



# Intro to High Throughput Computing

Greg Thain

Center for High Throughput Computing

# Welcome

- › Today's schedule of talks

# High Throughput Defined

$$\frac{\sum \text{Job Runtimes}}{\text{Wall Time}}$$

# More Correctly

$$\frac{\sum \text{Completed Job Runtime}}{\text{Wall Time}}$$

# Even more Correctly

$$\left( \frac{\sum \text{Completed Job Runtime}}{\text{Wall Time}} \right)^*$$



Subject to some notion of fairness

# Over a long period of time

## › Tension between:

- Finding lots of machines
  - Putting the minimum conditions on them
- Finding lots of jobs
  - That can run in as many places as possible

# The Goal

!

Project	Total	CAE	CHTC	CS	OSG
Total	862,121	8,012	413,570	39,144	159,909
User1	222,121	7,558	36,714	29,984	134,485
User2	80,821	0	142,323	0	118
User3	71,943	0	8,184	1,905	914

# HPC





# HTC



# Seven Principles

- › HTCondor manages jobs
- › HTCondor manages worker machines
- › HTCondor manages data
- › HTCondor is scalable and secure
- › HTCondor runs on the networks you have
- › HTCondor supports workflows
- › HTCondor is monitorable

# HTCondor manages jobs

- › A Job is like money in the bank
- › Saved liked a database
- › Survives crashes, networking glitches
- › Has lifetime log
- › Has own policy
- › Many types of jobs



# HTCondor manages machines

- › The machine's owner is King
  - Owner's policy trumps all
  - Owner may not be keyboard user
- › When job is gone, all trace removed
- › Condor should not be able to kill machine
- › Condor "knows" what resources a machine has

# HTCondor manages data

- › No need for shared filesystem
- › Transfers sandboxes
- › Transfers are managed, queued
- › Condor knows the size of sandboxes
- › Supports 3<sup>rd</sup> party transfers

# HTCondor: scalable and secure

- › Central manager stateless and lightweight
- › State at the edges
- › Supports 200,000 concurrent jobs
- › Supports: SSL, Kerberos, GSI, NTSPi, CLAIMTOBE, Host IP, password
- › Uses the libraries you have
- › Session based security

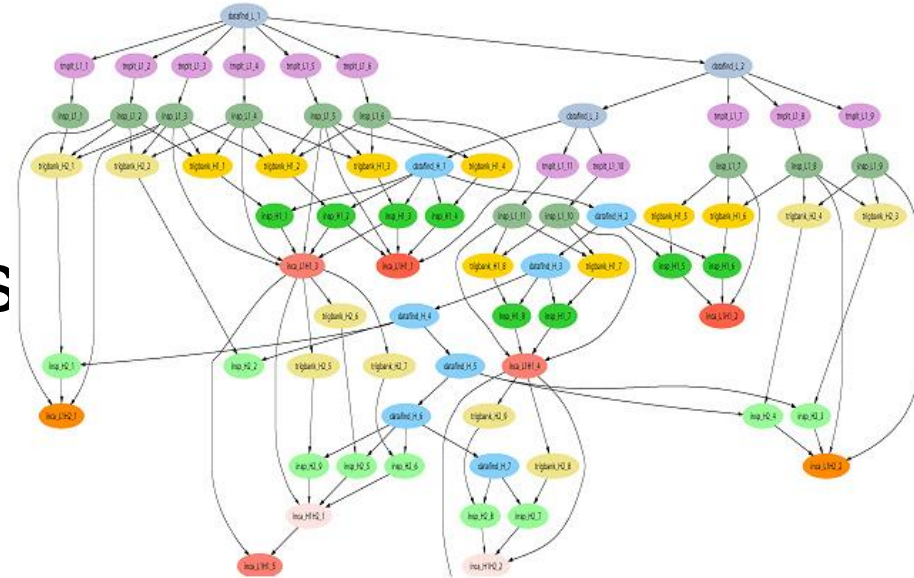


# HTCondor runs on the network you have

- › Can work around firewalls: CCB
- › Can use a single inbound port: shared\_port
- › Execute machines can be outbound only
- › Collector uses either UDP or TCP
- › Support for IPv6 only: dual stack soon
- › Can install condor w/o network people involved

# HTCondor supports workflows

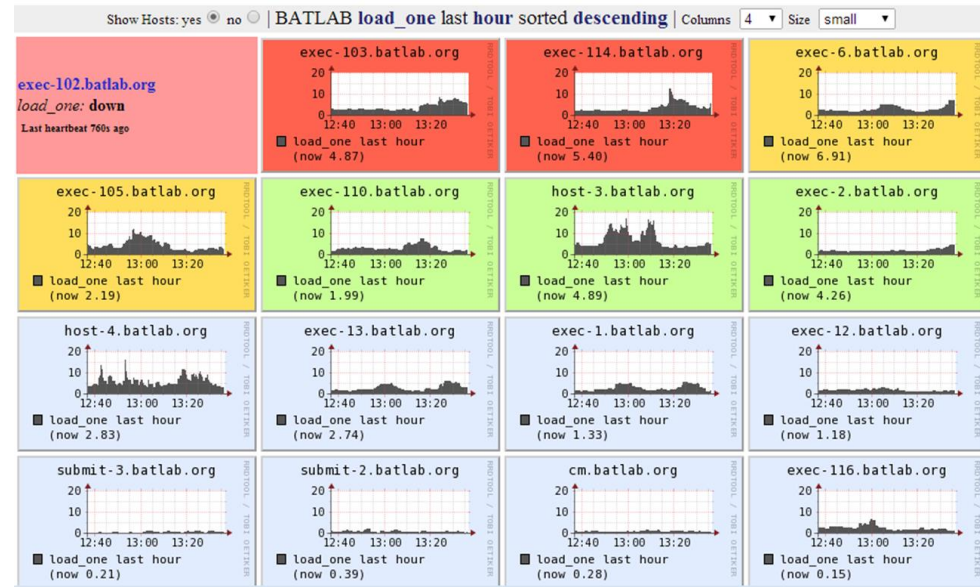
- › DAGman
- › Workflow, not just bag of tasks
- › DAGs can be huge
- › Retry, pre, post scripts



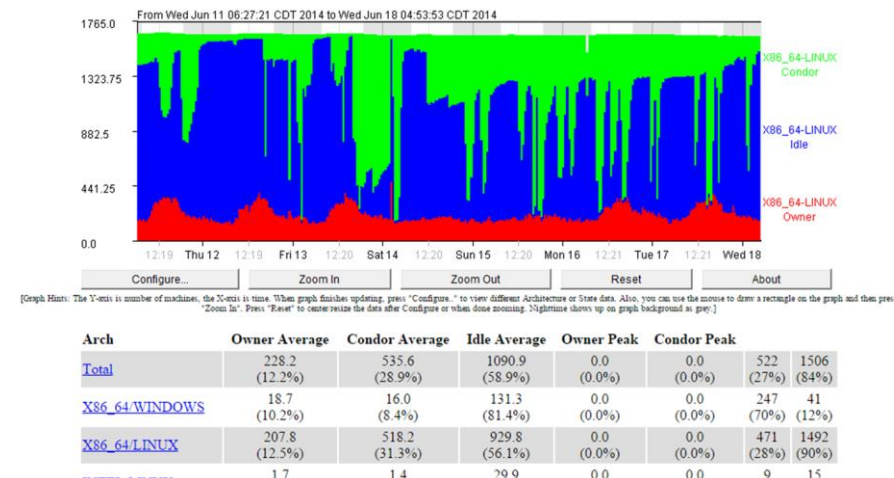


# HTCondor is monitorable

- › Statistics built in
- › Condor View
- › Ganglia
- › User logs



UW-Madison Comp Sci Condor Pool Machine Statistics for Week



# Thank you

# Now, on to the Good Stuff

- › We are here for your questions