

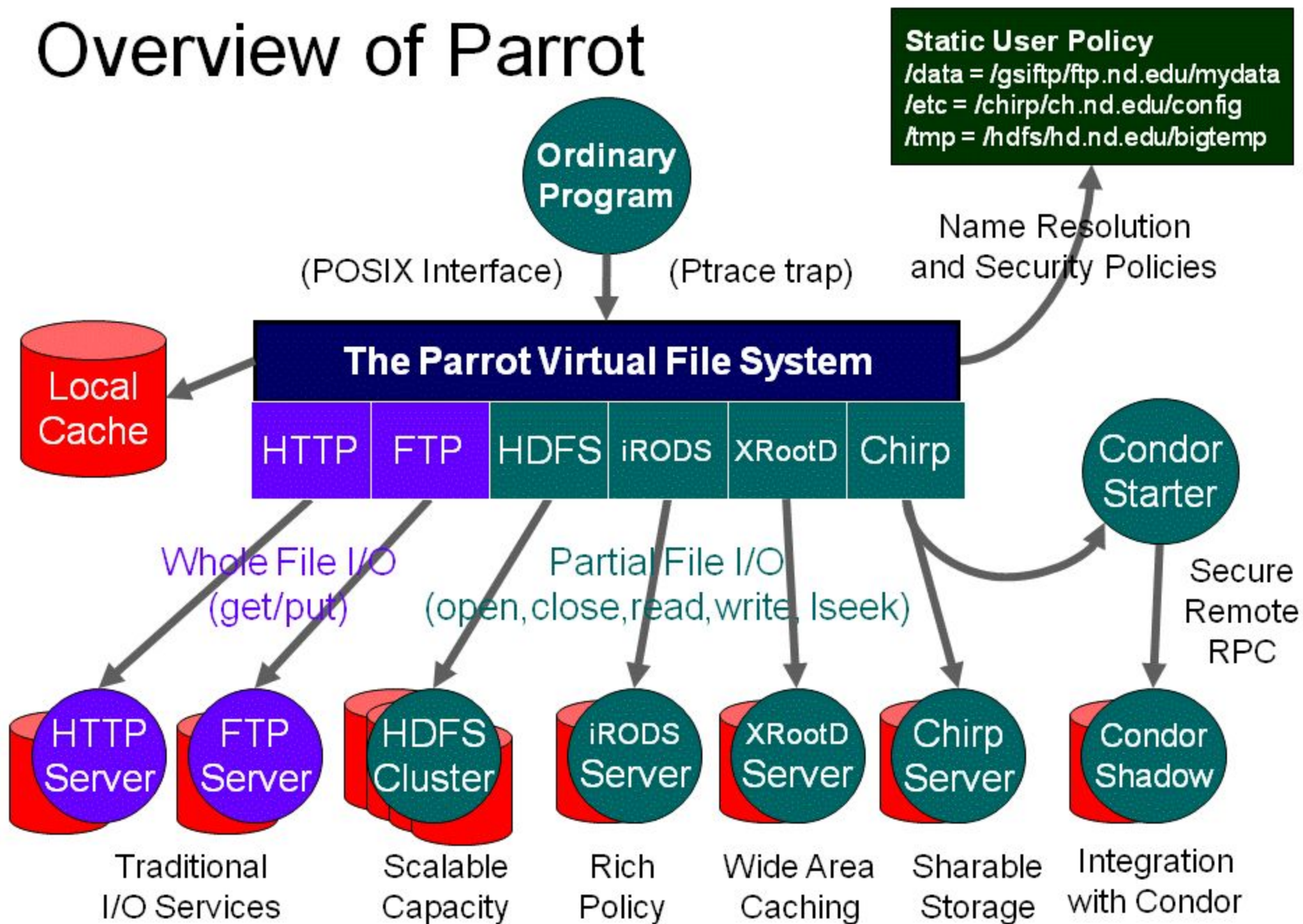
Using Parrot in Scientific Workflows

Tim Shaffer
University of Notre Dame

tshaffe1@nd.edu



Overview of Parrot



Misbehaving Tasks

Problem: a large number of temp files are accumulating on workers. Some tasks don't clean up properly before exiting.

Enter **Parrot**:

Set up each task with a private `/tmp`, now it's easy to identify/clean up what a task left behind.

```
tshaffe1@cclws19:\~
File Edit View Search Terminal Help
sh-4.1$ ls /tmp
concept-generic.png  mount.list          py1.png             sandbox2.png
concept-generic.svg  orbit-gdm           py2.png             sandbox3.png
foo                  orbit-tshaffe1      py3.png             seahorse-4VIb32
foo2                 parrot.213185        py4.png             seahorse-JYSDPc
hsperfdata_condor    parrot-poster.gif    py5.png             seahorse-uIUK1d
kde-tshaffe1          private.01           py6.png             tmp.png
keyring-bC1B23        pulse-j10IeCGIGEdv  py7.png             virtual-tshaffe1.lUjwS5
keyring-P19Vjf         pulse-mLj04nvS9dzI  root
krb5cc_213185_0lGP5o  pulse-zQR5WAFKE1yG  sandbox1.png
sh-4.1$
```

```
tshaffe1@cclws19:\~
File Edit View Search Terminal Help
sh-4.1$ ls /tmp
concept-generic.png  mount.list          py1.png             sandbox2.png
concept-generic.svg  orbit-gdm           py2.png             sandbox3.png
foo                  orbit-tshaffe1      py3.png             seahorse-4VIb32
foo2                 parrot.213185        py4.png             seahorse-JYSDPc
hsperfdata_condor    parrot-poster.gif   py5.png             seahorse-uIUK1d
kde-tshaffe1         private.01           py6.png             tmp.png
keyring-bC1B23        pulse-j10IeCGIGEdv  py7.png             virtual-tshaffe1.lUjwS5
keyring-P19Vjf        pulse-mLj04nvS9dzI  root
krb5cc_213185_0lGP5o  pulse-zQR5WAFKE1yG  sandbox1.png
sh-4.1$ parrot_run -m mount.list sh -l
sh-4.1$
```

```
tshaffe1@cclws19:\~
File Edit View Search Terminal Help
sh-4.1$ ls /tmp
concept-generic.png  mount.list          py1.png             sandbox2.png
concept-generic.svg  orbit-gdm           py2.png             sandbox3.png
foo                  orbit-tshaffe1     py3.png             seahorse-4VIb32
foo2                 parrot.213185       py4.png             seahorse-JYSDPc
hsperfdata_condor   parrot-poster.gif   py5.png             seahorse-uIUK1d
kde-tshaffe1         private.01          py6.png             tmp.png
keyring-bC1B23       pulse-j10IeCGIGEdv  py7.png             virtual-tshaffe1.lUjwS5
keyring-P19Vjf       pulse-mLj04nvS9dzI root
krb5cc_213185_0lGP5o pulse-zQR5WAFKE1yG sandbox1.png
sh-4.1$ parrot_run -m mount.list sh -l
sh-4.1$ ls /tmp
messy_script.sh
sh-4.1$
```

```
tshaffe1@cclws19:\~
File Edit View Search Terminal Help
sh-4.1$ ls /tmp
concept-generic.png  mount.list          py1.png             sandbox2.png
concept-generic.svg  orbit-gdm           py2.png             sandbox3.png
foo                  orbit-tshaffe1      py3.png             seahorse-4VIb32
foo2                 parrot.213185       py4.png             seahorse-JYSDPc
hsperfdata_condor   parrot-poster.gif   py5.png             seahorse-uIUK1d
kde-tshaffe1        private.01          py6.png             tmp.png
keyring-bC1B23      pulse-j10IeCGIGEdv  py7.png             virtual-tshaffe1.lUjwS5
keyring-P19Vjf      pulse-mLj04nvS9dzI  root
krb5cc_213185_0lGP5o pulse-zQR5WAFKE1yG  sandbox1.png
sh-4.1$ parrot_run -m mount.list sh -l
sh-4.1$ ls /tmp
messy_script.sh
sh-4.1$ ./messy_script.sh
writing setup files
making a mess of /tmp
doing other stuff
done!
exiting without cleaning up
sh-4.1$
```



```
tshaffe1@cclws19:\~
File Edit View Search Terminal Help
sh-4.1$ parrot_run -m mount.list sh -l
sh-4.1$ ls /tmp
messy_script.sh
sh-4.1$ ./messy_script.sh
writing setup files
making a mess of /tmp
doing other stuff
done!
exiting without cleaning up
sh-4.1$ ls /tmp
file1    file20  file32  file44  file56  file68  file8    file91
file10   file21  file33  file45  file57  file69  file80   file92
file100  file22  file34  file46  file58  file7    file81   file93
file11   file23  file35  file47  file59  file70  file82   file94
file12   file24  file36  file48  file6    file71  file83   file95
file13   file25  file37  file49  file60  file72  file84   file96
file14   file26  file38  file5    file61  file73  file85   file97
file15   file27  file39  file50  file62  file74  file86   file98
file16   file28  file4    file51  file63  file75  file87   file99
file17   file29  file40  file52  file64  file76  file88   messy_script.sh
file18   file3    file41  file53  file65  file77  file89   more_things
file19   file30  file42  file54  file66  file78  file9    options
file2    file31  file43  file55  file67  file79  file90   setup
sh-4.1$
```



```
tshaffe1@cclws19:\~
File Edit View Search Terminal Help
messy_script.sh
sh-4.1$ ./messy_script.sh
writing setup files
making a mess of /tmp
doing other stuff
done!
exiting without cleaning up
sh-4.1$ ls /tmp
file1    file20  file32  file44  file56  file68  file8    file91
file10   file21  file33  file45  file57  file69  file80   file92
file100  file22  file34  file46  file58  file7    file81   file93
file11   file23  file35  file47  file59  file70  file82   file94
file12   file24  file36  file48  file6    file71  file83   file95
file13   file25  file37  file49  file60  file72  file84   file96
file14   file26  file38  file5    file61  file73  file85   file97
file15   file27  file39  file50  file62  file74  file86   file98
file16   file28  file4    file51  file63  file75  file87   file99
file17   file29  file40  file52  file64  file76  file88   messy_script.sh
file18   file3    file41  file53  file65  file77  file89   more_things
file19   file30  file42  file54  file66  file78  file9    options
file2    file31  file43  file55  file67  file79  file90   setup
sh-4.1$ exit
logout
sh-4.1$
```

```
tshaffe1@cclws19:\~
File Edit View Search Terminal Help
file11 file23 file35 file47 file59 file70 file82 file94
file12 file24 file36 file48 file6 file71 file83 file95
file13 file25 file37 file49 file60 file72 file84 file96
file14 file26 file38 file5 file61 file73 file85 file97
file15 file27 file39 file50 file62 file74 file86 file98
file16 file28 file4 file51 file63 file75 file87 file99
file17 file29 file40 file52 file64 file76 file88 messy_script.sh
file18 file3 file41 file53 file65 file77 file89 more_things
file19 file30 file42 file54 file66 file78 file9 options
file2 file31 file43 file55 file67 file79 file90 setup
sh-4.1$ exit
logout
sh-4.1$ ls /tmp
concept-generic.png orbit-gdm py3.png sandbox5.png
concept-generic.svg orbit-tshaffe1 py4.png sandbox6.png
foo parrot.213185 py5.png seahorse-4VIb32
foo2 parrot-poster.gif py6.png seahorse-JYSDPc
hsperfdata_condor private.01 py7.png seahorse-uIUK1d
kde-tshaffe1 pulse-j10IeCGIGEdv root tmp.png
keyring-bC1B23 pulse-mLj04nvS9dzI sandbox1.png virtual-tshaffe1.lUjwS5
keyring-P19Vjf pulse-zQR5WAFKE1yG sandbox2.png
krb5cc_213185_0lGP5o py1.png sandbox3.png
mount.list py2.png sandbox4.png
sh-4.1$
```

Bonus: keep tasks from snooping around

They probably don't need access to

- `/home`
- `/dev`
- `/sys`
- `/proc`, maybe others

Alternatively, use a more fine-grained approach, e.g. "only allow a Makeflow job to write to the outputs it specified".

Portable Applications

It's hard to know what will be available at the execution site.

- missing libraries
- different filesystem layout (e.g. `/bin` vs. `/usr/bin`, or packages installed under `/opt`)
- libraries compiled with features missing
- bad `ld.so` (*really!*)

Portable Applications

Bundle all dependencies, and use Parrot to set up the filesystem.

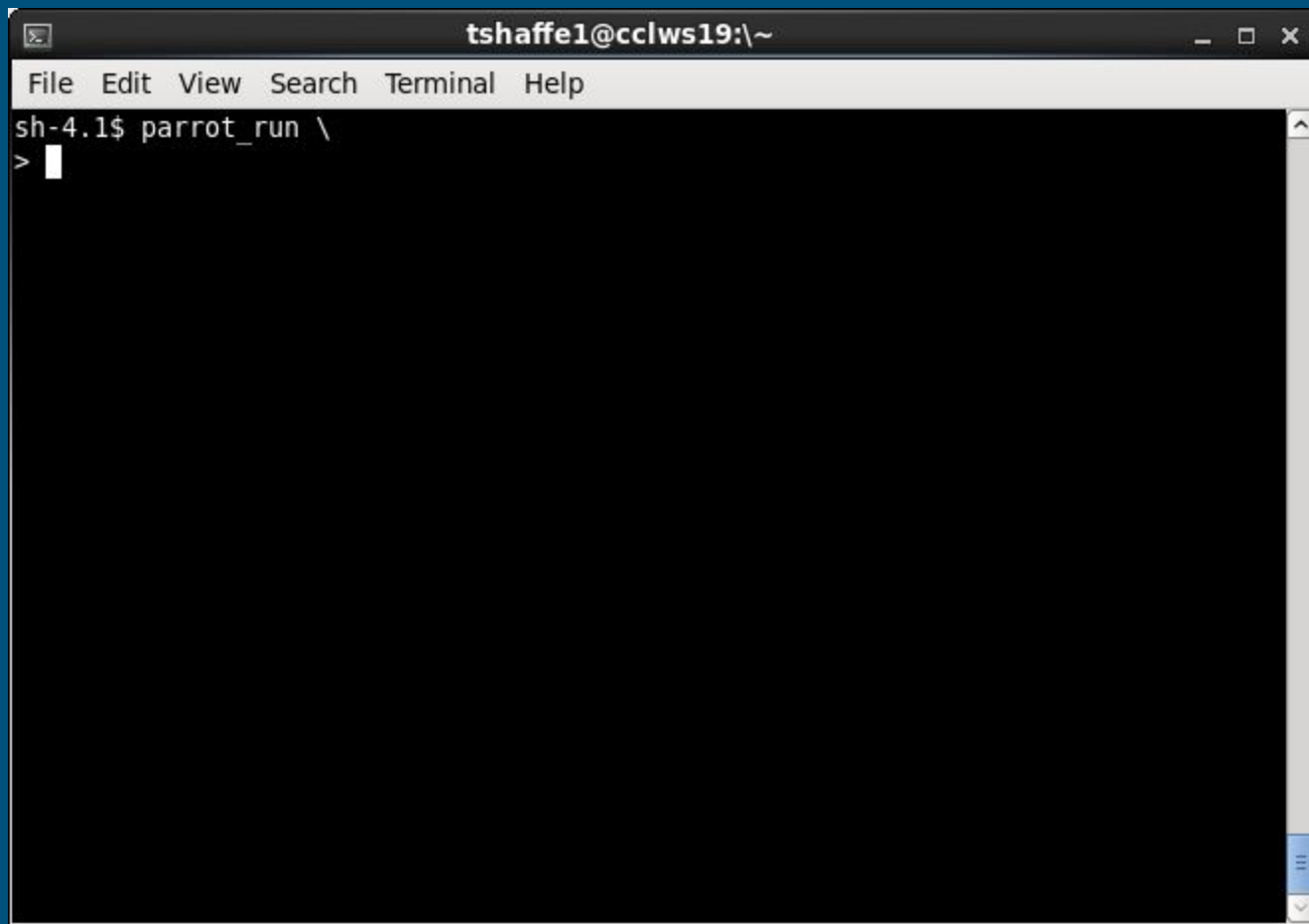
The app sees a consistent, known-good system configuration.

Parrot can automatically detect dependencies and make a package

Example: Portable Python

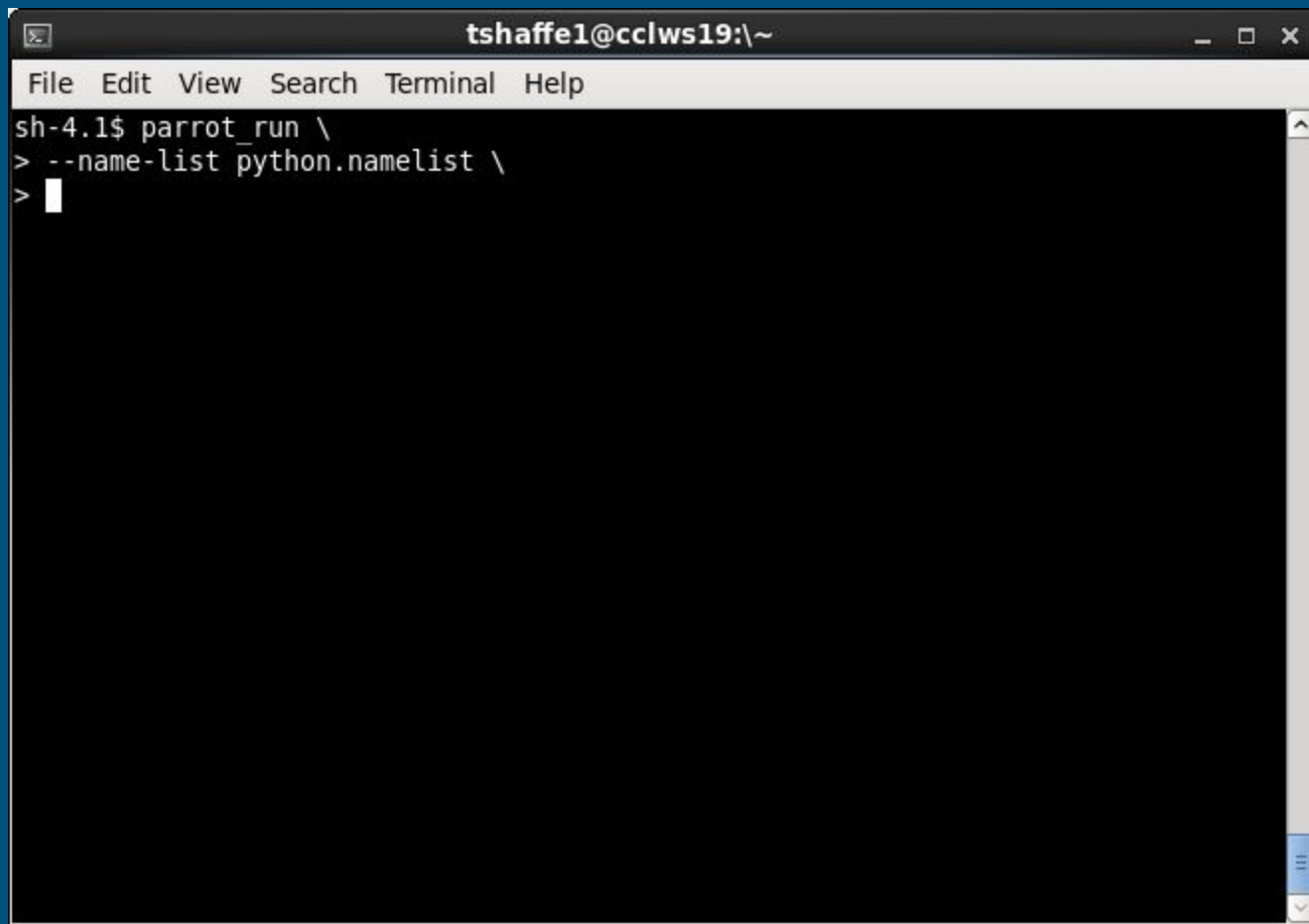
Copying the `python` binary to another computer won't work: we need libraries and dependencies

- bzip2
- db
- expat
- filesystem
- gdbm
- glibc
- iana-etc
- libffi
- linux-api-headers
- openssl
- perl
- python
- tzdata
- zlib



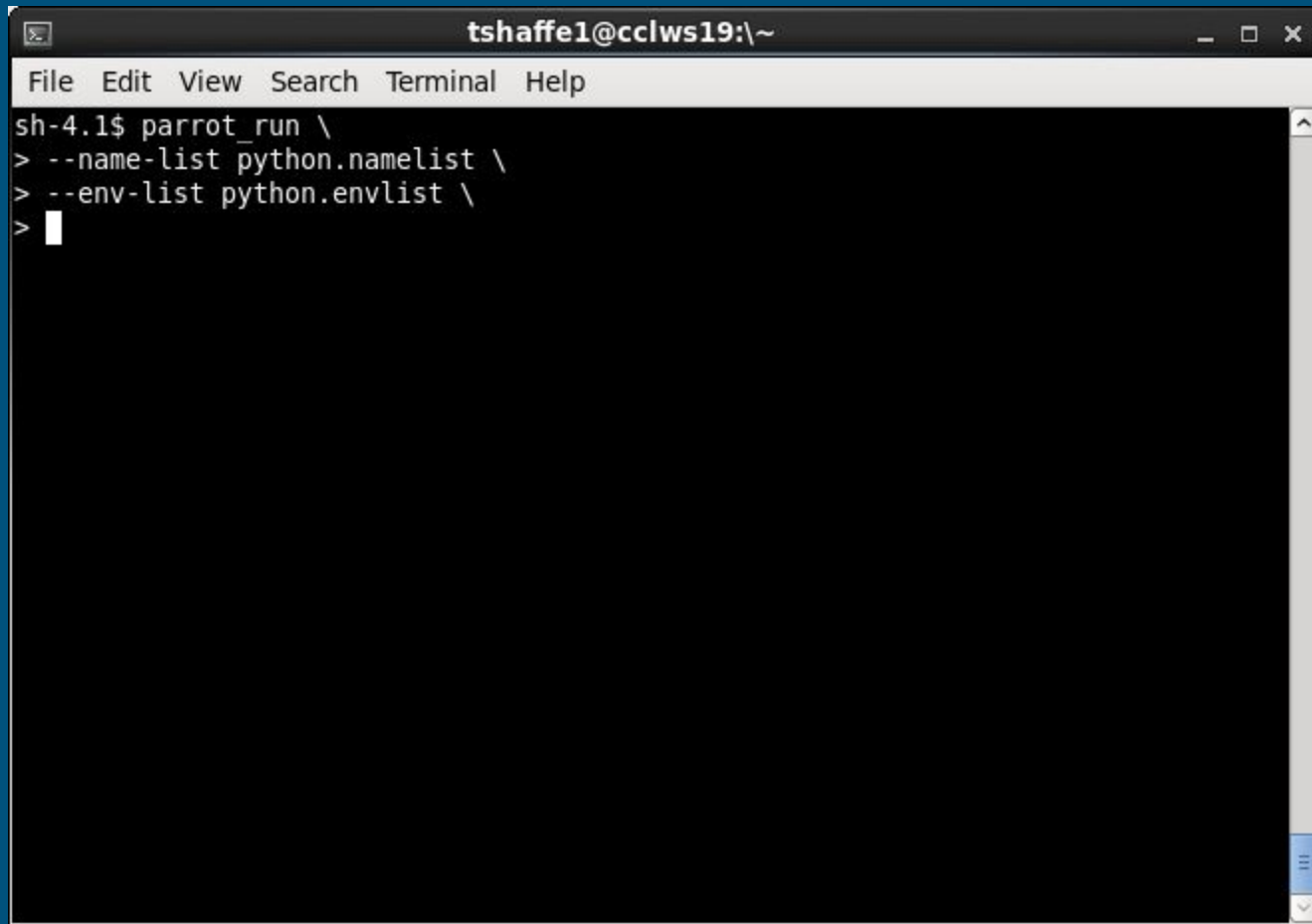
A terminal window titled "tshaffe1@cclws19:\~" with a menu bar containing "File", "Edit", "View", "Search", "Terminal", and "Help". The terminal content shows a shell prompt "sh-4.1\$" followed by the command "parrot_run \" and a new line starting with ">" and a cursor.

```
tshaffe1@cclws19:\~
File Edit View Search Terminal Help
sh-4.1$ parrot_run \"
> 
```

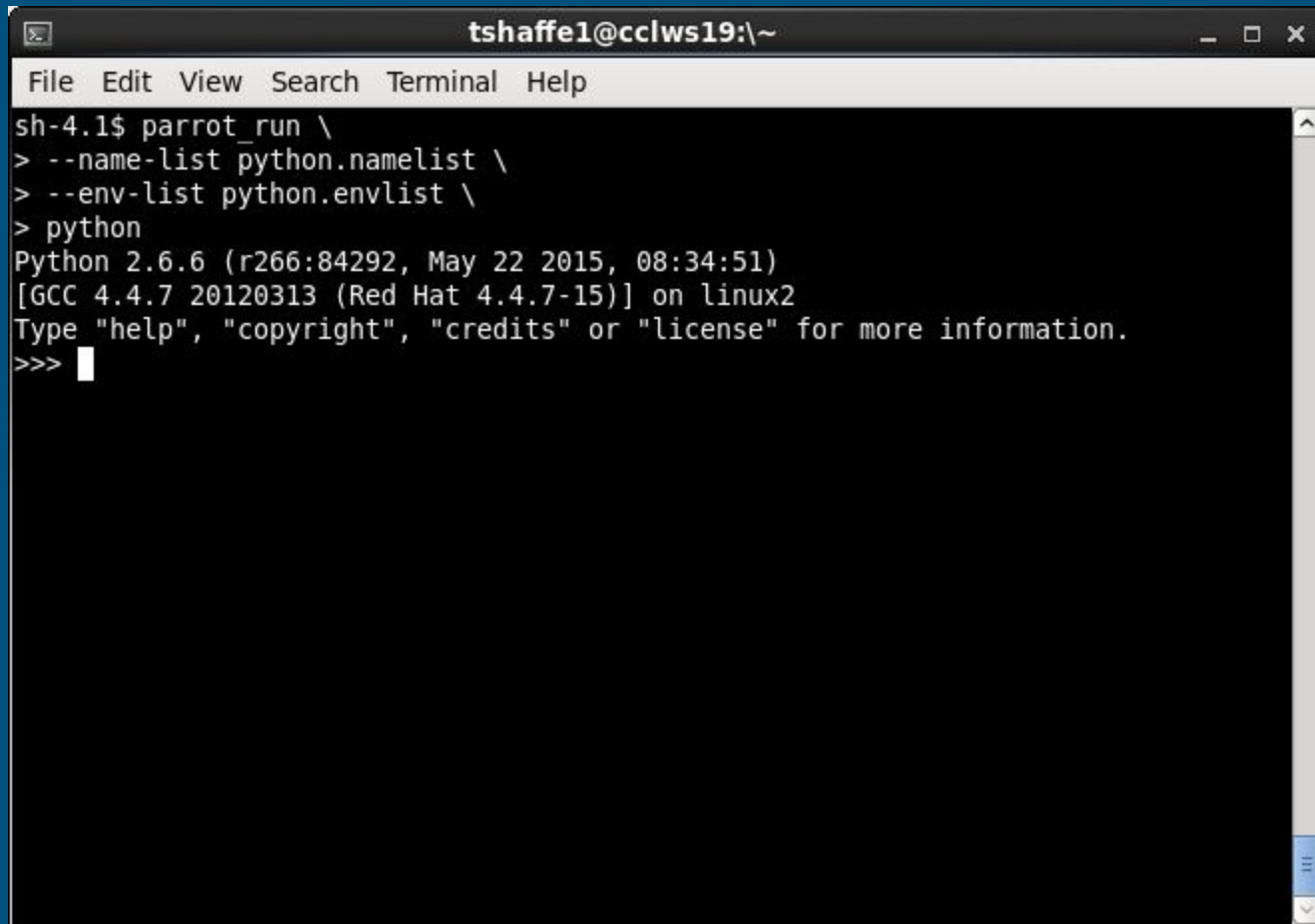
A terminal window titled "tshaffe1@cclws19:\~" with a menu bar containing "File", "Edit", "View", "Search", "Terminal", and "Help". The terminal content shows a shell prompt "sh-4.1\$" followed by the command "parrot_run \", then a new line with "> --name-list python.namelist \"", and a final line with "> " and a cursor. The window has standard OS controls (minimize, maximize, close) in the top right and a scrollbar on the right side.

```
tshaffe1@cclws19:\~  
File Edit View Search Terminal Help  
sh-4.1$ parrot_run \  
> --name-list python.namelist \  
> 
```



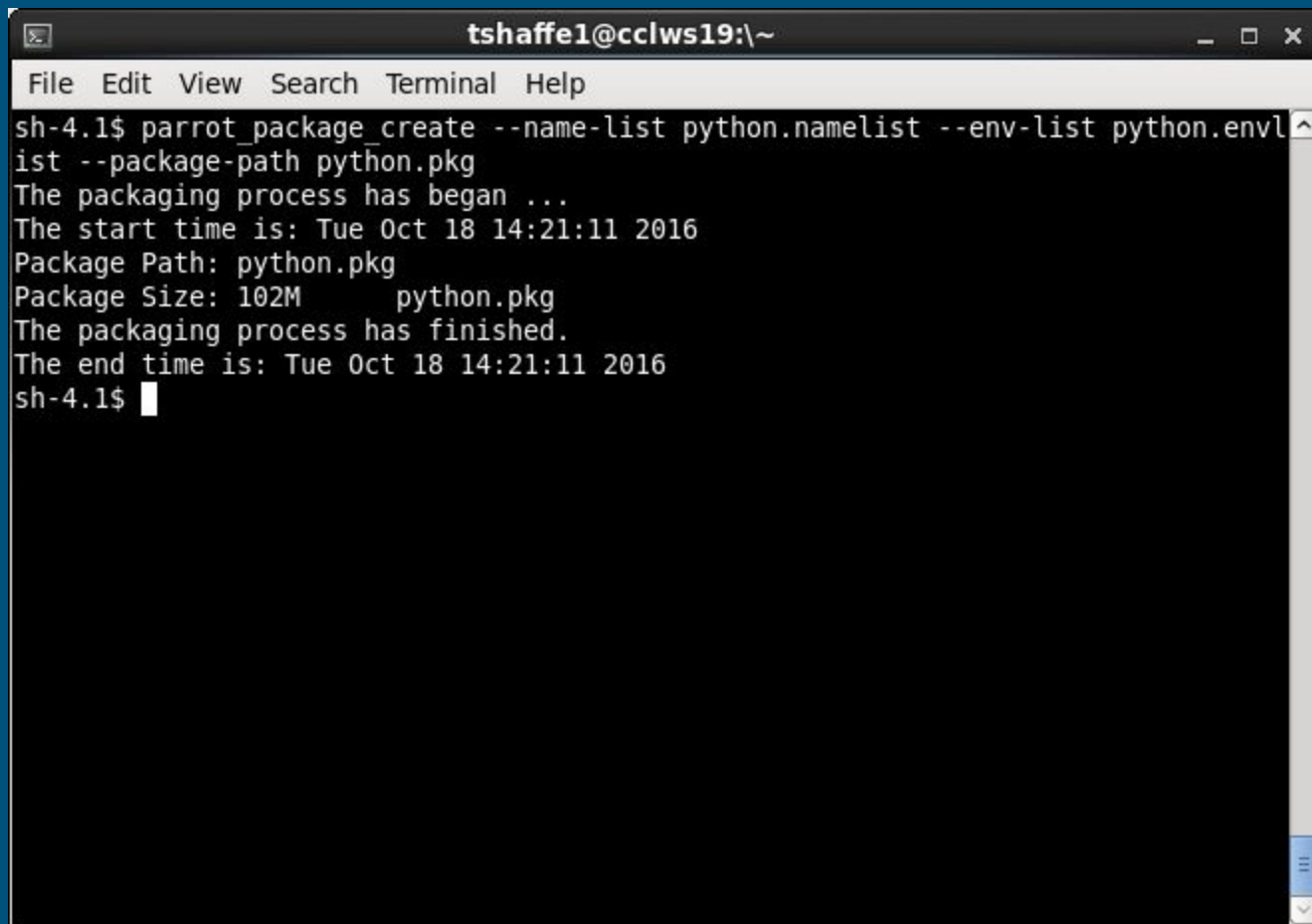
A terminal window titled "tshaffe1@cclws19:\~" with a menu bar containing "File", "Edit", "View", "Search", "Terminal", and "Help". The terminal content shows a shell prompt "sh-4.1\$" followed by the command "parrot_run \". Below this, there are three lines of input, each starting with a greater-than sign ">". The first two lines are "--name-list python.namelist \" and "--env-list python.envlist \". The third line is a single greater-than sign ">\" followed by a white cursor block. The terminal has a black background and a scrollbar on the right side.

```
tshaffe1@cclws19:\~
File Edit View Search Terminal Help
sh-4.1$ parrot_run \
> --name-list python.namelist \
> --env-list python.envlist \
> 
```



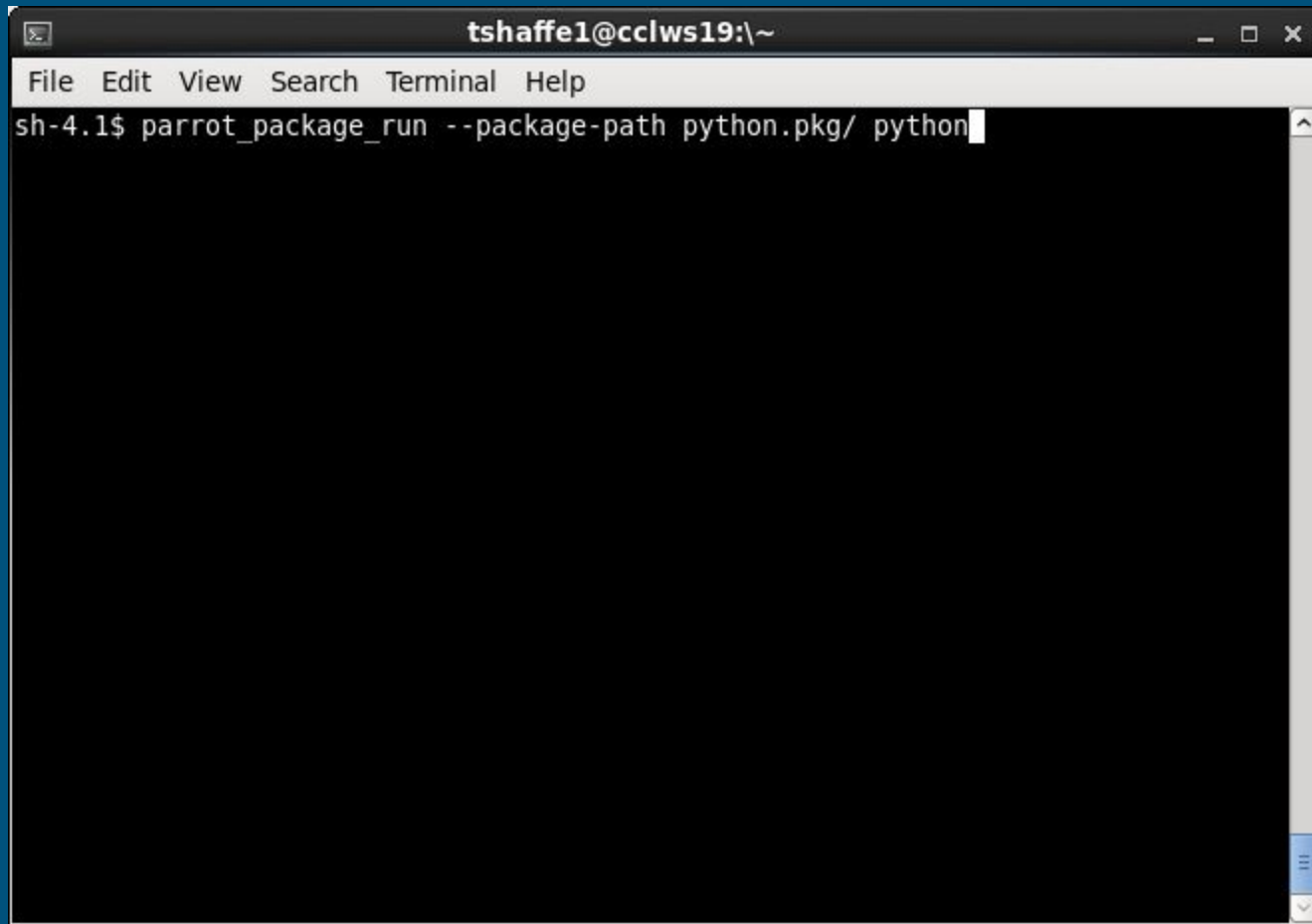
A terminal window titled "tshaffe1@cclws19:\~" with a menu bar containing "File", "Edit", "View", "Search", "Terminal", and "Help". The terminal shows the following commands and output:

```
sh-4.1$ parrot_run \  
> --name-list python.namelist \  
> --env-list python.envlist \  
> python  
Python 2.6.6 (r266:84292, May 22 2015, 08:34:51)  
[GCC 4.4.7 20120313 (Red Hat 4.4.7-15)] on linux2  
Type "help", "copyright", "credits" or "license" for more information.  
>>> 
```

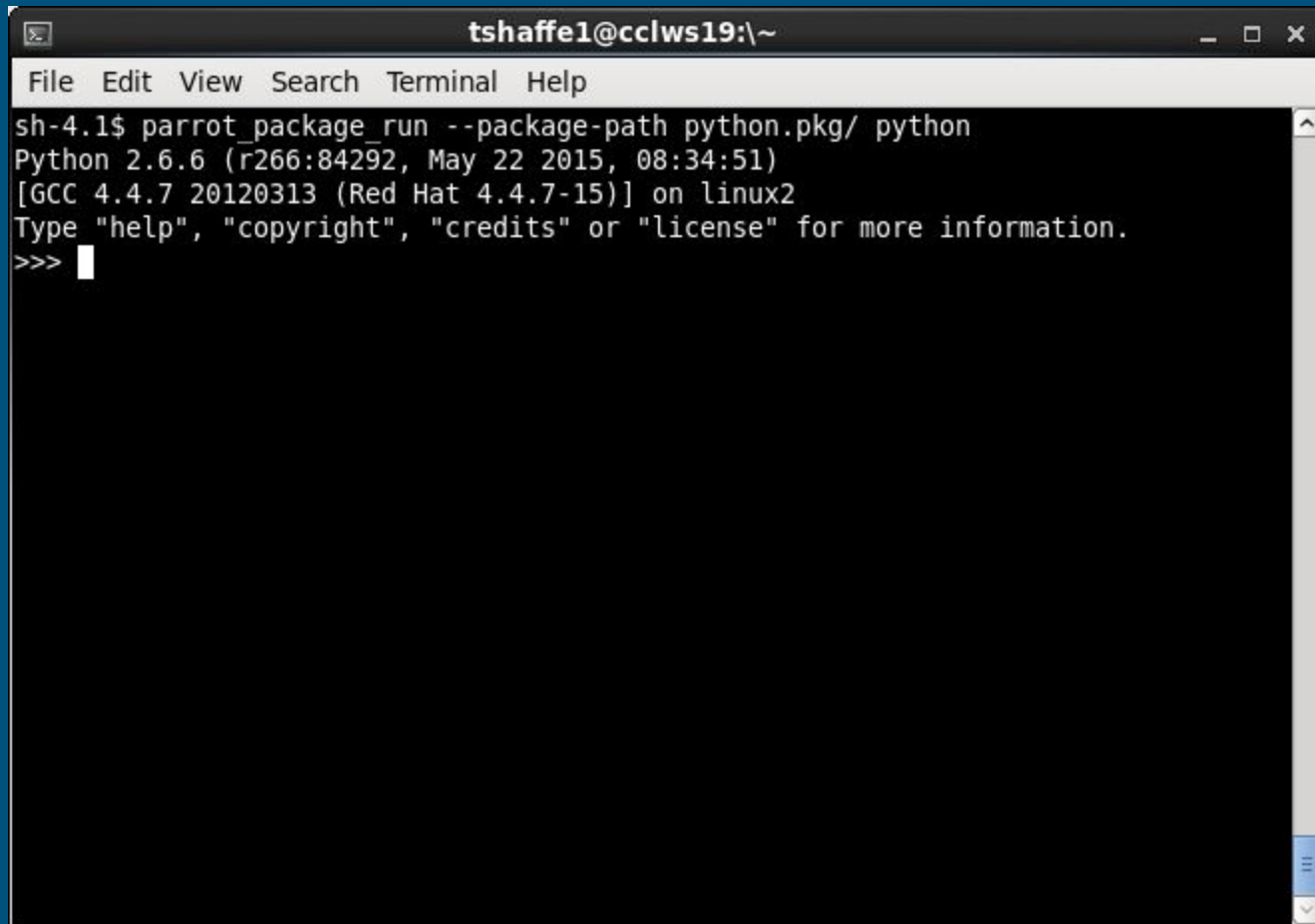



A terminal window titled "tshaffe1@cclws19:\~" with a menu bar containing "File", "Edit", "View", "Search", "Terminal", and "Help". The terminal shows the execution of the command "parrot_package_create --name-list python.namelist --env-list python.envlist --package-path python.pkg". The output indicates the packaging process has begun, shows the start time as "Tue Oct 18 14:21:11 2016", the package path as "python.pkg", the package size as "102M", and that the process has finished. The end time is also "Tue Oct 18 14:21:11 2016". The prompt "sh-4.1\$" is visible at the bottom.

```
tshaffe1@cclws19:\~
File Edit View Search Terminal Help
sh-4.1$ parrot_package_create --name-list python.namelist --env-list python.envlist --package-path python.pkg
The packaging process has begun ...
The start time is: Tue Oct 18 14:21:11 2016
Package Path: python.pkg
Package Size: 102M      python.pkg
The packaging process has finished.
The end time is: Tue Oct 18 14:21:11 2016
sh-4.1$
```


A terminal window with a dark title bar and a light menu bar. The title bar contains the text 'tshaffe1@cclws19:\~' and standard window control icons. The menu bar includes 'File', 'Edit', 'View', 'Search', 'Terminal', and 'Help'. The terminal area is black with white text. The prompt 'sh-4.1\$' is followed by the command 'parrot_package_run --package-path python.pkg/ python' and a cursor.

```
tshaffe1@cclws19:\~
File Edit View Search Terminal Help
sh-4.1$ parrot_package_run --package-path python.pkg/ python
```

A terminal window titled "tshaffe1@cclws19:\~" with a menu bar containing "File", "Edit", "View", "Search", "Terminal", and "Help". The terminal output shows a shell prompt "sh-4.1\$" followed by the command "parrot_package_run --package-path python.pkg/ python". The output of the command is the Python version and build information: "Python 2.6.6 (r266:84292, May 22 2015, 08:34:51) [GCC 4.4.7 20120313 (Red Hat 4.4.7-15)] on linux2". It then prompts the user to type "help", "copyright", "credits", or "license" for more information. The prompt ">>>" is followed by a cursor.

```
tshaffe1@cclws19:\~
File Edit View Search Terminal Help
sh-4.1$ parrot_package_run --package-path python.pkg/ python
Python 2.6.6 (r266:84292, May 22 2015, 08:34:51)
[GCC 4.4.7 20120313 (Red Hat 4.4.7-15)] on linux2
Type "help", "copyright", "credits" or "license" for more information.
>>> 
```

Remote Dependencies

Parrot can make remote resources available through the normal filesystem interface.

Rather than bundling all dependencies (which could be *far* more than needed on large projects), let Parrot fetch them on demand.

Programs see extra latency on initial access, but only retrieve the parts they actually use.

CVMFS

CernVM Filesystem (CVMFS) takes this approach to distribute experiment software.

Large, frequently updated codebase accessed daily from grid sites all over the world.

No need to explicitly install packages; just start running things, and dependencies are loaded as needed.



File System Buffers  CemVM-FS
Hard Disk Cache

 CemVM-FS "Repository"
(All Releases Available)

CVMFS on HPC

High performance computing (HPC) resources might not have an open internet connection and FUSE.

For the former, we can run an HTTP proxy on the login node.

Since Parrot supports CVMFS, just send a Parrot executable, no FUSE or setuid programs required.

CVMFS on HPC

Experiments are highly dependent on CVMFS to deliver software.

Long-running, compute-bound tasks don't suffer much performance penalty under Parrot.

With Parrot, take advantage of any worker with a working kernel, no need for cluster admins to install extra software.

Questions?

tshaffe1@nd.edu

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