



Freedom vs. Stability: Facilitating Research Training while Supporting Scientific Research

Ben Tovar and Douglas Thain



UNIVERSITY OF
NOTRE DAME

Who we are

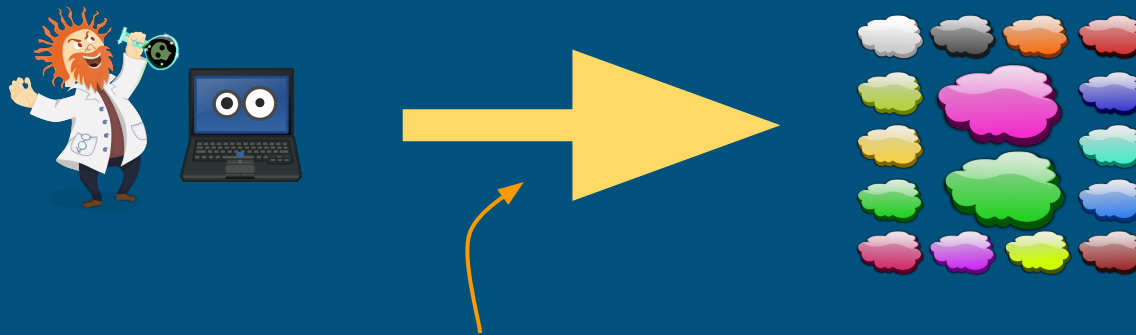


Scientist says:

"This example runs on my laptop, but I need much more for the real application. It would be great if we can run $O(10K)$ tasks like this on this cloud/grid/cluster I have heard so much about."

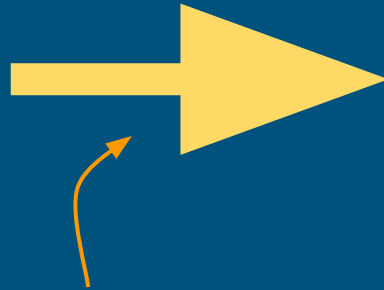


Who we are



The Cooperative Computing Lab
Computer Science and Engineering
University of Notre Dame

Who we are



The Cooperative Computing Lab
Computer Science and Engineering
University of Notre Dame

Who we are

- We *collaborate with people* who have large scale computing problems in science, engineering, and other fields.
- We *operate computer systems* on the order of 20k cores: clusters, clouds, grids.
- We *conduct computer science research* in the context of real people and problems.
- We *develop open source software* for large scale distributed computing.

CCTools

- Open source, GNU General Public License.
- Compiles in 1-2 minutes, installs in \$HOME.
- Runs on Linux, Solaris, MacOS, Cygwin, FreeBSD, ...
- Interoperates with many distributed computing systems.
 - Condor, SGE, Torque, Globus, iRODS, Hadoop...
- Components:
 - Makeflow – A portable workflow manager.
 - Work Queue – A lightweight distributed execution system.
 - All-Pairs / Wavefront / SAND – Specialized execution engines.
 - Parrot – A personal user-level virtual file system.
 - Chirp – A user-level distributed filesystem.

Personnel: Situation 2 years ago

1 faculty

5-6 P.h.D students

1-2 undergrad students

Challenge

"It is a miracle education survives formal education"

--

A. Einstein



vs.

"I'm generally a very pragmatic person: that which works, works."

--

L. Torvalds



Challenges

Flexibility / Training

Computer science research training.

Nurture novel ideas.

Train in programming and software engineering.

New students know script programming, but no system programming (e.g., C and kernel).

VS.

Stability / Quality

We failed if nobody uses our software.

Critical bugs trump everything.

Garbage collection anyone?

Stable APIs and protocols.

We cannot break the trust of our users.

Personnel: Planned situation

1 faculty

5-6 P.h.D students

1-2 undergrad students

+ 1 software engineer

Challenges

Flexibility / Training

Computer science research training.

Nurture novel ideas.

Train in programming and software engineering.

New students know script programming, but no system programming (e.g., C and kernel).

VS.

use the software engineer here



Stability / Quality

We failed if nobody uses our software.

Critical bugs trump everything.

Garbage collection anyone?

Stable APIs and protocols.

We cannot break the trust of our users.

Personnel: Current situation

1 faculty

5-6 P.h.D students

1-2 undergrad students

+ 1 software/research engineer

My Life as a Spring

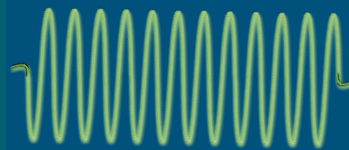
Flexibility / Training

Per year, students serve as 'first responder', and 'outreach coordinator'.

Immediately, new students are assigned low priority and easy bugs.

Main point of contact for new students.

One shared crazy master branch.



½ of my job

Stability / Quality

First responder to 'segfault fires'.

New features only if of immediate use to our customers.

Establish policies for clean and safe code.

Only commiter to the stable release branch.

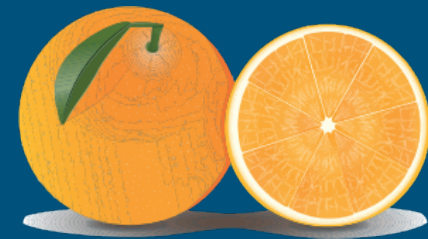
My other half

Research/Academia

Salary comes from project grants.

Develop novel research and implement software solutions as a research scientist.

Once in a while serve as a lecturer.



Contact

Ben Tovar
btovar@nd.edu