Computing on a budget, a system's engineers prospective ...

Christopher Harrison
HTCondor Week 2016
Systems Engineer
Biostatistics and Medical Informatics
School of Medicine and Public Health
UW - Madison

Get to know your users (Think Journalism)

- What problems are my users solving?
- Who is doing the actual development?
    - Closely collaborate with the development people
- When are the deadlines for projects?
- Where are the projects going? (not always known)
- How are the users expecting to solve their problems?
    - Can early intervention lead to better outcomes?

GOAL: Understand and Profile computational jobs of the users to understand requirements

Does this mean I need to talk to my users?

YES

UW - Biostatistics computational challenges

- Users are working on protected data sets (HIPAA)
- Genetic algorithms (Memory and I/O intensive jobs)
- CPU is not the bottleneck
- Data storage, I/O and memory are bottleneck
- Address data storage:
  - FREEBSD
  - ZFS
  - NFS
  - 500TB

Think outside the box

Do we really need new?

Do we need systems with warranties?

Pets
Dotty, my dog

VS

Cattle
rawhide-0(1-1023)

Chairman Mao, "People die every day"

Chris Harrison, "Computers die every day"

Venture into used equipment (ebay)

Purchase #1

42 x Dell C1100

72GB DDR3

Dual quad nehalem w/ HT

4 3.5" drive bays

Price per: $460

Purchase #2

26 x CISCO UCS C250 M2

192GB DDR3

Dual hex core Westmere

10x 2.5" drive bays

Price per: $1799

Keeping track at home

Pro

We added:

~1300 CPU cores (w/ HT), ~8 TB of ddr3 mem

Treat systems as cattle, discard if dead

Tripled Condor pool at yearly replacement cost

Cost per unit ~8x less than new

Con

Harder to procure, simplified bid process

No replacements on failures

Used equipment, more sysadmin time req.

Useful life of individual machines is less the new

Replace equipment more frequently

Questions?

• • •

Christopher Harrison
harrison@biostat.wisc.edu