

Using Parrot to access CVMFS repositories

Ben Tovar
University of Notre Dame

btovar@nd.edu



Who we are

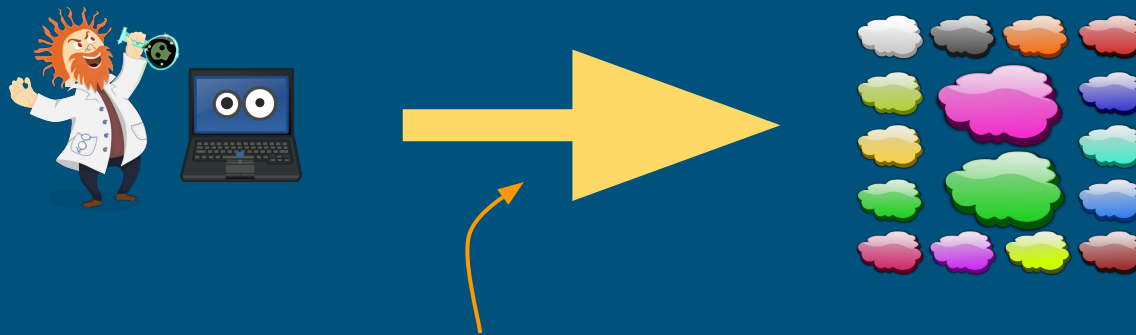


Scientist says:

"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."

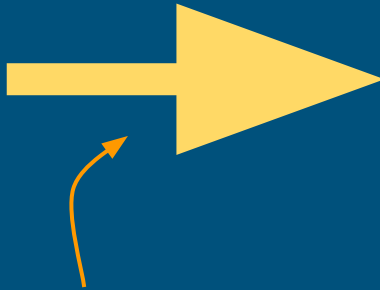


Who we are



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

Cooperative Computing Lab

Director



[Douglas Thain](#)

Staff



[Benjamin Tovar](#)
Research Software Engr



Graduate Students



[Patrick Donnelly](#)



[Peter Ivie](#)



[Haiyan Meng](#)
First Responder



[Nicholas Hazekamp](#)
Outreach Coordinator



[Nathaniel Kremer-Herman](#)

Not shown, grad students: Tim Shaffer , Chao Zheng

CCL Objectives

- Harness all the resources that are available: desktops, clusters, clouds, and grids.
- Make it easy to scale up from one desktop to national scale infrastructure.
- Provide familiar interfaces that make it easy to connect existing apps together.
- Allow portability across operating systems, storage systems, middleware...
- Make simple things easy, and complex things possible.
- **No special privileges required.**

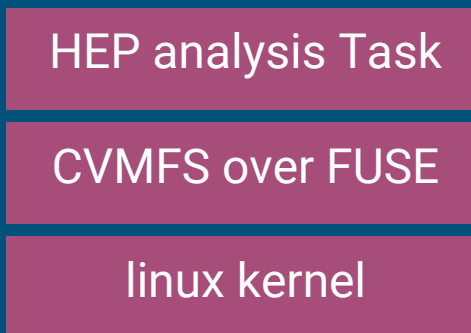
CCTools



- 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.

CVMFS for Deploying HEP Software Stack

Analysis software is distributed via CVMFS, a read-only filesystem over HTTP.



Get file from cache, or CVMFS repository.

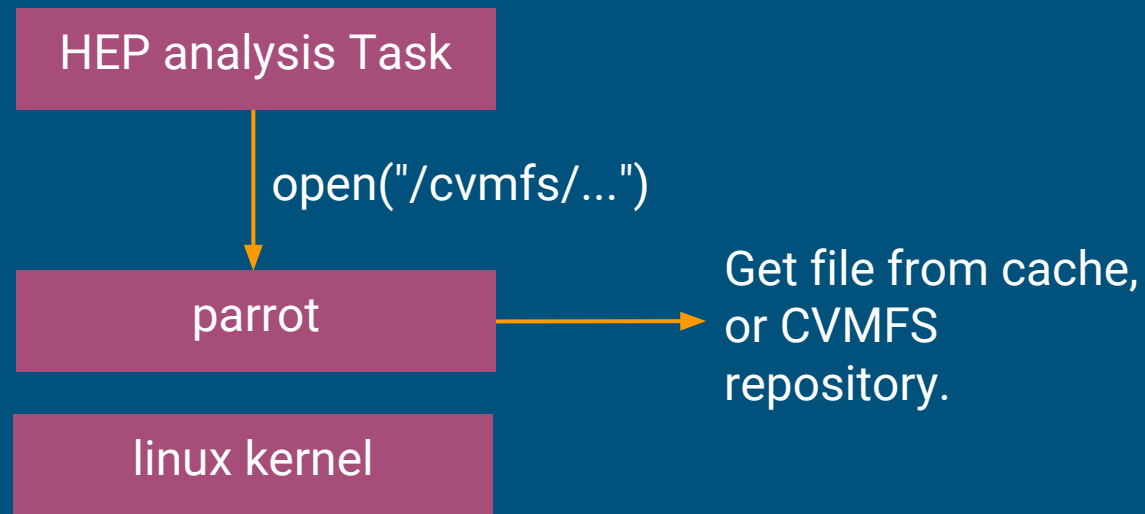
With FUSE, the remote software is local as far as the task is concerned.

Parrot and CVMFS: Main Idea

Run CVMFS based applications without setting up the nodes where they run.

```
lancre ~ > ls -lad /cvmfs/cms.cern.ch/slc6_amd64_gcc530
ls: cannot access '/cvmfs/cms.cern.ch/slc6_amd64_gcc530': No such file or
directory
lancre ~ >
lancre ~ > parrot_run ls -lad /cvmfs/cms.cern.ch/slc6_amd64_gcc530
drwxr-xr-x 1 root root 3 Mar  4 09:30 /cvmfs/cms.cern.ch/slc6_amd64_gcc530
lancre ~ >
lancre ~ > parrot_run bash
magrat@lancre:~$ ls -lad /cvmfs/cms.cern.ch/slc6_amd64_gcc530
drwxr-xr-x 1 root root 3 Mar  4 09:30 /cvmfs/cms.cern.ch/slc6_amd64_gcc530
magrat@lancre:~$ cp /cvmfs/cms.cern.ch/README .
magrat@lancre:~$
```

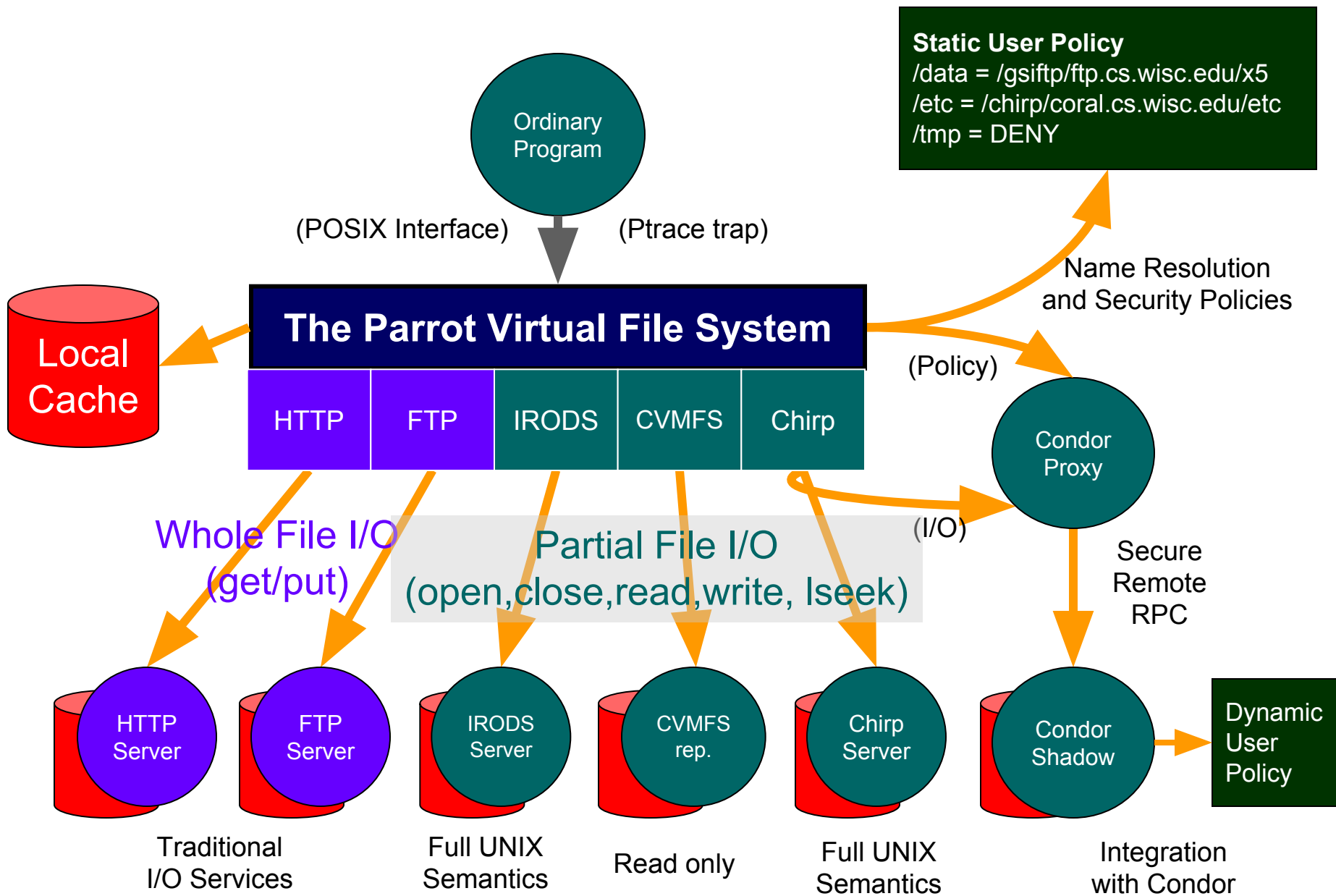
How



Parrot is a tool for attaching existing programs to remote I/O systems through the filesystem interface.

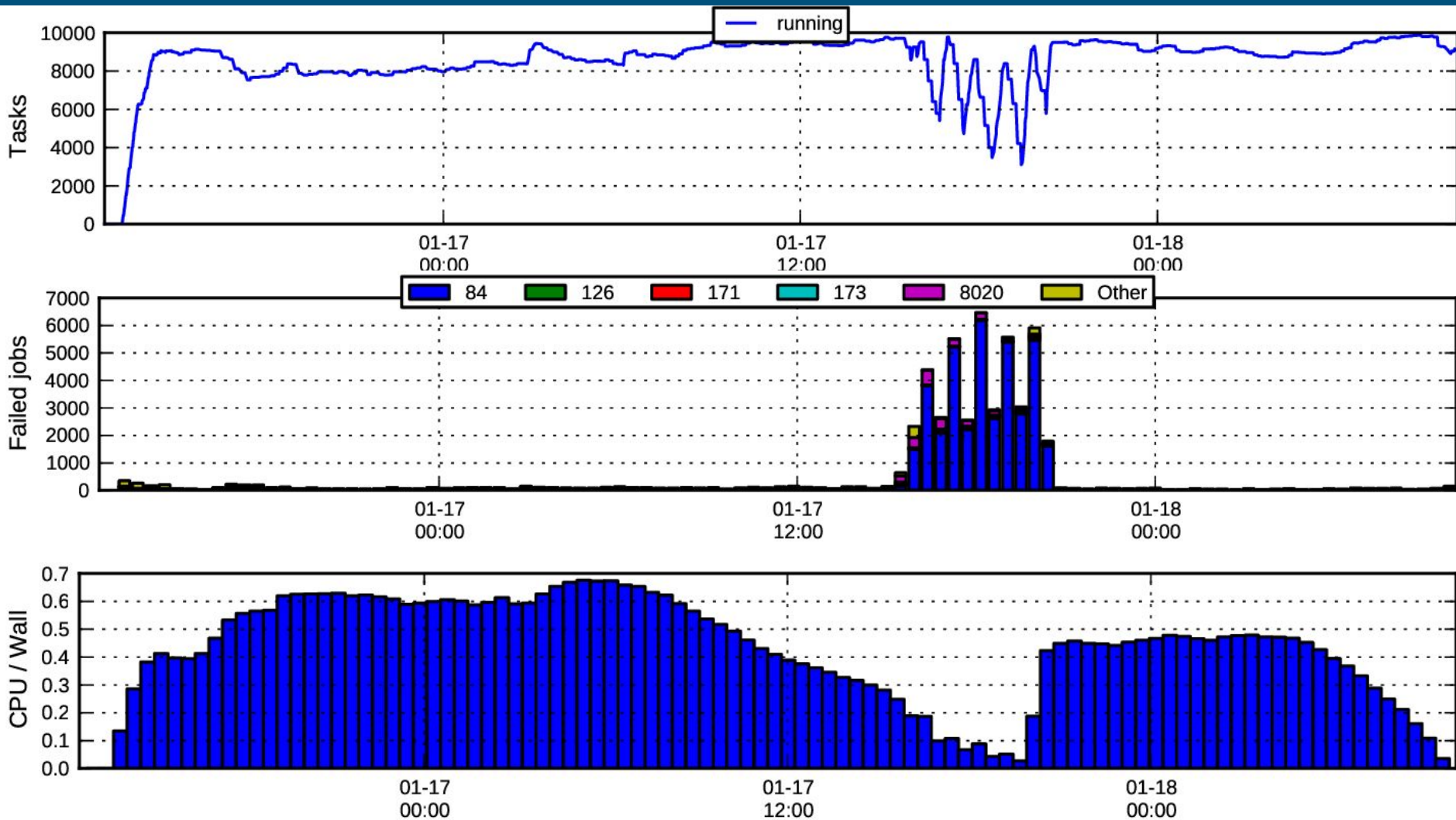
Why?

- You may not own the machines (e.g. opportunistic resources like Condor)
- You may not have admin. privileges on the machines.
- Easier to move a mountain, than to convince your sys admin to install a kernel module.
- You are running in a container, and the host system does not have CVMFS.
- The machine may have limited, or no external connectivity at all.



Parrot in CMS (ND Lobster, last year results)

This year O(25k) cores on non-dedicated resources.



ND CMS + CCTools + libCVMFS + CRC ~ Lobster

Anna Woodard

Matthias Wolf

Kenjy Hurtado

Charles Mueller

Nil Valls

Kevin Lannon

Michael Hildreth

Ben Tovar

Patrick Donnelly

Douglas Thain

Jakob Blomer

Dan Bradley

Rene Meusel

Paul Brenner

Serguei Fedorov











Lobster is a user-level system for deploying data intensive high-throughput application on non-dedicated resources.

(parrot-cvmfs and CRC not required...)

condor.cse.nd.edu

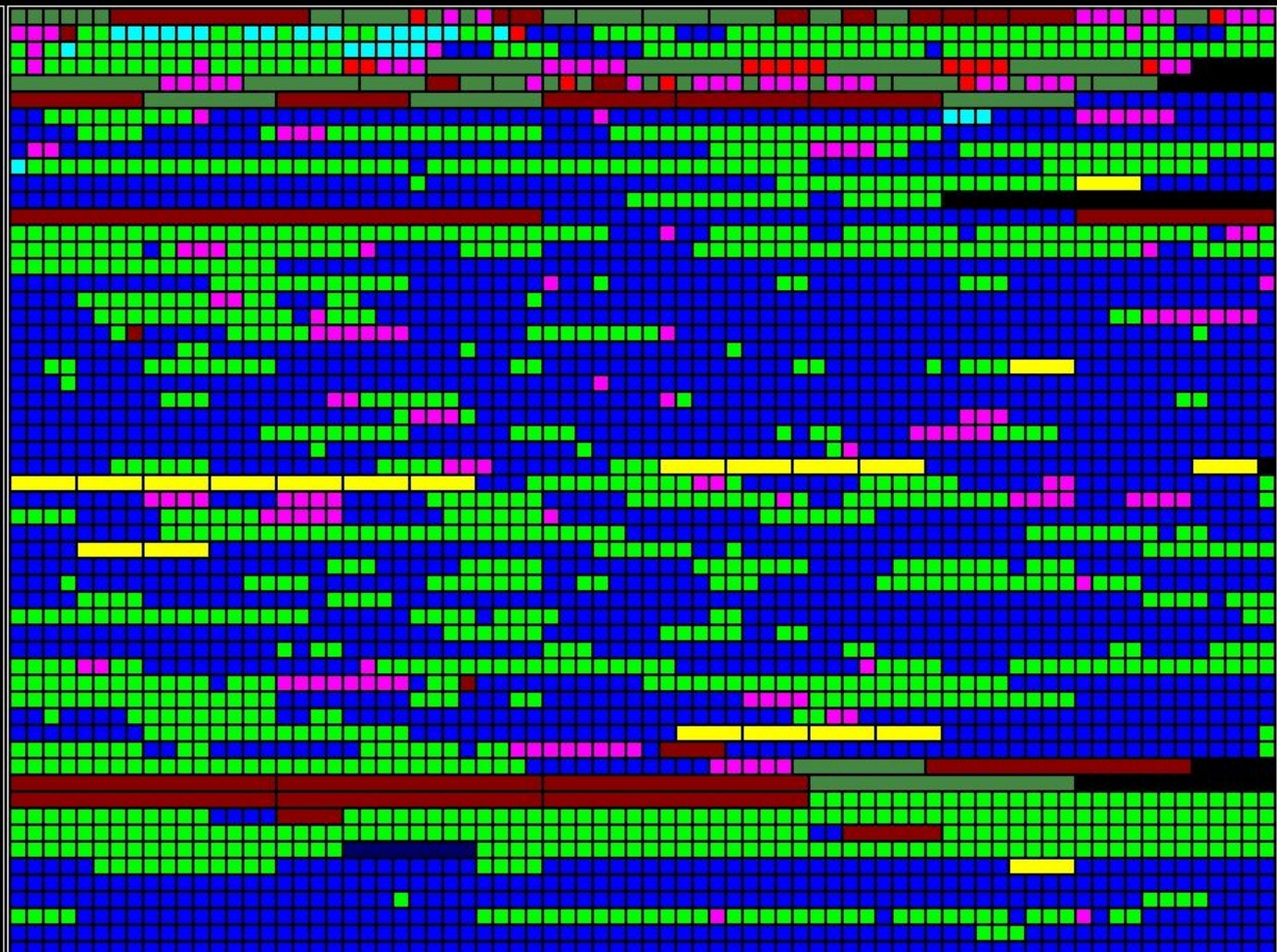
Notre Dame Condor Status

Slots Cores

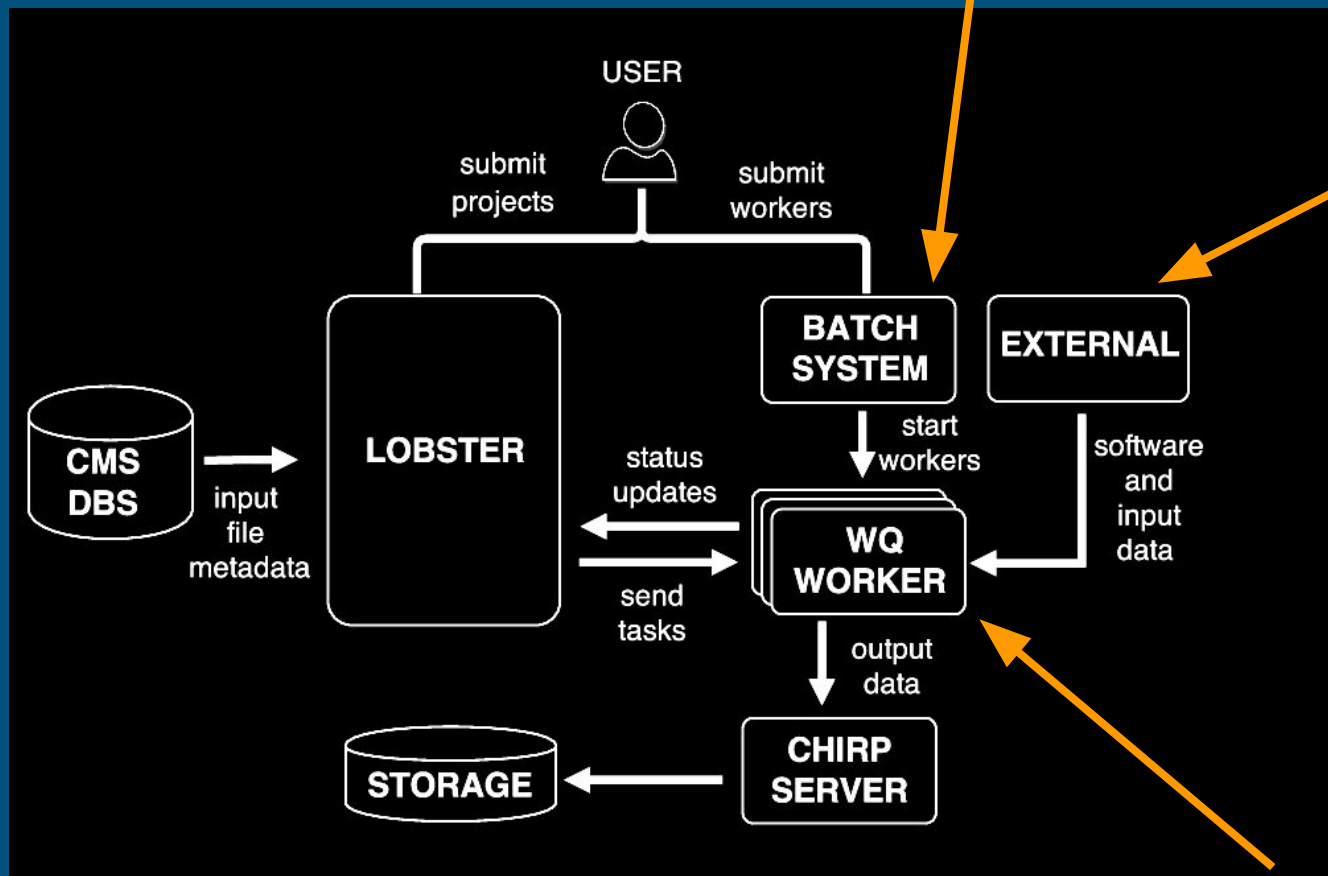
| | | |
|---|------|------|
|  dmitche6@nd.edu | 4195 | 4195 |
|  mzhu4@nd.edu | 1764 | 1764 |
|  awoodard@nd.edu | 89 | 356 |
|  rbixler@nd.edu | 182 | 182 |
|  apaul2@nd.edu | 39 | 39 |
|  nblancha@nd.edu | 18 | 18 |
|  Unclaimed | 69 | 319 |
|  Matched | 1 | 8 |
|  Preempting | | |
|  Owner | 48 | 475 |
| Total | 6405 | 7356 |

Display Options

Sort: [users](#) [machines](#)
Show: [users](#) [states](#)
Size: [bigger](#) [smaller](#)
Scale: [none](#)
[cores](#)



Lobster

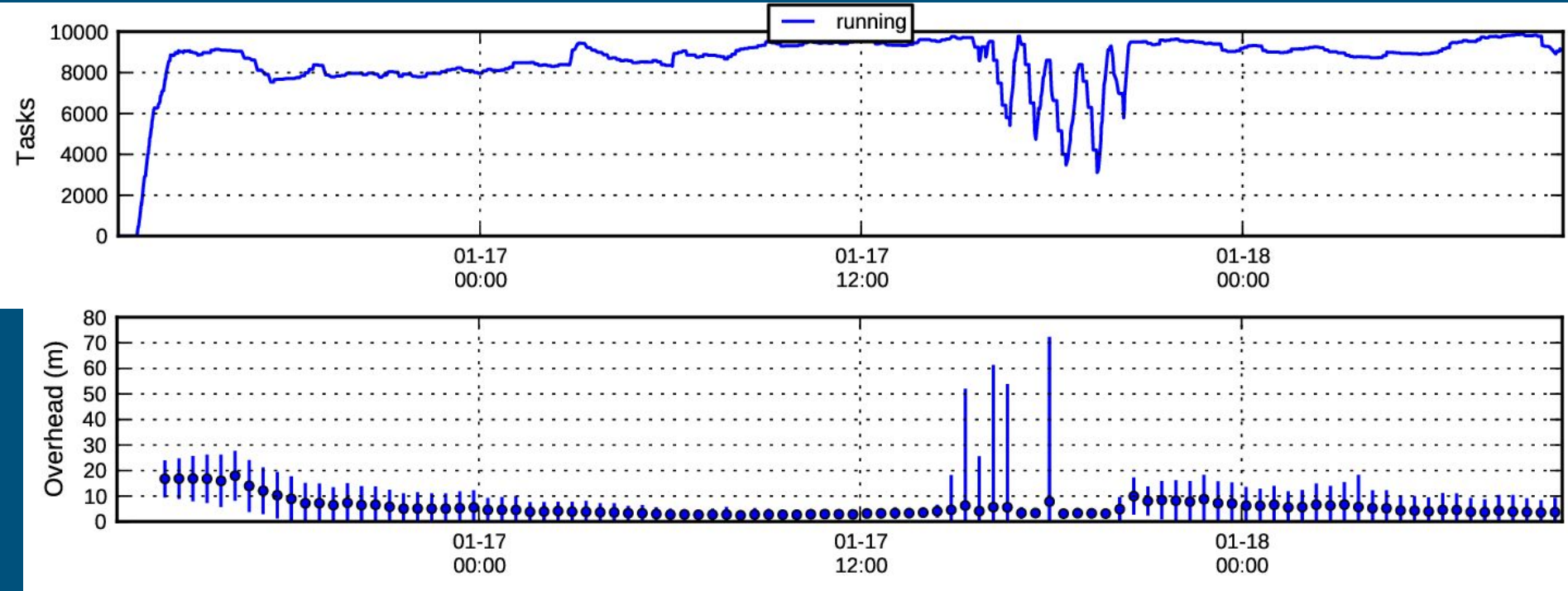


Non-dedicated resources through condor

CVMFS access through parrot

Parrot deployed as just another job input file

Measuring overheads

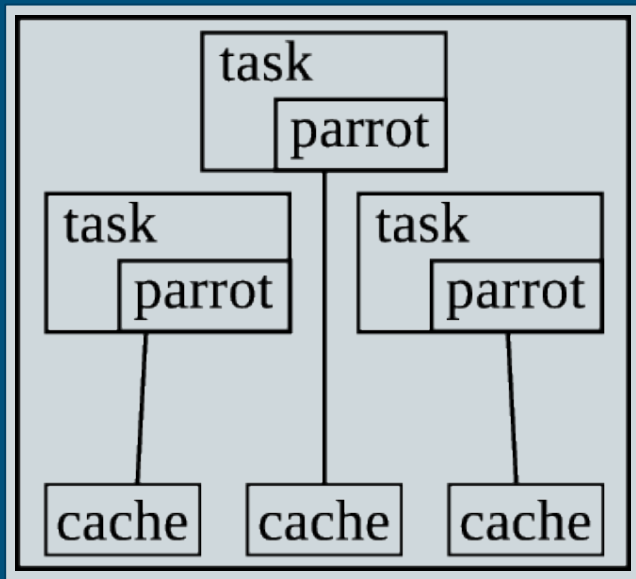


(a maximum of 4 tasks per worker/condor job)

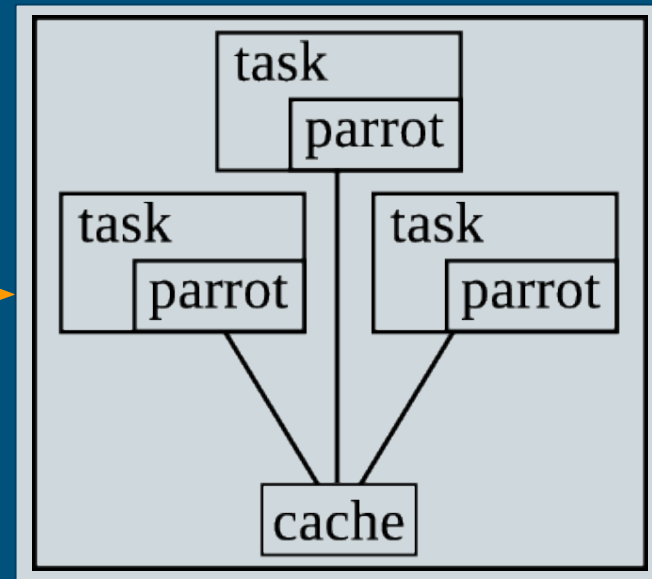
Efficient access to the same data



Using libcvmf's **alien cache** with parrot.

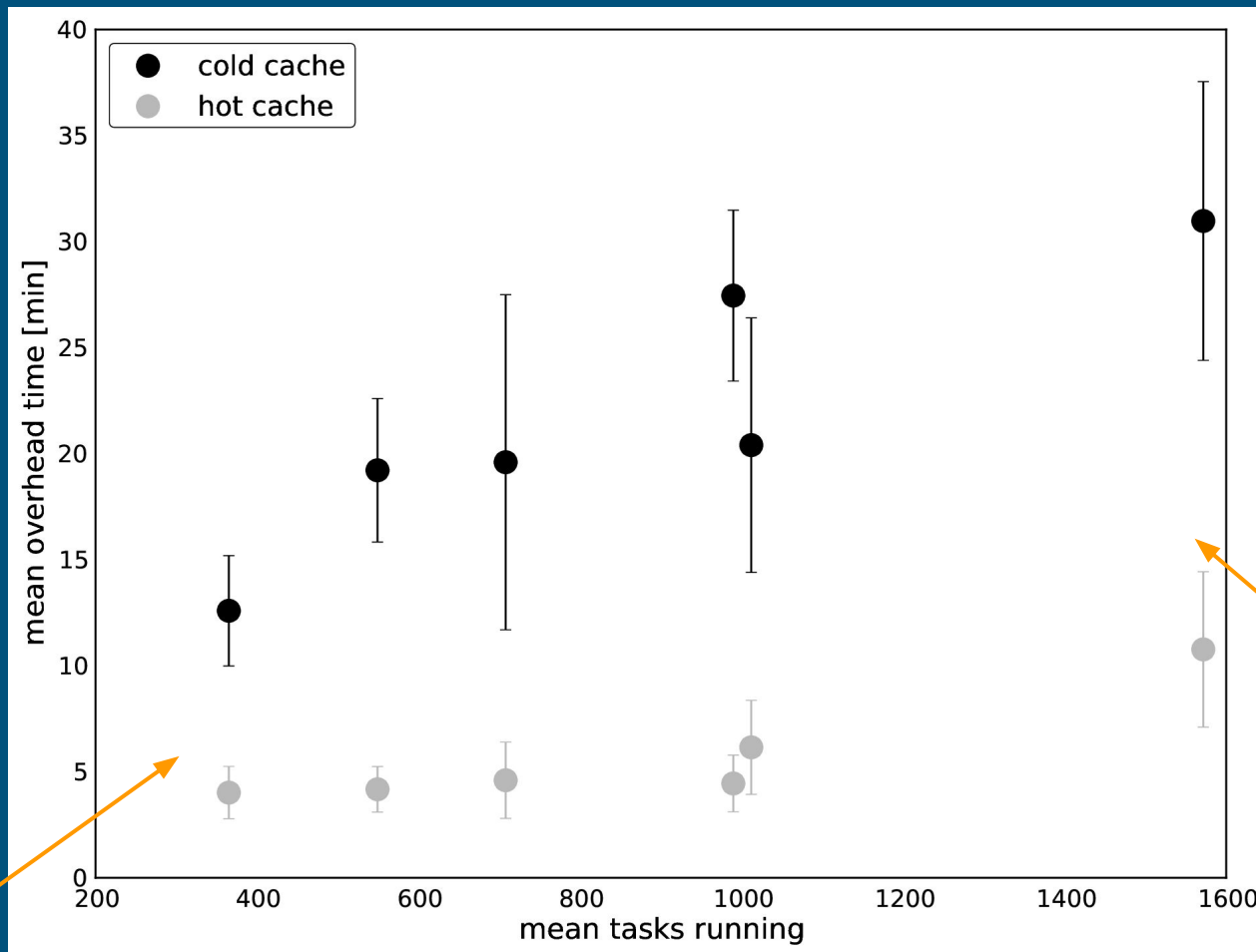


local cache per parrot



alien cache per node

Measuring overheads



few tasks,
overhead
mostly from
parrot.

many tasks,
overhead
from other
parts of
lobster

Parrot in Atlas (Rodney Walker)

Rodney is using 'alien cache' to the extreme.

- LMU-München nodes have very limited outside connectivity. **No connectivity to CERN.**
- Making local copies of repositories was error prone, as CVMFS paths are not relocatable.
- Rodney has CVMFS releases of interest as an alien cache on GPFS, accessible by all parrot instances. (300 nodes, O(40K) nodes)).
- Size of alien cache is about 1TB.
- Atlas applications run non-the-wiser, as if they had access to CERN for CVMFS data.

CernVM as Docker container with parrot

Work by Jakob Blomer and Tom Boccali.

Technology preview!

<https://cernvm.cern.ch/portal/docker>

```
docker run -it my_cernvm /init ls -lad /cvmfs/...
```

parrot's dream use

parrot_run

a whole
workflow

Parrot Troubles (just last week...)

batrick commented 4 days ago

The Cooperative Computing Lab member

@annawoodard reported an assertion failure in Parrot:

```
tsetparrot: pfs_table.cc:157[HEAD:3b27b153]: Assertion 'fd == rfd ||
((0 <= rfd && rfd < pointer_count) && pointers[rfd] == __ null)'
failed.
```

An excerpt from a debug log:

```
2016/03/04 14:52:19.22 parrot_run[18430] <child:20111> debug: diverting to openat(4094, `pfs@
2016 setp
2016 20110 is a racing thread with 19174. 20110 is attempting to set FD_CLOEXEC on a file descriptor (shared
with 19174). The newly cloned process 20111 may or may not have the change to the fd table by
20110.
```

The reason for the assertion failure is that Parrot expects its version of the file descriptor table to be in sync with the kernel. Unfortunately, this is tricky to fix. I'm not yet sure what the right approach is.



batrick added a commit to batrick/cctools that referenced this issue 13 hours ago



Add wait barriers to serialize fd table changes. ...

61e8f1a



batrick closed this in #1183 13 hours ago

parrot's recommended use

a whole
workflow

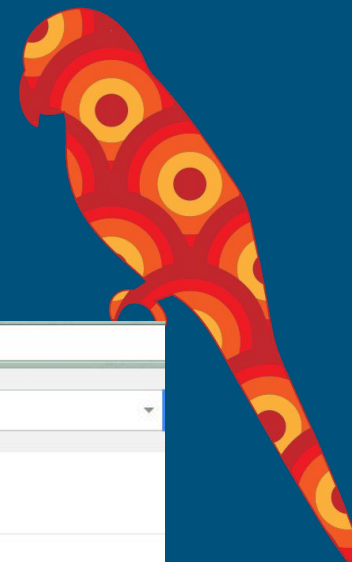
parrot_run

parrot_run

parrot_run

parrot has to mimic the kernel and de facto behaviour of glibc. It is a good way to discover the skeletons in the closet of the kernel and glibc. Thus, it is better to localize its use.

Questions



The screenshot shows a web browser window with the address bar containing `https://groups.google.com/forum/#!forum/cctools-nd`. The page features the Google logo and a search bar. Below the search bar, there are navigation buttons: 'NEW QUESTION' (in red), a refresh icon, 'Mark all as read', and a 'Filters' dropdown. The main content area is titled 'Cooperative Computing Tools' and is marked as 'Shared publicly'. It shows '20 of 20 topics (18 unread)' and an 'Apply to join group' button. A paragraph of text describes the group's purpose: 'Questions, bug reports, discussion, and announcements about the Cooperative Computing Tools, and other software. Don't forget to read the manuals! Code patches and detailed bugs may also be posted. Please respond to each question in a timely way.' Below this, there is a link to a retired listserve archive. Two topic entries are visible: 'Infinite retry of large file transfer (11)' by matthew...@gmail.com (11 posts, 2 views) and 'Unable to Connect Workers to WorkQueue Master (9)' by jkin...@nd.edu (9 posts, 2 views). The left sidebar contains navigation options: 'My groups' (Home, Starred), 'Favorites' (with a tip to click a star icon), and 'Recently viewed' (listing 'comp.unix.shell' and 'comp.text.tex').

btovar@nd.edu

<http://ccl.cse.nd.edu>

<http://ccl.cse.nd.edu/downloads>

<http://ccl.cse.nd.edu/community/forum>

<https://github.com/cooperative-computing-lab/cctools>