Ganeti Advanced Features Taster

Brian Candler
Network Startup Resource Center
brian@nsrc.org

Cluster upgrade

- This is much simplified for ganeti 2.10+
- Install the new version on all nodes
  - `apt-get dist-upgrade`
- Then issue one command on master:
  - `gnt-cluster upgrade --to 2.11`
- Later, remove the old version from all nodes
- Or revert:
  - `gnt-cluster upgrade --to 2.10`

Job submission

- Add "--dry-run" for dummy run (test arguments)
- Add "--submit" for job to be processed in the background; command returns immediately
  - e.g. `gnt-instance add --submit …`
- `gnt-job list`
- `gnt-job info <num>`
- `gnt-job watch <num>`

Instance allocator (hail)

- If you omit "-n" then most operations will pick the least-loaded node(s) automatically
  - exception: converting plain instance to drbd
- Pluggable
  - could implement your own allocation policy

Cluster balancer (hbal)

- Same algorithm as hail
- `hbal -m <cluster.name>`
  - Makes suggests for how to move instances to achieve better balance
- `hbal -C -m <cluster.name>`
  - Tells you what commands to run to achieve it

Tags

- Can set "tags" on cluster, nodes and instances
- Used to group instances
  - `gnt-instance start --tags foo`
- Used to control other features
  - e.g. "exclusion tags" so that two instances providing a redundant service don't end up being placed on the same node
    - `gnt-cluster add-tags htools:iextags:service`
    - `gnt-instance add-tags vm1 service:dns`
    - `gnt-instance add-tags vm2 service:dns`

ganeti-watcher

- Run from cron every 5 minutes
- Periodically checks cluster and re-starts failed instances, missing DRBD disks
- Can be surprising
  - "halt -p" within an instance, then it gets restarted! (but see next slide)
- Can disable it temporarily for debugging
  - `gnt-cluster watcher pause 1h`

User-initiated shutdown

- New feature in ganeti 2.11
- Detects when user intentionally halted their VM, and sets it to "USER_down" state
  - The watcher will not restart it
- Not enabled by default. To turn it on:
  - `gnt-cluster modify --user-shutdown=true`
  - `gnt-cluster modify -H kvm:user_shutdown=true`

Auto-repair (harep)

- Disabled by default
- Enable by setting instance tags
  - `ganeti:watcher:autorepair:<type>`
- Examples:
  - disk replacement
  - automatic restart (failover) of instance on secondary if primary dies

Node groups

- Maybe nodes are not fully-connected?
  - e.g. separate replication networks in separate buildings; different public networks
- You can create nodes within "node groups"
  - `gnt-group add <groupname>`
  - `gnt-node add -g <groupname> <nodename>`
- By default, instances will only migrate to other nodes in the *same* node group

Network abstraction

- Create using `gnt-network add ...`
- Network has pool (subnet) of available IPs
  - Can automatically assign IPs to instances
  - Passes the IP, netmask and gateway to instance creation scripts
  - Hence fully automated configuration of networking
- Bind network to real interface (per node-group)
- Instance NIC linked to named network
  - `--net 0:network=<name>,ip=pool`

OpenVSwitch

- Connect an instance to an access port
  - `--net 0:mode=openvswitch,link=br0,vlan=10`
- Connect an instance to a trunk port with multiple tagged VLANs
  - `--net 0:mode=openvswitch,link=br0,vlan=:10:20:30`
- Avoids explosion of separate bridges

Custom OS parameters

- `gnt-instance add \`
  `-o debootstrap+default \`
  `-O filesystem=ext4` `<vm>`
- (Also available through RAPI as "osparams")
- Passed as environment variables to the OS creation scripts, e.g.
  - OSP_FILESYSTEM="ext4"
- See `man ganeti-os-interface`

And more for you to discover :-)