Science & Technology Facilities Council

Grid job submission using HTCondor

Andrew Lahiff

Overview

- Introduction
- Examples
  - Direct submission
  - Job Router
  - Match making

Introduction

- You don’t need EMI WMS or DIRAC to submit jobs to the grid
- HTCondor can submit to:
  - CEs (ARC, CREAM, UNICORE, Globus, …)
  - Batch systems (HTCondor, PBS, LSF, SGE)
  - Clouds (EC2, Google Compute Engine)
  - BOINC
- Grid manager daemon manages grid jobs
  - Started automatically by the schedd, one instance per user
  - Grid ASCII Helper Protocol (GAHP) server
    - Runs under the grid manager
    - Communicates with the remote service

Introduction

- This is not new
  - People were doing this 10 years ago
  - Europe moved on to Torque / Maui, LCG CE, glite-WMS, …
  - ATLAS & CMS pilot submission done using HTCondor-G
- Features
  - Input & output files can be transferred as necessary
  - Can monitor the job using standard HTCondor commands, e.g. condor_q
  - Can use MyProxy for proxy renewal
  - Very configurable
    - Polices for dealing with failures, retries, etc
    - Resubmission of jobs to other sites under specified conditions, e.g. idle for too long

Direct submission

Science & Technology
Facilities Council

Direct submission

- Simplest method – specify directly the CE you want your jobs to run on
- Works with the “out-of-the-box” HTCondor configuration
  - i.e. after “yum install condor; service condor start”

Example 1

```
-bash-4.1$ cat grid3.sub
universe = grid
grid_resource = nordugrid t2arc01.physics.ox.ac.uk
executable = test.sh
Log=log.$(Cluster).$(Process)
Output=out.$(Cluster).$(Process)
Error=err.$(Cluster).$(Process)
should_transfer_files = YES
when_to_transfer_output = ON_EXIT
x509userproxy=/tmp/x509up_u33133
queue 1


-bash-4.1$ condor_submit grid3.sub
Submitting job(s).
1 job(s) submitted to cluster 16240.
```

Example 1

- Check job status

```
-bash-4.1$ condor_q


-- Submitter: lcgui03.gridpp.rl.ac.uk : <130.246.180.41:45033> : lcgui03.gridpp.rl.ac.uk
 ID      OWNER            SUBMITTED     RUN_TIME ST PRI SIZE CMD
16240.0   alahiff         5/30 15:42   0+00:00:00 I  0   0.0  test.sh

1 jobs; 0 completed, 0 removed, 1 idle, 0 running, 0 held, 0 suspended




-bash-4.1$ condor_q -grid


-- Submitter: lcgui03.gridpp.rl.ac.uk : <130.246.180.41:45033> : lcgui03.gridpp.rl.ac.uk
 ID       OWNER          STATUS     GRID->MANAGER    HOST       GRID_JOB_ID
16240.0   alahiff        INLRMS:R   nordugrid->[?] t2arc01.phys N84MDm8BdAknD0VB
```

Science & Technology Facilities Council

Example 1

- Log file

```
-bash-4.1$ cat log.16240.0
000 (16240.000.000) 05/30 15:42:27 Job submitted from host: <130.246.180.41:45033>
...
027 (16240.000.000) 05/30 15:42:36 Job submitted to grid resource
    GridResource: nordugrid t2arc01.physics.ox.ac.uk
    GridJobId: nordugrid t2arc01.physics.ox.ac.uk
N84MDm8BdAknD0VBFmzXO77mABFKDmABFKDmpmMKDmABFKDmBX8aZn
...
001 (16240.000.000) 05/30 15:57:35 Job executing on host: nordugrid t2arc01.physics.ox.ac.uk
…
```

Example 2

```
-bash-4.1$ cat grid3.sub
universe = grid
executable = test.sh
Log=log.$(Cluster).$(Process)
Output=out.$(Cluster).$(Process)
Error=err.$(Cluster).$(Process)
should_transfer_files = YES
when_to_transfer_output = ON_EXIT
x509userproxy=/tmp/x509up_u33133

grid_resource = nordugrid arc-ce01.gridpp.rl.ac.uk
queue 1
grid_resource = nordugrid heplnv146.pp.rl.ac.uk
queue 1
grid_resource = nordugrid t2arc01.physics.ox.ac.uk
queue 1
grid_resource = cream https://cccreamceli05.in2p3.fr:8443/ce-
cream/services/CREAM2 sge long
queue 1
```

Example 2

```
-bash-4.1$ condor_submit grid3.sub
Submitting job(s)....
4 job(s) submitted to cluster 16241.


-bash-4.1$ condor_q -grid


-- Submitter: lcgui03.gridpp.rl.ac.uk : <130.246.180.41:45033> : lcgui03.gridpp.rl.ac.uk
 ID       OWNER          STATUS     GRID->MANAGER    HOST       GRID_JOB_ID
16241.0   alahiff        INLRMS:Q   nordugrid->[?] arc-ce01.gri rLvKDmUsfAknCIXD
16241.1   alahiff        INLRMS:R   nordugrid->[?] heplnv146.pp uLXMDmQsfAknOGRD
16241.2   alahiff        INLRMS:R   nordugrid->[?] t2arc01.phys piXMDmQsfAknD0VB
16241.3   alahiff        REALLY-RUN cream->sge/long cccreamceli CREAM723043637
```

Science & Technology Facilities Council

Example 2

```
-bash-4.1$ condor_status -grid
Name                                NumJobs Allowed  Wanted RunningJobs IdleJobs

cream https://cccreamceli05.in2p3.        2       0       0           0        2
nordugrid arc-ce01.gridpp.rl.ac.uk        2       0       0           0        2
nordugrid heplnv146.pp.rl.ac.uk           2       0       0           0        2
nordugrid t2arc01.physics.ox.ac.uk        2       0       0           0        2
```

Science & Technology Facilities Council

Job router

Science & Technology
Facilities Council

Job router

- Optional daemon
  - Transforms jobs from one type to another according to a configurable policy
  - E.g. can transform vanilla jobs to grid jobs
- Configuration
  - Limits on jobs & idle jobs per CE
  - What to do with held jobs
  - When to kill jobs
  - How to detect failed jobs
  - Black-hole throttling
  - Can specify arbitrary script which updates routes
    - E.g. create routes for all CMS CEs from the BDII

Job router

- Most basic configuration: just a list of CEs

```
JOB_ROUTER_ENTRIES = \
   [ GridResource = "nordugrid arc-ce01.gridpp.rl.ac.uk"; \
     name = "arc-ce01.gridpp.rl.ac.uk"; ] \
   [ GridResource = "nordugrid arc-ce02.gridpp.rl.ac.uk"; \
     name = "arc-ce02.gridpp.rl.ac.uk"; ] \
   [ GridResource = "nordugrid arc-ce03.gridpp.rl.ac.uk"; \
     name = "arc-ce03.gridpp.rl.ac.uk"; ] \
   [ GridResource = "nordugrid arc-ce04.gridpp.rl.ac.uk"; \
     name = "arc-ce04.gridpp.rl.ac.uk"; ]
```

Job router

- Example – submit 20 jobs
  - Checking status: summary by route:

```
-bash-4.1$ condor_router_q -S
   JOBS ST Route                 GridResource
      5  I arc-ce01.gridpp.rl.ac.uk nordugrid arc-ce01.gridpp.rl.ac.uk
      1  R arc-ce02.gridpp.rl.ac.uk nordugrid arc-ce02.gridpp.rl.ac.uk
      4  I arc-ce02.gridpp.rl.ac.uk nordugrid arc-ce02.gridpp.rl.ac.uk
      1  R arc-ce03.gridpp.rl.ac.uk nordugrid arc-ce03.gridpp.rl.ac.uk
      4  I arc-ce03.gridpp.rl.ac.uk nordugrid arc-ce03.gridpp.rl.ac.uk
      1  R arc-ce04.gridpp.rl.ac.uk nordugrid arc-ce04.gridpp.rl.ac.uk
      4  I arc-ce04.gridpp.rl.ac.uk nordugrid arc-ce04.gridpp.rl.ac.uk
```

Job router

- Checking status: individual jobs

```
-bash-4.1$ condor_q -grid


-- Submitter: lcgvm21.gridpp.rl.ac.uk : <130.246.181.102:48749> :
lcgvm21.gridpp.rl.ac.uk
 ID      OWNER          STATUS     GRID->MANAGER    HOST         GRID_JOB_ID
2130.0   alahiff        IDLE       nordugrid->[?] arc-ce03.gri HL9NDmGVt9jnzEJD
2131.0   alahiff        PREPARING  nordugrid->[?] arc-ce04.gri 18kLDmMVt9jnE6QD
2132.0   alahiff        INLRMS:Q   nordugrid->[?] arc-ce01.gri 5VKKDmHVt9jnCIXD
2133.0   alahiff        IDLE       nordugrid->[?] arc-ce02.gri oKSLDmHVt9jnc1XD
2134.0   alahiff        IDLE       nordugrid->[?] arc-ce03.gri gpeKDmHVt9jnzEJD
2135.0   alahiff        PREPARING  nordugrid->[?] arc-ce04.gri 7iFMDmMVt9jnE6QD
2136.0   alahiff        INLRMS:Q   nordugrid->[?] arc-ce01.gri 5fmKDmHVt9jnCIXD
2137.0   alahiff        PREPARING  nordugrid->[?] arc-ce04.gri KTeMDmMVt9jnE6QD
2138.0   alahiff        IDLE       nordugrid->[?] arc-ce02.gri 7QBLDmHVt9jnc1XD
2139.0   alahiff        IDLE       nordugrid->[?] arc-ce03.gri sZPKDmHVt9jnzEJD
…
```

Science & Technology Facilities Council

Match making

Science & Technology
Facilities Council

Match making

- Add ClassAds corresponding to grid resources
  - Jobs can then be matched to them, just like jobs are matched to worker nodes
- Could use simple cron to update from BDII
  - Not everything from BDII (!), just the basic essentials
    - GlueCEUniqueID, GlueHostMainMemoryRAMSize, …

Match making

- Can run queries to see what CEs exist

```
$ condor_status -constraint 'regexp("ox",GlueCEUniqueID) == True' -autoformat GlueCEUniqueID
t2arc01.physics.ox.ac.uk:2811/nordugrid-Condor-gridAMD
t2ce02.physics.ox.ac.uk:8443/cream-pbs-shortfive
t2ce04.physics.ox.ac.uk:8443/cream-pbs-shortfive
t2ce06.physics.ox.ac.uk:8443/cream-pbs-shortfive
```

- Can then do things like this to select sites in JDLs

```
requirements = RegExp("gridpp.rl.ac.uk", GlueCEUniqueId)
```

- Looks similar to basic EMI WMS functionality, but
  - Much simpler
  - More transparent
  - Easier to debug

Science & Technology Facilities Council

Conclusion

- With the eventual decommissioning of EMI WMS, non-LHC VOs need alternatives
- Do all non-LHC VOs really make use of advanced functionality in EMI WMS?
  - Some VOs submitting to RAL clearly don't
  - Some others do use it but might not really need to
- More users becoming familiar with HTCondor as
  - Sites in the UK migrating to it
  - CERN may migrate it
- Submission to the grid is no different to local submission

Conclusion

- Lots of possibilities
  - Submission to specific CEs from UI with HTCondor installed
    - Already works today
  - Route jobs from user's UI (job router, flocking or remote submission) to a central WMS-like service based on HTCondor?
  - A service similar to OSG Connect?
    - HTCondor submit UI where people can login & submit jobs to many resources
  - …