

# Running Interpreted Jobs

Condor Project  
Computer Sciences Department  
University of Wisconsin-Madison



# Overview

- > Many folks running Matlab, R, etc.
- > Interpreters complicate Condor jobs
- > Let's talk about best practices.

# What's R?

```
#!/usr/bin/R  
X <- c(5, 7, 9)  
cat (X)
```



What could possibly go wrong?

# Submit file

```
universe = vanilla
executable = foo.r
output = output_file
error = error_file
log = log
queue
```

# What's so hard?

```
#!/usr/bin/R
```

What if /usr/bin/R isn't there?

```
#!/usr/bin/env R
```

isn't good enough -- Condor doesn't set the PATH for a Condor job.

# Pre-staging: One (not-so-good) solution

If you control the site, pre-stage R

```
#!/software/R/bin/R
```

> Fragile!

# Pre-staging: If you must...

"test and advertise"

Use a Daemon ClassAd hook like:

```
STARTD_CRON_JOBLIST = R_INFO
STARTD_CRON_R_INFO_PREFIX =
STARTD_CRON_R_INFO_EXECUTABLE = \ $
    (STARTD_CRON_MODULES)/r_info
STARTD_CRON_R_INFO_PERIOD = 1h
STARTD_CRON_R_INFO_MODE = periodic
STARTD_CRON_R_INFO_RECONFIG = false
STARTD_CRON_R_INFO_KILL = true
STARTD_CRON_R_INFO_ARGS =
```

# R\_info script contents

```
#!/bin/sh
if [[ -d /path/to/r/bin &&
      /path/to/R/bin/R -version > /dev/null ]]
then
    echo "has_r = true"
fi
```

What about multiple installations of R ?

# Pre-staging is bad

- > Limits where your job can run
- > Must be an administrator to set up
- > Difficult to change
  - Pre-staged files can change unexpectedly
    - Upgrade, new system installation, disk problems, ...

# Solution: take it with you

- > Bundle up the whole runtime
- > Transfer the bundle with the job
- > Wrapper script unbundles and runs
- > Downsides:
  - Extra time overhead to unbundle
  - Not so good for short\* jobs

# Benefits

- > Can run anywhere\*:
  - Flocked, Campus Grids, OSG, etc.
- > Each job can have own runtime version/configuration.

# Revised submit file

```
universe = vanilla
executable = wrapper.sh
output = output_file
error = error_file
transfer_input_files = runtime.tar.gz,
    foo.r
should_transfer_files = true
when_to_transfer_output = on_exit
log = log
queue
```

# wrapper.sh

```
#!/bin/sh  
tar xzf runtime.tar.gz  
./bin/R foo.r
```

# Downside: Those Huge Runtimes

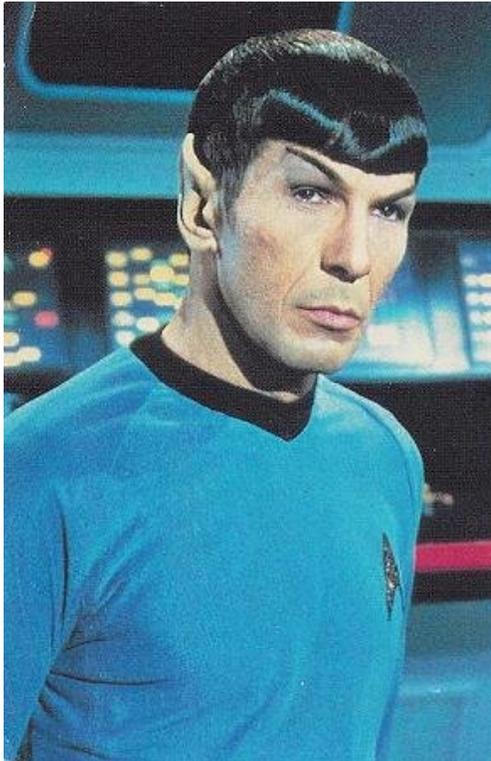
- > Full R, matlab runtime 100 Mb
  - Adds up when running thousands of jobs
- > Trivia: How long to transfer 100 Mb?
  - Is this really a problem?

# Mitigating Huge Runtimes

1. Trim the bundle down  
(identify unneeded files with strace)
2. Second, perhaps  $> 1$  task per job

Finally, cache with Squid

# Users, not admins



# http://condor-wiki.cs.wisc.edu

CondorWiki: Wiki Index

https://condor-wiki.cs.wisc.edu/index.cgi/wiki

**CONDOR**  
High Throughput Computing

[Help] [Home] [Login] [Reports] [Search] [Timeline]  
Not logged in

**Wiki Index**

[Contents] [Diff] [History] [Text]

## Welcome to the Condor Flightworthy Development Wiki

### User FAQs

- [Administrator HOWTO Recipes](#)
- [How to make a contribution](#)
- [How to provide a contrib module](#)
- [List of available contrib modules](#)
- [Running DAGMan itself as a Condor-C job](#)
- [Using SOAP with Condor](#) - includes a Python/suds example
- [Ports and Where They Work](#)
- [The Ports Table](#)

### People who maintain and author Condor

- [Developer Home Pages](#)
- [Condor Contributors Past and Present](#)

### Really Important Stuff

- [What branch should receive my code?](#)
- [Work plan for upcoming releases](#)
- [Nightly build results](#) ([Built-Test Blog](#))
- [GitWeb](#) (and deprecated [Bonsai](#))
- [Info on how gitweb is updated](#) and which branches are mirrored to gitweb
- [Download Analysis](#)
- [Coverity-found bugs](#)
- Flightworthy meetings (in 4259)
  - Tuesdays @ 3:00
  - Fridays @ 2:15
  - Todd's bridge: 712-432-0175, access code: 733599#
- [Effort Tracking](#)

### Wisdom

- [Release Punch-List](#)
- [Creating GitTrac Accounts](#)
- [Git Wisdom](#)
- [Version History How To](#)
- [#1377](#) How our download pages work.
- [#1397](#) How our nightly builds work.
- [Daemon Meditations - Control flows in daemons](#)
- [Adding Configuration File Parameters](#)
- [Repository Backups](#)
- [Profiling Memory Usage](#)
- [Profiling Code Execution](#)
- [SyscallMode](#) (Feb. 1999)
- [JavaUniverse](#) (Dec. 2001)
- [Creating a new daemon](#) (Feb. 2002)
- [MPI Wisdom](#) (July 2002)
- [Ins and Outs of VMWARE Jobs](#) (Nov 2009)
- [Coding Guidelines/Rules when using Imake in the Condor build system](#) (May 2004)
- details about [condor\\_stats](#) (Jan 2004)
- [Adding a new component to build tree](#) (Jan 2004)
- [Adding a new external package](#)
- [CondorError data structure](#) (Jan 2006)
- [Autoclustering and SIGNIFICANT\\_ATTRIBUTES](#) (Jan 2006)
- [Performing Code Reviews](#)
- [Stork](#) (June 2006)
- [Debugging Smashed Stacks](#) (Aug 2006)
- [Windows Development and Debugging](#)
- [GCC Inline Assembly Tutorial](#)
- [HOWTO quickly setup GSI Security for testing](#)
- [Magic Numbers in Condor and Grid Software](#)
- [Architecture Manuals](#)
- [Child Management \(VMS\)](#)

Building and Testing Condor



# http://condor-wiki.cs.wisc.edu

www.cs.wisc.edu/Condor



# Using HTTP/Squid

- Change wrapper to manually wget
- Set env http\_proxy to squid source
  - OSG\_SQUID\_LOCATION in OSG
  - Otherwise, set with Daemon ClassAd hooks and \$\$
- Cut runtime.tar.gz from transfer\_input\_files, add  
wget -retry-connrefused -waitretry=10  
your\_http\_server
- To the wrapper script - note retries

Don't use curl!

- Or set -H pragma

# Matlab complications

- > Licensing...
  - Octave (?)
  - Matlab compiler!
- > Matlab parallel toolkit
  - HTPC

# Cross Platform submit

- > Many grids > 1 platform:
  - Unix vs. Windows; 32 vs 64 bit
- > Huge benefit of High Level language:
  - Write once, run, ... well...
- > Use Condor \$\$ to expand:

```
executable = wrapper.$$  
(OPSYS).bat
```

- > Condor will expand OPSYS to LINUX or WINNT<XX>
- > Write both wrappers, make sure to wget correct runtime

# Summary

Many folks running lots of interpreted jobs

Transferring runtime along beneficial, but requires set up

Cross platform submits can be huge win