NHC node health check

David Akin
Senior Systems Analyst
OU Supercomputing Center for Education and Research

Why NHC

- Currently no standard -- Most sites use custom, home-grown scripts
  - Often site-specific
  - Usually lacking portability
- Unreliable execution, reporting, parent performance
- Need a simple, robust framework easy to understand/apply.

What NHC does

Simply:

Prevent jobs from running on unhealthy nodes

Scheduler Integration

- NHC knows which scheduler, so knows the commands needed to find out
  - What user's jobs are running on a node
  - What hosts belong to which queues
  - How to close off a node

What does this do for me?

Benefits

- Clean left over job processes and files.
- Log out users who don't belong on a node anymore
- Make sure the resources are available for the job.
- Make sure node isn't underperforming and sick

Key Features

- 100% native bash framework
- Compatible with RHEL4+
- Single config, infinite targets
- Match config file directives via glob, regex, or pdsh-like range
- Flexible, unrestrictive syntax
- Per-run data cache for speed
- Control via CLI or config
- Run via RM, cron, pdsh, or all TORQUE/PBS, SLURM, SGE, LSF
- Detached mode for low delay
- Built-in watchdog timer
- Unit tests for driver script and every built-in check
- Works with LDAP, NIS, SMB
- 42 checks already built in for hardware, processes, commands, filesystems, jobs, and more
- More checks to come
- Contribute your own checks or ideas for new checks!

Installation

1. Download NHC:
http://warewulf.lbl.gov/downloads/releases/
2. Install RPM (or build and install from tarball)
3. Edit configuration file (default: /etc/nhc/nhc.conf)
4. Configure launch mechanism:

- crond – Consider using sample script nhc.cron
- TORQUE – $node_check_script & $node_check_interval
- SLURM – HealthCheckProgram & HealthCheckInterval
- SGE – Load sensor: load_sensor & load_thresholds
- IBM Platform LSF – lsb.queues: PRE_EXEC & POST_EXEC

Our NHC config (pt 1) /etc/nhc.conf

```
# Filesystem checks
*  || oscer_check_fs_health all
*  || check_fs_mount_rw /tmp
*  || check_fs_free /tmp 10M
# Process checks
*  || check_ps_unauth_users log kill
*  || check_ps_userproc_lineage log syslo
```

Our NHC config (pt 2) /etc/nhc.conf

```
# Hardware checks
*  || check_hw_mem 1024 1073741824
{c[001-310]}  || check_hw_ib 40
{c[001-310]}  || oscer_check_ipath_contexts

# Site-specific-checks
*  || oscer_check_nic_speed em1 1000Mb/s
*  || oscer_check_home_accesible
*  || oscer_check_scratch_accesible
```

NHC environment config

/etc/sysconfig/nhc

```
. /lsf/conf/profile.lsf
NHC_RM=lsf
NHC_AUTH_USERS="roger eddie phil" #admins
MAX_SYS_UID=499 #anything below UID is daemon
HOSTNAME=`hostname –s`
TIMEOUT=600  #timeout of nhc before giving up
LSF_USER_AUTH_RETRY_INTERVAL=120 #site-specific
```

Extra Examples

- Verify that the rpcbind service is alive

  ```
  check_cmd_output -t 1 -r 0 -m '/is running/'
  /sbin/service rpcbind status
  ```

- Search for HTTP daemon IPv4 listening socket and restart if missing:

  ```
  check_net_socket -n "HTTP daemon" -p tcp -s LISTEN
  -l '0.0.0.0:80' -d httpd -e 'service httpd start'
  ```