SYNNEFO: A COMPLETE CLOUD STACK OVER GOOGLE GANETI. VANGELIS KOUKIS
TECHNICAL LEAD, SYNNEFO
Running a public cloud: ~okeanos

History
  - Design started late 2010
  - Production since July 2011

Numbers
  - Users: > 3500
  - VMs: > 5500 currently active
  - More than 160k VMs spawned so far, more than 44k networks
Running a public cloud: ~okeanos

Our choices

- Build own AWS-like service (Compute, Network, Storage)
- Persistent VMs
- Everything open source
- Production-quality IaaS
- Super-simple UI

How?
Running a public cloud: ~okeanos

The tough stuff

- Stability
- Persistent VMs: VMs are not cattle, they are pets
- Commodity hardware
- Scalability
- Manageability: Gradual rollout of upgrades and new features
Running a public cloud: ~okeanos

Our approach

- Synnefo
- Google Ganeti
- DRBD
- Ceph
- OpenStack APIs
~okeanos VMs

CloudOpen
Europe 2013
vkoukis@grnet.gr
## Cluster vs Cloud

<table>
<thead>
<tr>
<th>OPENSTACK</th>
<th>SYNNEFO</th>
<th>UI</th>
</tr>
</thead>
<tbody>
<tr>
<td>OPENSTACK</td>
<td>OPENSTACK</td>
<td>API</td>
</tr>
<tr>
<td>OPENSTACK</td>
<td>SYNNEFO</td>
<td>CLOUD</td>
</tr>
<tr>
<td>OPENSTACK</td>
<td>GANETI</td>
<td>CLUSTER</td>
</tr>
<tr>
<td>LIBVIRT</td>
<td></td>
<td>NODE</td>
</tr>
<tr>
<td>KVM</td>
<td>KVM</td>
<td>HYPervisor</td>
</tr>
</tbody>
</table>
Live demo!

Login, view/upload files
Unified image store: Images as files
View/create/destroy servers from Images
...on multiple storage backends
...on Archipelago, for thin, super-fast creation
...with per-server customization, e.g., file injection
View/create/destroy virtual networks
Interconnect VMs, with NIC hotplugging
Snapshot a VM’s disk into an Image, in seconds
Create a virtual cluster from this Image
...from the command-line, and in Python scripts
Google Ganeti

Mature, production-ready VM cluster management
- used for Google’s corporate infrastructure

Multiple storage backends out of the box
- LVM, DRBD
- Files on local or shared directory
- RBD (Ceph/RADOS)

External Storage Interface for SAN/NAS support
Ganeti cluster = masterd on master, noded on nodes

Easy to integrate into existing infrastructure
- Remote API over HTTP, pre/post hooks for every action!
Architecture
Identity: Astakos

Identity Management, Resource Accounting and SSO
- Platform-wide service
- Simple service- (Cyclades, Pithos) and user-facing APIs
- Multiple authentication methods per user
- Fine-grained per-user, per-resource quota

A single dashboard for users
- View/modify profile information and active authentication methods
- Easy, integrated reporting of per-resource quotas
- Project management: View/Join/Leave projects
- Manage API access and retrieve authentication tokens
Identity: Astakos

Supported 3\textsuperscript{rd}-party providers
- Shibboleth / AAI Federation
- Google
- Twitter
- LinkedIn
Compute/Network/Image/Volume: Cyclades

Thin Compute layer over Ganeti
- Python/Django
- Supports *multiple* Ganeti clusters, for scaling
- OpenStack APIs

Networking
- No restrictions on deployment – it’s the *Ganeti* side
- IPv4/IPv6 public networks, complete isolation among VMs
- Thousands of private networks, private L2 segments over single VLAN
- Software-Defined Networking, pilots with VXLAN integration
Compute/Network/Image/Volume: Cyclades
Interaction with Ganeti

Support for all Ganeti storage templates

External Storage Interface for SAN/NAS support

Networking =  gnt-network +
            snf-network (KVM ifup scripts) +
            nfdhcpd (custom NFQUEUE-based DHCP server)

Asynchronous operation
- Effect path: Receive API requests, enqueue requests over RAPI
- Update path: Receive asynchronous notifications, update DB
Every file is a collection of blocks
Content-based addressing for blocks
Partial file transfers, deduplication, efficient syncing
Independent of storage backend (NFS, RADOS, ...)
OpenStack Object Storage API plus extensions
An integral part of Synnefo
- Single store for Files, VM Images
- Uses common backend with Archipelago (Volumes)
Storage: Pithos
Spawn

my own Ubuntu

Freeze
Clone

Snapshot

Ubuntu + user data
CloudOpen
Europe 2013
vkoukis@grnet.gr

RADOS
Monitor nodes
Object Storage nodes

block I/O

Volume Composer

Mapper and Blocker

Archipelago

object I/O
Volumes: Archipelago
Volumes: Archipelago

Unified storage for Files, Images ↔ Volumes
Thin layer over the actual storage cluster
Storage backend agnostic
- NFS, RADOS, ...
Efficient syncing / sharing of Images as files on the storage service
Zero-copy cloning of volumes from Images
Features

Integrated Compute and Storage
- Spawning VMs from custom Images
- Images on Storage service
- Thin provisioning from existing Images
- Volume clones and snapshots independently of underlying hw

Storage
- Efficient syncing with native clients

Identity
- Multiple login methods, Shibboleth-based federated logins
Features

Quotas
- Per-user, per-resource quotas
- Usage reports on UI

Projects
- Granting extra virtual resources to members
- UI to create/approve/join projects
- Current use case: virtual labs, R&D project proposals
Experience

Operations

- Rolling hardware and software upgrades
  - kernel, Ganeti, RADOS, Synnefo
  - with no VM downtime
- Node evacuations with live VM migrations
- Cross-datacenter move, Intel → AMD, no VM downtime
- On-the-fly migration from NFS-backed storage to RADOS
- IP renumbering of all VMs
Experience

Scalability
- From few physical hosts to multiple racks
  - dynamic addition of Ganeti clusters

Diverse workloads
- Different network and storage backends
- Choice exposed to the user
Try it out!

http://www.synnefo.org
Thank you!