Synnefo - Feature # 505

Status:	Closed	**Priority:**	Medium
Author:	Giorgos Gousios	**Category:**	logic
Created:	05/10/2011	**Assignee:**	Giorgos Gousios
Updated:	06/03/2011	**Due date:**	
Subject:	Μηχανισμός επανυπολογισμού κατάστασης VM		

Description

Κατά τη λειτουργία του συστήματος μπορεί να υπάρξουν στιγμές κατά τις οποίες η κατάσταση ενός VM στη βάση του συστήματος είναι μη συγχρονισμένη με την κατάσταση του ίδιου VM στο Ganeti. Ο μηχανισμός αυτός θα πρέπει να ελέγχει περιοδικά (?) ή όταν αντιμετωπιστεί κάποιο πρόβλημα και να συγχρονίζει τις 2 καταστάσεις.

History

#1 - 05/26/2011 12:10 pm - Giorgos Gousios

Κάτι που νομίζω θέλει συζήτηση και ίσως αλλαγή τώρα που έχουμε αλλάξει τον τρόπο που ενημερώνουμε το status του VM, είναι τα πεδία που χρειάζονται στη βάση και οι καταστάσεις που υποστηρίζουμε.

Το OpenStack 1.1 (Apr 25, 2011), υποστηρίζει τα παρακάτω.

ACTIVE, BUILD, REBUILD, SUSPENDED, QUEUE_RESIZE, PREP_RESIZE, RESIZE, VERIFY_RESIZE, PASSWORD, RESCUE, REBOOT, HARD_REBOOT, DELETE_IP, and UNKNOWN.

Στο Ganeti, το οποίο πλέον θεωρούμε ώς το authoritative source για την κατάστασή μας, δεν έχω βρεί κάπου ομαδοποιημένες τις καταστάσεις που μπορεί να υπάρξει ένα VM. Γενικά φαίνεται να υποστηρίζει τουλάχιστο τις παρακάτω τερματικές καταστάσεις, όπως φαίνεται από lib/query.py:

ERROR_nodeoffline, ERROR_nodedown, ERROR_wrongnode, running, ERROR_up, ERROR_down, ADMIN_down

Από όσο καταλαβαίνω από τις συζητήσεις μας και τον κώδικα που έχω δει, το ganeti στην ουσία δεν κρατάει κάπου το state του μηχανήματος και δεν έχει κάποιο state machine για το τι επιτρέπεται. Το κάνει infer από το αποτέλεσμα της τελευταίας δουλειάς που έχει τρέξει για το VM.

Το Σύννεφο προς το παρόν στην ουσία κάνει mirror τη συμπεριφορά του ganeti (6 πεδία που κρατάνε το job status για το τελευταίο job που ξεκίνησε ο χρήστης, όπου job στην ουσία μια ενέργεια που τροποποιεί την κατάσταση του vm στο ganeti) και 2 flags (suspended και deleted) που κρατάνε κάποιες καταστάσεις που ο συνδιασμός των 6 πεδίων δεν είναι ικανός να αναπαραστήσει (ή κάτι άλλο που δεν καταλαβαίνω). Επιπλέον, ο μηχανισμός που κάνει infer την κατάσταση από τα σύνολο 8 πεδία που έχουμε στη βάση μπορεί να δώσει στο API προς απάντηση είναι οι εξής καταστάσεις: UNKNOWN, REBOOT, BUILD, ERROR, STOPPED, ACTIVE, DELETED. Επιπλέον, το UI έχει κάποιες ενδιάμεσες μεταβατικές καταστάσεις (BUILDING, REBOOTING κτλ) που δεν αντικατοπτρίζονται πουθενά.

Νομίζω ότι το όλο σύστημα μπορεί να γίνει πολύ ποιο απλό, αν κάνουμε τις 2 παρακάτω παραδοχές:

- Οι καταστάσεις που υποστηρίζει το σύστημα είναι μόνο αυτές που υποστηρίζει το API
- Το οποιοδήποτε backend είναι υπεύθυνο να μας ενημερώνει για την οποιαδήποτε καινούργια καταστάση (πάλι από το set των καταστάσεων που υποστηρίζουμε μόνο).

Έτσι, στην ουσία, το Συννεφο απλά καταγράφει την τελευταία γνωστή κατάσταση και ενημερώνει τον όποιο API client για αυτή. Είναι δουλειά του backend (ganeti, vmware, ?) να ενημερώνει το frontend σχετικά με την τρέχουσα κατάσταση γιατί εκεί υπάρχει γνώση των όποιων εσωτερικών καταστάσεων και των μεταβάσεών τους.

Καταλαβαίνω ότι το μέχρι τώρα design είχε βασιστεί στην ιδέα ότι το backend ειναι αποκλειστικά ganeti, αλλά νομίζω ότι το σύστημα θα ήταν πολύ ποιο απλό όπως το περιγράφω.

#2 - 05/26/2011 12:14 pm - Panagiotis Louridas

Giorgos Gousios wrote:

> *Κάτι που νομίζω θέλει συζήτηση και ίσως αλλαγή τώρα που έχουμε αλλάξει τον τρόπο που ενημερώνουμε το status του VM, είναι τα πεδία που χρειάζονται στη βάση*
> *Νομίζω ότι το όλο σύστημα μπορεί να γίνει πολύ ποιο απλό, αν κάνουμε τις 2 παρακάτω παραδοχές:*
> *- Οι καταστάσεις που υποστηρίζει το σύστημα είναι μόνο αυτές που υποστηρίζει το API*

Εννοείς το OpenStack API ή το RSAPI;

> *- Το οποιοδήποτε backend είναι υπεύθυνο να μας ενημερώνει για την οποιαδήποτε καινούργια καταστάση (πάλι από το set των καταστάσεων που υποστηρίζουμε μόνο).*

Καμμία διαφωνία εδώ, αφού μόνο αυτό ξέρει την πραγματική κατάσταση.

#3 - 05/26/2011 01:26 pm - Giorgos Gousios

> *- Οι καταστάσεις που υποστηρίζει το σύστημα είναι μόνο αυτές που υποστηρίζει το API*
>
> *Εννοείς το OpenStack API ή το RSAPI;*

Όποιο υποστηρίζουμε, το OpenStack αν δεν κάνω λάθος.

Να προσθέσω στα παραπάνω ότι στο σύστημα ως έχει, ο υπολογισμός της κατάστασης θα πρέπει να μετακινηθεί σε κάποιο κεντρικό σημείο στην πλευρά του backend. Επίσης, πρέπει να σχεδιάσουμε τα ελάχιστα μηνύματα που ενδιαφέρουν το σύστημά μας και να βάλουμε τα hooks στο ganeti να μεταφράζουν τα job status του σε αυτά που ακούμε εμείς. Προς το παρόν, αυτό είναι πολύ δουλεια και αλλάζει πολύ το σύστημα για να πάει στην v0.5. Προτείνω να αφήσουμε το ticket ανοιχτό και να το κοιτάξουμε σοβαρά μετά το release.

#4 - 05/26/2011 07:56 pm - Giorgos Gousios

- % Done changed from 0 to 60

Δυστυχώς, δεν γίνεται να προχωρήσουμε εύκολα με αυτό το feature καθώς το ganeti επιστρέφει λάθος JSON

- μονά εισαγωγικά, τα διπλά είναι δικιά μου προσθήκη για να μπορέσω να το περάσω από pretty printer
- μαργαριτάρια όπως το "dry_run":None, "static":False

φαντάζομαι ότι με κάποιο μη έγκυρο τρόπο έχουν κάνει κάποιο εσωτερικό data structure serialize. Αν υπάρχει κάποιος εύκολος τρόπος για να κάνω de-serialize, παρακαλώ ενημερώστε με.

```
1{
2  "status":"success",
3  "ops":[
4    {
5      "dry_run":None,
6      "instances":[
7        "gousiosg-1"
8      ],
9      "priority":0,
```

```
    "debug_level":0,
    "OP_ID":"OP_INSTANCE_QUERY_DATA",
    "static":False
   }
 ],
 "end_ts":[
   1306427618,
   616642
 ],
 "start_ts":[
   1306427618,
   413629
 ],
 "summary":[
   "INSTANCE_QUERY_DATA"
 ],
 "received_ts":[
   1306427618,
   388367
 ],
 "opresult":[
   {
     "gousiosg-1":{
       "config_state":"up",
       "network_port":12359,
       "serial_no":2,
       "be_instance":{
         "auto_balance":True,
         "vcpus":1,
         "memory":1024
       },
       "be_actual":{
         "auto_balance":True,
         "vcpus":1,
         "memory":1024
       },
       "name":"gousiosg-1",
       "pnode":"store67",
       "hypervisor":"kvm",
       "disks":[
         {
           "logical_id":[
             "ganeti",
             "37d717a4-7f86-4dbd-9834-31630e1e8948.disk0"
           ],
           "sstatus":None,
           "dev_type":"lvm",
           "pstatus":[
             "/dev/ganeti/37d717a4-7f86-4dbd-9834-31630e1e8948.disk0",
             254,
             17,
             None,
             None,
```

```
            False,
            1
          ],
          "mode":"rw",
          "physical_id":[
            "ganeti",
            "37d717a4-7f86-4dbd-9834-31630e1e8948.disk0"
          ],
          "children":[

          ],
          "iv_name":"disk/0",
          "size":2000
        }
      ],
      "uuid":"10182022-2110-4240-b0be-9eb4ef5da114",
      "hv_actual":{
        "nic_type":"paravirtual",
        "use_chroot":False,
        "vnc_x509_path":"",
        "vnc_bind_address":"127.0.0.1",
        "usb_mouse":"",
        "migration_downtime":30,
        "security_model":"none",
        "cdrom_image_path":"",
        "boot_order":"disk",
        "vhost_net":False,
        "disk_cache":"default",
        "kernel_path":"",
        "acpi":True,
        "vnc_x509_verify":False,
        "vnc_tls":False,
        "use_localtime":False,
        "security_domain":"",
        "serial_console":False,
        "kvm_flag":"",
        "vnc_password_file":"",
        "disk_type":"paravirtual",
        "kernel_args":"ro",
        "root_path":"/dev/vda1",
        "initrd_path":"",
        "mem_path":""
      },
      "hv_instance":{

      },
      "nics":[
        [
          None,
          "aa:00:00:fc:99:26",
          "bridged",
          "br0"
        ]
```

```
116         ],
117         "snodes":[
118
119         ],
120         "disk_template":"plain",
121         "mtime":1306417204.811044,
122         "run_state":"up",
123         "os_instance":{
124
125         },
126         "os":"debootstrap+default",
127         "os_actual":{
128
129         },
130         "ctime":1306417137.888135
131       }
132     }
133   ],
134   "opstatus":[
135     "success"
136   ],
137   "oplog":[
138     [
139
140     ]
141   ],
142   "id":"18160"
143}
144
```


#5 - 05/27/2011 10:40 am - Giorgos Gousios

Βοηθάει βέβαια το να μην αντιμετωπίζεις τα return types από Python calls σαν return types απο HTTP calls :-) Credits to vkoukis for the heads up.

#6 - 06/03/2011 04:42 pm - Vangelis Koukis

- Status changed from New to Closed

Έχει υλοποιηθεί μηχανισμός reconciliation.
Ενεργοποιείται με λήψη απαντήσεων σε αιτήσεις OP_INSTANCE_QUERY_DATA, που στέλνονται μαζικά από mgmt command του Django, περιοδικά.

Ο διαχειριστής μπορεί κατά βούληση (gnt-instance info) να κάνει το ίδιο, αν χρειαστεί.