



Submit HTCondor jobs using Python

John (TJ) Knoeller
HTCondor Week UK 2018

Overview

- › The old
- › The new
- › The future
- › Show and tell?

Multiple submission APIs

- › Several python APIs for submitting jobs
 - Schedd.submit
 - Schedd.submitMany
 - Submit.queue
 - Submit.queue_with_itemdata
- › Python helper modules (coming soon)
 - ezsubmit, htmap

First there was Schedd.submit

- › Schedd.submit, Schedd.submitMany
 - Lowest level interface, the Schedd "native" API
 - Accepts only job classads
 - First version of python bindings had **only** this API
 - Hard to use correctly
 - Minimal error checking
 - easy to submit incomplete, incorrect or invalid jobs
 - Makes it hard for us to improve the native API

Next came Submit.queue

- › Submit.queue(*transaction*, *num_jobs*)
 - A "condor_submit" level interface
 - Accepts of dict of submit language key:value pairs
 - Can do "QUEUE 10" but not "QUEUE matching *.dat"
 - Added in 8.6
 - Submit **files** are still awkward to use from python

Submit example (works in 8.6)

```
sub = htcondor.Submit({
    'executable' : '/bin/echo',
    'arguments' : 'hello from job $(ClusterId).$(ProcId)'
})
num = 10
schedd = htcondor.Schedd() # get the local schedd
with schedd.transaction() as txn :
    cluster_id = sub.queue(txn, num)
print "Submitted %d jobs to cluster %d" % (num, cluster_id)
Submitted 10 jobs to cluster 23
```

Important Safety tip

```
# Don't try to do this. It may crash with Python 3
# Because Python 3 may delete the Schedd() object
# before executing sub.queue(). This is HTCondor bug #6721
with htcondor.Schedd().transaction() as txn :
    cluster_id = sub.queue(txn, num)

# this works (preferred if you use schedd more than once)
schedd = htcondor.Schedd()
with schedd.transaction() as txn :
    cluster_id = sub.queue(txn, num)
```

Submit class today

- › Submit class can be initialized from either
 - a dict of submit language key:value pairs
 - a string (the contents of a submit file)
 - Queue statement is optional (only 1 allowed)
- › `Submit.getQArgs()` / `Submit.setQArgs()`
 - get and set the queue arguments for `Submit.queue`
- › `Submit.queue_with_itemdata()`
 - queue from python iterator

Submit example (works in 8.7)

```
sub = htcondor.Submit("""
    executable = /bin/echo
    arguments = hello from job $(ClusterId).$(ProcId)
    queue 10
""")
with schedd.transaction() as txn :
    cluster_id = sub.queue(txn)
num = sub.getQArgs()
print "Submitted %s jobs to cluster %d" % (num, cluster_id)
Submitted 10 jobs into cluster 23
```

Submit example2 (works in 8.7)

```
with open("filename.sub") as fh :
    subtext = sub.read()

sub = htcondor.Submit(subtext)

with schedd.transaction() as txn :
    res = sub.queue_with_itemdata(txn,1,iter(sub.itemdata()))
print(res)
Submitted 10 jobs into cluster 23.0 :
Args = "Hello to you"
BufferBlockSize = 32768
...
```

queue vs queue_with_itemdata

- › Submit.queue
 - Accepts a count
 - Uses QArgs if no count provided
 - returns a clusterId
- › Submit.queue_with_itemdata
 - Accepts a count or item iterator or both
 - Use Submit.itemdata() as iterator or provide your own
 - Returns a SubmitResult class

SubmitResult is cluster and ClassAd

```
res = sub.queue_with_itemdata(txn,5,iter(sub.itemdata()))
print "cluster %d has %d procs" % (res.cluster(), res.num_procs())
cluster 23 has 10 procs

clusterad = res.clusterad()
print "job log is %s" % clusterad['Log']
job log is /home/alice/joblogs/23.log

# follow the log to wait for the job(s) to start/finish/etc
jel = htcondor.JobEventLog(clusterad['Log'])
for event in jel.follow(1000)
```

The future

- › Pure python helper modules
 - ezsubmit.py
 - htmap.py

ezsubmit.py

- › Simple helper class that submits a set of jobs and returns a handle that can be used to
 - Check job status
 - Wait for completion
 - Hold/Release/Remove the jobs

htmap.py

- › Turns a python map of a function into a set of HTCCondor jobs

```
@htmap.htmap
def double(z):
    return 2*z
```

```
job = double.map('double', [5, 'foo', 10, 'bar'])
print ("=== results as they complete ===")
for result in job: print(result)
```

Demo time...

HT
CENTER FOR
HIGH THROUGHPUT
COMPUTING

HTCCondor



Any Questions?