

Table of Contents

PandaTier3	1
Introduction	2
Site Prerequisites	3
Preparations	4
Setup	5
Pilot scheduler/monitor software installation.....	5
Setup of crons for pilot submission, pilot monitoring, and maintenance (pilot log cleanup and software update).....	5
Testing and validation.....	6
Usage	7
Further information	8

PandaTier3

Introduction

This page describes the Panda configuration and setup for ATLAS Tier3 sites of the 'T3g' type -- sites that are not full grid sites. Pilots are submitted directly to a local condor queue, from a pilot scheduler cron job running at the site. Pilots are monitored in the conventional autopilot manner, with pilot logs accessible from the Panda monitor. Job log tarballs are also accessible via the Panda monitor job pages in the usual way provided the Tier3 supports the necessary web links into their job output area.

All jobs run under the unix identity of the pilots. Consequently all job outputs are owned by this uid. Panda authenticates and traces all usage -- jobs can be submitted only by users with a valid ATLAS grid proxy -- so the actual user associated with any job can be determined via Panda monitoring.

Panda sites of this type are used as any other Panda site. Jobs are directed there using the pathena/prun --site option. The principal difference is a consequence of the 'off-grid' nature of these sites: they do not support DQ2, thus have no notion of datasets (at present), so input files are specified not via --inDS dataset name specification but via a file list specified with the --pfnList option. Note that remote file access is supported, eg. file specs such as xrootd://... or dcache:/pnfs... .

Site Prerequisites

These are the prerequisites:

- Outbound http, https must be supported. May be proxied. This includes worker nodes and the pilot submit host. All Panda communication is based on http.
- An apache web server configured to support browsing of the pilot run directories must be supported. This enables monitor access to job logs (condor logs and pilot stdout) essential for diagnostics and debugging. The service can be provided on port 80 or any other port, at the site's discretion.
- At present, only condor batch queues are supported. Others may be added depending on demand and available manpower. So a requirement is that a machine be provided to host cron jobs that run the pilot submitter and monitor based on the use of condor_submit and condor_q. This must be either the same machine as that hosting the apache service, or they must share a common disk area, such that condor run directories are accessible to both pilot scheduler and apache.
- Ensuring the pilot management crons (submission, monitoring, log cleanup and code updating) keep operating is a site responsibility, because they run locally at the site. (For conventional Panda sites these crons run from a central pilot submit host, and submit to the site using CondorG, but CondorG is not a supported service at T3gs). The crons are generally low-to-zero maintenance.
- An account must be provided from which pilots will be submitted on the submit host. Note that this account will be the owner of output files produced by Panda user jobs at the site.
- The cron mechanism must be functional in the pilot submission account.
- svn must be supported on the submit host to obtain the Panda software.
- A Panda expert should be given login access to the submit host/account, at least for initial setup/debug. Sustained access is recommended so you can get help with problems if needed.
- A disk storage area accessible from worker nodes, and writable by the pilot submission account, must be provided for job outputs. Outputs will be deposited there with the following directory structure: Year/UserName/DatasetName/FileName. Space management and cleanup of this area is the responsibility of the site. In order for log tarballs to be accessible in the Panda monitor, as for conventional Panda sites, this area must be accessible from the apache server.

Preparations

Once you have established the prerequisites described above, provide the following information to the Panda team:

- Specifications for the pilot submit host configuration:
 - ◆ machine name of the submit host (eg. atl001.phy.duke.edu)
 - ◆ directory path where pilot logs are to be written (eg. /scratch/shared/pilots)
 - ◆ web URL that maps to this directory (eg. <http://atl001.phy.duke.edu:9099/pilots>) (not a functioning URL)
 - ◆ name pattern for the worker nodes (e.g. atl[0-9]+\phy\.duke\.edu)
- Specifications for input/output data management:
 - ◆ directory which will be the default path for input files (eg. /atlas/shared/data)
 - ◆ directory where job outputs will be deposited (under this directory will be the Year/UserName/DataSet... structure)
 - ◆ web URL mapping to the job output directory (eg. <http://atl001.phy.duke.edu:9099/outputs>) (not a functioning URL)
- Specifications for Panda site configuration:
 - ◆ Preferred Panda site name (eg. ANALY_DUKE)
 - ◆ Condor job definition that submits to the queue (eg. see the [Duke queue configuration](#))
 - ◆ Site access policy. Access can be limited to a site-managed list of users if desired. Default access is any ATLAS user can submit jobs to the site. Sites can configure their own access policy, access list and usage rights using puserinfo. If you will use puserinfo to restrict access you need to define a responsible person to manage access rights, and the Panda team will set this up in the DB.

Setup

The Panda database maintainer (schedconfig@gmailNOSPAMPLEASE.com) will set up your site in the central Panda system based on the info above, and will also give you rights to modify the configuration yourself for later adjustments. Please make yourself familiar with the instructions.

Panda setup at the site involves:

- Pilot scheduler/monitor software installation
- Setup of crons for pilot submission, pilot monitoring, and maintenance (pilot log cleanup and software update)
- Testing and validation

Pilot scheduler/monitor software installation

- login to the pilot submitter host/account and create an area to install the Panda code

```
mkdir pilots; cd pilots
svn co http://svnweb.cern.ch/guest/panda/autopilot/trunk autopilot
svn co http://svnweb.cern.ch/guest/panda/monitor monitor
svn co http://svnweb.cern.ch/guest/panda/panda-server/current/pandaserver panda-server
```

Setup of crons for pilot submission, pilot monitoring, and maintenance (pilot log cleanup and software update)

- install panda_setup.sh in the home directory of the account. The script should contain:

```
#!/bin/bash
## PANDA_HOME should point to where you checked out the Panda code, see above
export PANDA_HOME=$HOME/pilots
## PANDA_LOGS should point to the directory that maps to the web URL you support for web access t
export PANDA_LOGS=/export/share/pilot
export SCHEDULER_LOGS=$PANDA_LOGS/scheduler
export CRON_LOGS=$PANDA_LOGS/cron
export PYTHONPATH=$PANDA_HOME/monitor:$PANDA_HOME/panda-server/current/pandaserver:$PYTHONPATH
```

- install panda_manage.sh in the panda code directory, ~/pilots. The script should contain

```
#!/bin/bash
source $HOME/panda_setup.sh
python $PANDA_HOME/autopilot/cleanSpace.py > /dev/null
cd $PANDA_HOME/autopilot
svn update
cd $PANDA_HOME/monitor
svn update
cd $PANDA_HOME/panda-server
svn update
```

- install crons to manage pilot submission, monitoring, and maintenance in crontab, e.g. (substitute your Panda site name)

```
5 0,6,12,18 * * * ~/pilots/autopilot/pilotCron.sh --queue=ANALY_DUKE --pandasite=ANALY_DUKE --pi
5 0,6,12,18 * * * ~/pilots/autopilot/pilotCron.sh --monitor --nocheck > ~/.pilotMon.txt
0 0,6,12,18 * * * ~/pilots/panda_manage.sh > ~/.pilotManage.txt
```

Testing and validation

Once the pilotCron.sh crons are running (you can run the commands interactively to get started if you want) you should see a listing of your pilots on a URL of this form:

http://panda.cern.ch?tp=pilots&accepts=ANALY_DUKE. If not, there is a problem with your submitter. Make sure you have requested that the submit host be added to the `submithosts` table.

As a basic job, set up pathena/prun on lxplus (instructions here). After setup, try:

```
cat >purepython.py << EOF
import sys
print sys.argv
f = open('out.dat', 'w')
f.write('hello')
f.close()
sys.exit(0)
EOF
```

```
prun --exec "python purepython.py" --outDS user.yourusername.PurePythonTest4
--site=YOUR_SITEID
```

If that runs, you can do further testing with standard Workbook analysis test jobs. The submission method is the same except for the requirement that you specify a list of input files rather than a dataset. Full paths, please. Use the `--pfnList` option for the file list (text file).

Report problems to the Panda team (Alden Stradling for starters).

The scheduler service submitting the pilots should also be listed there; you can access its log by clicking its ID number and then the URL on the service information page.

The page also reports the queues serving the site. (For Tier 3s, the queue and the site have the same name.)

Click on the queue name to see the queue configuration, eg.

http://panda.cern.ch?tp=queue&id=ANALY_ANLASC. Take note of the parameter `nqueue`, it is the 'queue depth', the number of pilots in a queued state that the scheduler maintains. The higher `nqueue` is, the higher the pilot throughput is. Adjust it as you like based on your usage and requirements. It can be adjusted with a curl command, see the instructions on the shift instructons wiki (search on `nqueue`). You have to have a valid proxy when you issue the command.

Usage

Once pilots are flowing users can use pathena/prun/ganga as usual to submit Panda jobs to the site. Job outputs will be found in the location specified in the configuration. The configured path prefix appears as the 'se' parameter on the queue configuration page. Outputs are deposited under the path prefix following the convention /Prefix/Year/UserName/DatasetName/FileName.

Further information

See Torre's talk in the US ATLAS Tier 3 meeting [\[7\]](#), ANL, June 2010.

Major updates:

- TorreWenaus - 09-Jun-2010
- TorreWenaus - 19-May-2010
- TorreWenaus - 05-Mar-2010

Responsible: TorreWenaus

Never reviewed

This topic: PanDA > PandaTier3

Topic revision: r15 - 2014-11-20 - TWikiAdminUser



Copyright &© 2008-2020 by the contributing authors. All material on this collaboration platform is the property of the contributing authors.

Ideas, requests, problems regarding TWiki? Send feedback