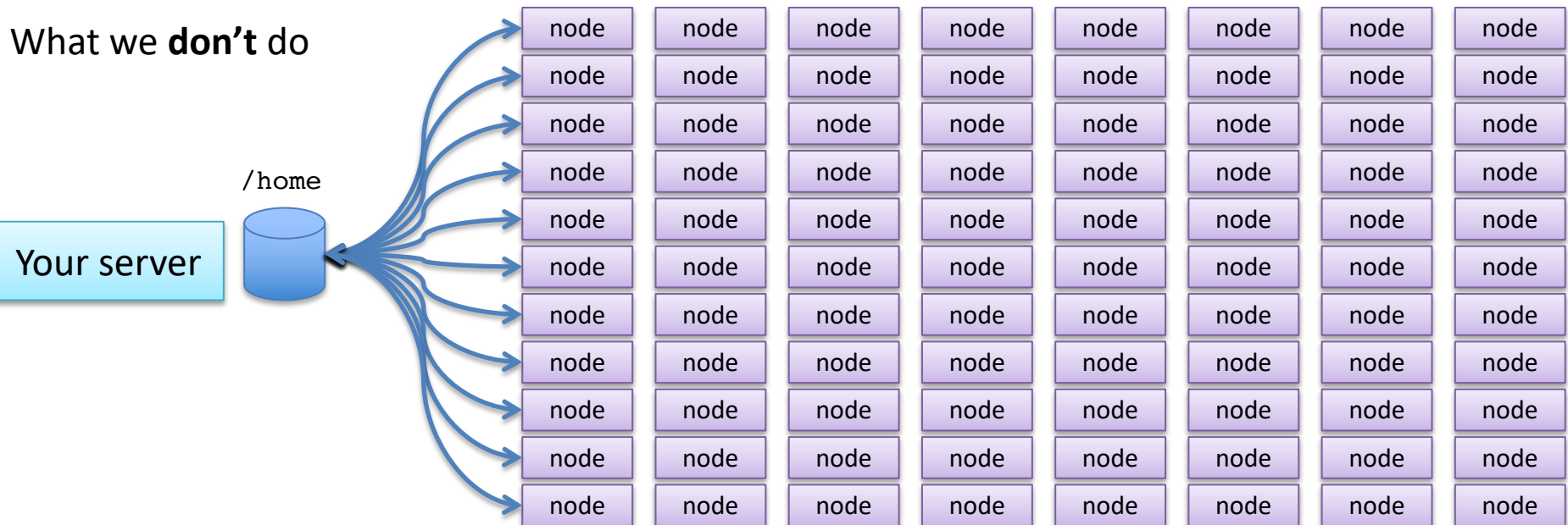
When you can't let condor transfer your files,
here are disk-sharing methods outside of condor:

- NFS – used at Nevis
- CVMFS – Fermilab and CERN
- Grid, BlueArc – only used at Fermilab

Resource Planning

- Condor can't do *everything* for you.
- Think about input files (including programs) and output files and how they'll be accessed.
- Think about disk space. “df -h” and “du -shx *” can help.
- Fun fact: The particle-physics Condor pools **can't** see your home directory!
- Moral: Let condor transfer your files... when possible.

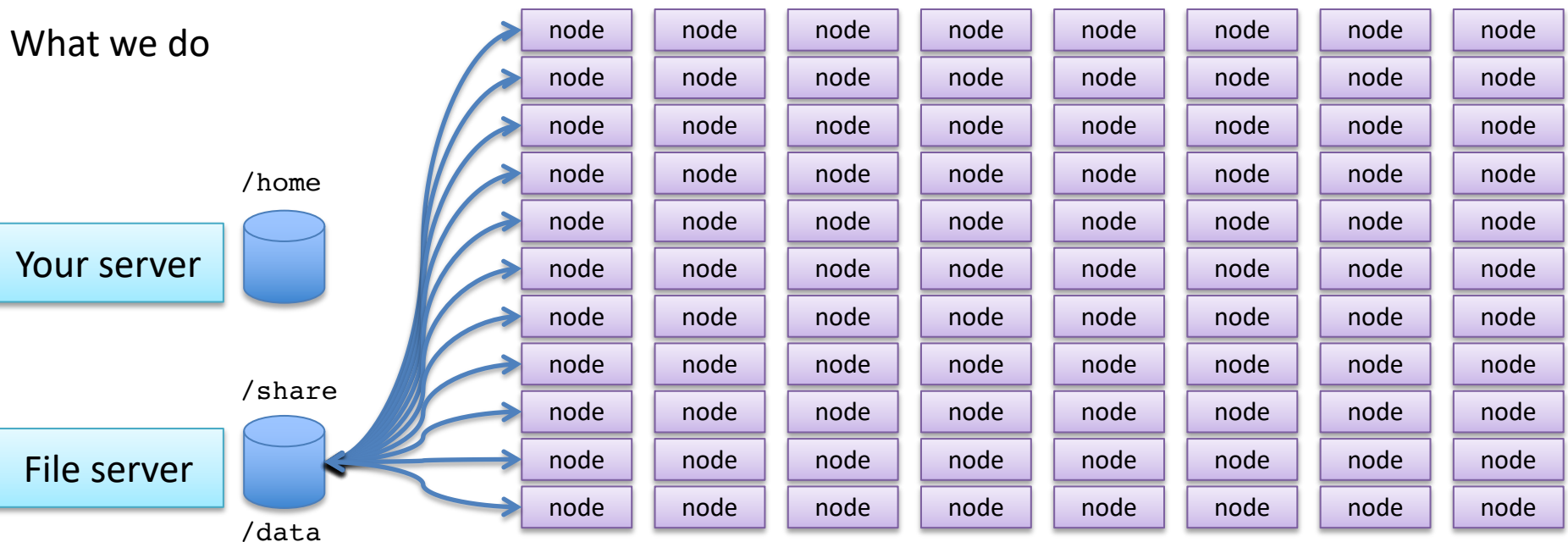
What we **don't** do



Resource Planning

- Condor can't do *everything* for you.
- Think about input files (including programs) and output files and how they'll be accessed.
- Think about disk space. "`df -h`" and "`du -shx *`" can help.
- Fun fact: The particle-physics Condor pools **can't** see your home directory!
- Moral: Let condor transfer your files... when possible.

What we do



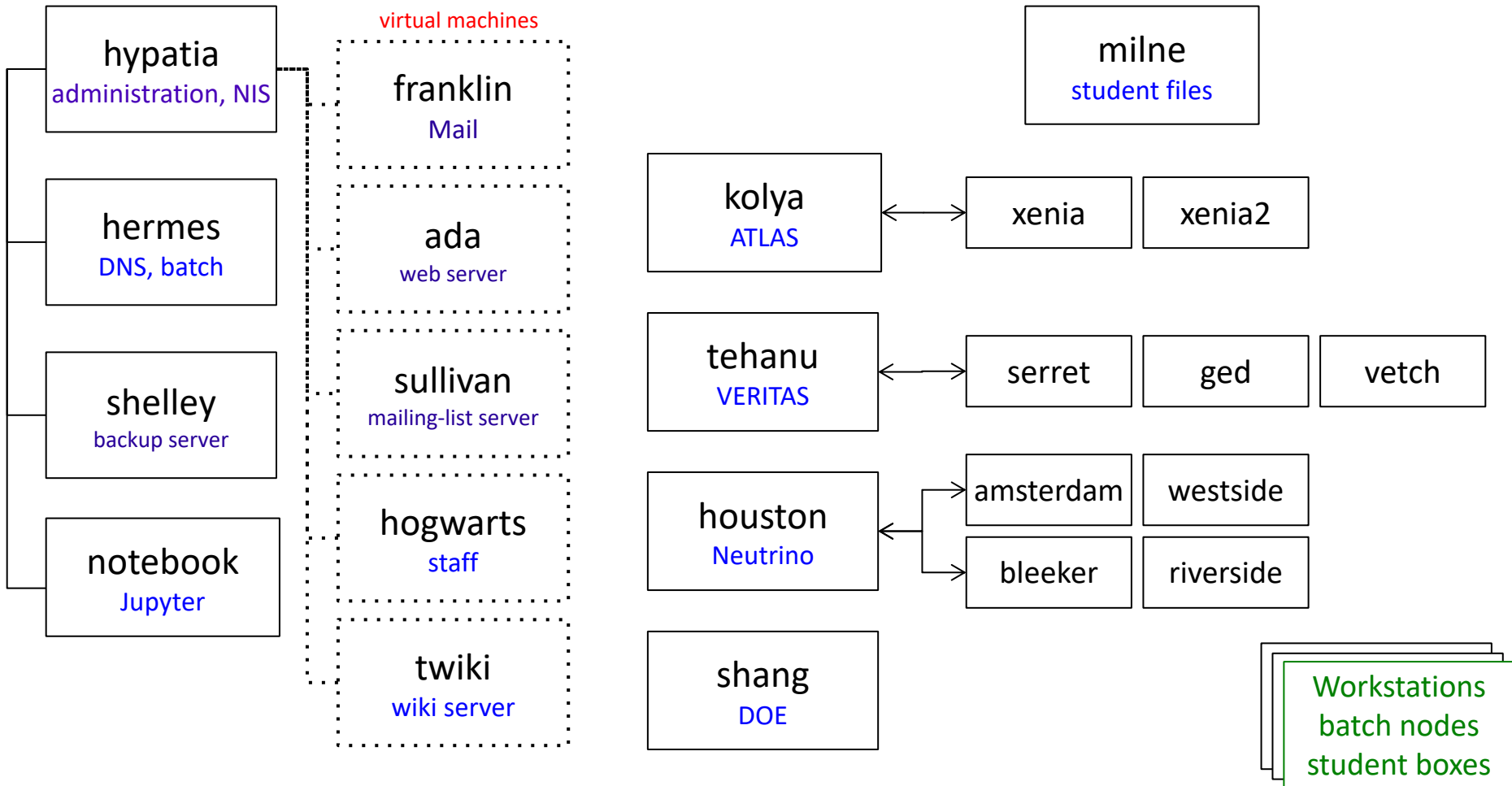
Particle-Physics Computer Systems

Linux Cluster

Administrative servers

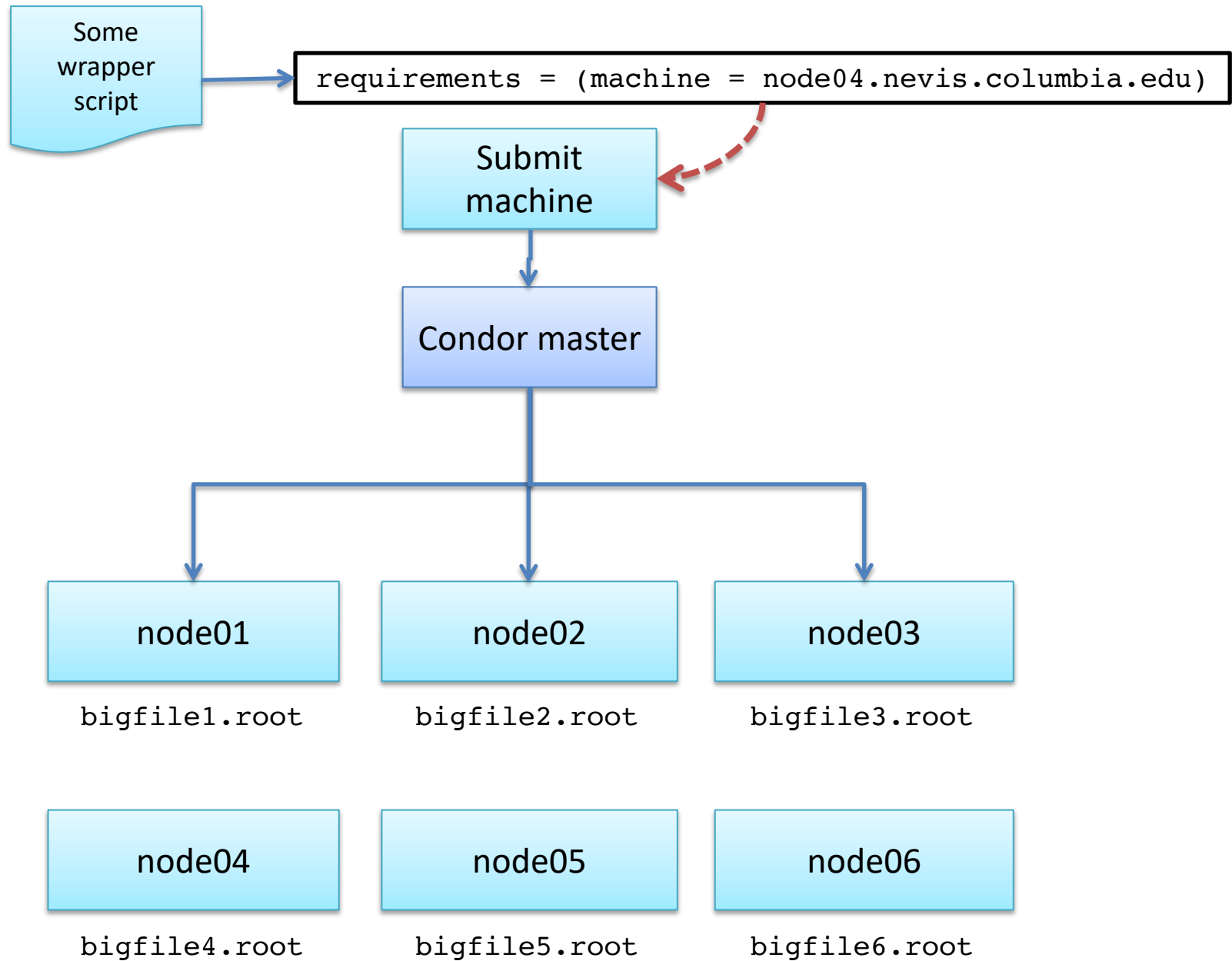
Workgroup/Login servers

File servers



<<http://www.nevis.columbia.edu/linux/>>
<<http://www.nevis.columbia.edu/linux/cluster-names.html>>

Bringing the job to the data



Final tips

- Split up your task so each condor job takes 20-60 minutes
- If your job must be preempted, it will have to run from the beginning on the same machine that cancelled the job
- Test your job with one process before submitting it for 10,000 processes!

Resources

Main web page

<http://research.cs.wisc.edu/htcondor/>

Quick start

<https://htcondor.readthedocs.io/en/latest/users-manual/quick-start-guide.html>

Full manual

<https://htcondor.readthedocs.io/en/latest/>

Nevis particle-physics condor guide

<https://twiki.nevis.columbia.edu/twiki/bin/view/Nevis/Condor>

Basic Condor@Nevis tutorial

<http://www.nevis.columbia.edu/~seligman/root-class/>