Python Cluster Documentation

Release 0.6.1

Michael Dacre <mike.dacre@gmail.com>

Jun 18, 2016

CONTENTS

Submit jobs to slurm or torque, or with multiprocessing.

Author	Michael D Dacre <mike.dacre@gmail.com>
License	MIT License, property of Stanford, use as you wish
Version	0.6.1

Allows simple job submission with *dependency tracking and queue waiting* with either torque, slurm, or locally with the multiprocessing module. It uses simple techiques to avoid overwhelming the queue and to catch bugs on the fly.

Contents:

CHAPTER

ONE

BASIC USAGE

1.1 Installation

This module will work with Python 2.7+ on Linux systems.

To install, use the standard python method:

```
git clone https://github.com/MikeDacre/python-cluster
cd python-cluster
python ./setup.py install --user
```

In general you want user level install even if you have sudo access, as most cluster environments share /home/<user> across the cluster, making this module available everywhere.

Note: While the name is *python-cluster* you will import it just as *cluster*:

```
import cluster
```

1.1.1 Prerequisites

The only external module that I use in this software is dill. It isn't 100% required but it makes function submission much more stable.

If you choose to use dill, it must be installed cluster wide.

1.1.2 Function Submission

In order to submit functions to the cluster, this module must import them on the compute node. This means that all of your python modules must be available on every compute node. To avoid pain and debugging, you can do this manually by running this on your loggin node:

```
freeze --local | grep -v '^\-e' | cut -d = -f 1 > module_list.txt
```

And then on the compute nodes:

```
cat module_list.txt | xargs pip install --user
```

This will ensure that all of your modules are installed globally.

In general it is a good idea to install modules as *–user* with pip to avoid this issue.

1.2 Simple Usage

1.2.1 Setting Environment

To set the environement, set queue.MODE to one of ['torque', 'slurm', 'local'], or run get_cluster_environment().

1.2.2 Simple Job Submission

At its simplest, this module can be used by just executing submit(<command>), where command is a function or system command/shell script. The module will autodetect the cluster, generate an intuitive name, run the job, and write all outputs to files in the current directory. These can be cleaned with clean_dir().

To run with dependency tracking, run:

```
import cluster
job  = cluster.submit(<command1>)
job2 = cluster.submit(<command2>, dependencies=job1)
exitcode, stdout, stderr = job2.get()  # Will block until job completes
```

1.2.3 Functions

The submit function works well with python functions as well as with shell scripts and shell commands.

However, in order for this to work, *cluster* ends up importing your original script file on the nodes. This means that all code in your file will be executed, so anything that isn't a function or class must be protected with an:

```
if __name__ == '__main__':
```

protecting statment.

If you do not do this you can end up with multi-submission and infinate recursion, which could mess up your jobs or just crash the job, but either way, it won't be good.

1.2.4 File Submission

If you want to just submit a file, that can be done like this:

```
from cluster import submit_file
submit_file('/path/to/script', dependencies=[7, 9])
```

This will return the job number and will enter the job into the queue as dependant on jobs 007 and 009. The dependencies can be omitted.

1.3 Queue Management

This module provides simple queue management functions

To generate a queue object, do the following:

```
import cluster
q = cluster.Queue(user='self')
```

This will give you a simple queue object containg a list of jobs that belong to you. If you do not provide user, all jobs are included for all users. You can provide *qtype* to explicitly force the queue object to contain jobs from one queing system (e.g. local or torque).

To get a dictionary of all jobs, running jobs, queued jobs, and complete jobs, use:

```
q.jobs
q.running
q.complete
q.queued
```

Every job has a number of attributes, including owner, nodes, cores, memory.

1.4 Advanced Usage

1.4.1 Keyword Arguments

To make submission easier, this module defines a number of keyword arguments in the options.py file that can be used for all submission and Job() functions. These include things like 'cores' and 'nodes' and 'mem'.

The following is a complete list of arguments that can be used in this version:

```
Used in every mode::
cores:      Number of cores to use for the job
            Type: int; Default: 1
modules:    Modules to load with the `module load` command
            Type: list; Default: None
filedir:    Folder to write cluster files to, must be accessible to the compute
            nodes.
            Type: str; Default: .
dir:        The working directory for the job
            Type: str; Default: path argument
suffix:     A suffix to append to job files (e.g. job.suffix.qsub)
            Type: str; Default: cluster
outfile:    File to write STDOUT to
            Type: str; Default: None
errfile:    File to write STDERR to
            Type: str; Default: None

Used for function calls::
imports:    Imports to be used in function calls (e.g. sys, os) if not provided,
            defaults to all current imports, which may not work if you use complex
            imports. The list can include the import call, or just be a name, e.g.
            ['from os import path', 'sys']
            Type: list; Default: None

Used only in local mode::
threads:    Number of threads to use on the local machine
            Type: int; Default: 8

Options that work in both slurm and torque::
nodes:      Number of nodes to request
            Type: int; Default: 1
features:   A comma-separated list of node features to require
            Type: list; Default: None
time:       Walltime in HH:MM:SS
```

```
            Type: str; Default: 12:00:00
mem:        Memory to use in MB (e.g. 4000)
            Type: ['int', 'str']; Default: 4000
partition:  The partition/queue to run in (e.g. local/batch)
            Type: str; Default: None
account:    Account to be charged
            Type: str; Default: None
export:     Comma separated list of environmental variables to export
            Type: str; Default: None

Used for slurm only::
begin:      Start after this much time
            Type: str; Default: None
```

Note: Type is enforced, any provided argument must match that python type (automatic conversion is attempted), the default is just a recommendation and is not currently used. These arguments are passed like regular arguments to the submission and Job() functions, eg:

```
Job(nodes=1, cores=4, mem='20MB')
```

This will be interpretted correctly on any system. If torque or slurm are not available, any cluster arguments will be ignored. The module will attempt to honor the cores request, but if it exceeds the maximum number of cores on the local machine, then the request will be trimmed accordingly (i.e. a 50 core request will become 8 cores on an 8 core machine).

Adding your own keywords

There are many more options available for torque and slurm, to add your own, edit the options.py file, and look for CLUSTER_OPTS (or TORQUE/SLURM if your keyword option is only availble on one system). Add your option using the same format as is present in that file. The format is:

```
('name', {'slurm': '--option-str={}', 'torque': '--torque-option={}',
          'help': 'This is an option!', 'type': str, 'default': None})
```

You can also add list options, but they must include 'sjoin' and 'tjoin' keys to define how to merge the list for slurm and torque, or you must write custom option handling code in `cluster.options.options_to_string()`. For an excellent example of both approaches included in a single option, see the 'features' keyword above.

I happily accept pull requests for new option additions (any any other improvements for that matter).

1.4.2 Profiles and the Config File

To avoid having to enter all keyword arguments every time, profiles can be used. These profiles can store any of the above keywords and drastically simplify submission. For example:

```
job = submit(my_function, profile='large')
```

Instead of:

```
job = submit(my_funtion, nodes=2, cores=16, mem='64GB', partition='bigjobs',
             features=['highmem'], export='PYTHONPATH')
```

These profiles are saved in a config file at ~/.python-cluster and can be editted in that file directly, or using the below functions. To edit them in the file directly, you must make sure that the section is labelled 'prof_<name>' where <name> is whatever you want it to be called. e.g.:

```
[prof_default]
nodes = 1
cores = 16
time = 24:00:00
mem = 32000
```

Note: a default profile must always exist, it will be added back if it does not exist.

Alternatively, the functions `cluster.config_file.set_profile()` and `cluster.config_file.get_profile()` can be used:

```
cluster.config_file.set_profile('small', {'nodes': 1, 'cores': 1,
                                          'mem': '2GB'})
cluster.config_file.get_profile('small')
```

To see all profiles run:

```
config_file.get_profile()
```

Other options are defined in the config file, including the maximum number of jobs in the queue, the time to sleep between submissions, and other options. To see these run:

```
cluster.config_file.get()
```

You can set options with:

```
cluster.config_file.set()
```

The defaults can be directly edited in `config_file.py` , they are clearly documented.

1.4.3 Job Files

All jobs write out a job file before submission, even though this is not necessary (or useful) with multiprocessing. In local mode, this is a *.cluster* file, in slurm is is a *.cluster.sbatch* and a *.cluster.script* file, in torque it is a *.cluster.qsub* file. 'cluster' is set by the suffix keyword, and can be overridden.

To change the directory these files are written to, use the 'filedir' keyword argument to Job or submit.

NOTE: This directory *must* be accessible to the compute nodes!!!

All jobs are assigned a name that is used to generate the output files, including STDOUT and STDERR files. The default name for the out files is STDOUT: name.cluster.out and STDERR: name.cluster.err. These can be overwridden with keyword arguments.

All Job objects have a `clean()` method that will delete any left over files. In addition there is a clean_job_files script that will delete all files made by this package in any given directory. Be very careful with the script though, it can clobber a lot of work all at once if it is used wrong.

1.4.4 Dependecy Tracking

Dependency tracking is supported in all modes. Local mode uses a unique queueing system that works similarly to torque and slurm and which is defined in jobqueue.py.

To use dependency tracking in any mode pass a list of job ids to submit or submit_file with the *dependencies* keyword argument.

1.4.5 Logging

I use a custion logging script called logme to log errors. To get verbose output, set logme.MIN_LEVEL to 'debug'. To reduce output, set logme.MIN_LEVEL to 'warn'.

1.5 Code Overview

There are two important classes for interaction with the batch system: Job and Queue. The essential flow of a job submission is:

```
job = Job(command/function, arguments, name)
job.write()  # Writes the job submission files
job.submit() # Submits the job
job.wait()   # Waits for the job to complete
job.stdout   # Prints the output from the job
job.clean()  # Delete all of the files written
```

You can also wait for many jobs with the Queue class:

```
q = Queue(user='self')
q.wait([job1, job2])
```

The jobs in this case can be either a Job class or a job number.

1.6 Issues and Contributing

If you have any trouble with this software add an issue in https://github.com/MikeDacre/python-cluster/issues

If you want to help improve it, please fork the repo and send me pull requests when you are done.

CHAPTER

TWO

SCRIPTS

This package contains a few little helper scripts to make your life easier. These are not required in order to use the cluster library.

2.1 clean_job_files

```
usage: clean_job_files [-h] [-d DIR] [-s SUFFIX] [-q {torque,slurm,local}]
                       [-n] [-v VERBOSE]

Clean all intermediate files created by the cluster module from this dir.

================================================================================

        AUTHOR: Michael D Dacre, mike.dacre@gmail.com
  ORGANIZATION: Stanford University
       LICENSE: MIT License, property of Stanford, use as you wish
       CREATED: 2016-34-15 15:06
 Last modified: 2016-06-16 10:42

   DESCRIPTION: Uses the cluster.job.clean_dir() function

       CAUTION: The clean() function will delete **EVERY** file with
                extensions matching those these::
                    .<suffix>.err
                    .<suffix>.out
                    .<suffix>.sbatch & .cluster.script for slurm mode
                    .<suffix>.qsub for torque mode
                    .<suffix> for local mode
                    _func.<suffix>.py
                    _func.<suffix>.py.pickle.in
                    _func.<suffix>.py.pickle.out

================================================================================

optional arguments:
  -h, --help            show this help message and exit
  -d DIR, --dir DIR     Directory to clean
  -s SUFFIX, --suffix SUFFIX
                        Directory to clean
  -q {torque,slurm,local}, --qtype {torque,slurm,local}
                        Limit deletions to this qtype
  -n, --no-confirm      Do not confirm before deleting (for scripts)
```

```
  -v VERBOSE, --verbose VERBOSE
                        Show debug information
```

- search

CHAPTER

THREE

API DOCUMENTATION

The following documentation is primarily built from the docstrings of the actual source code and can be considered an API reference.

3.1 Queueing

The most import thing is the *Queue()* class which does most of the queue mangement. In addition, *get_cluster_environment()* attempts to autodetect the cluster type (torque, slurm, normal) and sets the global cluster type for the whole file. Finally, the *wait()* function accepts a list of jobs and will block until those jobs are complete.

3.2 Job Management

Job management is handeled by the *Job()* class, full instructions on using this class are above, in particular review the 'Keyword Arguments' section above.

The methods of this class are exposed by a few functions that aim to make job submission easier. The foremost of these is *submit()* which can take as little as a single command and execute it. *make_job()* and *make_job_file()* work similarly but just return a Job object, or write the file and then return the Job object respectively. *clean()* takes a list of Job objects and runs their internal *clean()* methods, deleting all written files.

There are two additional functions that are completely independent of the Job object: *submit_file()* and *clean_dir()*. *submit_file()* uses similar methods to the Job class to submit a job to the cluster, but it does not involve the job class at all, instead just submitting an already created job file. It can do dependency tracking in the same way as a job file, but that is all. *clean_dir()* uses the file naming convention established in the Job class (and defined separately here) to delete all files in a directory that look like they could be made by this module. It has an autoconfirm feature that can be activated to avoid accidental clobbering.

3.3 Options

All keyword arguments are defined in dictionaries in the *options.py* file, alongside function to manage those dictionaries. Of particular importance is *option_help()*, which can display all of the keyword arguments as a string or a table. *check_arguments()* checks a dictionary to make sure that the arguments are allowed (i.e. definied), it is called on all keyword arguments in the package.

The way that option handling works in general, is that all hardcoded keyword arguments must contain a dictionary entry for 'torque' and 'slurm', as well as a type declaration. If the type is NoneType, then the option is assumed to be a boolean option. If it has a type though, *check_argument()* attmepts to cast the type and specific idiosyncracies are handled in this step, e.g. memory is converted into an integer of MB. Once the arguments are sanitized *format()* is

called on the string held in either the 'torque' or the 'slurm' values, and the formatted string is then used as an option. If the type is a list/tuple, the 'sjoin' and 'tjoin' dictionary keys must exist, and are used to handle joining.

The following two functions are used to manage this formatting step.

option_to_string() will take an option/value pair and return an appropriate string that can be used in the current queue mode. If the option is not implemented in the current mode, a debug message is printed to the console and an empty string is returned.

options_to_string() is a wrapper around *option_to_string()* and can handle a whole dictionary of arguments, it explicitly handle arguments that cannot be managed using a simple string format.

3.4 Config File

Profiles are combinations of keyword arguments that can be called in any of the submission functions. They are handled in the *config_file.py* file which just adds an abstraction layer on top of the builtin python ConfigParser script.

The config file also contains other options that can be managed with the *get()* and *set()* functions. Profiles are wrapped in a *Profile()* class to make attribute access easy, but they are fundamentally just dictionaries of keyword arguments. They can be created with *cluster.config_file.Profile({kewywds})* and then written to a file with that class' *write()* method. The easiest way to interact with profiles is with the *get_profile()* and *set_profile()* functions. These make it very easy to go from a dictionary of keywords to a profile.

Profiles can then be called with the *profile=* keyword in any submission function or Job class.

3.5 Local Queue Implementation

The local queue implementation is based on the multiprocessing library and is not intended to be used directly, it should always be used via the Job class because it is somewhat tempramental. The essential idea behind it is that we can have one JobQueue class that is bound to the parent process, it exclusively manages a single child thread that runs the *job_runner()* function. The two process communicate using a *multiprocessing.Queue* object, and pass *cluster.jobqueue.Job* objects back and forth between them.

The Job objects (different from the Job objects in *job.py*) contain information about the task to run, including the number of cores required. The job runner manages a pool of *multiprocessing.Pool* tasks directly, and keeps the total running cores below the total allowed (default is the system max, can be set with the threads keyword). It backfills smaller jobs and holds on to larger jobs until there is enough space free.

This is close to what torque and slurm do, but vastly more crude. It serves as a stopgap to allow parallel software written for compute clusters to run on a single machine in a similar fashion, without the need for a pipeline alteration. The reason I have reimplemented a process pool is that I need dependency tracking and I need to allow some processes to run on multiple cores (e.g. 6 of the available 24 on the machine).

The *job_runner()* and *Job* objects should never be accessed except by the JobQueue. Only one JobQueue should run at a time (not enforced), and by default it is bound to *cluster.jobqueue.JQUEUE*. That is the interface used by all other parts of this package.

3.6 Logme

This is a package I wrote myself and keep using because I like it. It provides syslog style leveled logging (e.g. 'debug'->'info'->'warn'->'error'->'critical') and it implements colors and timestamped messages.

The minimum print level can be set module wide at runtime by changing *cluster.logme.MIN_LEVEL*.

3.7 Other Functions

Some other wrapper functions are defined in *run.py*, these are just little useful knick-knacks that make function submission and queue management possible.

3.8 Indices and tables

- genindex
- search
- modindex
- search