Follow the Sun through the Clouds: Application Migration for Geographically Shifting Workloads

Robbert van Renesse
Cornell University

Joint work with Zhiming Shen, Qin Jia, Gur-Eyal Sela, Ben Rainero, Weijia Song, Hakim Weatherspoon
Infrastructure as a Service (IaaS) Clouds

- Offer on-demand virtual machines (VMs)
- Pay-as-you-go: charge according to used hours
- Provide useful services such as auto-scaling and failure recovering
Handling Geographically Shifting Workloads

Follow the sun
Handling Geographically Shifting Workloads

Follow the sun

- Lack of homogeneous interface
- Lack of privileged control
- Lack of infrastructure support
- Lack of common resource management
Supercloud Overview

• Application migration as a service across cloud providers and availability zones
  • Support ALL major virtualization platforms and ALL major public cloud providers

• Live migration without changing IP addresses or breaking TCP connections

• Automatic scheduling framework
  • Optimize metrics such as average perceived latency

• Provide cross-cloud storage and networking solution
Supercloud Overview

• Computation
  • Nested hypervisor: Xen-Blanket
  • Support all major platforms

• Network
  • SDN overlay
  • Support migration with public IP

• Storage:
  • Geo-replicated storage
  • Optimized for serving VM images

• Resource management
  • OpenStack platform
Nested Virtualization

Xen-Blanket
• Second Layer Hypervisor
• Uniformity
DROSTE CACAO

netto
453 g
Supercloud Networking

• Goal:
  • Inter-connection
  • Optimized routing
  • Supporting migration
• VPN overlay
• Full-mesh tunnels
• Frenetic SDN controller
• Transparent VM migration
• Public IP address support
VM Migration with Public IP Address
VM Migration with Public IP Address

54.172.26.213

52.69.94.195
Centralized VM Image Storage

Long latency; Low throughput
Geo-Replicated VM Image Storage

Challenges:
- Strong consistency requirement
- Long latency and low throughput in WAN
Decouple Consistency and Data Propagation

- Version number
- Location of the latest block

- Local version number

Consistency

Data Propagation

Data View Layer

Data Store Layer

VM

Local Meta-Data

Propagation Manager

Global Meta-Data

Cloud 1

NFS/iSCSI

Cloud 2

VM

Local Meta-Data

Propagation Manager

Global Meta-Data

Back-End Storage

On-demand fetch

Pro-active data propagation

Strong Consistency

Eventual Consistency

15
Global Meta-Data Propagation

• Challenge:
  • Long latency

• Observation:
  • Single writer
  • No read-write sharing

• Relaxed consistency model
  • Close-to-open consistency

• Propagation policy
  • Commit locally
  • Flush to centralized controller when closing
Evaluation: ZooKeeper Migration

• Application level vs. VM level migration

<table>
<thead>
<tr>
<th></th>
<th>ZooKeeper Dynamic Reconfiguration</th>
<th>Supercloud VM migration</th>
</tr>
</thead>
</table>
| Code complexity          | • Add/remove nodes: 6700+ lines of code change  
                          | • Leader rotation: not supported yet | No code change          |
| Transparency             | Clients need to be notified       | Completely transparent   |
| Performance              | Several seconds of downtime due to state synchronization and leader election | Little performance impact |
Comparing ZooKeeper Migration Mechanisms

- Initially: Asia 1, US 2
- 2-step reconfiguration:
  - Asia + 1, US -1
- 3-step reconfiguration:
  - Asia +2, US -2
  - Asia -1, US +1
- Supercloud
  - Migrate the leader from US to Asia
Follow the Sun

• Experimental Setup
  • Global ZooKeeper deployment in US and Asia
  • MSN trace
  • Comparing different deployments
    • **US Ensemble**: all ZooKeeper nodes in the US
    • **Global Ensemble**: majority in US, one node in Asia
    • **Dynamic Ensemble**: using Supercloud VM migration
Follow the Sun

(a) US Ensemble

(b) Global Ensemble

(c) Dynamic Ensemble
Supercloud Scheduler

• Decides placement and migration automatically
• Requires run-time monitoring and performance models for cloud resources
Performance Modeling for IaaS Clouds

Click on the resource type to check:
Memory System, Disk and File system, Network, Comprehensive Benchmarks

1. Histograms
2. By Provider
   - Amazon EC2
     - m3.medium
     - m3.large
     - c4.large
   - Google Compute Engine
     - n1-standard-1
     - n1-standard-2
   - Microsoft Azure
     - Standard DS1
     - Standard DS2
   - Fractus
     - Fractus Bare Metal
     - Fractus Virtual Machine
   - Local
     - Dell OptiPlex 760

Sequential Read/Write Throughput of google vm N1S1 (unknown CPU/unknown memory)

Throughput (GB/s)

Working Set Size (KB)

L1 Cache 32KB  L2 Cache 256KB  L3 SmartCache 45MB

28MB Effective L3 Cache
Partners in crime

• NIST ANTD (Advanced Network Technologies Division):
  Monitoring and Security
  • Abdella Battou
  • Fred de Vaulx
  • Lotfi Benmohamed
  • Charif Mahmoudi

• Cornell Aristotle Project and XSEDE
  Academic cloud sharing and bursting
  • David Lifka (Cornell CIO)
  • ...
Conclusion

• Supercloud: application migration for geographically shifting workloads
  • Crossing heterogeneous cloud providers
  • Automatic scheduling
  • Geo-replicated image storage
  • Wide-area SDN
• Visit our workshop tomorrow morning (Thursday)
• We’ll also present exciting cloud performance comparison studies
• More at http://supercloud.cs.cornell.edu

Thank You. Questions?