SYNNEFO: A COMPLETE CLOUD PLATFORM OVER GOOGLE GANETI WITH OPENSTACK APIs
VANGELIS KOUKIS, TECH LEAD, SYNNEFO
Synnefo cloud platform

An all-in-one cloud solution

- Written from scratch in Python
- Manages multiple Google Ganeti clusters of VMs
- Uses Archipelago to unify all cloud storage resources
- Exposes the OpenStack APIs to end users

Production since 2011

- Came out of the ~okeanos public cloud service
Synnefo cloud platform

A complete cloud platform

- Identity Service
- Object Storage Service
- Compute Service
- Network Service
- Image Service
- Volume Service
Unified view of storage resources

Files
- User files, with Dropbox-like syncing

Images
- Templates for VM creation

Volumes
- Live disks, as seen from VMs

Snapshots
- Point-in-time snapshots of Volumes
Services Overview
Live demo! (screenshots at end of presentation)

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
Identity Service

Identity Management, Resource Accounting and SSO

- Multiple authentication methods per user
  - LDAP, AD, Local username/password, Federated (Shibboleth)
  - Google, Twitter, LinkedIn
- Fine-grained per-user, per-resource quota
- Exposes the OpenStack APIs (Keystone) to users
Identity Service

A single dashboard for users

- View/modify profile information
- Set/unset active authentication methods
- Easy, integrated reporting of per-resource quotas
- Project management: View/Join/Leave projects
- Manage API access and retrieve authentication tokens
Compute/Network/Image/Volume Service

Layer over multiple Ganeti clusters
- Python/Django implementation
- Exposes the OpenStack APIs (Nova, Neutron, Glance, Cinder)

A thin translation layer
- From user (API) requests
- To VM operations on multiple Ganeti clusters

Ganeti clusters are distinct entities
- May be geographically remote
- Admin always has direct access for troubleshooting
Compute/Network/Image/Volume Service

Networking

- Fully pluggable, for integration with existing infrastructure
  - VLAN pool, or MAC-prefix-based filtering on single VLAN
  - VXLAN for all-IP datacenter-wide networking
  - Open vSwitch support
- IPv4/IPv6 public networks, complete isolation among VMs
- Tens of thousands of private networks over single VLAN
- Floating (“elastic”) IPv4 addresses, shared among VMs
- NIC hotplugging for dynamic IP attachment/detachment
- No need for NAT setup
Compute/Network/Image/Volume Service

Image Handling
- Spawning VMs from custom Images
- Images treated as Files on Storage service
- System and User Images, fine-grained sharing, custom ACLs

Images for all major Operating Systems
- Debian, Ubuntu, RHEL, CentOS, Fedora, ArchLinux, openSUSE, Gentoo
- NetBSD, FreeBSD, OpenBSD
Storage service

A single location for user Files, VM Images, and Snapshots
Exposes the OpenStack Object Storage API (Swift)
  - plus extensions, for sharing and syncing
Rich sharing, with fine-grained Access Control Lists
Hash-based (sha256) deduplication for individual blocks
Partial file transfers, efficient syncing (Dropbox-like)
Backed by Archipelago
# Layered design

<table>
<thead>
<tr>
<th>Client</th>
<th>OpenStack</th>
<th>Synnefo</th>
<th>UI</th>
</tr>
</thead>
<tbody>
<tr>
<td>vCloud</td>
<td>OpenStack</td>
<td>OpenStack</td>
<td>API</td>
</tr>
<tr>
<td>vCloud</td>
<td>OpenStack</td>
<td>Synnefo</td>
<td>CLOUD</td>
</tr>
<tr>
<td>vCenter</td>
<td>OpenStack</td>
<td></td>
<td></td>
</tr>
<tr>
<td>vSphere</td>
<td>libvirt</td>
<td>Ganeti</td>
<td>CLUSTER</td>
</tr>
<tr>
<td>ESXi</td>
<td>KVM / XEN</td>
<td>KVM / XEN</td>
<td>NODE</td>
</tr>
</tbody>
</table>

**Software:**
- OpenStack
- Synnefo
- Ganeti
- libvirt
- KVM / XEN

**Platforms:**
- vCloud
- vCenter
- vSphere
- ESXi

**Hypervisors:**
- KVM / XEN

**API:**
- UI

**Technologies:**
- CLOUD
- CLUSTER
- NODE
- HYPervisor
Google Ganeti

Mature, production-ready VM cluster management
- developed by Google, for all of Google’s corporate infra
- as open source VMware alternative
- scalable over commodity hw
- in production inside Google since 2006

Easy to integrate into existing infrastructure
- Remote API over HTTP, pre/post hooks for every action!
Google Ganeti

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

External Storage Interface for SAN/NAS support
Support for Archipelago
Archipelago

Storage Virtualization System
- Powering storage in Synnefo

Decouples storage resources from storage backends
- Files / Images / Volumes / Snapshots

Unified way to provision, handle, and present resources
Decouples logic from actual physical storage
- Software-Defined Storage
Interaction with Archipelago

A common storage substrate for Synnefo
Everything is a resource on Archipelago
The same resource is exposed as
- A File through the API of the Storage Service
- An Image through the API of the Image Service
- A live disk / VM Volume through the API of the Volume Service
- A Snapshot through the API of the Volume Service

All data remain in one place
No copying of data around
Cloud Storage with Archipelago

Storage backend 1 (e.g., Ceph)

Storage backend 2 (e.g., GlusterFS)

Storage backend 3 (e.g., NFS over NAS)
Composing Resources with Archipelago
Archipelago logic

Thin provisioning, with clones and snapshots
  - Independent from the underlying storage technology

Hash-based data deduplication

Pluggable architecture
  - Multiple endpoint (northbound) drivers
  - Multiple backend (southbound) drivers

Multiple storage backends
  - Unified management
  - with storage migrations
**Linux block driver**

**Northbound interface**

**Volume Composer**

**Mapper**

**Archipelago Core**

**Ceph/RADOS driver**

**Southbound interface**

**RADOS**

- Monitor nodes
- Object Storage nodes

block I/O

Northbound interface

object I/O

Block I/O interface

Southbound interface

Object Storage nodes

Northbound interface
Archipelago interfaces
Running Archipelago
Comparison to OpenStack?
Why Synnefo?  
A: Enterprise VMs at Cloud scale.

The best of both worlds

- Enterprise, persistent, stable VMs, live migrations (VMware-like)
  - Key technologies: Ganeti
- Over commodity hardware, no SAN needed
  - Key technologies: DRBD, Archipelago, Ceph
- At Cloud scale, accessible over Cloud APIs (OpenStack-like)
  - Key technologies: Synnefo

Storage virtualization with Archipelago

- Common storage pool for everything
  - User files, Images (VM templates), live VM volumes, Snapshots
  - Zero-copy thin cloning / snapshotting for super-fast provisioning
  - Over commodity hardware, no SAN needed
  - Less than 30 sec for a VM to be fully up and running
- Independent of the actual data store
- Pluggable storage: NFS/NAS, Ceph, Gluster, even SAN all at once
  - With inter-backend data moves
Why Synnefo? C: Easier to run at scale.

Distinct management domains: Synnefo and Ganeti
- Management of self-contained Ganeti clusters
- Distinct Synnefo and Ganeti upgrade cycles
- Independent upgrades with no VM downtime

Limited access to DBs, decentralized VM state
- Only Synnefo workers need access to DBs
- No access from Ganeti nodes
  ★ Reduces impact of possible VM breakout
  ★ Boosts scalability to thousands of nodes
- Easier to firewall, easier to handle security-wise

Physical node management
- Dynamically add/remove/drain/set offline physical nodes
- Dynamically add/remove/drain/rebalance whole Ganeti clusters
- Evacuate failing nodes with live VM migrations, no VM downtime

Recovery from failure
- Built-in reconciliation mechanisms
- Able to recover from Synnefo/Ganeti downtime
  * Ensures in-sync state across components

Easier to contain failure
- Outages contained inside smaller domains
  * inside a node, or inside a Ganeti cluster
The ~oceanos use case

Production since July 2011

Numbers

- Users: > 10000
- VMs: > 10000 currently active
- More than 350k VMs spawned so far, more than 100k networks

Physical Infrastructure

- 13 Ganeti Clusters, spanning a whole DC
- 1PB of raw storage capacity
Try it out!

http://www.synnefo.org
Thank you!
Screenshots.
### Public IPv4 Network

<table>
<thead>
<tr>
<th>Name</th>
<th>IPv4</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>snf-6189.vm.oceanos.grnet.gr</td>
<td>83.212.96.147</td>
<td></td>
</tr>
<tr>
<td>snf-20546.vm.oceanos.grnet.gr</td>
<td>83.212.105.230</td>
<td></td>
</tr>
</tbody>
</table>

### Public IPv6 Network

<table>
<thead>
<tr>
<th>Name</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>Connections (2)</td>
<td></td>
</tr>
</tbody>
</table>
New Network +

Public IPv4 Network
- Connections (2)

Public IPv6 Network
- Connections (2)
- snf-20546.vm.oceanos.grnet.gr
  - Firewall (Off)
- snf-6189.vm.oceanos.grnet.gr
  - Firewall (Off)
Upload

My Files
- pithos
- mp3
- photos_public
- pics
- presentations-public
- ThunderBird FileLink
- trash

Used: 901.5MB of 100GB (1%)

Shared with me

Shared by me

Groups

Name | Size | Last Modified
--- | --- | ---
00201_lakejipe_1920x1200.jpg (view) | 198.6 KB | 12/11/2012 11:07 AM
00388_fallintennessee_1920x1200.jpg (view) | 402.7 KB | 12/11/2012 11:07 AM
00423_polynesian_1920x1200.jpg (view) | 610.4 KB | 12/11/2012 11:07 AM
00649_almostnightfall_1920x1200.jpg (view) | 488.3 KB | 12/11/2012 11:07 AM
00785_bodega_gulch_1920x1200.jpg (view) | 405.9 KB | 12/11/2012 11:07 AM
01392_dreambeach_1920x1200.jpg (view) | 1008.8 KB | 12/11/2012 11:08 AM
01407_harboursunset_1920x1200.jpg (view) | 814.3 KB | 12/11/2012 11:08 AM
1537_grassysunset_1920x1200.jpg (view) | 1.6 MB | 12/11/2012 11:08 AM
If you are a student, professor or researcher you can login using your academic account.

ACADEMIC LOGIN

vkoukis@grnet.gr

SUBMIT

Forgot your password?
<table>
<thead>
<tr>
<th>Resource</th>
<th>Usage Information</th>
<th>Usage (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Storage Space</td>
<td>901.48 MB out of 100.00 GB Storage Space</td>
<td>1%</td>
</tr>
<tr>
<td>System Disk</td>
<td>80.00 GB out of 300.00 GB System Disk</td>
<td>27%</td>
</tr>
<tr>
<td>CPUs</td>
<td>5 out of 22 CPUs</td>
<td>23%</td>
</tr>
<tr>
<td>RAM</td>
<td>5.00 GB out of 26.00 GB RAM</td>
<td>19%</td>
</tr>
<tr>
<td>Virtual Machines</td>
<td>2 out of 12 Virtual Machines</td>
<td>17%</td>
</tr>
<tr>
<td>Private Networks</td>
<td>0 out of 15 Private Networks</td>
<td>0%</td>
</tr>
<tr>
<td>Public IPs</td>
<td>3 out of 9 Public IPs</td>
<td>33%</td>
</tr>
</tbody>
</table>