GANETI@LUFTHANSA CMS, GANETICON2015

JOERG JASPERT

JOERG@NSB-SOFTWARE.DE

TABLE OF CONTENTS

LUFTHANSA GROUP

Big group of companies

- Lufthansa
- Austrian
- Brussels Airlines
- Germanwings
- Swiss
- [...]
- about 540 subsidiaries and equity investmens

- Lots of different IT Systems and groups

LUFTHANSA PASSAGE, CMS

- This only talks about **Lufthansa Passage**
  - And only about **CMS**
- I am **not** a Lufthansa Employee
  - Small company of consultants (NSB GmbH)
  - We offer about everything a company needs for their IT
  - specialised for Airline industry
- I am a Debian Developer

CMS

- CMS AKA Crew Management Systems
- Big mixup of IT
  - Linux (Debian, RedHat, SuSE)
  - HPUX
  - Windows
  - Loads of Open Source Software
  - Loads of propietrary Software, custom and self developed/maintained Software
- Whatever the Crew needs, from planning to actual flying (requests of flights to the Pilots laptop connected to the airplane)
- I am mostly with Debian (and very few RedHat) machines

SERVICES WE OFFER

- Login
  - RSA SecurID
  - SSH Jumphost (Access from external)
  - Reverse Proxy with Auth
  - Transfer Server
- Infrastructure (DNS, Firewall, Proxy, Syslog, …)
- Mail (all of the game)
- Webservers
- Application Servers
- Virtual Machines for other application developers

GANETI

Finally…

Ganeti @ Lufthansa CMS (GanetiCon 2015 ; Joerg Jaspert <joerg@nsb-software.de>)

THE PAST

In the past most services ran on real hardware.

Only a few had been virtualized - mostly based on Xen.

Less than 10 hosts with Xen, 2 with KVM (libvirt).

More than 200 have been real hardware.

THE PRESENT

Migration project (since 2013) to switch to entirely new Hardware.

To make it easier, also change the whole IP layout below us.

And even made simpler by a complete change of policies (security/runtime/...).

And of course: Virtualize as much as possible!

NEW SETUP

2 datacentres (HA requirements).

Big storage in the background (HP 3PAR x 2).

Cisco based Networking (one Nexus 5500 per DC, everything cross-connected).

Some HP Enclosures, filled with HP Blades 465c Gen8.

No local storage, all off into SAN.

Enclosures connect Blades using "HP VirtualConnect" Modules.

As few rack mounted systems as possible.

As many machines as needed - virtual please.

HARDWARE

- Rack mounted
  - 8 rack mounted machines, for monitoring/managing.
  - 2x AMD Opteron 6308 (8-cores total), 128GB RAM, 2x1GBit Network, 2x300GB SAS, 4x1TB SATA, QLogic SAN
- Blades
  - 15 of them (for my area)
  - 13x AMD Opteron 6376 (32-cores, 256G RAM, 10GBit Net, 2x300GB SAS each)
  - 2x AMD Opteron 6320 (16.cores, 64G RAM, 10GBit Net, 2x300GB SAS each)

HP 3PAR SAN STORAGE

Some TB of disk space in lots of spindles.

A hell of a broken API (Mix of REST and SSH).

A hell of a broken firmware. Easy to crash the REST Interface.

-> Wrote an **extstorage** Interface for Ganeti (GPL, but not yet published)

HP 3PAR EXTSTORAGE

- Uses the 3Par REST/SSH "api"
- Mix of Python and Shell code
- Currently requires 2 3Pars
  - Creates LUNs on both
  - Exports them to Hosts
  - Multipath has its fun
  - Put mirror RAID on top of multipath devices
- Ganeti uses /dev/mdXY
  - Usually one for system, one for swap
  - But not limited, some VMs have half a dozen
  - Turns out python had a bug in the os.major/minor calls, breaking when you got too many devices. Oops.

NETWORK

Enclosure connected to switch with 4x10GBit

All Blades get 10GBit via "Virtual Connect" Modules from the Enclosure

Lots of different VLANs in use (Prod/Test/Int/Admin/Backup/… seperation)

Blades use OpenVSwitch (currently not centrally controlled)

Ganeti just "attaches" to one of the tagged interfaces in OpenVSwitch

-> **NO** Routing/Firewall on the Blade!

-> Blade "invisible" for the VM!

OPENSTACK

- First tried Openstack
  - It does sound much more like a candidate
- **WAY** too complicated
- Does not support proper HA for VMs
- Did i say complicated already?

GANETI

- Started with 2.10 (or was it 2.9?)
- Now on 2.15 (since last Thursday) (love gnt-cluster upgrade)
- Exclusively running on Debian hosts
- Mostly Debian VMs too
- Very few RedHat VMs
  - Wrote an own OS definition for RedHat, to kickstart-install them

3 CLUSTERS

DMZ	5 machines (1 to move to CMS) - all VMS with connections to "Outside CMS"
CMS	6 machines. Will get 1 from DMZ, and 2 new small Blades. everything that doesn't need the DMZ
Test	2 machines. Need a place to test and play with new features.

PROD / TEST / INT

Most everything is additionally seperated into "prod"/"test"/"int", plus a few more VLans.

Blades live in an own "Management" vlan, not directly accessible from anything except very few machines. Thats provided via 2 own (bonded) interfaces, untagged.

Openvswitch on the Blades manages the needed VLans for the VMs.

Seperate Backup VLAN connecting every VM with Backup hosts. Special firewall rules ensure seperation there.

ENVIRONMENT

- Using ganeti-instance-debootstrap or own RedHat Kickstart
- OS definitions only do most basic setup, everything else handled by Puppet
- Own module in puppet to setup Ganeti on our Blades
- Ganeti may not do any SSH "fuckup"
  - Host keys generated and distributed centrally from Puppet
  - /root /.ssh /authorized_keys may not exist (and even if, would be ignored)
  - DSA keys forbidden by policy
  - Puppet ensures SSH access between nodes

ENVIRONMENT

- Script "addvm" to setup new VMs
- Ideally everything can just be reinstalled
- Ganeti manages (currently) around 200 VMs

FUTURE

- Currently all Ganeti Nodes are master candidates
- Thinking of moving to one ganeti "control" node outside blades, on a non vm_capable host
- But 3 clusters and not enough different rack-mounted machines to use
  - Maybe "merging" them and using node groups?
- More use of the h* tools (now that luxid isn't broken anymore)
- Get better in using cluster/node/vm tags to avoid services being run too "near" to each other (same RZ or even same enclosure)

WISHLIST

- Completely remove dependency on LVM if used disk layouts do not need it (say, extstorage).
- SSH Setup: Way more configurable. Keytype, SSH port, maybe only few machines, …
  - Example: regen of SSL keys seems to use hardcoded -o Port=22
- Keep it able to run on older releases (Debian wheezy)
- In DRBD/LVM Setups: Ignore LVs/DRBDs not managed by Ganeti - it may not be exclusively managing them

QUESTIONS?

Ganeti @ Lufthansa CMS (GanetiCon 2015 ; Joerg Jaspert <joerg@nsb-software.de>)