Using EC2 with HTCondor

Todd L Miller
Outline

› Introduction
› Submitting an EC2 job (user tutorial)
› New features and other improvements
› John Hover talking on Thursday about
  Dynamic cloud-based clusters with
  HTCondor (admins)
EC2: A Cloud and a Protocol

› A cloud service offered by Amazon
› The protocol used to access that service
› Software supporting that protocol:
  • Amazon’s
  • OpenStack
  • Eucalyptus
  • Nimbus
Why EC2?

- Immediate access to additional computational resources
  - Amazon (especially) has a lot of cores
  - No purchasing or installation process
  - Using VMs required
    - Good if you’ve already got a VM
    - Can be troublesome if you don’t
Required Fields

› Service endpoint
  • (https://ec2.amazonaws.com)

› AMI ID and instance type
  • (ami-0256b16b, m1.small)

› Access Key ID

› Secret Access Key
universe = grid
executable = EC2 test job
grid_resource = ec2 https://ec2.amazonaws.com/

ec2_ami_id = ami-0256b16b
ec2_instance_type = m1.small
ec2_access_key_id = ~/Amazon/accessKeyFile
ec2_secret_access_key = ~/Amazon/secretKeyFile
grid resource “protocols”

› http[s]
  • Use for Amazon’s EC2 as well as OpenStack and other cloud software.

› euca3[s]
  • Use for services running Eucalyptus 3.

› x509
  • Use your X.509 proxy instead of your Amazon-style credentials to authenticate (and secure the connection).
# ec2_keypair = test-keypair
ec2_keypair_file = ~/kp-$(Cluster).$(Process)

ec2_spot_price = 0.011

queue
Spot Pricing

› Your bid is the most you’ll pay, but you could pay less

› May not start immediately, and you could wait forever

› May end at any time, but you always have two minutes’ warning
Spot Instance Jobs

- HTCondor doesn’t support
  - group scheduling
  - multiple-instance bids
  - persistent bids
- The job model remains the same
  - simpler to think about
  - queue management commands stay the same
  - extra status information
EC2 Job Submit File (pg 3)

c2_user_data, ec2_user_data_file]
c2_security_groups
c2_elastic_ip
c2_vpc_subnet, ec2_vpc_ip
c2_availability_zones
c2_ebs_volumes
c2_tag_<name>, ec2_tag_names
Dealing with Problems

› Check the HoldReason
  • Sometimes plain English (more or less)
  • Sometimes quotes server responses; look for the `<message>` example error text` </message>`.

› Check the EC2 GAHP log
  • `condor_config_val EC2_GAHP_LOG`

› Check the grid manager log
  • `condor_config_val GRIDMANAGER_LOG`
SSH keypair names
• `SSH_collector:6921_submitter#1.0#1339408`
• Changed to generate spaces, instead.

Handle nonstandard states
• `SHUTOFF, STOPPED`
• treat as “successful errors” – kill them, but report normal termination
Other Improvements

› Batched status updates
› Bug fixes
  • Tickets #2274, #3289, #3061, #3060, #3507, #3492, #3388, #3477, #3367, #3287, #3387, and #2823
Thank You

› Any questions?
Eucalyptus Compatibility

› Doesn’t handle a standard parameter (InstanceInstantiatedShutdownBehavior)
› Work-around enabled by “euca3” protocol