Unter den Wolken

Bare-Metal Provisioning

Felix Krohn <felix@kro.hn>
@FelixKrohn
gpg: 0xC0ED0C8D4AF209DE

Vortragsthema

- Versprechen von Cloud: *Dematerialisierung as a Service*
- Aber: *"There is no cloud. It's just someone else's computer"*
- Automatisiertes Verteilen & Konfigurieren von Anwendungen erhält viel Aufmerksamkeit (Docker, Puppet, Chef, Docker, Saltstack, Docker, Docker, ...)
- Bare-Metal verschwindet im *SEP-Feld* [1]
- andere Gründe auf bare-metal zu setzen *statt* VMs: spezielle (I/O-intensive) Workloads wie Storage, Big Data, HPC ...

[1] `[...] it relies on people's natural predisposition not to see anything they don't want to, weren't expecting, or can't explain.` (Douglas Adams, 1982)

OVH.de
Innovation is Freedom

Vortragsthema

TL;DR: Thema

Wie bekomme ich die Kiste soweit daß ich mit CM-tools arbeiten kann?

Zur Person

- Felix 'kro' Krohn
- Hier in FuWa: CN WS 03/04 bis '08
- Systemadministrator bei OVH seit 2008
- Aufgabenbereich: Unix/Linux Installationen & Infrastruktur
- *"Product Manager Operating Systems"*
- Infrastruktur: Installer, Rescue-Systeme, Mirrors, ...
- Werkzeuge: Perl, Shell, Unix/Linux-Toolchain, Geduld & Vorschlaghammer

Über OVH

- >250.000 Server in derzeit 17 Rechenzentren
- >5.5 TBit/s Gesamtanbindung an 32 POPs
- ca. 1300 Mitarbeiter auf 3 Kontinenten
- vom CPU-Kühler bis zum RZ alles aus einer Hand
- fast 500 Domain-Endungen
- VPS ab 3.50, Dedicated Server ab 5.99 EUR
- zahlreiche Lösungen für IaaS, PaaS, SaaS
- stolzer UnFUCK-Sponsor ;)

Distributionen bei OVH

- Knapp 90 verschiedene Distributionen verfügbar
- Wachsende Zahl an Plattformen (i386, x86_64, armhf, power8, ppc64el, ...
- Darf's noch etwas sein? - Feedback oder Wünsche gerne auch direkt an mich melden!
- Vier Betriebssystem-Familien:
  - Linux
  - Windows
  - BSD
  - Solaris
- Drei Kategorien:
  - Basis
  - Hosting
  - Virtualisierung

OVH.de
Innovation is Freedom

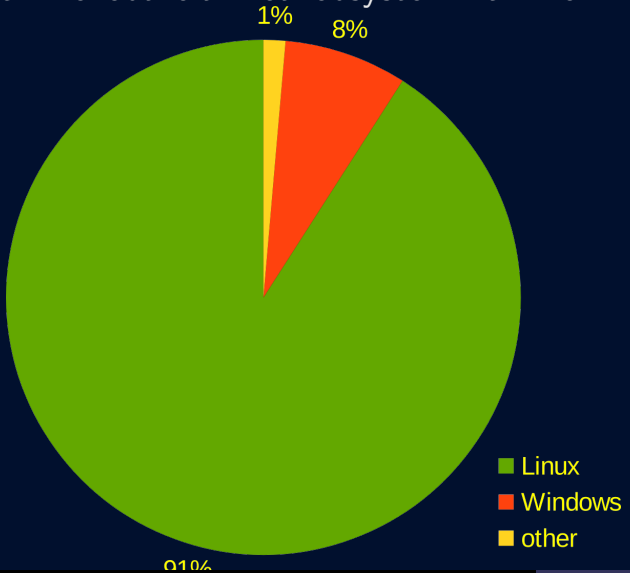
Distributionen bei OVH

- Knapp 90 verschiedene Distributionen verfügbar
- Wachsende Zahl an Plattformen (i386, x86_64, armhf, power8, ppc64el, ...
- Darf's noch etwas sein? - Feedback oder Wünsche gerne auch direkt an mich melden!
- Vier Betriebssystem-Familien:
  - Linux
  - Windows
  - BSD
  - Solaris
- Drei Kategorien:
  - Basis
  - Hosting
  - Virtualisierung

Distributionen bei OVH

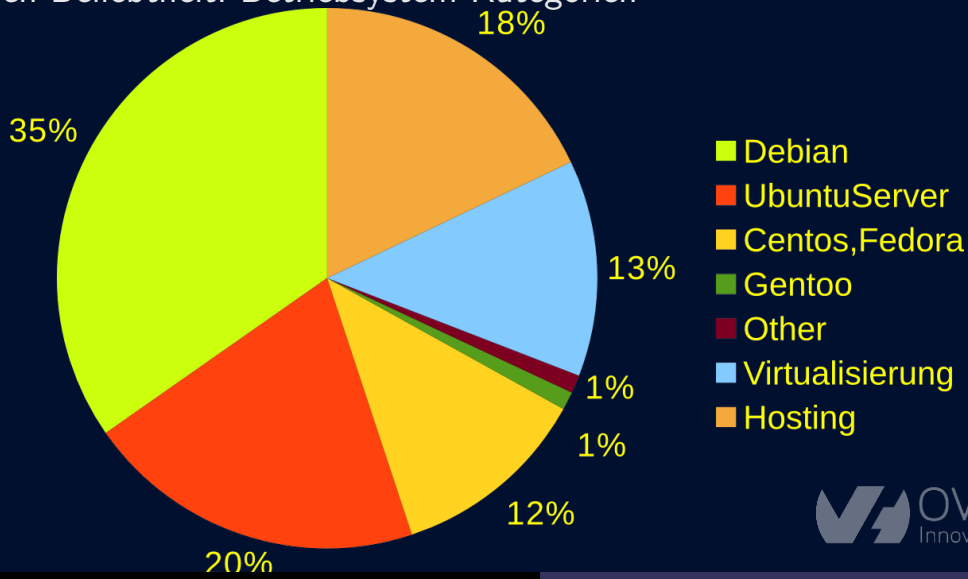
- Knapp 90 verschiedene Distributionen verfügbar
- Wachsende Zahl an Plattformen (i386, x86_64, armhf, power8, ppc64el, ...
- Darf's noch etwas sein? - Feedback oder Wünsche gerne auch direkt an mich melden!
- Vier Betriebssystem-Familien:
  - Linux
  - Windows
  - BSD
  - Solaris
- Drei Kategorien:
  - Basis
  - Hosting
  - Virtualisierung

Nach Beliebtheit: Betriebsystem-Familien



OVH.de
Innovation is Freedom

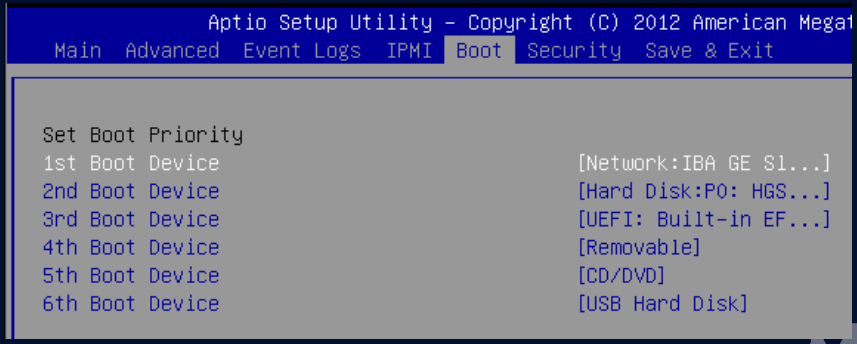
Nach Beliebtheit: Betriebsystem-Kategorien



OVH.de
Innovation is Freedom

BIOS - setup

- BIOS / EFI
- Bootreihenfolge: Netzwerk zuerst



BIOS - DHCP

```
Intel(R) Boot Agent GE v1.3.71
Copyright (C) 1997-2010, Intel Corporation

CLIENT MAC ADDR: 00 25 90 AC 8A 6E  GUID: 00000000 0000 0000 0000 002590AC8A6E
DHCP..\
```

- DHCP: IP, gateway ... next-server und filename
- *next-server*: Adresse des (üblicherweise) tftp-servers
- *filename*: Network Bootstrap Program (NBP) ist erstes "eigenes" binary
- => tftp://next-server/$filename

pxelinux.0

```
Intel(R) Boot Agent GE v1.3.71
Copyright (C) 1997-2010, Intel Corporation

CLIENT MAC ADDR: 00 25 90 AC 8A 6E  GUID: 00000000 0000 0000 0000 002590AC8A6E
CLIENT IP: 178.33.226.37  MASK: 255.255.255.0  DHCP IP: 91.121.126.137
GATEWAY IP: 178.33.226.254

PXELINUX 3.52 2007-09-25  Copyright (C) 1994-2007 H. Peter Anvin
UNDI data segment at:   000BF570
UNDI data segment size: 63B0
UNDI code segment at:   00085920
UNDI code segment size: 5190
PXE entry point found (we hope) at 9592:0106
My IP address seems to be B221E225 178.33.226.37
ip=178.33.226.37:91.121.126.137:178.33.226.254:255.255.255.0
TFTP prefix:
Trying to load: pxelinux.cfg/00000000-0000-0000-0000-002590ac8a6e
Trying to load: pxelinux.cfg/01-00-25-90-ac-8a-6e
Trying to load: pxelinux.cfg/B221E225
boot: _
```

pxelinux.cfg

pxelinux.cfg/B221E225

```
default netboot
timeout 30

label netboot
kernel bzImage-4.4.8-amd_64-generic
append root=/dev/ram0 initrd=initramfs.gz rw myOption=foobar

label HDD
localboot 0
```

pxelinux.0

```
Intel(R) Boot Agent GE v1.3.71
Copyright (C) 1997-2010, Intel Corporation

CLIENT MAC ADDR: 00 25 90 AC 8A 6E  GUID: 00000000 0000 0000 0000 002590AC8A6E
CLIENT IP: 178.33.226.37  MASK: 255.255.255.0  DHCP IP: 91.121.126.137
GATEWAY IP: 178.33.226.254

PXELINUX 3.52 2007-09-25  Copyright (C) 1994-2007 H. Peter Anvin
UNDI data segment at:   0008F570
UNDI data segment size: 63B0
UNDI code segment at:   00085920
UNDI code segment size: 5190
PXE entry point found (we hope) at 9592:0106
My IP address seems to be B221E225 178.33.226.37
ip=178.33.226.37:91.121.126.137:178.33.226.254:255.255.255.0
TFTP prefix:
Trying to load: pxelinux.cfg/00000000-0000-0000-0000-002590ac8a6e
Trying to load: pxelinux.cfg/01-00-25-90-ac-8a-6e
Trying to load: pxelinux.cfg/B221E225
boot: netboot_
```

iPXE script

https://ipxe.mydomain.foo/ipxe.cgi?server

```
#!ipxe
echo netboot
kernel https://mydomain.foo/netboot/bzImage-4.4.8-amd_64-generic
root=/dev/ram0 initrd=initramfs.gz
initrd https://mydomain.foo/netboot/initramfs.gz
boot
```

Was nun?

- vom Netz gebootetes System kann für discovery/inventarisierung/tests genutzt werden
- vom Netz gebootetes System kann für weiteres setup genutzt werden
- . . . oder auch stand-alone (*"stateless"*)
- (Beispiele für letzteres: coreOS, SmartOS)

(zu) weiterführende Themen

- IPMI
- UEFI
- Secure Boot
- non-Linux Systeme
- non-Intel Plattformen

- Komplettpaket (alle benötigten Dienste & Configs via puppet)
- unterstützt VMs und bare metal
- Ruby
- Web-GUI (RoR oder apache + mod_passenger)
- *Smart Proxies* zur Arbeitsverteilung/Dezentralisierung via restful API
- Discovery: bootet server & inventarisiert
- Installation durch Os-native Automatismen (kickstart etc)
- Eng mit Puppet verwoben

OpenStack Ironic

- Kein Witz: *Ironic*
- Witz: *"bear metal"*
- OpenStack Treiber für bare-metal
- Keine standalone-Komponente
- Sinnvoll verwendbar (nur) mit OpenStack Nova, Glance, ...
- Dienste-Konfiguration nicht vorgegeben

MaaS - Metal as a Service

- http://maas.io
- Hersteller: Canonical
- Ubuntu-zentrisch, unterstützt aber auch andere OS (gegen $$$)
- beansprucht Exklusivzugriff auf alle Hardware im eigenen Netz
- Anbindung an den “Juju charms” Appstore

Cobbler

- stammt aus dem rpm-Umfeld
- Automatisierung von Installationen mit Kickstart
- Unterstützt auch nicht-rpm Distributionen (Debian/Ubuntus preseed-Dateien)
- Kümmert sich um Konfiguration von PXE, TFTP und DHCP

OPNFV

- Open Platform for Network Functions Virtualisation
- OPNFV “Brahmaputra”
- https://www.opnfv.org/software

Warum?

- (kein) NIH-Syndrom(?)
- heterogene Infrastrukturen (Hardware)
- heterogene Infrastrukturen (Software)
- Gemisch an Deployment-Tools und manuellen installationen in einem Netz
- Bedarf an aktuellen Entwicklungen (ISC Kea, fbtftp, . . .)

Einkaufsliste

- DHCP: ISC DHCPd: gut abgehangen
- DHCP: ISC Kea DHCP Server: moderner, aber noch nicht alle features
- TFTP-server: verschiedene verfügbar
- PXE-binary (PXElinux aus dem syslinux-paket oder ipxe)
- Web- und/oder NFS-Server und/oder iSCSI
- Hardware-Inventarisierung

Hinweis in eigener Sache

- Teil-Einkauf: Umsetzung mit gemieteten Servern möglich
- ipxe-Skripte via API
- Eigenes Netzwerk ("vRack") ermöglicht Betrieb von DHCP etc

GET	/me/ipxeScript	List of all your IPXE scripts
POST	/me/ipxeScript	Add an IPXE script
GET	/me/ipxeScript/{name}	Get this object properties
DELETE	/me/ipxeScript/{name}	Remove this IPXE Script

Image-Formate: Qual der Wahl

- .raw, .qcow, .vmdk, .vhd, ...
- Gemeinsamkeit: Image enthält komplett *alles*
- d.h. partitionierung, filesysteme, bootloader, ...
- keine flexibilität ohne Aufwand
- Alternativ: nur Dateisystem-Inhalt kopieren (tar)

- Ordentliches Säubern im image notwendig!
- (z.B. SSL keys, SSH hostkeys, UUIDs, ...)
- Konservieren von Berechtigungen, UIDs, Änderungsdaten, ...
- Konservieren von extended attributes [2]

[2] looking at you, GNU tar!

Schlußfolgerungen

- Fazit:

  Es kommt drauf an.

Unter den Wolken...

...sind die Freiheiten nicht begrenzter als drüber :)

Schlußfolgerungen

- Fazit:

  Es kommt drauf an.

Unter den Wolken...

...sind die Freiheiten nicht begrenzter als drüber :)

Unter den Wolken

Bare-Metal Provisioning

Felix Krohn <felix@kro.hn>
@FelixKrohn
gpg: 0xC0ED0C8D4AF209DE