

Python Usage Metrics on Blue Waters

Colin MacLean (NCSA/University of Illinois)
Email: cmaclean@illinois.edu

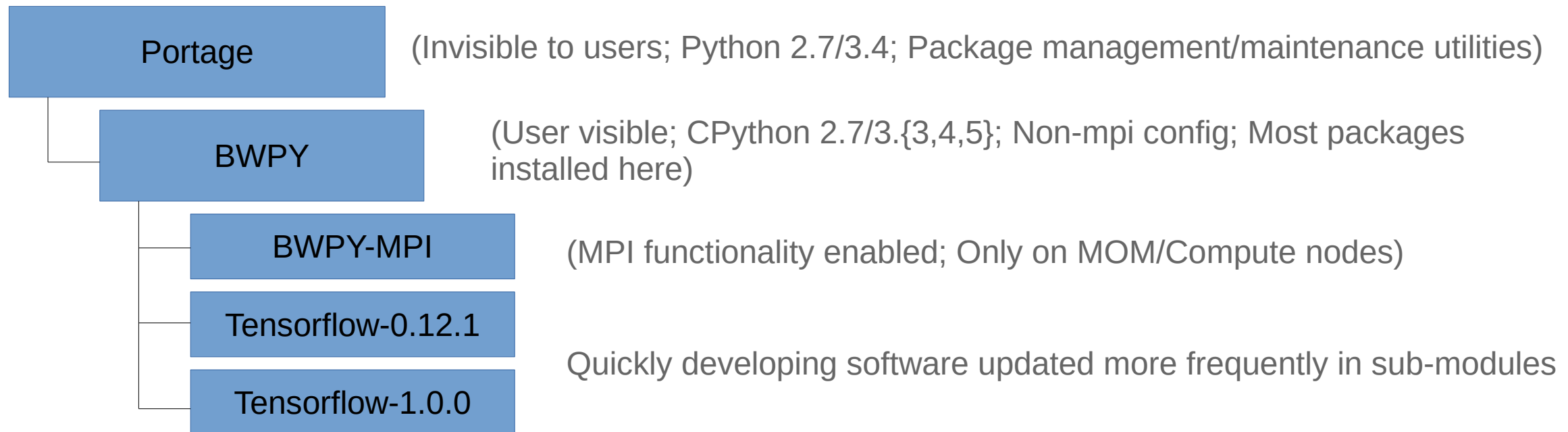


Python on Blue Waters

- Python on Blue Waters is maintained via a patched version of Gentoo Prefix
 - Over 650 packages installed (230 Python)
 - Includes non-Python dependencies
 - Uses Cray MPICH and libsci
 - Patches to Gentoo Portage enable chaining of prefixes
 - Eselect module for selecting Environment Modules
 - Gentoo Portage made aware of Environment Modules provided dependencies
 - Most ebuilds work out of the box

Use of Gentoo Prefix on Blue Waters

Blue Waters Python prefix hierarchy:



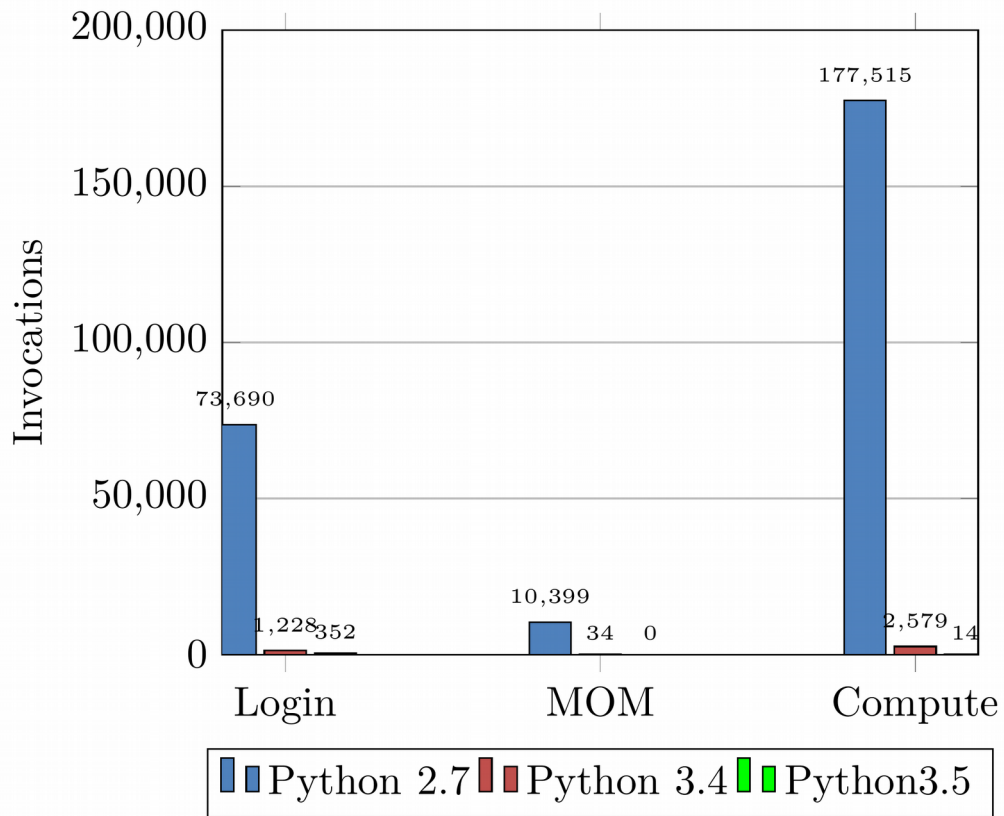
Gentoo Prefix Environment

- Gentoo Prefix management runs bash with `$EPREFIX/etc/profile`
- No `LD_LIBRARY_PATH` to prevent breaking other software
- `CPATH/LIBRARY_PATH` for correct library location ordering
- Uses `RUNPATH`
 - May use custom `ld-linux.so` in future

Module Usage Logging

- Added exit hook in sitecustomize.py
- Errors in sitecustomize.py aren't fatal
- Logs timestamp, user, host, job, number of nodes, list of modules loaded at exit
- Only sees modules loaded at exit. Exec'd code missed.
- Compute node jobs counted once per aprun
- Collection started January 13th, 2017

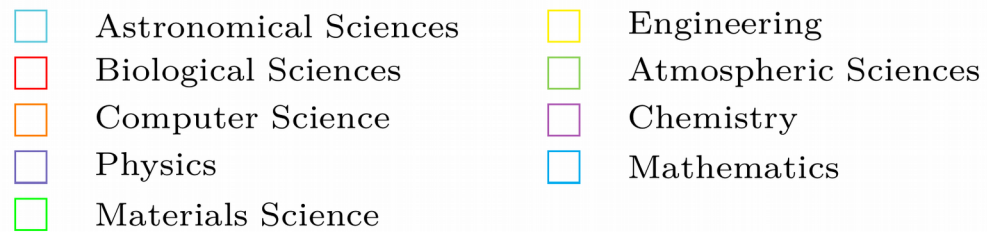
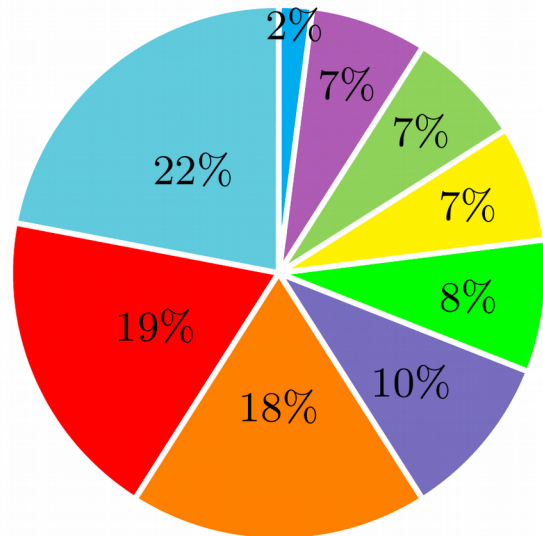
Python Versions



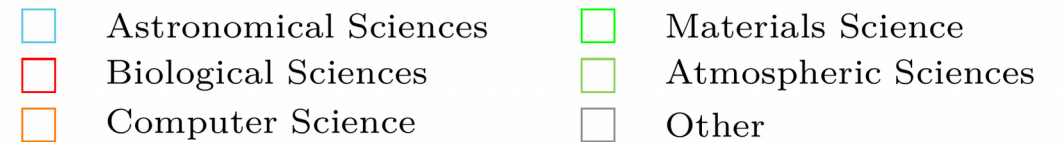
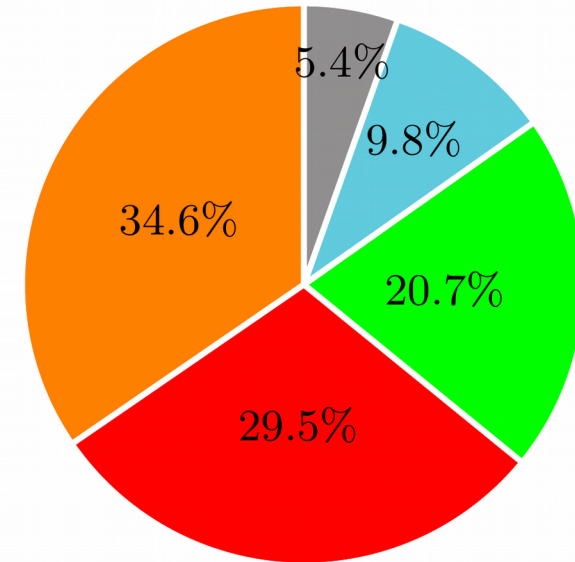
- BWPY supports CPython 2.7, 3.3, 3.4, 3.5, Pypy and Pypy3
- Python 2.7 dominates,
- Python 3.3, Pypy, and Pypy3 not used
- ~270,000 invocations

Python Users

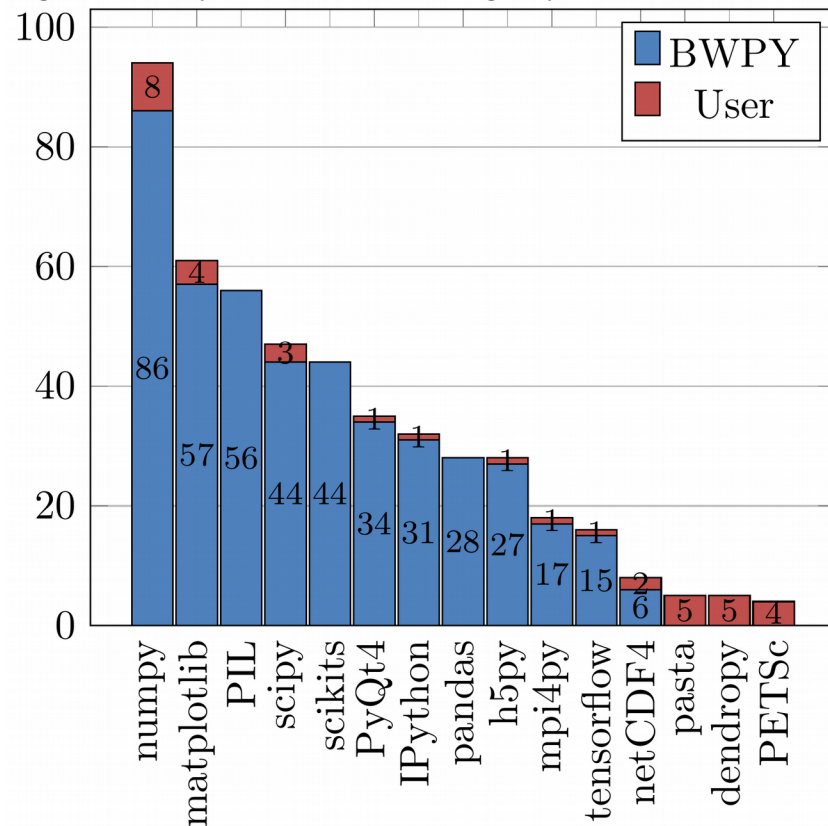
Users



Invocations

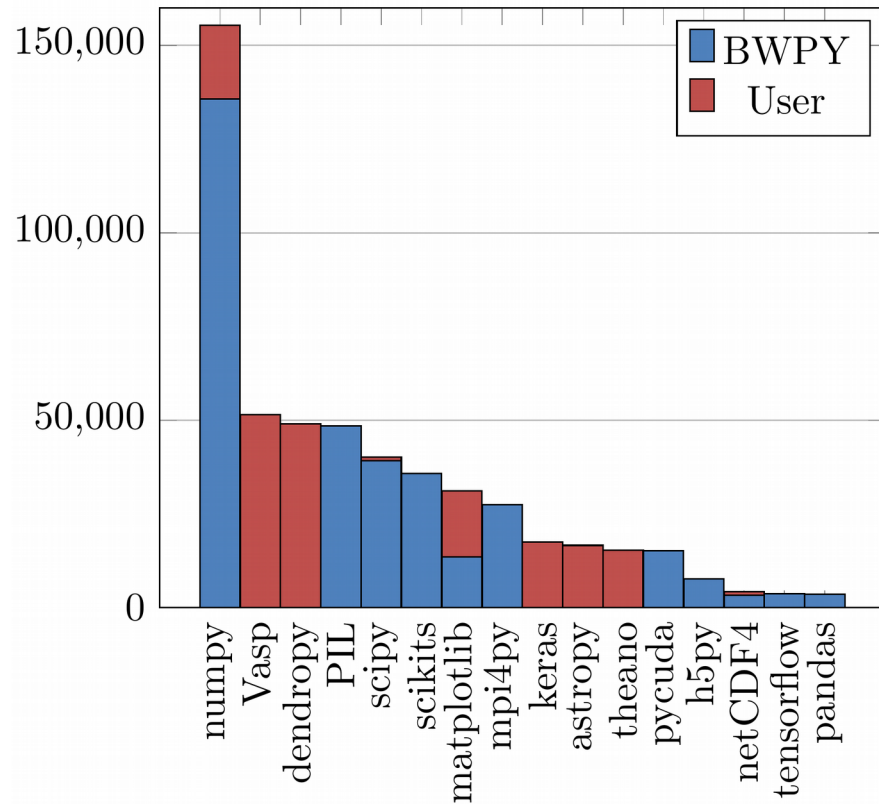


Module Usage: By User



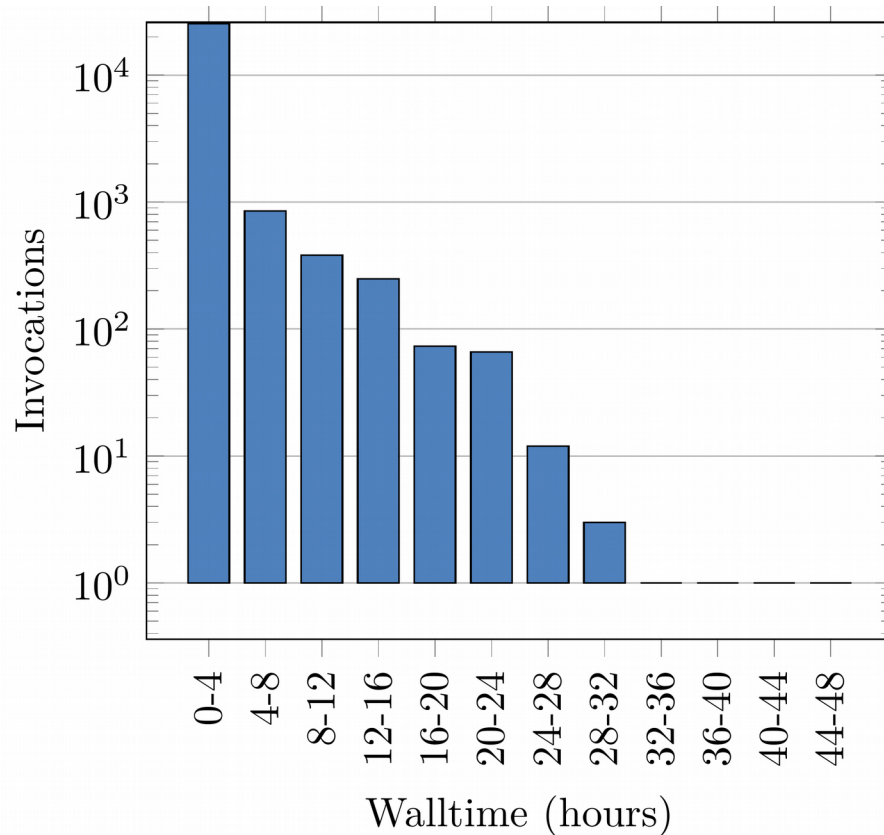
- 131 total users
- Unique module names imported: 663

Module Usage: Most Imported



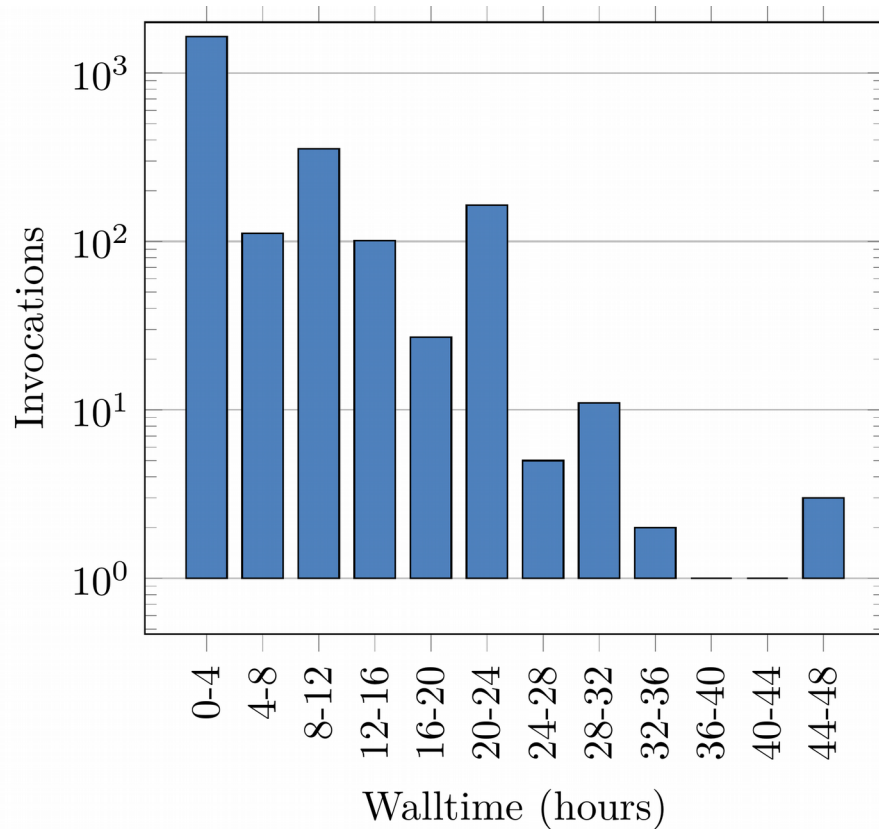
- 2,391,323 total imports
- Some user provided modules overrode BWPY modules, some weren't provided by BWPY

Compute Node Job Walltimes



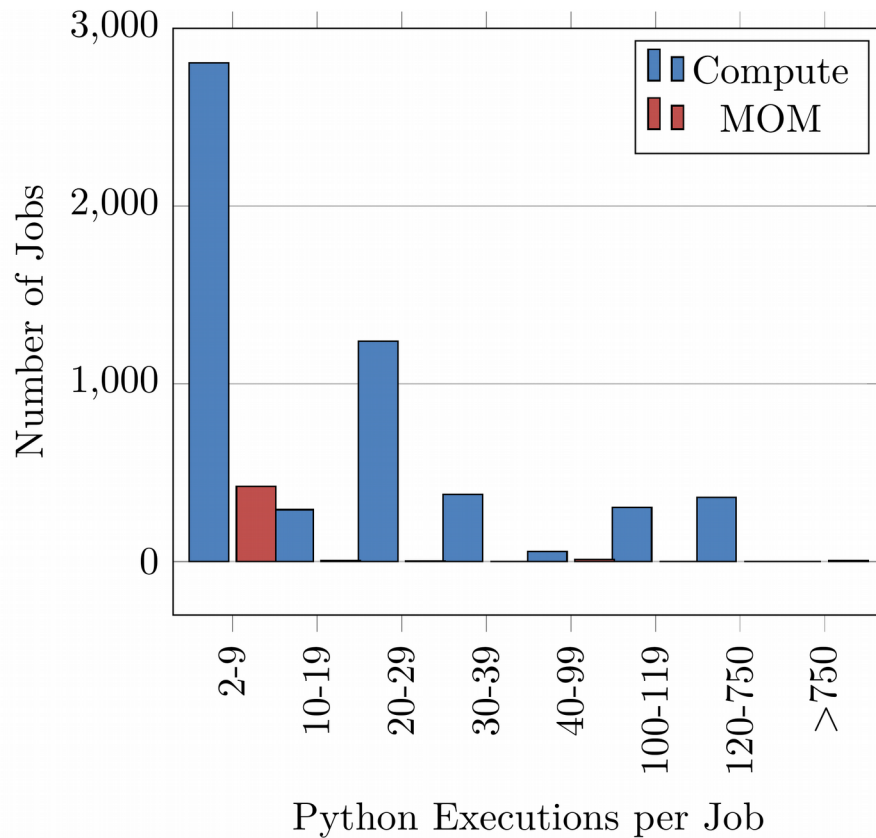
- 10893 jobs < 15 minutes
- 7004 jobs 15-30 minutes
- 6288 jobs 1/2-1 hour
- Only 162 jobs >16 hours

MOM Node Job Walltimes



- 424 jobs < 15 minutes
- 152 jobs 15-30 minutes
- 351 jobs 1/2-1 hour
- 719 jobs 1-4 hours
- 214 jobs >16 hours

Executions per Job



- Some jobs execute Python many times
- Detect and notify of this behavior?

Compute Job Sizes

# of nodes	Invocations
1	170488
2-32	8555
33-64	53
65-128	222
129-256	600
257-512	94
513-1024	71
1025-2048	25

- The vast majority of Python jobs were single node
- How does slow Python startup time affect job width?

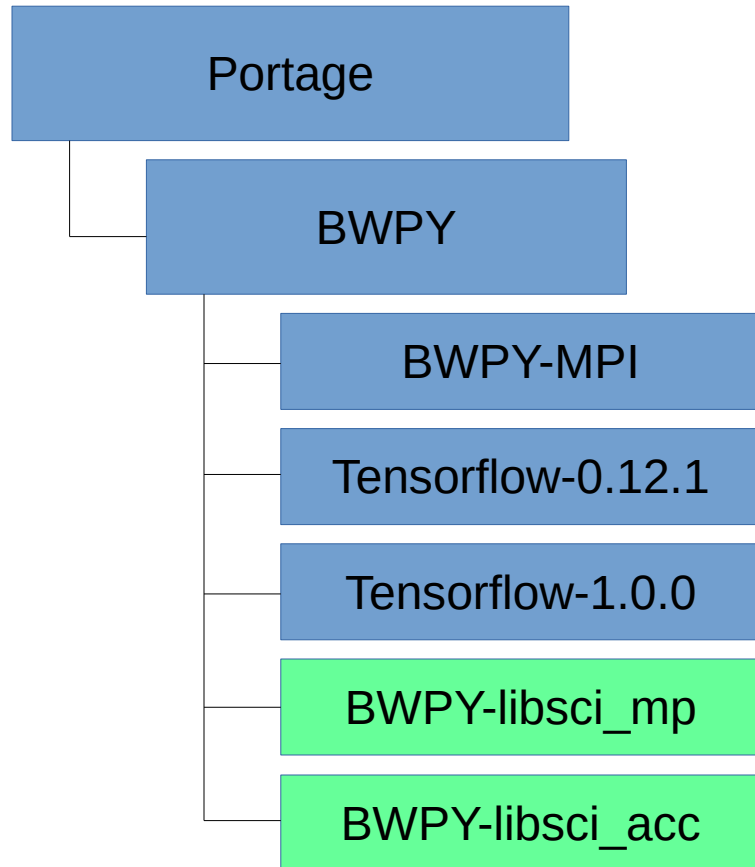
Python Usage

- Python used in many ways
 - Login node data analysis
 - MOM node job launching
 - For short compute jobs
 - Some HPC workloads
- Support for different situations
 - Separate MPI-enabled site-packages for non-Compute use

Conclusions

- Python used in many ways
 - Login node data analysis
 - MOM node job launching
 - For short compute jobs
 - Some HPC workloads
- Support for different situations
 - Separate MPI-enabled site-packages for non-Compute use

New Modules



- Adding new BLAS options due to extensive use of numpy
- May help work around CPython Global Interpreter Lock



Thank You!

