Overview

- Introduction
- Traditional clusters vs. Ganeti
- Design goals
- Cluster setup
- Instance failover example
- Usage in Google
- Open Source and Roadmap
What is virtualization?

• Abstraction of computer resources
  ▪ CPUs, memory, storage, network

• Advantages
  ▪ Consolidation, increase hardware utilization
  ▪ Transparent for user
  ▪ Flexibility

• Disadvantages
  ▪ Depending on application: performance losses

• Different types
  ▪ Paravirtualization
  ▪ Full virtualization

• Hypervisor
What is Ganeti and why should you use it?

- Software to manage clusters of virtual servers
  - Automation allows you to scale easily
  - Makes it simple to manage 10s of nodes and 100s of instances
- Combines virtualization and data replication
  - All integrated in a unified interface
  - Virtual systems are portable between nodes
- Hypervisor backends
  - Abstraction layer
  - Currently based on Xen, but others are possible
Terms

• Node
  ▪ Physical machine
  ▪ Xen Dom0

• Instance
  ▪ Virtual machine
  ▪ Xen DomU

• DRBD
  ▪ Distributed Replicated Block Device, http://www.drbd.org/
  ▪ Used for data replication

• LVM (Logical Volume Manager)
  ▪ Used to manage instances' volumes
Traditional high-availability cluster

Application data, SW

Heartbeat

system 1

application

system 2

OS disk

Application data, SW

Application failover (data, IP address)
Ganeti cluster

Virtual system failover

virt. system 3

virt. system 1
virt. system 2
virt. system 4

system 1
system 2
system 3
system 4

Master role:
Tests and control
Overview

- Introduction
- Traditional clusters vs. Ganeti
- Design goals
- Cluster setup
- Instance failover example
- Usage in Google
- Open Source and Roadmap
Design goals and principles

• Goals
  ▪ Increase availability
  ▪ Reduce hardware cost
  ▪ Increase flexibility
  ▪ Transparency

• Principles
  ▪ Not dependent on specific hardware (e.g. SAN)
  ▪ Scales linearly with the number of systems
  ▪ One node takes the master role
    • Failover is possible
Redundancy, Replication and Failover

• Redundancy
  ▪ Disks
  ▪ Memory
  ▪ → Primary & secondary node for each instance

• Replication
  ▪ Real time data replication for disks (primary → secondary)
  ▪ DRBD8

• Failover
  ▪ Instance failover
  ▪ Secondary failover (disk replica replacement)
Overview

• Introduction
• Traditional clusters vs. Ganeti
• Design goals
• Cluster setup
• Instance failover example
• Usage in Google
• Open Source and Roadmap
Ganeti commands

- Administration is done on the master node
- All commands have man pages and support interactive help
- `gnt-cluster`: Cluster commands
- `gnt-node`: Add, remove, list cluster nodes
- `gnt-instance`:
  - Add, remove instance
  - Failover instance, change secondary
  - Stop, start instance, change parameters
- `gnt-os`: Instance OS definitions
- `gnt-backup`: Instance export and import
Cluster creation

node1# gnt-cluster init mycluster
node1# gnt-node add node2
node1# gnt-node add node3
Listing nodes

node1# gnt-node list --human-readable

<table>
<thead>
<tr>
<th>Node</th>
<th>DTotal</th>
<th>DFree</th>
<th>MTotal</th>
<th>MNode</th>
<th>MFree</th>
<th>Pinst</th>
<th>Sinst</th>
</tr>
</thead>
<tbody>
<tr>
<td>node1.example.com</td>
<td>928.8G</td>
<td>432.3G</td>
<td>4.0G</td>
<td>512M</td>
<td>13.5G</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>node2.example.com</td>
<td>928.8G</td>
<td>430.9G</td>
<td>4.0G</td>
<td>512M</td>
<td>14.8G</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>node3.example.com</td>
<td>928.8G</td>
<td>434.1G</td>
<td>4.0G</td>
<td>512M</td>
<td>14.7G</td>
<td>1</td>
<td>4</td>
</tr>
</tbody>
</table>
Cluster creation

```
node1# gnt-instance add --node node1:node2 \>
   --disk-template drbd --os-type etch mail1
```
Listing instances

node1# gnt-instance list --human-readable

Instance                  OS     Primary_node      Status   Memory
mail1.example.com         etch   node1.example.com running   512M
www1.example.com          etch   node3.example.com running   512M
john.example.com          suse   node2.example.com running  1024M
build-foo.example.com.com centos node2.example.com running  2048M

node1# gnt-instance list -o name,vcpus,os --no-headers --separator=: 
mail1.example.com:2:etch
www1.example.com:1:etch
john.example.com:1:suse
build-foo.example.com:2:centos
Node failure

• Power loss, hardware failure, etc.
node1# gnt-instance failover --ignore-consistency mail1
Secondary node failover

node1# gnt-instance replace-disks --on-secondary \ > --new-secondary=node1 mail1
After failover

• “node3” can be replaced
Overview

• Introduction
• Traditional clusters vs. Ganeti
• Design goals
• Cluster setup
• Instance failover example
• Usage in Google
• Open Source and Roadmap
Ganeti usage in Google

- 20-node Ganeti cluster
- 64-bit node OS
- 80 virtual instances
- Used for internal systems
- **Not** used for google.com
- Not targeted for resource intensive systems
  - Yes: DNS, DHCP, NTP, etc.
  - No: Fileserver
Open Source

• Code location: http://code.google.com/p/ganeti/
• License: GPL v2
• August 2007
  ▪ Ganeti 1.2 Beta 1 and Open Source
• February 2008
  ▪ Ganeti 1.2.3
• Late 2008
  ▪ Ganeti 1.3
Roadmap

• Job queue
• Granular locking
• Remote cluster API
• File-based storage
• Live failover
• Multiple coexisting hypervisors
Questions & Answers