Ganeti

Scalable Virtualization with Ganeti

by Lance Albertson

About Me

OSU Open Source Lab

Server hosting for Open Source projects

Lead Systems Administrator / Architect

Gentoo developer / contributor

Jazz trumpet performer

What I will cover

Ganeti terminology, comparisons, & goals

Cluster & virtual machine setup

Dealing with outages

Web Interface to Ganeti

OSUOSL usage of ganeti

Future roadmap

State of Virtualization

Citrix XenServer

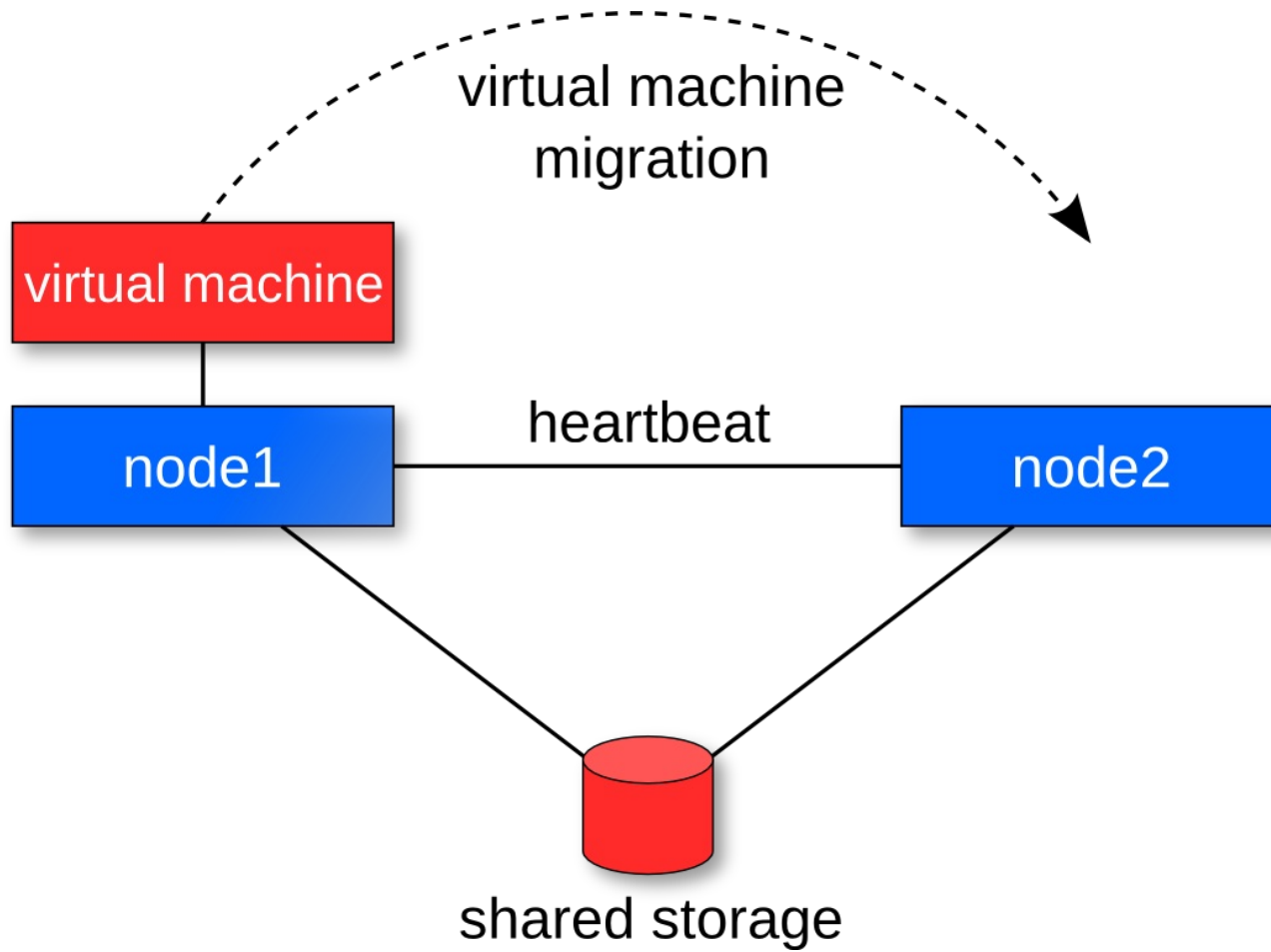libvirt: oVirt, virt-manager

Eucalyptus

VMWare

Open Stack

Issues

Overly complicated
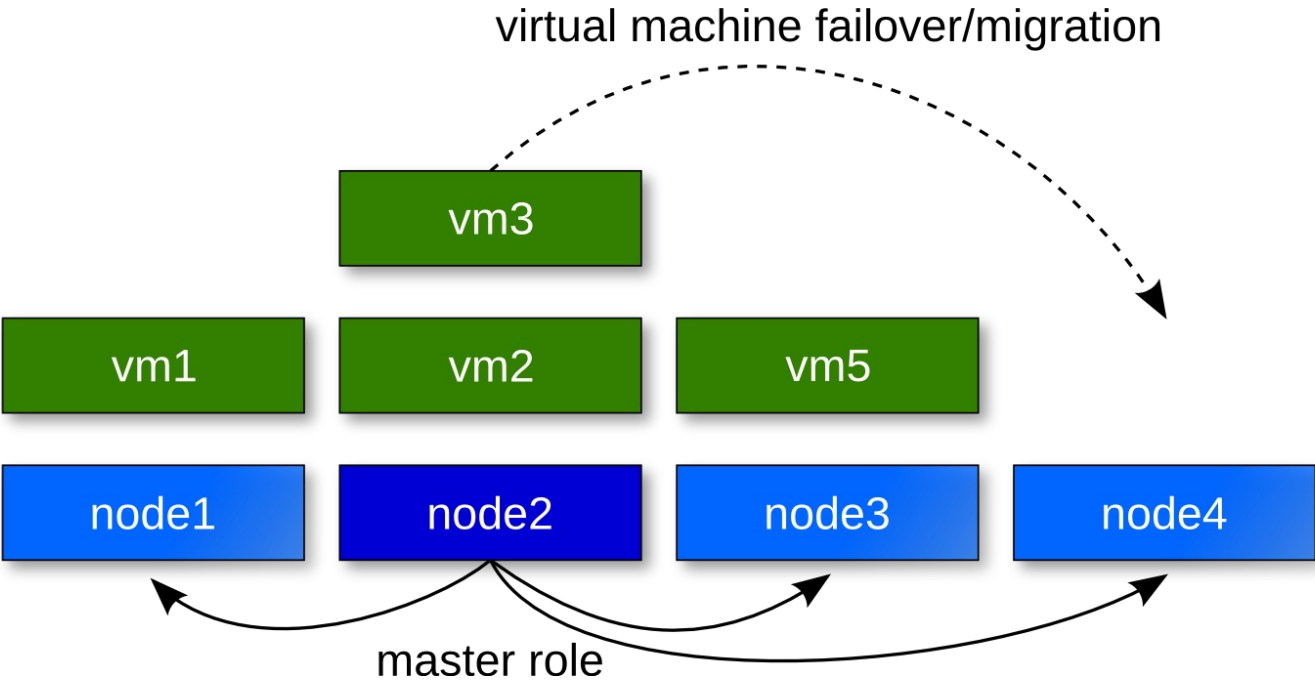
Multiple layers of software

Lack of HA Storage integration

Not always 100% open source

Traditional virtualization cluster

virtual machine migration

virtual machine

node1   heartbeat   node2

shared storage

Ganeti cluster

virtual machine failover/migration

vm3

vm1   vm2   vm5

node1   node2   node3   node4

master role

What is ganeti?

Software to manage a cluster of virtual servers

Project created and maintained by Google

Combines virtualization & data replication

Works with multiple hypervisors

Automates storage management

Automates OS deployment

Ganeti software requirements

Python

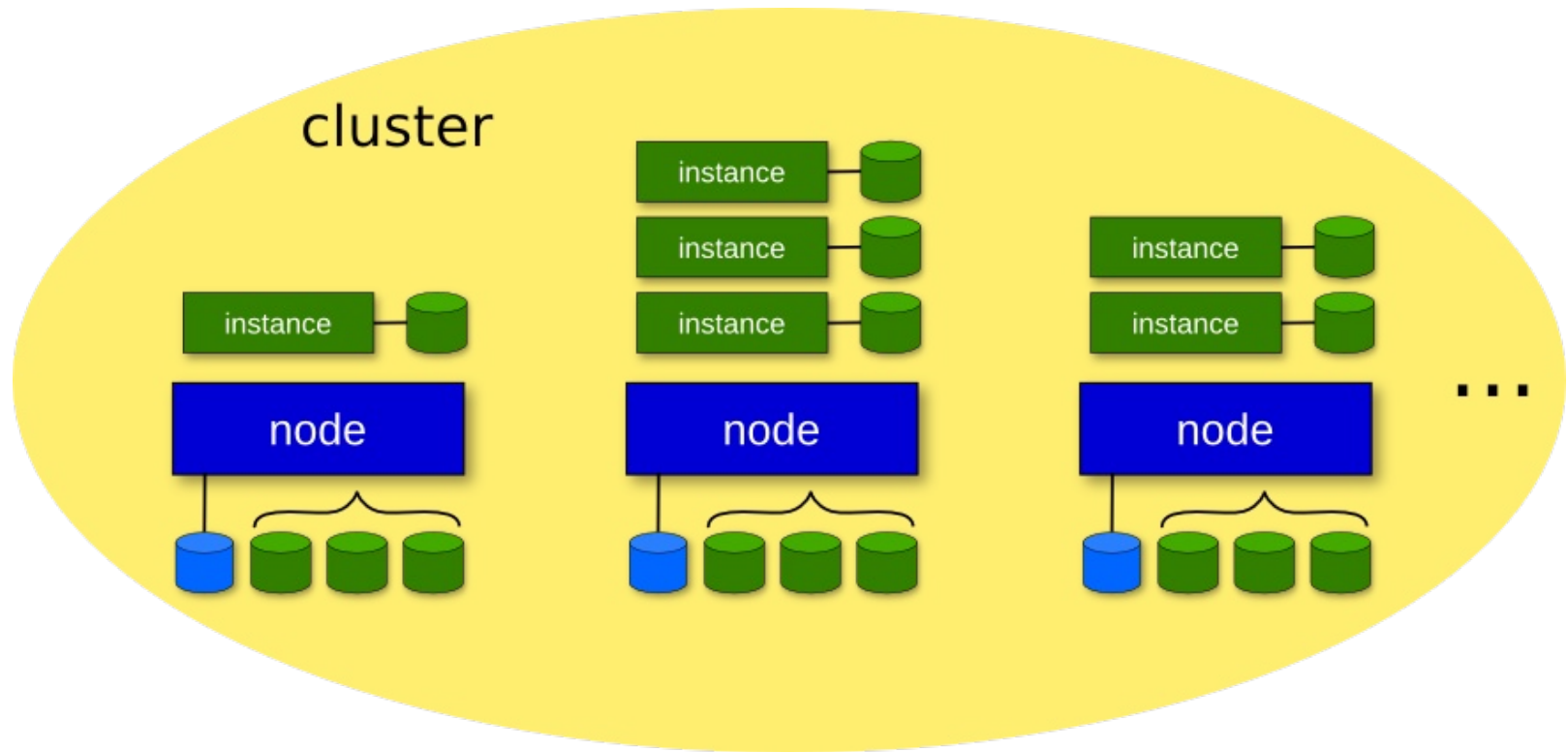various python modules

DRBD

LVM

KVM, Xen, or LXC*

Ganeti terminology

Cluster - group of nodes

Node - physical host

Instance - virtual machine, aka guest

Goals

Reduce hardware cost

Increase service availability

Simplify VM cluster management

Administration transparency

Principles

Not dependent on specific hardware

Scales linearly

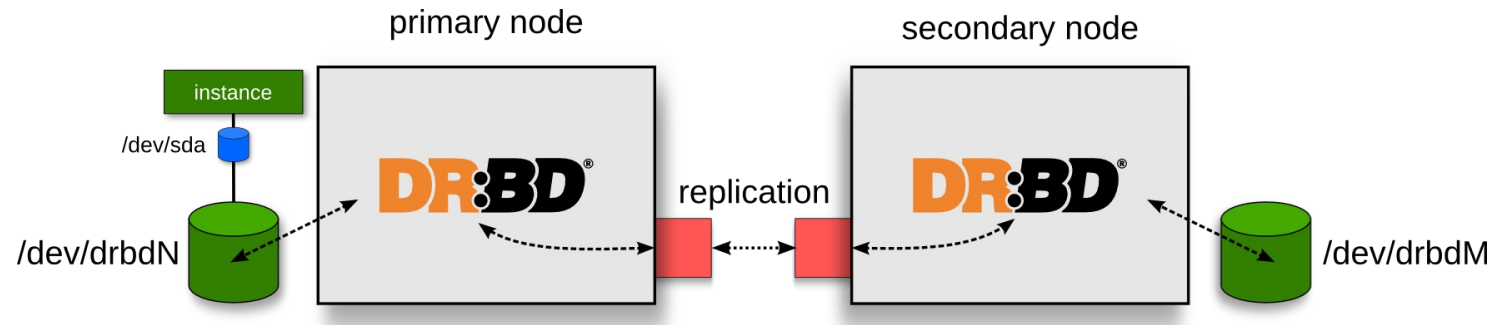Centralized administration

N+1 redundancy

Storage Options

LVM

LVM + DRBD (supports migration)

File based (raw, qcow2, etc)

Shared storage patch coming soon!

Storage: LVM + DRBD

Primary & secondary storage nodes

Each instance disk synced separately

Dedicated backend DRBD network

Allows instance failover & migration

Ganeti manages setup, starting/stopping DRBD devices

Ganeti administration

Command line based

Administration via single master node

All commands support interactive help

Consistent command line interface

gnt-<command>

Ganeti Commands

gnt-cluster

gnt-node

gnt-instance

gnt-backup

gnt-os

gnt-group (>= 2.4 only)

gnt-cluster

Cluster-wide configuration

Initialize & destroy cluster

Fail-over master node

Verify cluster integrity

gnt-node

Node-wide configuration/administration

Add & remove cluster nodes

Relocate all secondary instances from a node

List information about nodes

gnt-instance

Per-instance configuration/administration

Add, remove, rename, & reinstall instance

Serial console

Fail-over instance, change secondary

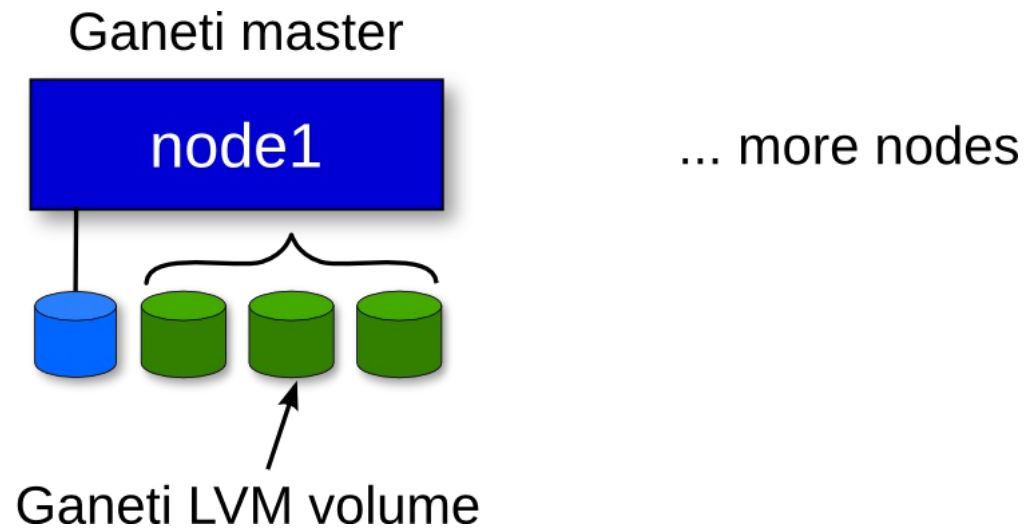Stop, start, migrate instance

List instance information

gnt-backup

Export instance to an image

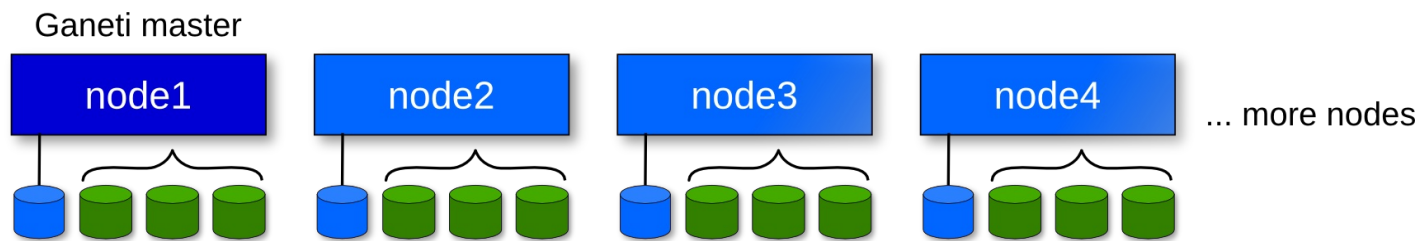Import instance from an exported image

Useful for inter-cluster migration

Cluster creation

Ganeti master

node1

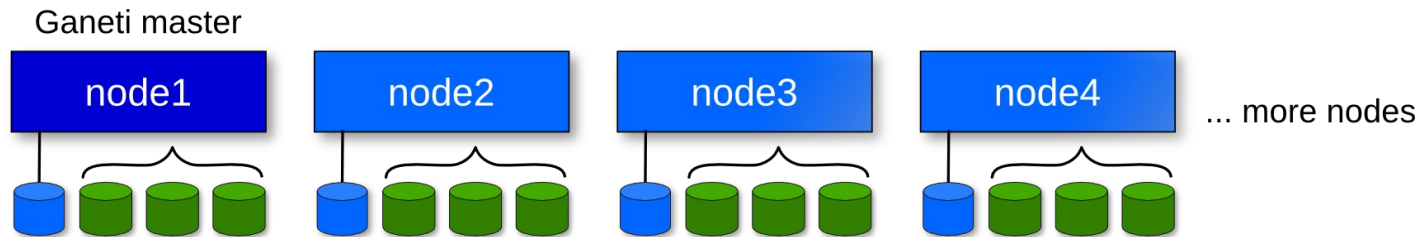... more nodes

Ganeti LVM volume

```
$ gnt-cluster init \
    --master-netdev=br42 \
    -g ganeti -s 10.1.11.200 \
    --enabled-hypervisors=kvm \
    -N link=br113 \
    -B vcpus=2,memory=512M \
    -H kvm:kernel_path=/boot/guest/vmlinuz-x86_64 \
    ganeti-cluster.osuosl.org
```
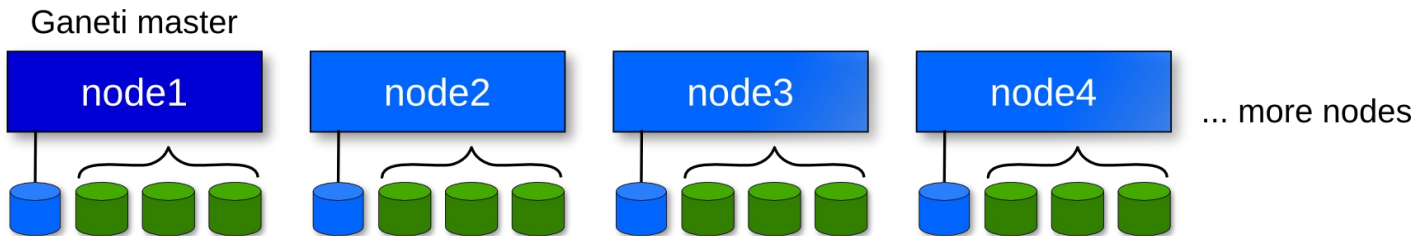
Adding nodes

Ganeti master

node1 node2 node3 node4 ... more nodes

```
$ gnt-node add -s 10.1.11.201 node2
```

Listing nodes

Ganeti master

| node1 | node2 | node3 | node4 | ... more nodes |

```
$ gnt-node list
Node          DTotal  DFree MTotal MNode MFree Pinst Sinst
g1.osuosl.bak 673.9G 163.8G  23.6G 16.8G  8.3G   18    18
g2.osuosl.bak 673.9G 149.2G  23.6G 16.1G 10.5G   18    17
g3.osuosl.bak 673.9G 120.5G  23.6G 16.3G  9.5G   18    18
g4.osuosl.bak 673.9G 100.0G  23.6G 16.4G  9.3G   17    18
```

Cluster verification

Ganeti master

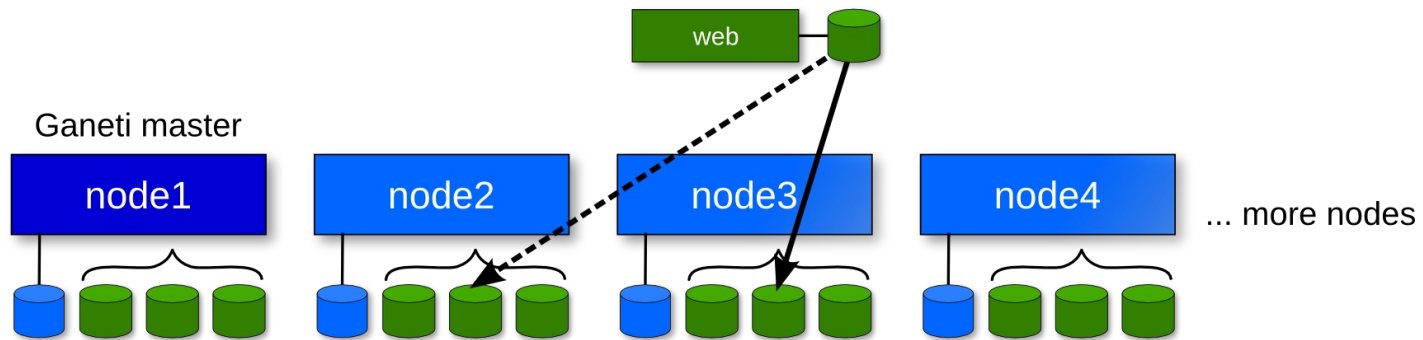node1 | node2 | node3 | node4 ... more nodes

```
$ gnt-cluster verify
Sun Feb 20 2011 * Verifying global settings
Sun Feb 20 2011 * Gathering data (4 nodes)
Sun Feb 20 2011 * Gathering disk information (4 nodes)
Sun Feb 20 2011 * Verifying node status
Sun Feb 20 2011 * Verifying instance status
Sun Feb 20 2011 * Verifying orphan volumes
Sun Feb 20 2011 * Verifying orphan instances
Sun Feb 20 2011 * Verifying N+1 Memory redundancy
Sun Feb 20 2011 * Other Notes
Sun Feb 20 2011 * Hooks Results
```

Cluster information

```
$ gnt-cluster info
Modification time: 2011-02-16 21:22:04
Master node: g1.osuosl.bak
Architecture (this node): 64bit (x86_64)
Tags: (none)
Default hypervisor: kvm
Enabled hypervisors: kvm
Hypervisor parameters:
  - kvm:
      boot_order: disk
      disk_type: paravirtual
      initrd_path:
      kernel_args: ro
      kernel_path: /boot/guest/vmlinuz-x86_64-hardened
      nic_type: paravirtual
      root_path: /dev/vda2
      serial_console: True
      vnc_bind_address: 0.0.0.0
OS-specific hypervisor parameters:
OS parameters:
Cluster parameters:
  - candidate pool size: 4
  - master netdev: br42
  - lvm volume group: ganeti
  - lvm reserved volumes: (none)
  - drbd usermode helper: /bin/true
  - file storage path: /var/lib/ganeti/export
  - maintenance of node health: False
  - uid pool:
  - default instance allocator: hail
  - primary ip version: 4
...
```

Creating an instance

Ganeti master

node1   node2   node3   node4   ... more nodes

web

```
$ gnt-instance add -t drbd -n node3:node2 \
$    -s 10G -o image+gentoo-hardened-cf \
$    --net 0:link=br42  web.example.org
* creating instance disks...
adding instance web.example.org to cluster config
 - INFO: Waiting for instance web.example.org to sync disks.
 - INFO: - device disk/0:  3.90% done, 205 estimated seconds remaining
 - INFO: - device disk/0: 29.40% done, 101 estimated seconds remaining
 - INFO: - device disk/0: 54.90% done, 102 estimated seconds remaining
 - INFO: - device disk/0: 80.40% done, 41 estimated seconds remaining
 - INFO: - device disk/0: 98.40% done, 3 estimated seconds remaining
 - INFO: - device disk/0: 100.00% done, 0 estimated seconds remaining
 - INFO: Instance web.example.org's disks are in sync.
* running the instance OS create scripts...
* starting instance...
```

List all instances

```
$ gnt-instance list
Instance    OS                    Primary_node   Status    Memory
ads         image+debian-lenny        phobos          running   1.0G
area51      image+debian-lenny         deimos          running   3.0G
code        image+debian-lenny         deimos          running   4.0G
db1         image+gentoo-hardened-cf phobos           running   4.0G
db2         image+gentoo-hardened-cf deimos           running   4.0G
demo        image+debian-lenny         deimos          running   512M
lists       image+gentoo-hardened-cf phobos           running   2.0G
mail        image+debian-lenny         deimos          running   1.0G
misc        image+debian-lenny         deimos          running   2.0G
testing     image+debian-lenny         phobos          running   2.0G
www         image+gentoo-hardened-cf phobos           running   2.0G
```

Other instance commands

```
$ gnt-instance console web

$ gnt-instance migrate web

$ gnt-instance failover web

$ gnt-instance reinstall -o image+ubuntu-lucid web

$ gnt-instance info web

$ gnt-instance list
```

Guest OS Installation

Bash scripts

Format, mkfs, mount, install OS

Hooks

OS Definitions

debootstrap

Disk image

Other OS-specific

ganeti-instance-image

Disk image based (filesystem dump or tarball)

Flexible OS support

Fast instance deployment (~30 seconds)

ganeti-instance-image

Setup serial for grub, grub2, & login prompt

Automatic networking setup (DHCP or static)

Automatic ssh hostkey regen

Add optional kernel parameters to grub

Primary node failure

Ganeti master

node1 node2 node3 node4 ... more nodes

web web

$ gnt-instance failover --ignore-consistency web

Secondary node failure


```
$ gnt-instance replace-disks --on-secondary \
    --new-secondary=node1 web
```

Ganeti htools

Automatic allocation tools

Cluster rebalancer - **hbal**

IAllocator plugin - **hail**

Cluster capacity estimator - **hspace**

hbal

```
$ hbal -m ganeti.osuosl.bak
Loaded 4 nodes, 63 instances
Initial check done: 0 bad nodes, 0 bad instances.
Initial score: 0.53388595
Trying to minimize the CV...
    1. bonsai          g1:g2 => g2:g1 0.53220090 a=f
    2. connectopensource g3:g1 => g1:g3 0.53114943 a=f
    3. amahi           g2:g3 => g3:g2 0.53088116 a=f
    4. mertan          g1:g2 => g2:g1 0.53031862 a=f
    5. dspace          g3:g1 => g1:g3 0.52958328 a=f
Cluster score improved from 0.53388595 to 0.52958328
Solution length=5
```

hspace

```
$ hspace --memory 512 --disk 10240 -m ganeti.osuosl.bak
HTS_INI_INST_CNT=63

HTS_FIN_INST_CNT=101

HTS_ALLOC_INSTANCES=38
HTS_ALLOC_FAIL_REASON=FAILDISK
```

hail

```
$ gnt-instance add -t drbd -I hail \
$    -s 10G -o image+gentoo-hardened-cf \
$    --net 0:link=br42  web.example.org \
 - INFO: Selected nodes for instance web.example.org
        via iallocator hail: gtest1.osuosl.bak, gtest2.osuosl.bak
* creating instance disks...
adding instance web.example.org to cluster config
 - INFO: Waiting for instance web.example.org to sync disks.
 - INFO: - device disk/0:  3.60% done, 1149 estimated seconds remaining
 - INFO: - device disk/0: 29.70% done, 144 estimated seconds remaining
 - INFO: - device disk/0: 55.50% done, 88 estimated seconds remaining
 - INFO: - device disk/0: 81.10% done, 47 estimated seconds remaining
 - INFO: Instance web.example.org's disks are in sync.
* running the instance OS create scripts...
* starting instance...
```

Ganeti Web Manager

http://code.osuosl.org/projects/ganeti-webmgr

Ganeti Web Manager

Django based front-end for Ganeti

OSUOSL funded project

Includes a permission & quota system

Uses Ganeti RAPI interface

HTML5 console using noVNC

Ganeti usage at OSUOSL

4-node production OSUOSL cluster

~71 virtual instances

qemu-kvm 0.12.x

64bit Gentoo Linux

Node details

4 x HP DL360 G4

24G RAM

630G - RAID5 6x146G 10K SCSI HDDs
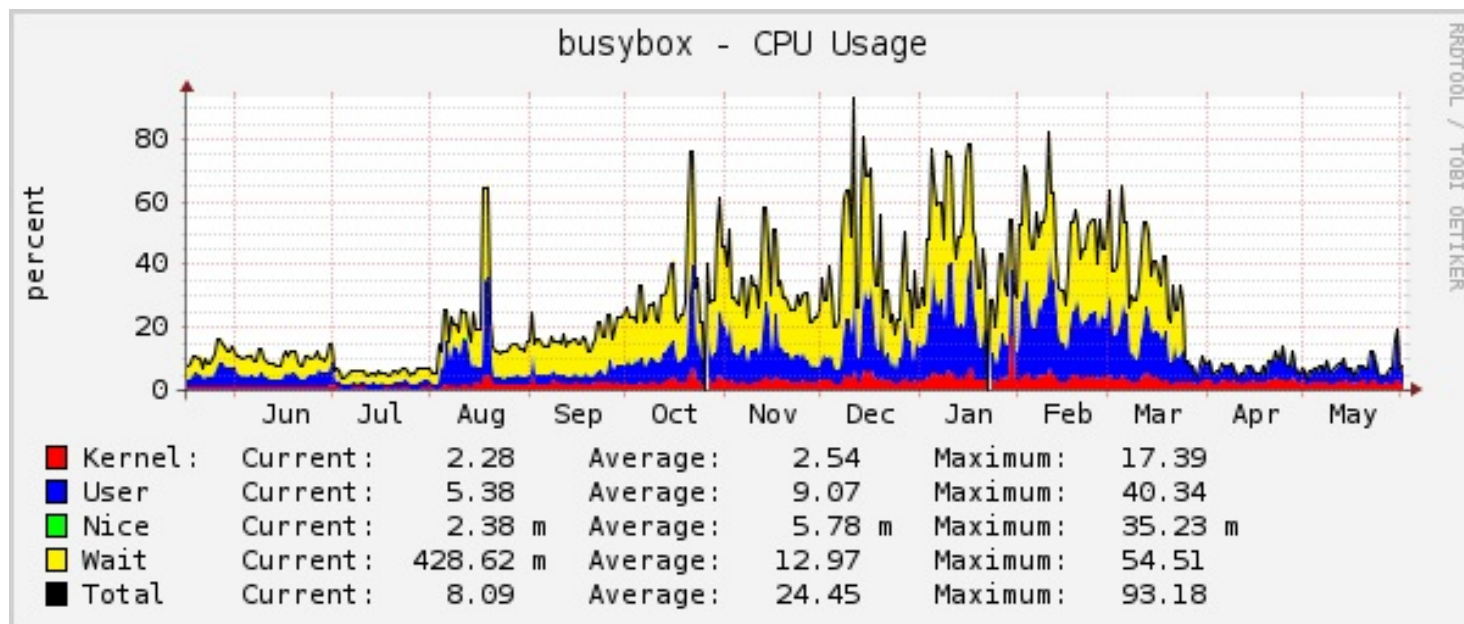
Project Ganeti clusters

OSGeo - 9 instances / 2 nodes
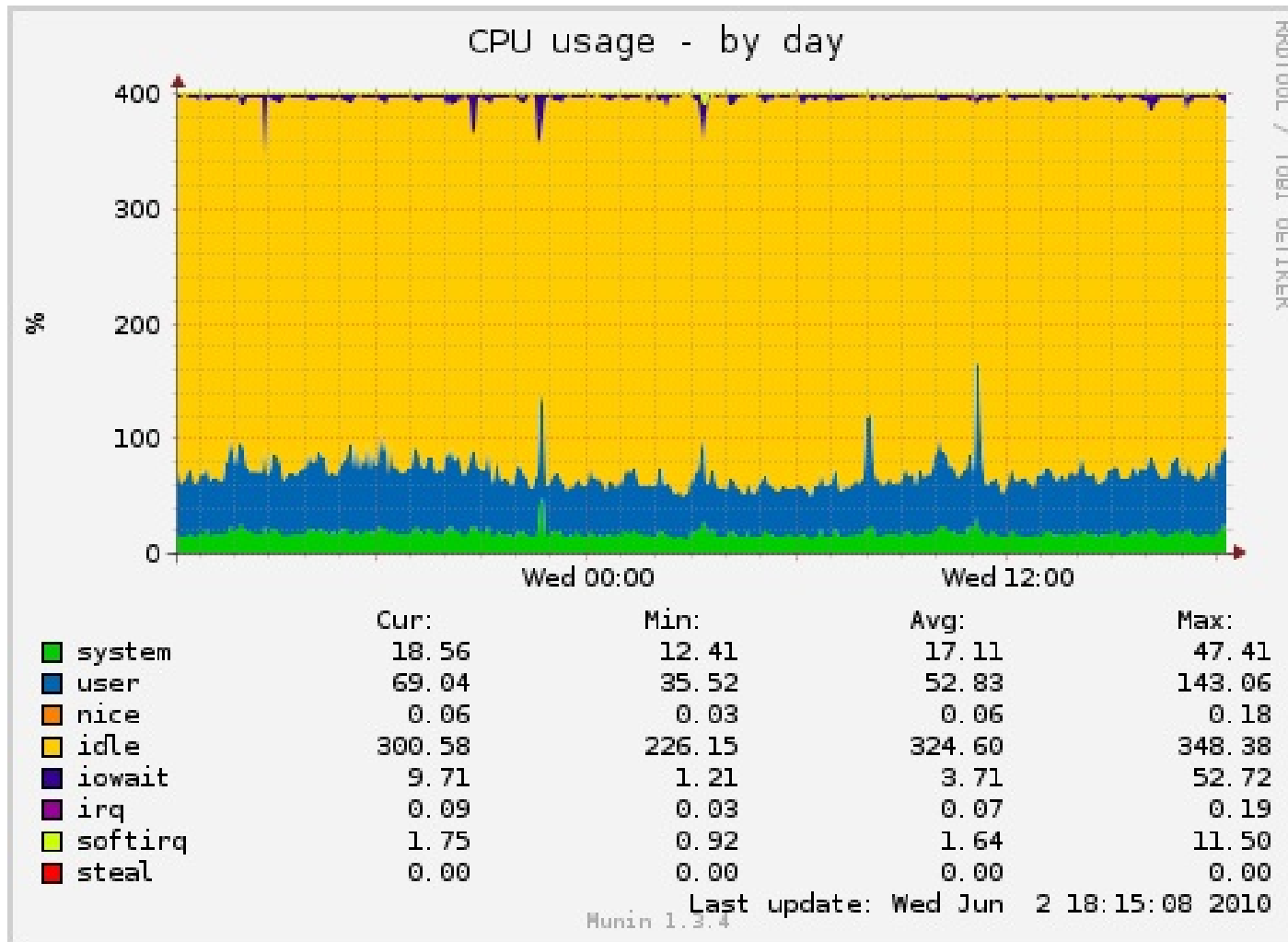
OSDV - 5 instances / 3 nodes

phpBB - 11 instances / 2 nodes
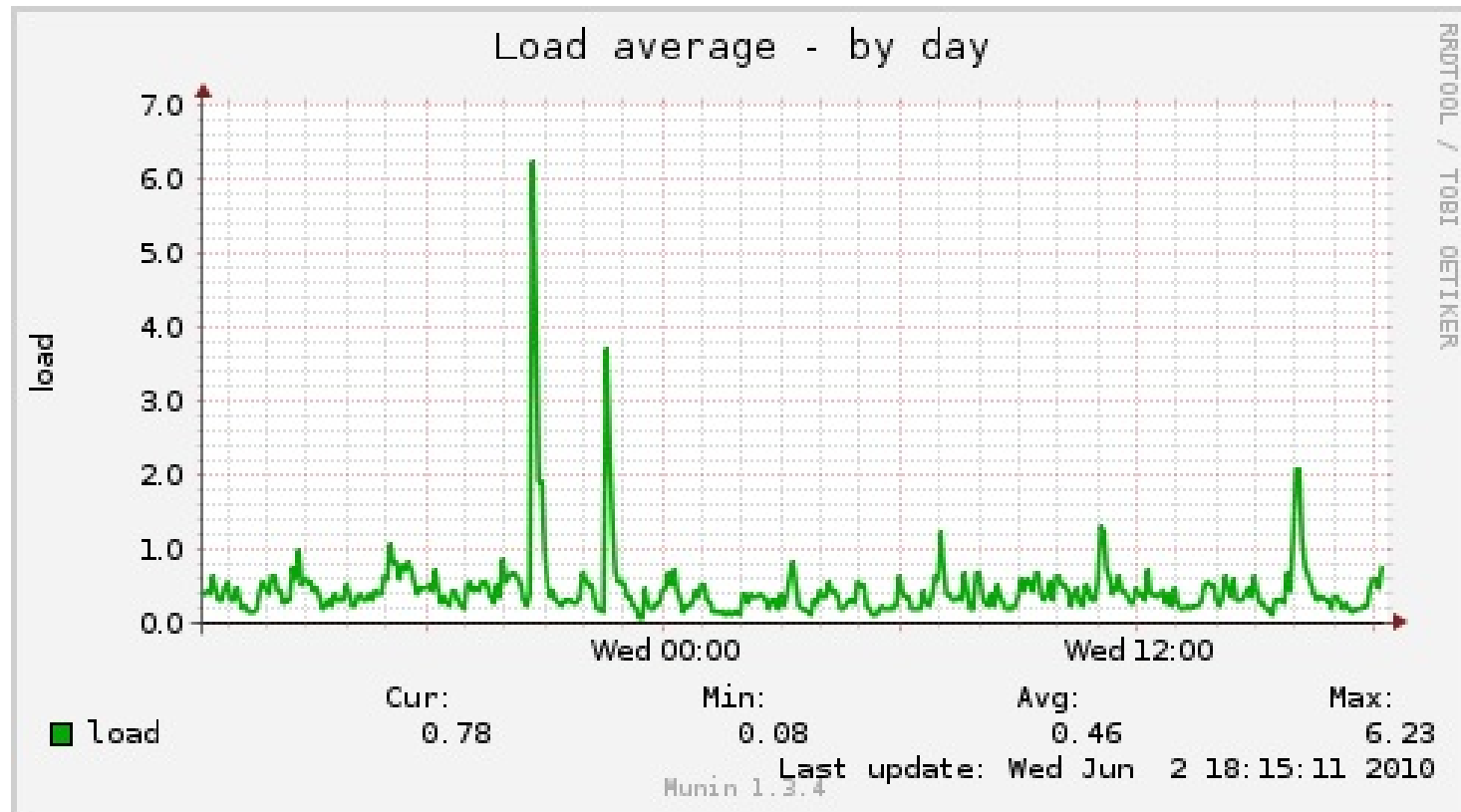
ORVSD - 11 instances / 2 nodes

Xen + iSCSI vs. kvm + DRBD

busybox - CPU Usage

			Current			Average			Maximum	
■	Kernel:	Current:	2.28	Average:	2.54	Maximum:	17.39
■	User	Current:	5.38	Average:	9.07	Maximum:	40.34
■	Nice	Current:	2.38 m	Average:	5.78 m	Maximum:	35.23 m
■	Wait	Current:	428.62 m	Average:	12.97	Maximum:	54.51
■	Total	Current:	8.09	Average:	24.45	Maximum:	93.18

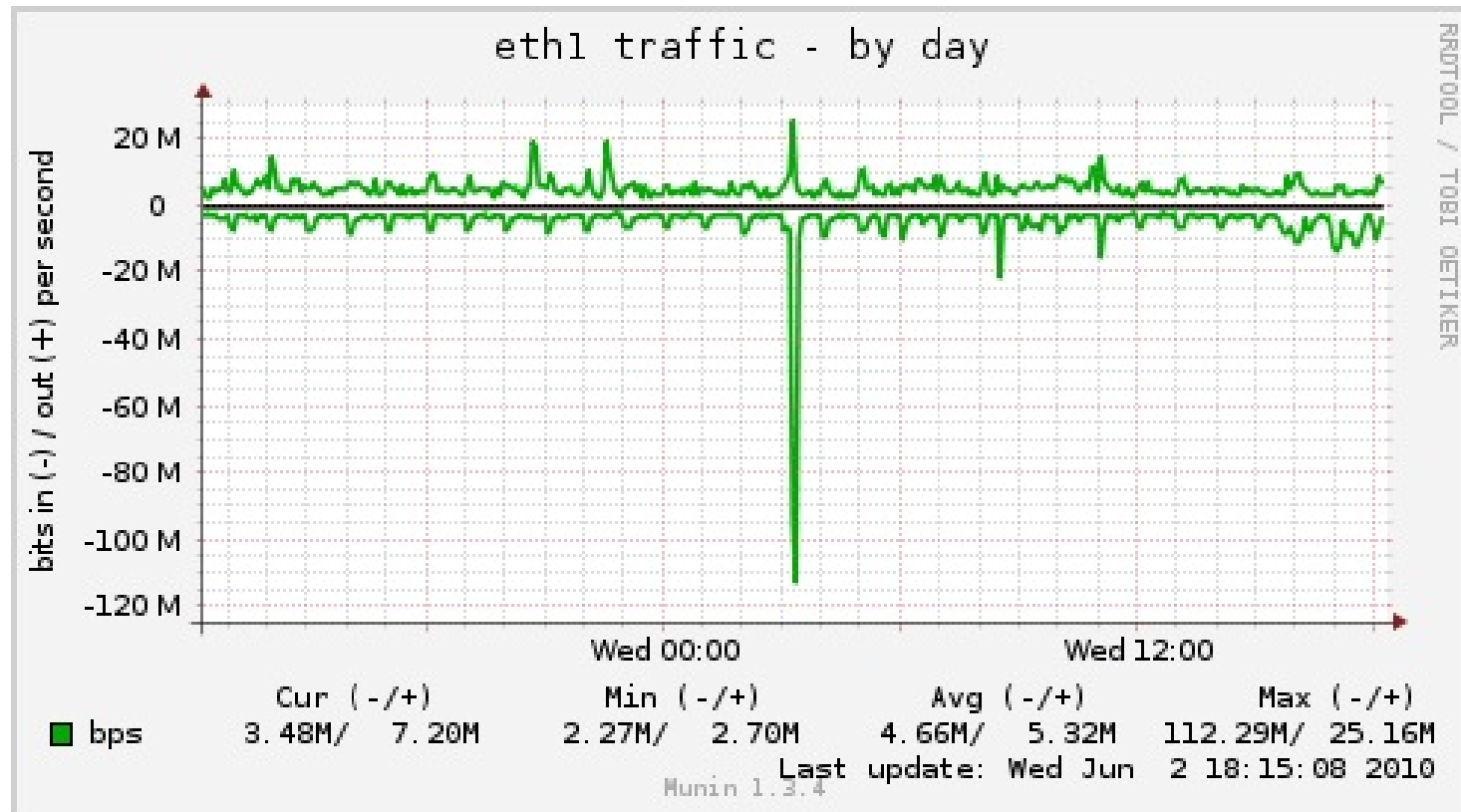RRDTOOL / TOBI OETIKER

Ganeti node CPU usage

Ganeti node LOAD

Ganeti node DRBD network

OSUOSL future ganeti plans

KSM (Kernel SamePage Merging)

Puppet integration

Web-based tools - In Progress

Open source

http://code.google.com/p/ganeti/

License: GPL v2

Ganeti 1.2.0 - December 2007

2.0.0 - May 2009, 2.1.0 March 2010

Ganeti 2.2.0 - Oct 2010 / 2.2.2 current

Ganeti 2.3.0 - Dec 2010 / 2.3.1 current

Ganeti roadmap

LXC support - 2.2

Inter-cluster instance moves - 2.2

KVM security - 2.2

IPv6 - 2.3

Privilege Separation - 2.2/2.3

Node Groups - 2.4

OOB node management - 2.4

Shared storage - 2.5?

Resources

http://code.google.com/p/ganeti/ - main project website

http://code.google.com/p/ganeti/downloads/ - Ganeti-FISL-2008.pdf

http://code.osuosl.org/projects/ganeti-image

http://code.osuosl.org/projects/ganeti-webmgr

Questions?

lance@osuosl.org

@ramereth on twitter

Ramereth on freenode

blog: http://www.lancealbertson.com

slides: http://tinyurl.com/scale9x-ganeti

Presentation made with showoff

(http://github.com/schacon/showoff)

Demo

Create instance

Migrate instance

Fail-over instance

Re-install instance

Ganeti Web Manager