Adding High Availability to Condor Central Manager Tutorial

Gabi Kliot
Computer Sciences Department
Technion - Israel Institute of Technology
Introduction to HA

- Multiple Collectors run simultaneously on each CM machine
- All submission and execution machines must be configured to report to all CMs
- HAD – HA daemon runs on each CM
- HAD makes sure a single Negotiator runs on one of the CMs
Basic Scenario

I'm alive

HAD

Collector

Idle CM

Workstation – Startd and Schedd

Leader HAD

Collector

Active CM

Workstation – Startd and Schedd

Negotiator

HAD

Collector

Idle CM

www.cs.wisc.edu/condor
HA mechanism must be explicitly enabled
HAD_LIST

- List of server machines, where the HA daemons (HAD) will be installed, configured and run.
- Each element in the list is composed of IP or hostname, and a port number, separated by a colon. Elements are separated from each other using commas.
- HAD_LIST should be identical on all CM machines.
- HAD_LIST should be identical (ports excluded) to the COLLECTOR_HOST list, and in the same order.
HAD_USE_PRIMARY

- One HAD could be declared as primary
- Primary HAD is always guaranteed to be elected as active CM, as long as it is alive
- After primary recovers, it will become active CM, substituting one of its backups
- In case HAD_USE_PRIMARY = true the first element in the HAD_LIST will be the primary HAD. In that case, the rest of the daemons will serve as a backups
- Default is false
HAD_CONNECTION_TIMEOUT

› An upper bound on the time (in seconds) it takes for HAD to establish a TCP connection
› Recommended value is 2 seconds
› Default is 5 seconds
› Effects Stabilization time - the time it takes for HA daemons to detect failure and fix it
› Stabilization time = $12 \times \#\text{CMs} \times \text{HAD\_CONNECTION\_TIMEOUT}$
HAD_ARGS

- HAD_ARGS = -p <HAD_PORT>
- HAD_PORT should be identical to the port defined in HAD_LIST for that host
- Allows master to start HAD on a specified command port
- No default value. This one is a must
Regular daemon configuration

- HAD - path to condor_had binary
- HAD_LOG - path to the log file
- MAX_HAD_LOG - maximum size of the log file
- HAD_DEBUG - logging level for condor_had
Influenced configuration variables

On both client (schedd + startd) and CM machines:

- `COLLECTOR_HOST` - list of CM machines
- `HOSTALLOW_NEGOTIATOR` - must include all CM machines
Influenced configuration variables

▶ Only on Schedd machines:
  • HOSTALLOW_NEGOTIATOR_SCHEDD - must include all CM machines

▶ Only on CM machines:
  • DAEMON_LIST - must include Collector, Negotiator, HAD
  • DC_DAEMON_LIST - must include Collector, Negotiator, HAD
  • HOSTALLOW_ADMINISTRATOR - CM machine must have administrative privileges (in order to turn Negotiator on and off)
Configuration Files
Deprecated variables

- #unset these variables - they are deprecated
- NEGOTIATOR_HOST=
- CONDOR_HOST=
condor_config.local.
ha_central_manager

- CENTRAL_MANAGER1 = cm1.wisc.edu
- CENTRAL_MANAGER2 = cm2.wisc.edu
- COLLECTOR_HOST = $(CENTRAL_MANAGER1),$(CENTRAL_MANAGER2)
condor_config.local.
ha_central_manager

- HAD_PORT = 51450
- HAD_LIST = $(CENTRAL_MANAGER1):$(HAD_PORT), $(CENTRAL_MANAGER2):$(HAD_PORT)
- HAD_ARGS = -p $(HAD_PORT)
- HAD_CONNECTION_TIMEOUT = 2
- HAD_USE_PRIMARY = true
- HAD = $(SBIN)/condor_had
- MAX_HAD_LOG = 640000
- HAD_DEBUG = D_COMMAND
- HAD_LOG = $(LOG)/HADLog
condor_config.local.
ha_central_manager

- DAEMON_LIST = MASTER, COLLECTOR, NEGOTIATOR, HAD
- DC_DAEMON_LIST = MASTER, COLLECTOR, NEGOTIATOR, HAD
- HOSTALLOW_NEGOTIATOR = $(COLLECTOR_HOST)
- HOSTALLOW_ADMINISTRATOR = $(COLLECTOR_HOST)
condor_config.local.

ha_client

- CENTRAL_MANAGER1 = cm1.wisc.edu
- CENTRAL_MANAGER2 = cm2.wisc.edu
- COLLECTOR_HOST = $(CENTRAL_MANAGER1),$(CENTRAL_MANAGER2)
- HOSTALLOW_NEGOTIATOR = $(COLLECTOR_HOST)
- HOSTALLOW_NEGOTIATOR_SCHEDD = $(COLLECTOR_HOST)
Disabling HA mechanism

- Remove HAD and NEGOTIATOR from DEAMON_LIST on all machines
- Leave one NEGOTIATOR in DEAMON_LIST on one machine
- `condor_restart CM machines`
- Or turn off running HA mechanism:
  - `condor_off -all -negotiator`
  - `condor_off -all -had`
  - `condor_on -negotiator on one machine`
Configuration Sanity Check script

- Checks that all HA-related configuration parameters of RUNNING pool are correct
  - HAD_LIST consistent on all CMs
  - HAD_CONNECTION_TIMEOUT consistent on all CMs
  - COLLECTOR_HOST consistent on all machines and corresponds to HAD_LIST
  - DAEMON_LIST contains HAD, COLLECTOR, NEGOTIATOR
  - HAD_ARGS is consistent with HAD_LIST
  - HOSTALLOW_NEGOTIATOR and HOSTALLOW_ADMINISTRATOR are set correct
Backward Compatibility

- Non-upgraded client machines will run fine as long as the machine that served as Central Manager before the upgrade is configured as primary CM.
- Non-upgraded client machines will of course not benefit from CM failover.
FAQ

➤ Reconfigure and restart all your pool nodes, not only CMs
➤ Run sanity check scrip
➤ Condor_off -neg will actively shut down the Neg. No HA is provided
➤ In case primary CM failed, it takes more time for tools to return results. This is since they query the Collectors in order of COLLECTOR_HOST
➤ More than one Neg can be noticed at the beginning for very short time