

Welcome to Notre Dame!

CCL Workshop 2013

www.nd.edu/~ccl/workshop/2013

The Cooperative Computing Lab

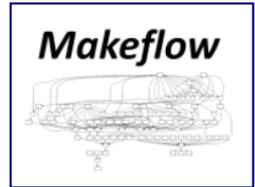
University of Notre Dame



<http://www.nd.edu/~ccl>

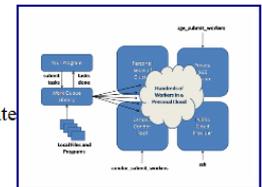
Makeflow

Makeflow is a workflow system for parallel and distributed computing that uses a language very similar to Make. Using Makeflow, you can write simple scripts that easily execute on hundreds or thousands of machines.



Work Queue

Work Queue is a system and library for creating and managing scalable master-worker style programs that scale up to thousands machines on clusters, clouds, and grids. Work Queue programs are easy to write in C, Python or Perl.



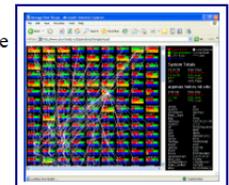
Parrot

Parrot is a transparent user-level virtual filesystem that allows any ordinary program to be attached to many different remote storage systems, including HDFS, iRODS, Chirp, and FTP.



Chirp

Chirp is a personal user-level distributed filesystem that allows unprivileged users to share space securely, efficiently, and conveniently. When combined with Parrot, Chirp allows users to create custom wide-area distributed filesystems.



Our goal:

Make it easy to scale up
real applications
from a single laptop
to national infrastructure,
enabling new discoveries.

The Cooperative Computing Lab

- We *collaborate with people* who have large scale computing problems in science, engineering, and other fields.
- We *operate computer systems* on the O(10,000) cores: clusters, clouds, grids.
- We *conduct computer science* research in the context of real people and problems.
- We *release open source software* for large scale distributed computing.

<http://www.nd.edu/~ccl>

CCL Team Changes



Ben Tovar, Ph.D.
Principal Research
Software Engineer



Li Yu
Financial Software
Bloomberg



Principal
Hadoop Developer
AT&T Foundry



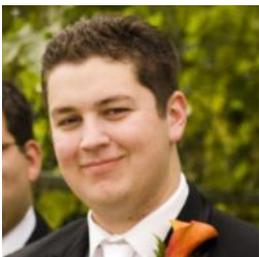
Dinesh Rajan



Patrick Donnelly



Peter Sempolinski



Casey Robinson



Peter Ivie



Haiyan Meng



Nick Hazekamp

Current Focus Areas

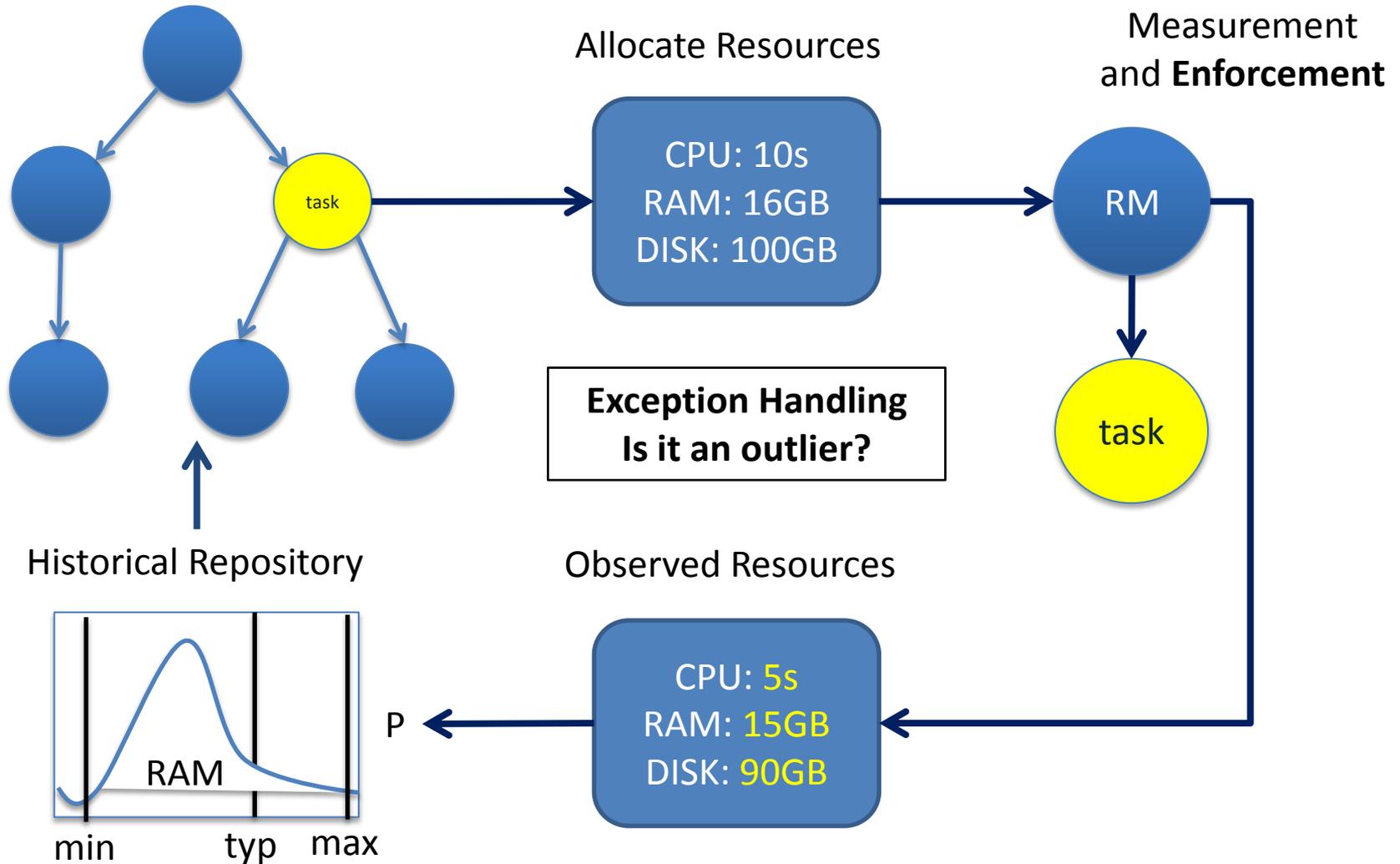
- Right Sizing Systems and Applications
 - I have this workload. How many machines (and what size) do I actually need to run it?
- Scaling Up without Blowing Up
 - 10M tasks on 10K workers?
- Data Intensive Applications
 - Move code to data and share across workers, across applications, across users....
- Evolving with Our Ecosystem:
 - Hadoop, Galaxy, GPUs, iRODS, OSG, XSEDE

dV/dt : Accelerating the Rate of Progress Towards Extreme Scale Collaborative Science

**Miron Livny (UW), Ewa Deelman (USC/ISI), Douglas Thain (ND),
Frank Wuerthwein (UCSD), Bill Allcock (ANL)**

... make it easier for scientists to conduct large-scale computational tasks that use the power of computing resources they do not own to process data they did not collect with applications they did not develop ...

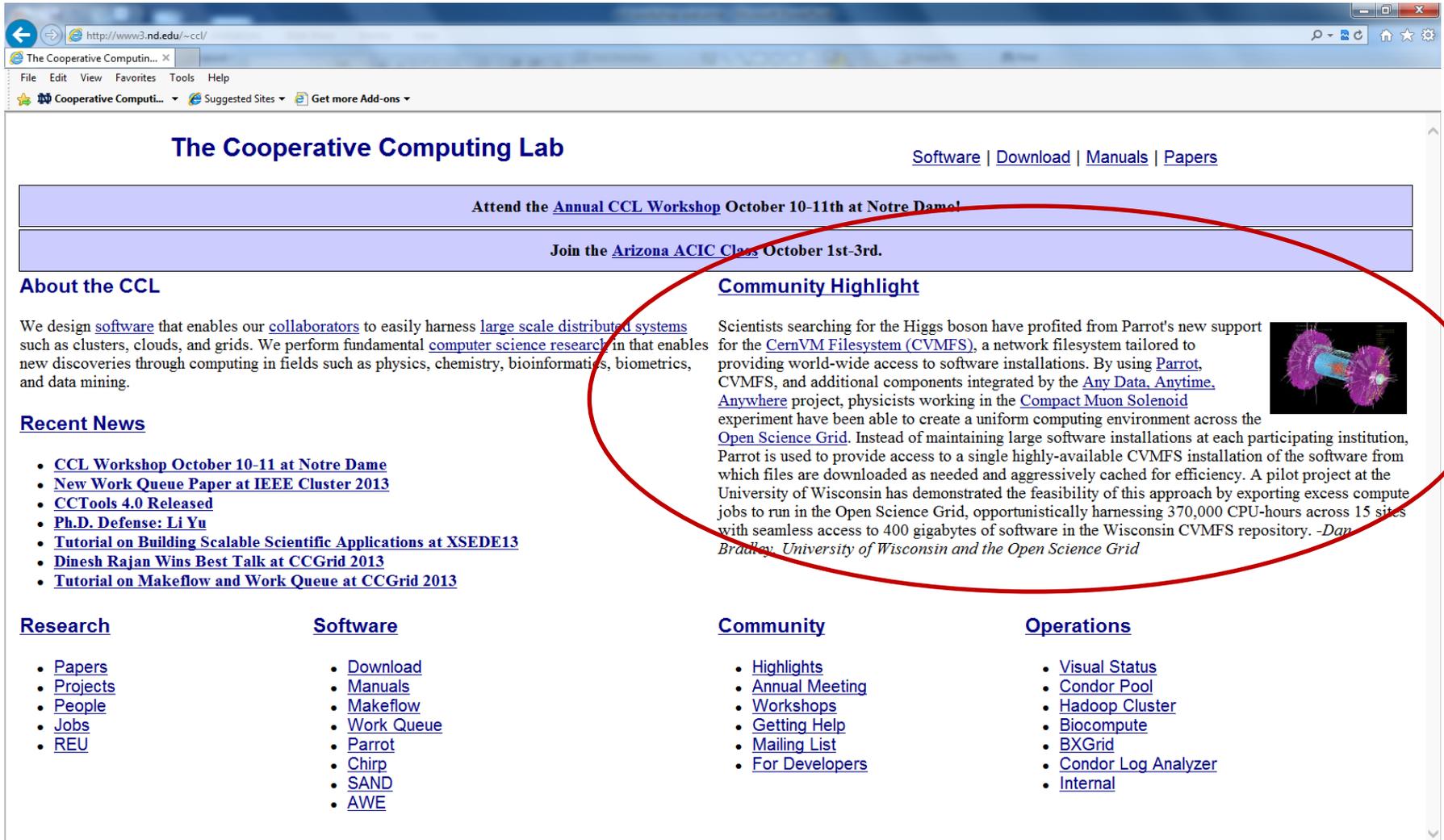
Resource Management Cycle



CCTools in Education

- UW – Eau Claire: IEEE Cluster 2013 Best Education Outreach and Training Paper:
 - Peter Bui, Travis Boettcher, Nicholas Jaeger, Jeffrey Westphal, “Using Clusters in Undergraduate Research”
- MF and WQ in University of Arizona Advanced Cyberinfrastructure Class (ACIC)
- WQ in Notre Dame Undergraduate Programming Paradigms Class

Your Highlight Here!



The screenshot shows a web browser window displaying the Cooperative Computing Lab website. A red oval highlights the 'Community Highlight' section, which contains text about the Higgs boson search and a small image of a particle detector. The website header includes navigation links for Software, Download, Manuals, and Papers. Below the header are two promotional banners for workshops and classes. The main content area is divided into sections for About the CCL, Recent News, Research, Software, Community, and Operations.

The Cooperative Computing Lab [Software](#) | [Download](#) | [Manuals](#) | [Papers](#)

Attend the [Annual CCL Workshop](#) October 10-11th at Notre Dame!

Join the [Arizona ACIC Class](#) October 1st-3rd.

About the CCL

We design [software](#) that enables our [collaborators](#) to easily harness [large scale distributed systems](#) such as clusters, clouds, and grids. We perform fundamental [computer science research](#) in that enables new discoveries through computing in fields such as physics, chemistry, bioinformatics, biometrics, and data mining.

Recent News

- [CCL Workshop October 10-11 at Notre Dame](#)
- [New Work Queue Paper at IEEE Cluster 2013](#)
- [CCTools 4.0 Released](#)
- [Ph.D. Defense: Li Yu](#)
- [Tutorial on Building Scalable Scientific Applications at XSEDE13](#)
- [Dinesh Rajan Wins Best Talk at CCGrid 2013](#)
- [Tutorial on Makeflow and Work Queue at CCGrid 2013](#)

Research

- [Papers](#)
- [Projects](#)
- [People](#)
- [Jobs](#)
- [REU](#)

Software

- [Download](#)
- [Manuals](#)
- [Makeflow](#)
- [Work Queue](#)
- [Parrot](#)
- [Chirp](#)
- [SAND](#)
- [AWE](#)

Community

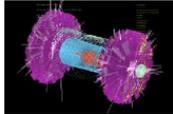
- [Highlights](#)
- [Annual Meeting](#)
- [Workshops](#)
- [Getting Help](#)
- [Mailing List](#)
- [For Developers](#)

Operations

- [Visual Status](#)
- [Condor Pool](#)
- [Hadoop Cluster](#)
- [Biocompute](#)
- [BXGrid](#)
- [Condor Log Analyzer](#)
- [Internal](#)

Community Highlight

Scientists searching for the Higgs boson have profited from Parrot's new support for the [CernVM Filesystem \(CVMFS\)](#), a network filesystem tailored to providing world-wide access to software installations. By using [Parrot](#), CVMFS, and additional components integrated by the [Any Data, Anytime, Anywhere](#) project, physicists working in the [Compact Muon Solenoid](#) experiment have been able to create a uniform computing environment across the [Open Science Grid](#). Instead of maintaining large software installations at each participating institution, Parrot is used to provide access to a single highly-available CVMFS installation of the software from which files are downloaded as needed and aggressively cached for efficiency. A pilot project at the University of Wisconsin has demonstrated the feasibility of this approach by exporting excess compute jobs to run in the Open Science Grid, opportunistically harnessing 370,000 CPU-hours across 15 sites with seamless access to 400 gigabytes of software in the Wisconsin CVMFS repository. -*Dan Bradley, University of Wisconsin and the Open Science Grid*



Agenda

- **9:00 Introduction**
 - Welcome and Introduction
 - What's new in the CCTools – CCL Team
- **10:15 Coffee Break**
 - Stash and Skeleton Key on the Open Science Grid, Rob Gardner, University of Chicago
 - Swift+Chirp for Synchrotron Beamline Data Analysis, Justin Wozniak, Argonne National Lab
 - Big Data and Data Science: The View from NIST, Geoffrey Fox, Indiana University
- **Noon - Lunch**
 - Scaling Up CMS Tier-3 Data Processing, Kevin Lannon
 - HotRAD: Harnessing distributed systems for population genetics sampling Scott Emrich
 - Data Management Challenges in the iPlant Collaboration, Nirav Merchant
 - Massively Parallel Molecular Dynamics Using Adaptive Weighted Ensemble, RJ Nowling
- **2:30 - Coffee Break**
 - Weather Data Processing with Makeflow and SLURM, Neil Best
 - Using Work Queue Inside and Outside the Classroom, Peter Bui
 - Systematic development of accurate model potentials using the Work Queue distributed computing environment, Lee-Ping Wang
 - Large Scale Image Processing for Biometrics, Joseph Thompson
- **4:00 Discussion and Adjourn**