



# Ganeti

Creating a low-cost clustered virtualization environment

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

OSUOSL usage of ganeti

Future roadmap

# Current solutions

Citrix XenServer

libvirt: oVirt, virt-manager

Eucalyptus

VMWare

Open Stack\*

# Issues

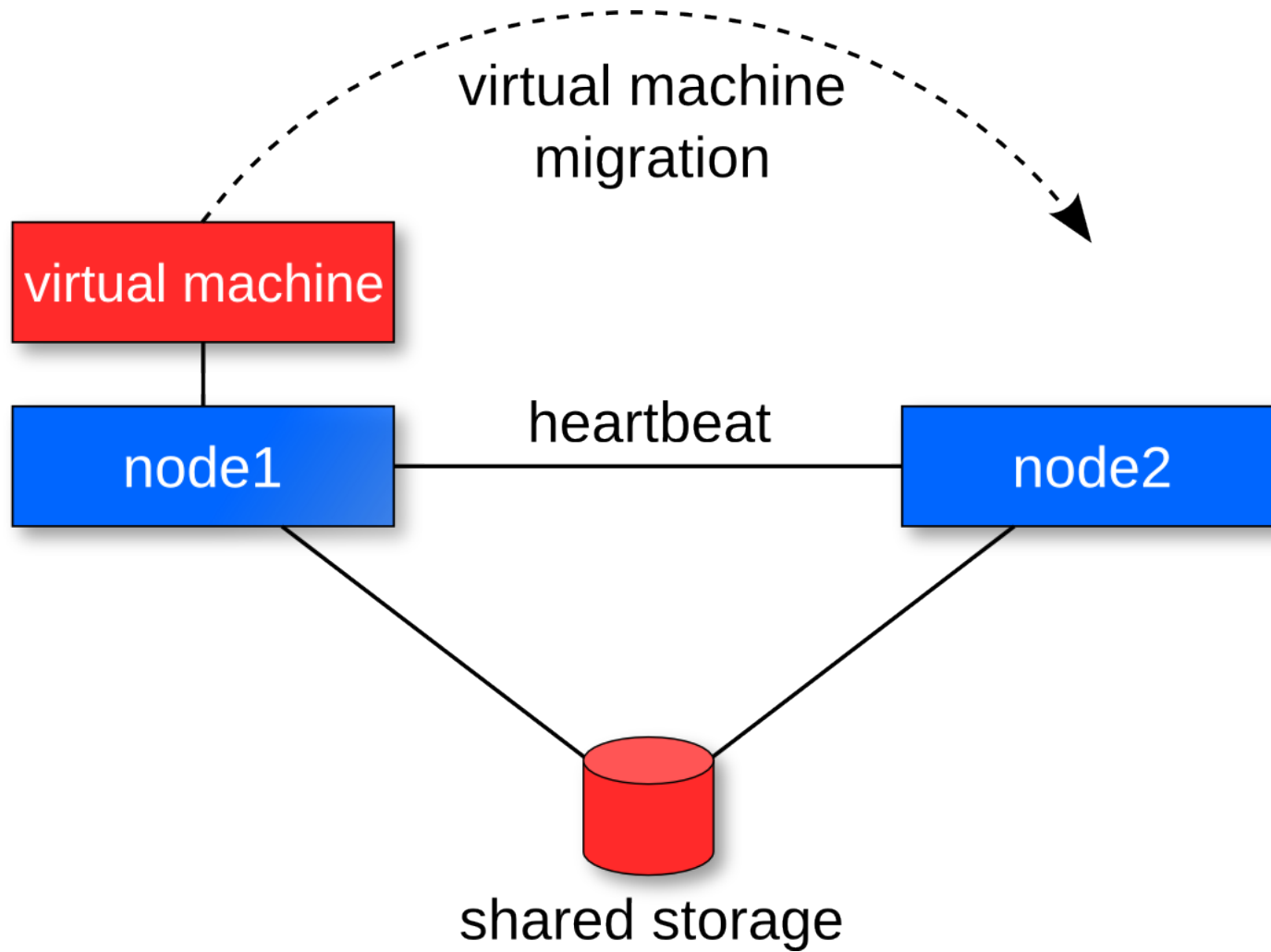
Overly complicated

Lack of HA Storage integration

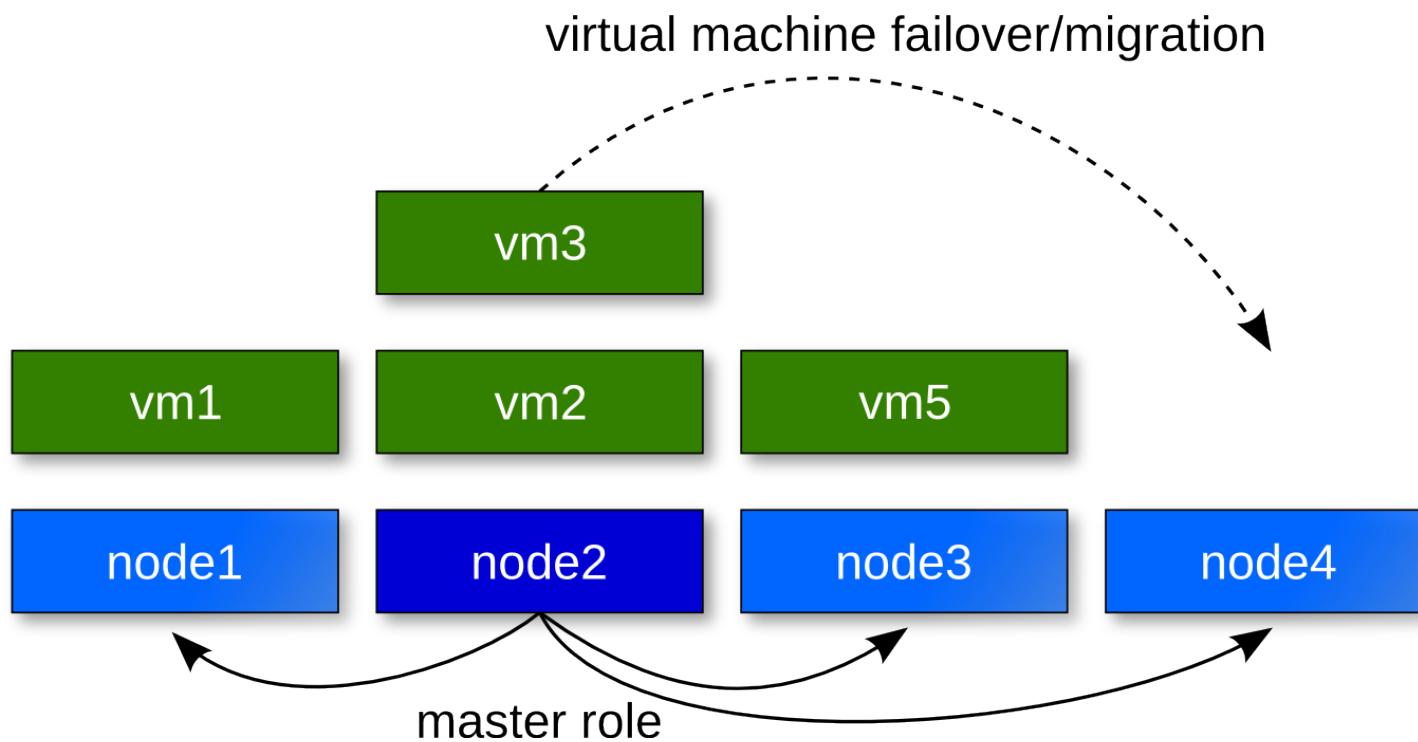
Not always 100% open source

Multiple layers of software

# Traditional virtualization cluster



# Ganeti cluster



# What is ganeti?

Software to manage a cluster of virtual servers

Combines virtualization & data replication

Automates storage management

Automates OS deployment

Project created and maintained by Google



# Ganeti software requirements

Python



simplejson



DRBD



LVM

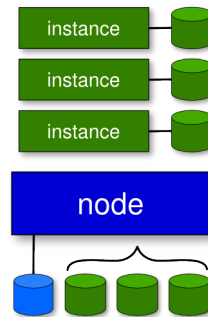


KVM and/or Xen

# Ganeti terminology

Node - physical host

Instance - virtual machine, aka guest



# Goals

Reduce hardware cost

Increase service availability

Increase management flexibility

Administration transparency

# Principles

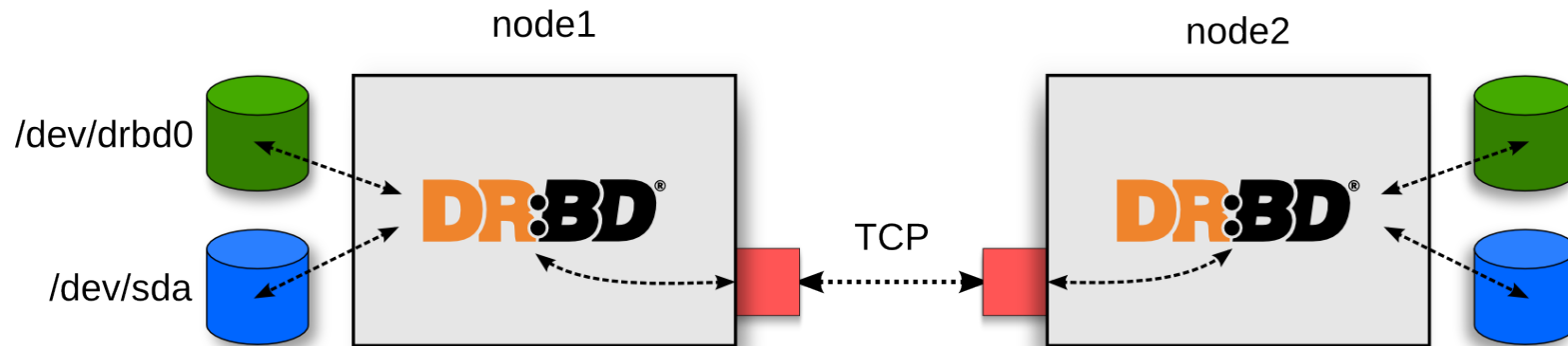
Not dependent on specific hardware

Scales linearly

Single node takes admin master role

N+1 redundancy

# Storage replication: DRBD



Primary & secondary storage nodes

Each instance LVM volume synced separately

Dedicated backend DRBD network

Allows instance failover & migration

# 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-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

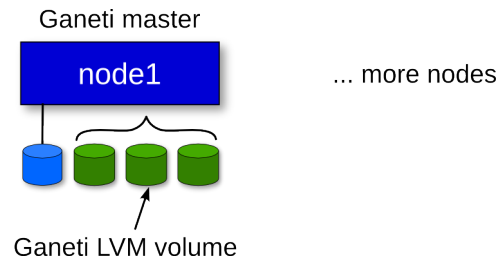
# gnt-backup

Export instance to an image

Import instance from an exported image

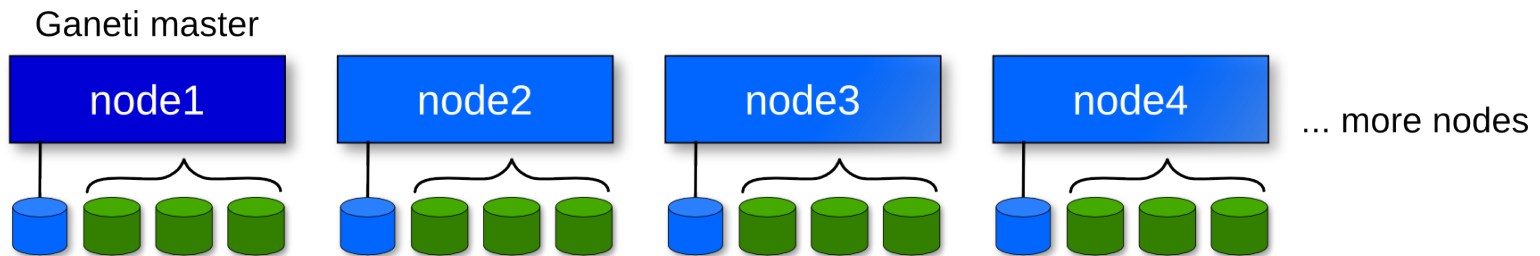
Useful for inter-cluster migration

# Cluster creation



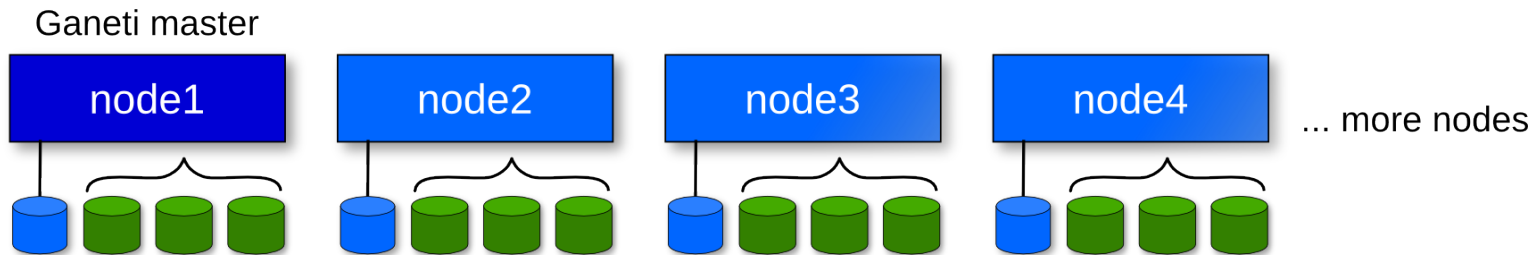
```
$ gnt-cluster init \  
  --master-netdev=br42 \  
  -g ganeti -s 10.1.11.200 \  
  --enabled-hypervisors=kvm \  
  -N link=br113 \  
  -B vcpus=2,memory=512M \  
  -H kvm:kernel_path=/boot/guest/vmlinuz-x86_64 \  
  ganeti-cluster.osuosl.org
```

# Adding nodes



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

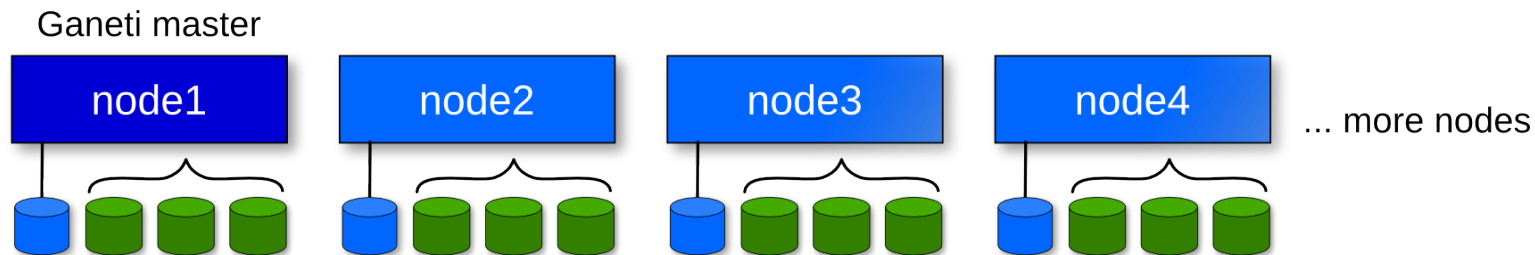
# Listing nodes



```
$ gnt-node list
```

Node	DTotal	DFree	MTotal	MNode	MFree	Pinst	Sinst
g1.osuosl.bak	673.9G	251.8G	23.6G	14.5G	14.0G	16	16
g2.osuosl.bak	673.9G	204.9G	23.6G	15.5G	14.2G	15	16
g3.osuosl.bak	673.9G	200.6G	23.6G	16.8G	13.3G	16	16
g4.osuosl.bak	673.9G	154.8G	23.6G	16.4G	15.4G	16	15

# Cluster verification



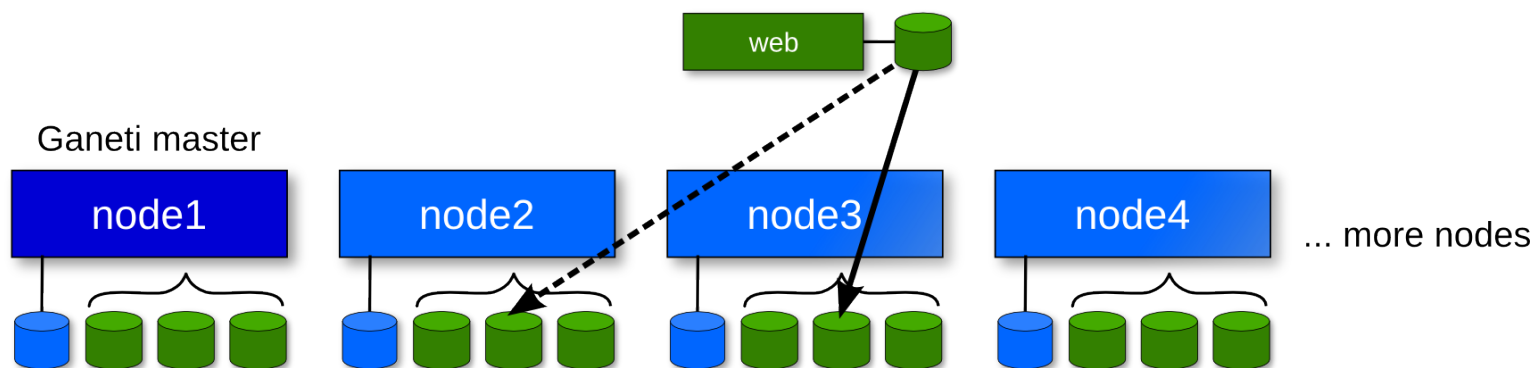
```
$ gnt-cluster verify
Wed Jun  2 17:31:07 2010 * Verifying global settings
Wed Jun  2 17:31:08 2010 * Gathering data (4 nodes)
Wed Jun  2 17:31:09 2010 * Verifying node status
Wed Jun  2 17:31:09 2010 * Verifying instance status
Wed Jun  2 17:31:09 2010 * Verifying orphan volumes
Wed Jun  2 17:31:09 2010 * Verifying oprhan instances
Wed Jun  2 17:31:09 2010 * Verifying N+1 Memory redundancy
Wed Jun  2 17:31:09 2010 * Other Notes
Wed Jun  2 17:31:09 2010 * Hooks Results
```

# Cluster information

```
$ gnt-cluster info
Cluster name: ganeti-test.osuosl.bak
Cluster UUID: a22576ba-9158-4336-8590-a497306f84b9
Creation time: 2010-04-08 00:08:29
Modification time: 2010-05-07 22:33:34
Master node: gtest1.osuosl.bak
Architecture (this node): 64bit (x86_64)
Tags: (none)
Default hypervisor: kvm
Enabled hypervisors: kvm
Hypervisor parameters:
- kvm:
  acpi: True
  boot_order: disk
  cdrom_image_path:
  disk_cache: default
  disk_type: paravirtual
  initrd_path:
  kernel_args: ro
  kernel_path: /boot/guest/vmlinuz-x86_64-hardened
  kvm_flag:
  migration_port: 8102
  nic_type: paravirtual
  root_path: /dev/vda2
  security_domain:
  security_model: none
  serial_console: True
  usb_mouse:
  use_localtime: False
  vnc_bind_address: 0.0.0.0
  vnc_password_file:
  ....
```



# Creating an instance



```
$ 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
monkeyhttpd image+ubuntu-lucid g2.osuosl running 512M
mozdev-stats image+manual g3.osuosl running 512M
mulgara image+manual g4.osuosl running 512M
musicbrainzvm image+manual g2.osuosl running 512M
myrtle image+manual g1.osuosl running 512M
olpc image+manual g3.osuosl running 512M
openberry image+manual g1.osuosl running 512M
openclipfont image+manual g4.osuosl running 512M
openht image+manual g4.osuosl running 512M
openmrs image+manual g1.osuosl running 512M
openvoting image+manual g2.osuosl running 512M
osi image+manual g4.osuosl running 256M
parrotvm image+manual g1.osuosl running 512M
pcc image+manual g1.osuosl running 512M
pdxplumbers image+manual g2.osuosl running 512M
polk image+manual g4.osuosl running 512M
puffin image+manual g3.osuosl running 256M
```

# 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

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

Disk image based (filesystem dump or tarball)

Flexible OS support

Fast instance deployment ( ~30 seconds)

# ganeti - instance - image

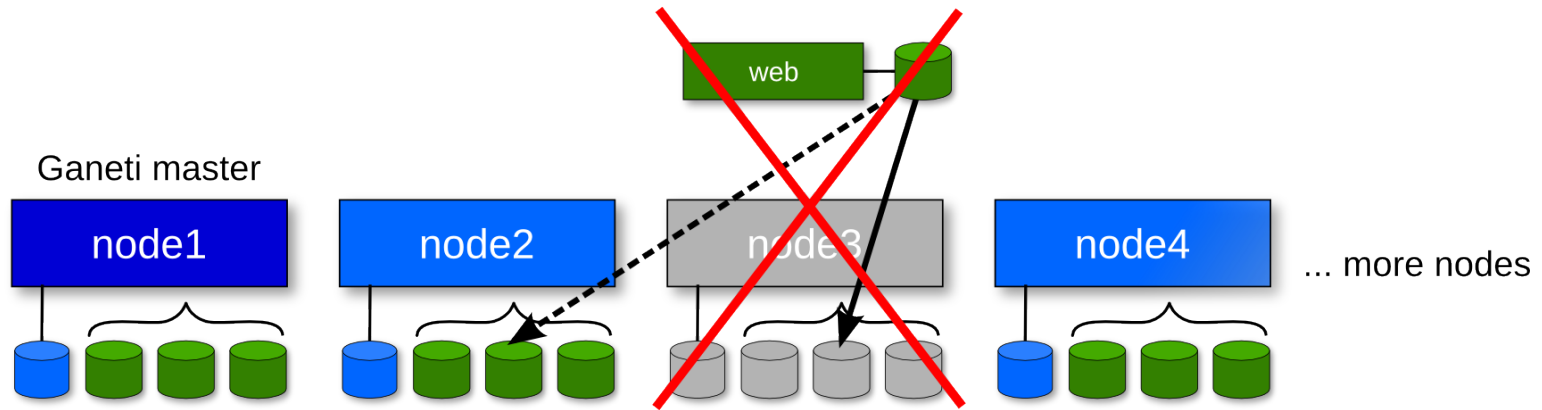
Setup serial for grub, grub2, & login prompt

Automatic networking setup (DHCP or static)

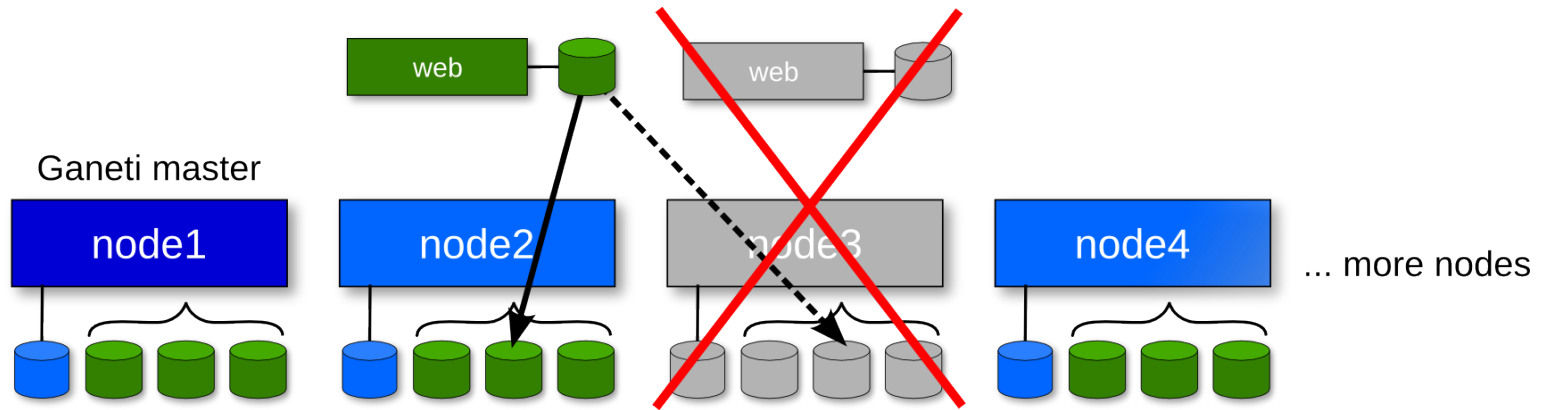
Automatic ssh hostkey regen

Add optional kernel parameters to grub

# Primary node failure



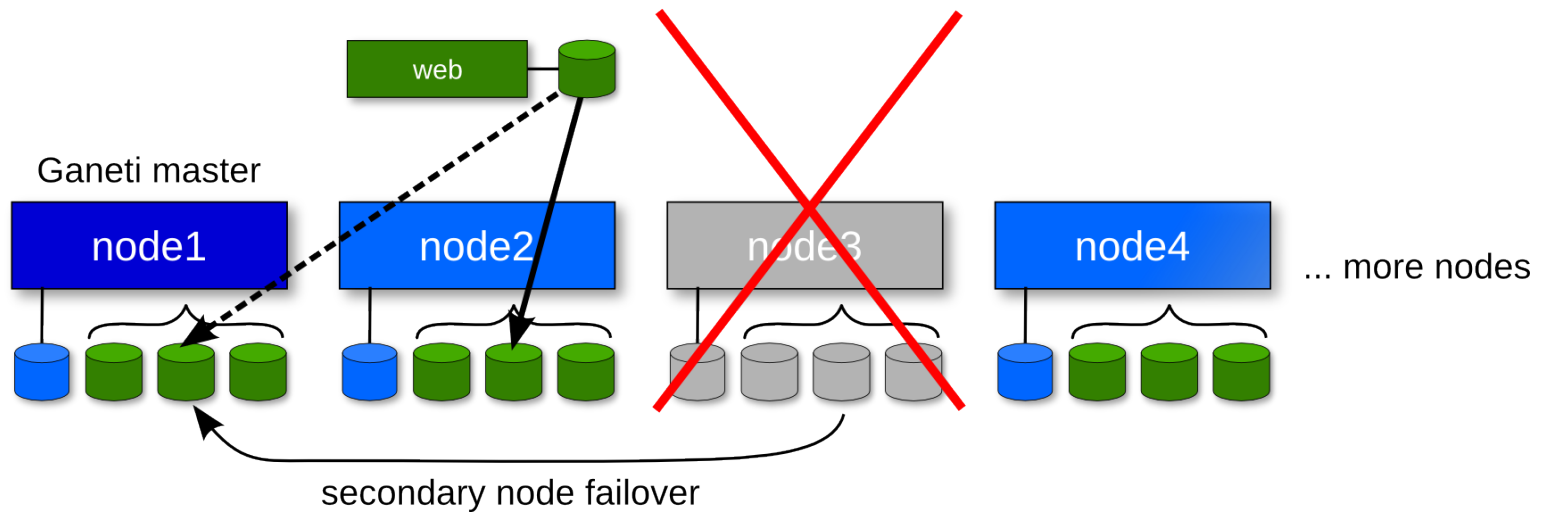
# Primary node failure



```
$ 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**

IAAllocator 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. connectopenso 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

**Ganeti Web** osladmin | [logout](#)

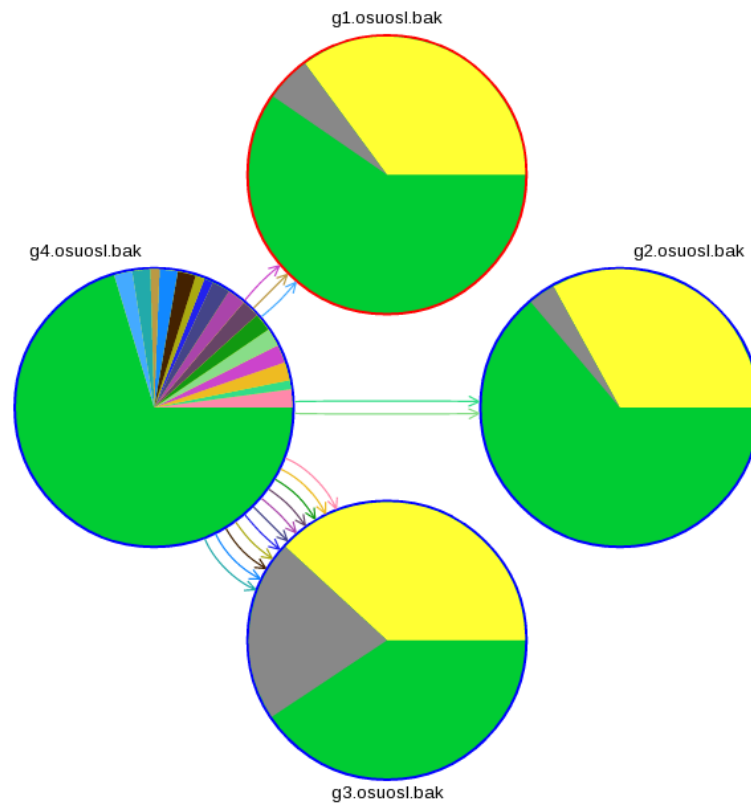
Cluster Overview | Nodes | Instances | **Visualization**

## Key

- Allocated Memory
- Unallocated Memory
- Master Node
- Master Candidate

## Instances

- arklinuxvm.osuosl.org
- flossfoundations.osuosl.org
- etherpad.osuosl.org
- polk.osuosl.org
- mulgara.osuosl.org
- openhht.osuosl.org
- twist.osuosl.org
- openclipfont.osuosl.org
- infamy.linuxlookup.com
- osi.osuosl.org
- sugarlabs.osuosl.org
- jaws.osuosl.org
- puppet.osuosl.org
- tillikum.osuosl.org
- linuxfund.osuosl.org
- webdav.osuosl.org



# Ganeti usage at OSUOSL

4-node production OSUOSL cluster

Project clusters (OSGeo, ORVSD, OSDV, phpBB, etc)

~64 virtual instances

qemu-kvm 0.11.x

64bit Gentoo Linux

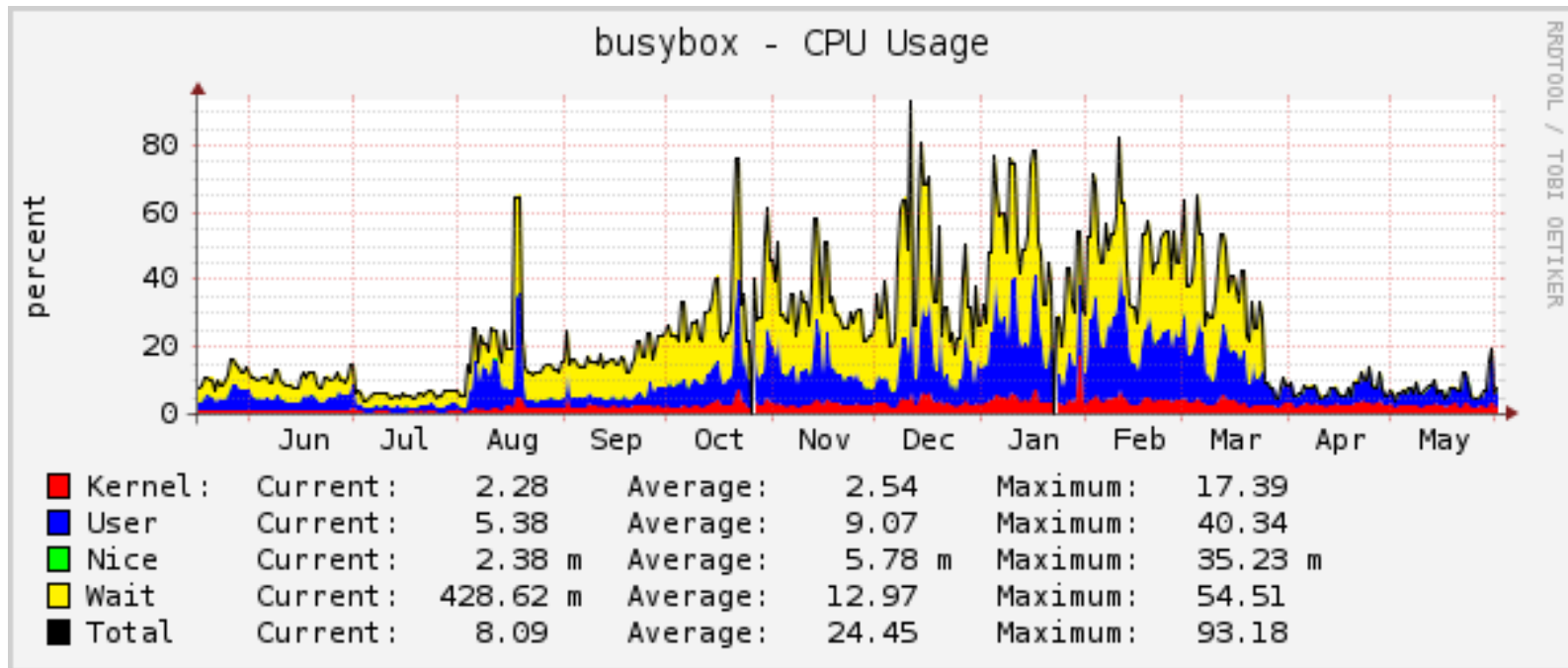
## Node details

DL360 G4

24G RAM

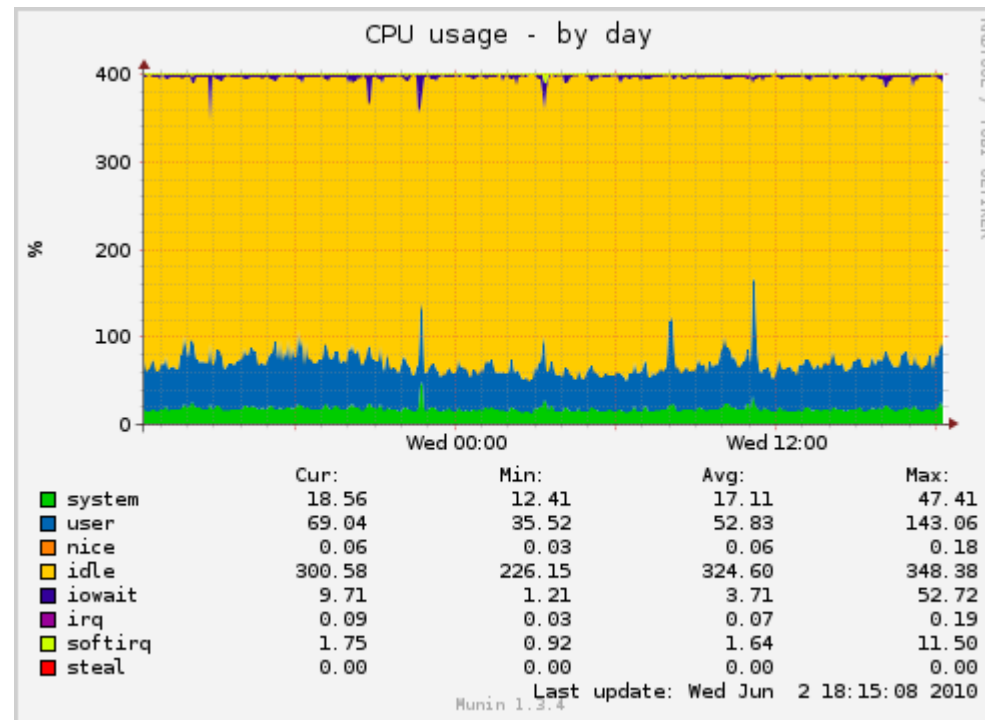
630G - RAID5 6x146G 10K SCSI HDDs

# Xen + iSCSI vs. kvm + DRBD

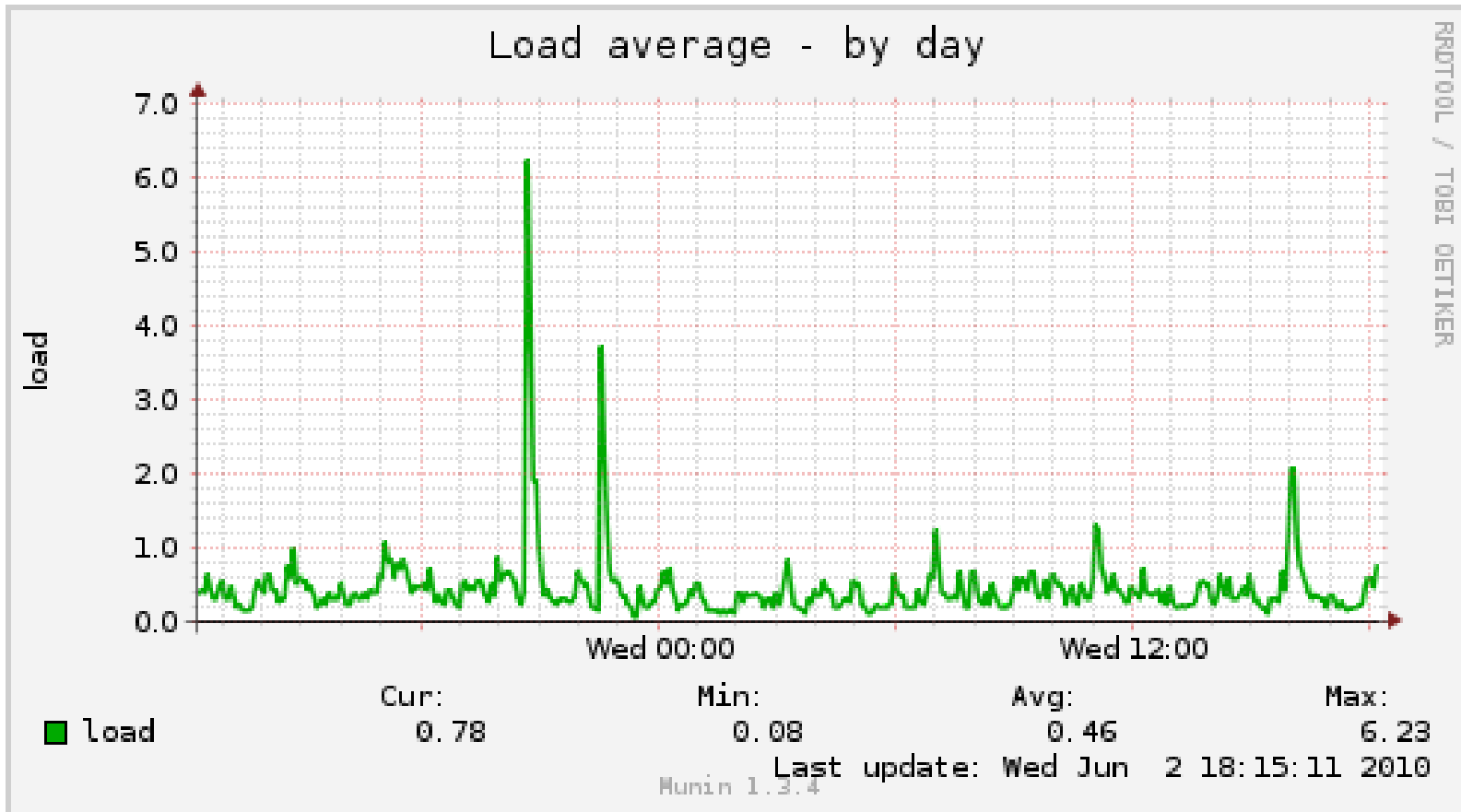




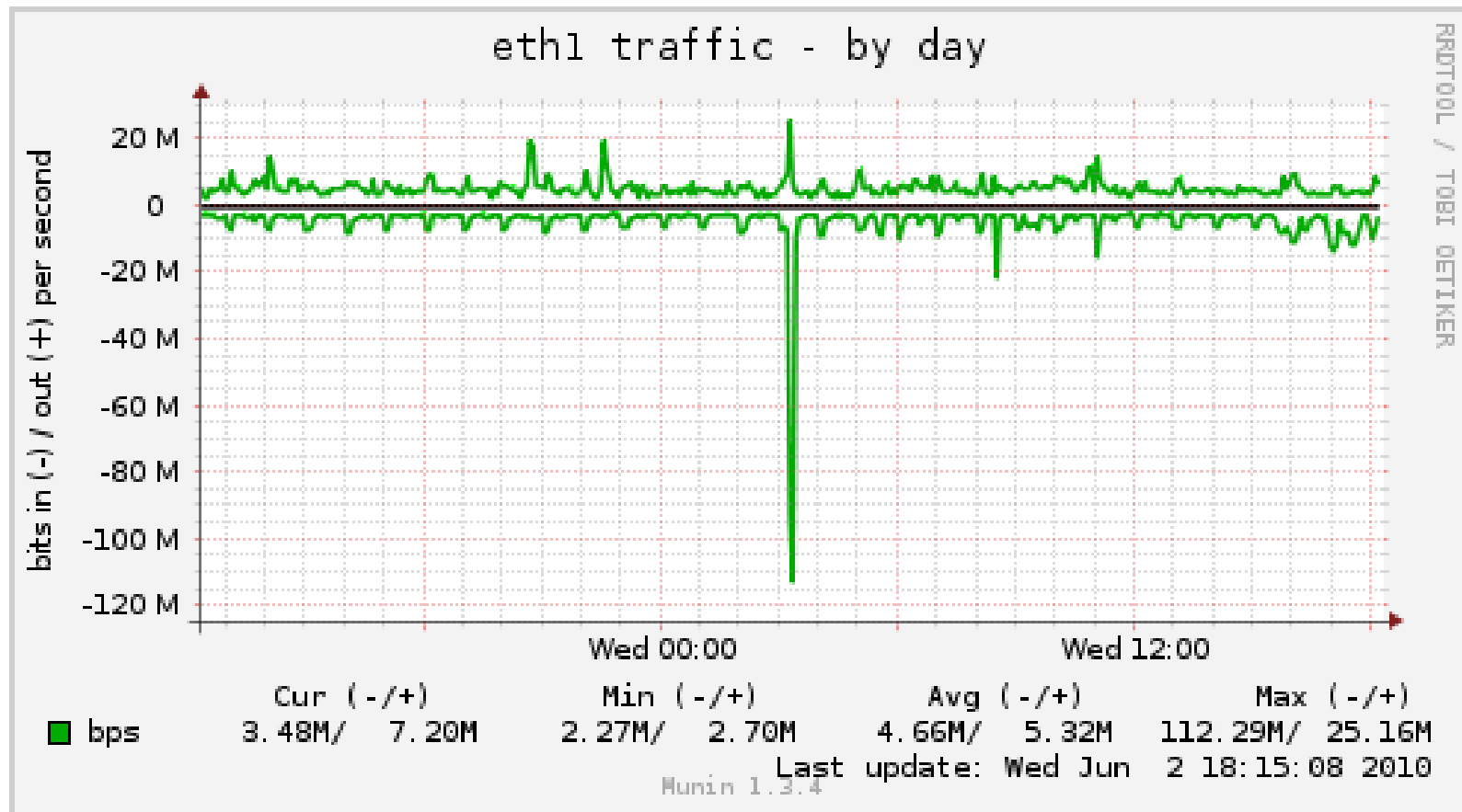
# Ganeti node CPU usage



# Ganeti node LOAD



# Ganeti node DRBD network



# OSUOSL future ganeti plans

KSM (Kernel SamePage Merging)

Upgrade to qemu-kvm 0.12.x

Migrate hosts from libvirt

Puppet integration

Web-based tools

libcloud

# Open source

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

License: GPL v2

Ganeti 1.2.0 - December 2007

Ganeti 2.0.0 - May 2009

Ganeti 2.1.0 - March 2010 / 2.1.6 current

Ganeti 2.2.0~beta0 - June 2010

# Ganeti roadmap

Inter-cluster instance moves

KVM security (currently in >= 2.1.2.1)

Cluster LVM support

LXC (Linux containers)

Job locking fixes

# 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>

# Questions?

lance@osuosl.org

@ramereth on twitter

Ramereth on freenode

blog: <http://www.lancealbertson.com>

slides: <http://tinyurl.com/linuxcon-ganeti>

Presentation made with showoff  
(<http://github.com/schacon/showoff>)



This work is licensed under a [Creative Commons Attribution-Share Alike 3.0 United States License](http://creativecommons.org/licenses/by-sa/3.0/).



# Demo

Create instance

Migrate instance

Fail-over instance

Re-install instance