GRNET Cloud Services

Offering Cloud Services to the Greek R&E Community

Yannis Mitsos[1] ymitsos@noc.grnet.gr,   Panos Louridas[1] louridas@grnet.gr

[1]Greek Research and Technology Network

September 2010

Concept

Enabling the Future Service-Oriented Internet

- ▶ A centralised services provisioning system
- ▶ Virtual working enviroment services
- ▶ Services targeted to the academic, scientific, research community...
- ▶ ... Provide support to production / administration services as well

Current Services

- ► GRNET Virtual Machine (ViMa) Service
- ► GRNET AAI Federation
- ► HellasGrid infrastructure
- ► Pithos online storage service

Coming Services

Based on our experience with IaaS we will offer additional SaaS facilities:

- ▶ Scientific Software as a Service
- ▶ Labs / Science 2.0 Environments as a Service
- ▶ Administration Software as a Service
- ▶ Database as a Service
- ▶ Storage as a Service
- ▶ Grid asa Service (cf. StratusLab)
- ▶ Private cloud as a Service
- ▶ Hydrib cloud as a Service
- ▶ Disaster recovery aaS
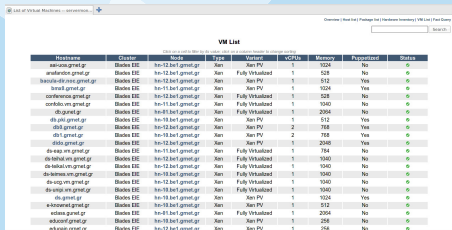- ▶ Data archiving / curation as a Service
- ▶ Backup as a Service

ViMa service

Scope

- ▶ Core services hosting
- ▶ Virtual Private Server (VPS) hosting for our clients

ViMa Service: Initial Deployment

What do we offer currently:

- Debian 5.0/Xen 3.2
- Manual administration of the VMs
- VM hardware node initially assigned in ldap.noc.grnet.gr
- Control dashboard with Django + libvirt + MySQL

ViMa Service: Initial Deployment

Problems

1. Xen instabilities when CPU in high load $\rightarrow$ kernel panics (fixed)
2. Xen scheduler races when CPU in high load + I/O loads: dom0 starvation
3. CPU starvation under high load when using SMP dom0 and high vCPU/CPU ratio
4. Unacceptable I/O performance!
5. Unsupported graceful shutdown of HVM guests
6. Unstable during live migrations (no-go) and problems with VM clocks (large drifts)
7. Lock-in in one kernel flavor

Minor improvements Xen 3.2 $\rightarrow$ 3.4

KVM seems to be the only solution

ViMa Service Redux

Redesigned ViMa based on our initial experiences to meet the needs of a production-level Infrastructure as a Service cloud.

Tests with KVM

Exhaustive tests with KVM over the last 8 months.

- ▶ Noticable improvements
- ▶ Major problems were tackled

KVM Issues

Problem

Filesystem corruption during consecutive migrations between two hosts

Solution

Caused by page-cache incoherency between the hardware nodes. KVM switch `cache=none` bypasses the host's page-cache and solves the problem.

KVM Issues

Problem

Instabilities in live migration of paravirtualized Linux instances with > 1 vCPUs

Solution

Caused by kvm-clock (paravirt) problems, fix exists in kernel 2.6.33+.
Workaround with `LD_PRELOAD` wrapper for the `ioctl()` that suppresses the kvm-clock capability advertisement.

grnet
Διασυνδέοντας την Έρευνα και την Εκπαίδευση

ViMa Service: Second Development

Prerequisites:

- Actual cluster management
- A firm hypervisor
- Web frontend for OOB access / power control
- Open architecture :)

Deployment:

- KVM
- Google's Ganeti project with GRNET NOC contributions
- Custom web frontend in Django (under development)

Results

- 10-node cluster in GRNET's main Data Center in production
- 150 VMs in KVM running core services (DNS, mail, radius, IdP, ldap etc.)
- Kernel 2.6.32 + qemu-kvm 0.11.1 with KSM support $\rightarrow \approx 1.2$ GB RAM de-duplicated
- Shared block storage over FC
- Working live-migrations for all O/Ses (Linux, Windows, FreeBSD)
- Working paravirtualized drivers for Linux, Windows
- No stability/performance problems have been identified

Ganeti: Shared-nothing Cluster

- No SPoF in the cluster
- Master node: cluster control, master IP address
- Master candidate: cluster configuration copy as master candidate
- Simple node: only as node controller
- If master fails: master failover to a master candidate

Ganeti: Storage Support

Four types of storage ("disk templates"):

- diskless: only used for special purpose operating systems or for testing
- file: the instance uses plain files as backend for its disks
- plain: the instance uses LVM devices as backend for its disks
- DRBD: 1 DRBD-device per guest disk on an LVM. Failover / migration capabilities

GRNET Extensions

- Shared block storage (SAN) support (in production)
- Shared file storage (NAS) support (planning)

Ganeti GRNET NOC Contribution

Ganeti modifications, submitted (Google)

- Shared block-storage support (working migrations and failover)
- Changes in the KVM backend:
  - Boot from HTTP + minor bugfixes
  - VNC forwarding
- Minor changes in the API

Cooperative upstream, keen in extending Ganeti's functionality

Conclusions

There are many supporting currents in our view of cloud computing:

- Economies of scale (favourable economic / market environment)
- Lack of sysadmin talent
- Quick to market (production)
- Long-term experience with providing quality services to the academia
- Excellent cooperation with the public sector / government, possible to exploit synergies (e.g., disaster recovery).