

# SmartOS

Zones + ZFS + KVM + DTrace = Awesome

Thomas Merkel <tm@core.io>

2014-04-20

# # whoami

- Frubar Network seit 2005
- Server Ninja / Partner bei SkyLime seit 2007
- Administration von ~450 Linux / Unix Server (Gentoo, Ubuntu, Debian, SmartOS)
- Fokus auf Monitoring, Webserver Cluster, Script Entwicklung

# Agenda

- Was ist SmartOS?
  - Zones
  - Crossbow
  - ZFS
  - KVM
- Thinking SmartOS
  - Installation
  - Tipps
  - Konfiguration der Globalen Zone
  - Datasets
  - Verwaltung von Zones und KVM Images
  - Paketmanagement in Zones
  - Update SmartOS
- Project FiFo
- Links

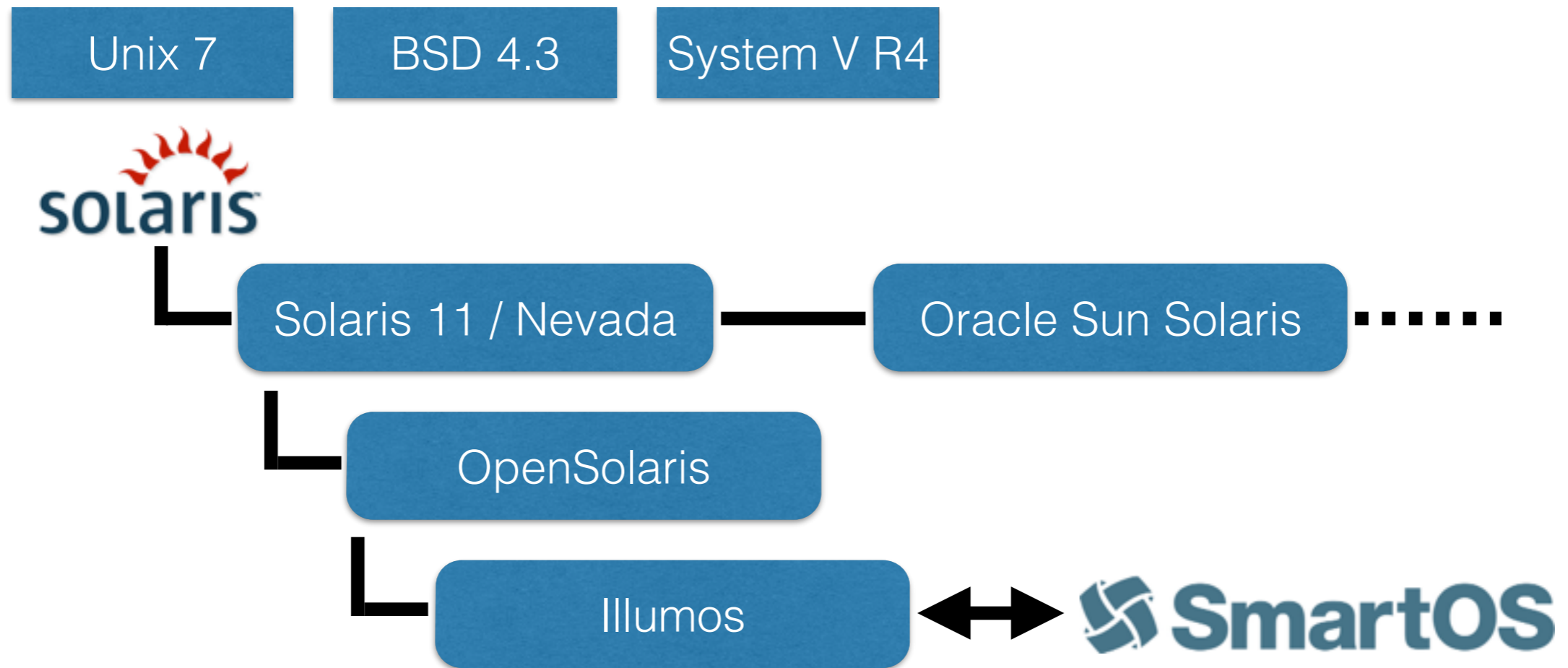




# Was ist SmartOS?



# Was ist SmartOS?



# Was ist SmartOS?

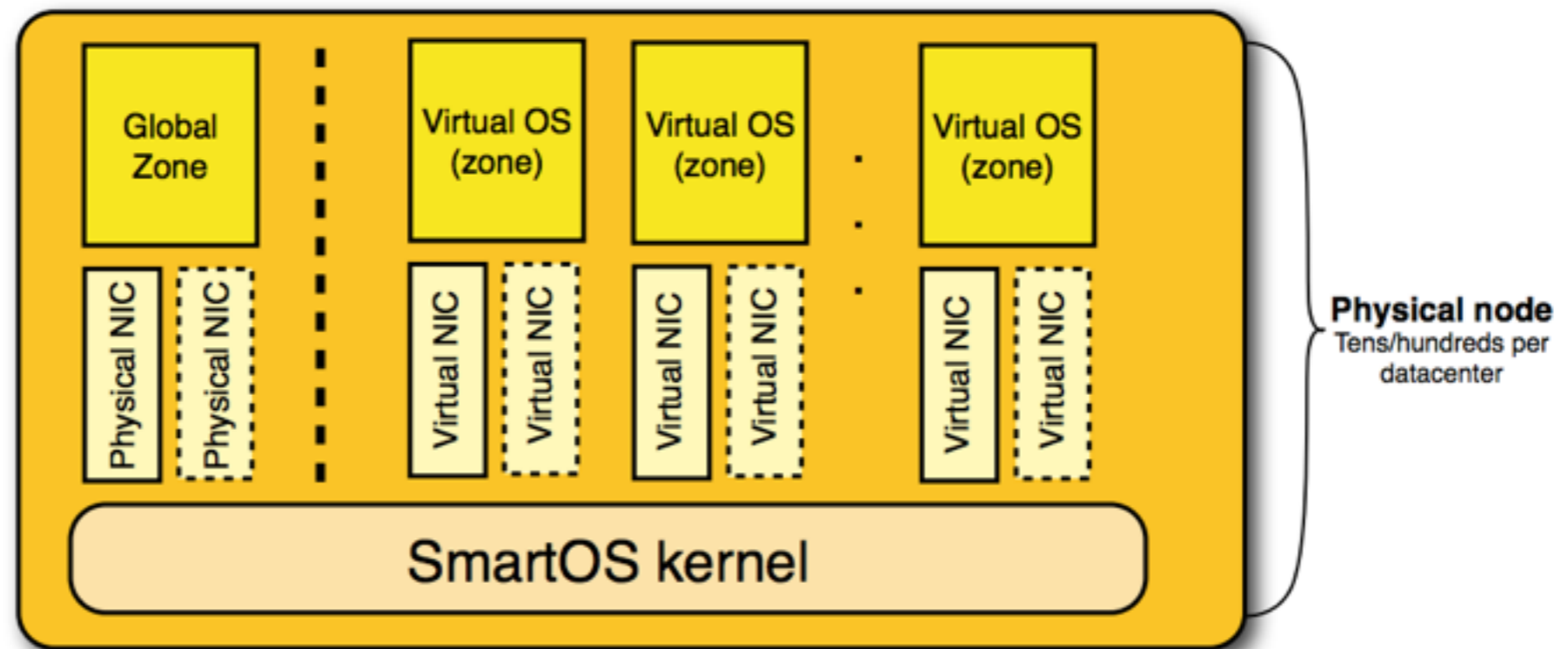
- Entwickelt von Joyent (Open Source)
- Solaris Features
  - Zones
  - Crossbow
  - ZFS
  - DTrace
- Fokus auf Virtualisierung
  - Boot via externer Medien (USB, Netboot)
  - Minimale Konfiguration
  - Verwaltungstools
- KVM Support

# Zones - OS Virtualisierung

- Selbstverwalteter Container
  - Konfiguration eigener Benutzer, Festplatten, Netzwerk und Dienste
  - „Gefühlt“ wie ein OS
- Isoliert
  - Zone sieht nur sich selbst
  - Globale Zone überwacht lokale Zonen
  - Eigenes Netzwerk und isoliertes Dateisystem
- Ressourcen Verwaltung
  - Arbeitsspeicher, Festplatte und Netzwerk I/O
  - CPU Verteilung

# Zones - OS Virtualisierung

- Minimaler Overhead
- Hardware muss nicht emuliert werden
- Dienste können aus der globalen Zone überwacht werden (z.B. via DTrace)





# Crossbow - Virtual NICs

- Erstellung von virtuellen Netzwerkinterfaces oder Switches
- Verbindung von VNICs zu
  - Physikalischen NICs
  - Virtuellen Switches
- Antispoofing
  - MAC Adressen
  - IP Adressen
  - DHCP
- Kontrolle der Bandbreite

# ZFS

- Copy on write Dateisystem
- Pool Storage
  - Partitionierung spielt keine Rolle
  - Verwaltung via Datasets
  - Live Änderungen an Quotas
- RAID
  - RAID 1 (Mirror)
  - RAIDZ, Z2, Z3, ... (Single, Double oder Triple parity)
  - RAID 0 (Striping)

# ZFS

- 128bit Checksum für alles
- Komprimierung
- Deduplizierung
- ZIL (Intent Log)
- L2ARC (Cache)
- Snapshot und Clone Support

# KVM

- Virtualisierung anderer Betriebssysteme
- Joyent unterstützt KVM nur bei Intel CPUs mit EPT
  - AMD Unterstützung durch Community
- Administrationsbefehle für KVM Image oder Zone sind gleich





# Thinking SmartOS



# Installation

- Download und `dd` des Images auf USB Stick
  - <http://wiki.smartos.org/display/DOC/Download+SmartOS>
- Server via USB Stick Booten und Installer folgen
  - Netzwerk Informationen, Zpool anlegen, Root Kennwort
- Zusätzliche Konfiguration falls nötig unter `/opt/custom` und `/usbkey/config`
  - Puppet, Bash Scripts, eigene SVCs
- Und los gehts :-)

# Tipps

- Globale Zone
  - RAM Disk nur zur Zonenadministration
  - Daten in **/etc** und **/root**: nicht dauerhaft
  - Daten in **/opt** und **/var**: dauerhaft
  - Basis Konfiguration in **/usbkey**
  - SMF Manifest Dateien in **/opt/custom/smf/** werden beim Booten geladen

# Konfiguration Global Zone

- /usbkey/config

```
# admin_nic is the nic admin_ip will be connected to for headnode.
admin_nic=00:26:b9:87:47:6c
admin_ip=80.190.131.134
admin_netmask=255.255.255.128
admin_network=...
admin_gateway=80.190.131.134

headnode_default_gateway=80.190.131.129

# create a virtual switch for an internal NAT solution.
etherstub="switch0"

dns_resolvers=8.8.8.8,8.8.4.4
dns_domain=srv.skylime.net

ntp_hosts=pool.ntp.org
```

# Konfiguration Global Zone

- /opt/custom/

```
├─ cfg
│   ├── datacenter
│   ├── global
│   │   ├── crontab
│   │   │   ├── fmadm
│   │   │   └─ zpool
│   │   ├── root
│   │   │   └─ root
│   │   └─ script
│   │       ├── 10-hostname.sh
│   │       ├── 15-ipv6.sh
│   │       └─ 20-sendmail.sh
├─ script
│   └─ postboot.sh
└─ smf
    └─ postboot.xml
```

# Konfiguration Global Zone

- /opt/custom/smf/postboot.xml

```
<?xml version="1.0"?>
<!DOCTYPE service_bundle SYSTEM [...]>
<service_bundle type='manifest' name='site:postboot'>
<service name='site/postboot' type='service' version='1'>
  <create_default_instance enabled='true' />
  <single_instance />
  <dependency name="network" grouping="require_all"
    restart_on="error" type="service">
    <service_fmri value="svc:/milestone/network:default"/>
  </dependency>
  <exec_method type='method' name='start' timeout_seconds='0'
    exec=' /opt/custom/script/postboot.sh '>
  </exec_method>
  [...]
</service>
</service_bundle>
```



# Datasets

- Bestehende Images für Zones oder KVM
- Community Images bei <http://datasets.at>

```
# imgadm avail
```

UUID	NAME	VERSION	OS	PUBLISHED
7241e29a-a07b-11e3-9a5c-53df1db058c4	base	13.4.1	smartos	2014-02-28
c3321aac-a07c-11e3-9430-fbb1cc12d1df	base64	13.4.1	smartos	2014-02-28
835e27b2-a47e-11e3-9eb6-e78ef6d1ee8f	postgresql	13.3.1	smartos	2014-03-05
398deede-c025-11e3-8b24-f3ba141900bd	standard64	13.4.1	smartos	2014-04-09
0fbd0a74-c028-11e3-ac6a-bf348e9f760c	elasticsearch	13.4.1	smartos	2014-04-09
edd43ca8-c0ee-11e3-a027-c3a9f2cd2f6a	stm-developer	13.4.1	smartos	2014-04-10
19daa264-c4c4-11e3-bec3-c30e2c0d4ec0	centos-6	2.6.1	linux	2014-04-15

```
[...]
```

# Zonenerstellung (1)

- Gewünschtes Dataset importieren

```
# imgadm import 398deede-c025-11e3-8b24-f3ba141900bd  
Importing image 398deede-c02... (standard 13.4.1) from "https://images.joyent.com"  
398deede-c025-11e3-8b24-f3ba141900bd [ ] 0% 2.83MB 592.83KB/s 19m28s
```

# Zonenerstellung (2)

- JSON Beschreibungsdatei erstellen
- Generator auf <http://datasets.at>

```
{  
  "brand": "joyent",  
  "image_uuid": "398deede-c025-11e3-8b24-f3ba141900bd",  
  "autoboot": true,  
  "alias": "nsa-box",  
  "hostname": "mirror-customer-information",  
  "dns_domain": "nsa.gov",  
  "resolvers": [ "8.8.8.8" ],  
  "max_physical_memory": 512,  
  "max_swap": 512,  
  "nics": [ { "nic_tag": "admin", "ip": "dhcp", "primary": true } ]  
}
```



# Zonenerstellung (3)

- Zone erstellen via **vmadm**

```
# vmadm create -f nsa-box.json  
Successfully created 59aaecb6-467d-4116-92bd-f31842d44c90
```

- Zone anzeigen

```
# vmadm list
```

UUID	TYPE	RAM	STATE	ALIAS
59aaecb6-467d-4116-92bd-f31842d44c90	OS	512	running	nsa-box

# KVM

- Das selbe wie bei Zones aber mit zusätzlichen Möglichkeiten
  - Erstellen einer leeren Instanz und booten von ISO
  - Verwendung bestehender Datasets
- QEMU läuft in einer minimalen Zone
  - Logdateien unter `/zones/UUID/root/tmp`



# Paketmanagement

- Kein Paketmanagement in der Globalen Zone
- Paketverwaltung erfolgt durch **pkgsrc** von NetBSD
- Binary Pakete (32bit und 64bit)
- Releases jedes Quartal (2014Q1)
  
- Layout
  - PREFIX=/opt/local
  - PKG\_SYSCONFDIR=/opt/local/etc
  - PKG\_DBDIR=/opt/local/pkg

# Paketmanagement

- Bootstrap **pkgsrc** falls nicht vorhanden

```
# cd /
# curl -k \
    http://pkgsrc.smartos.skylime.net/[...]/bootstrap-2014Q1-x86_64.tar.gz \
    | gzcata | tar -xf -
# echo http://pkgsrc.smartos.skylime.net/packages/SmartOS/2014Q1/x86_64/All/ \
    > /opt/local/etc/pkgin/repositories.conf
# pkg_admin rebuild
# pkgin -y update
```

# Paketmanagement

- Datenbank update

```
# pkgin update
pkg_summary.bz2
processing remote summary (http://pkgsrc.smartos.skylime.net/[...]/x86_64/All)...
updating database: 100%
```

- Suche

```
# pkgin search bash
bash-doc-2.05.2      Documentation for the GNU Bourne Again Shell
bash-completion-2.1 Programmable completion specifications for bash
bash-4.2nb3 =       The GNU Bourne Again Shell

=: package is installed and up-to-date
<: package is installed but newer version is available
>: installed package has a greater version than available package
```

# Paketmanagement

- Installation

```
# pkgin install nginx-1.5.0
calculating dependencies... done.

nothing to upgrade.
1 packages to be installed: nginx-1.5.0 (342K to download, 777K to install)

proceed ? [Y/n]
```

- Einige Pakete installieren SMF (Service Management Facility)
- Alternativ NetBSD rc.d-Scripte

# Update SmartOS

- Ganz einfach :-)
- USB Stick
  - **smartos-platform-upgrade** Script  
<https://github.com/calmh/smartos-platform-upgrade>
  - Ersetzt den alten **platform** Ordner mit neuem **platform** Ordner



# Update SmartOS

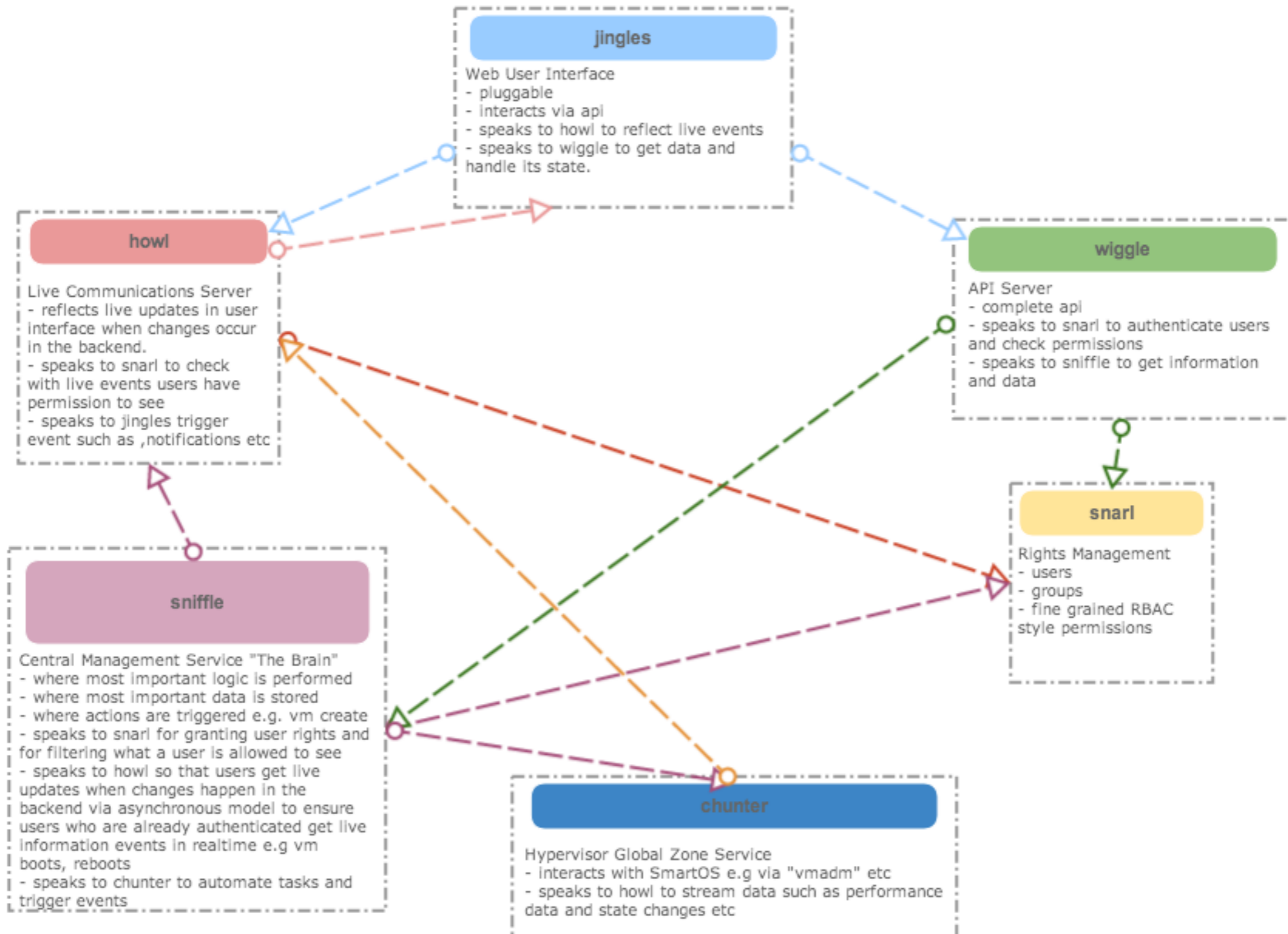
- TFTP Server
  - Platform Archiv herunterladen  
[https://us-east.manta.joyent.com/Joyent\\_Dev/public/SmartOS/platform-latest.tgz](https://us-east.manta.joyent.com/Joyent_Dev/public/SmartOS/platform-latest.tgz)
  - **IPXE Konfiguration** anpassen
- Server rebooten

# Projekt FiFo

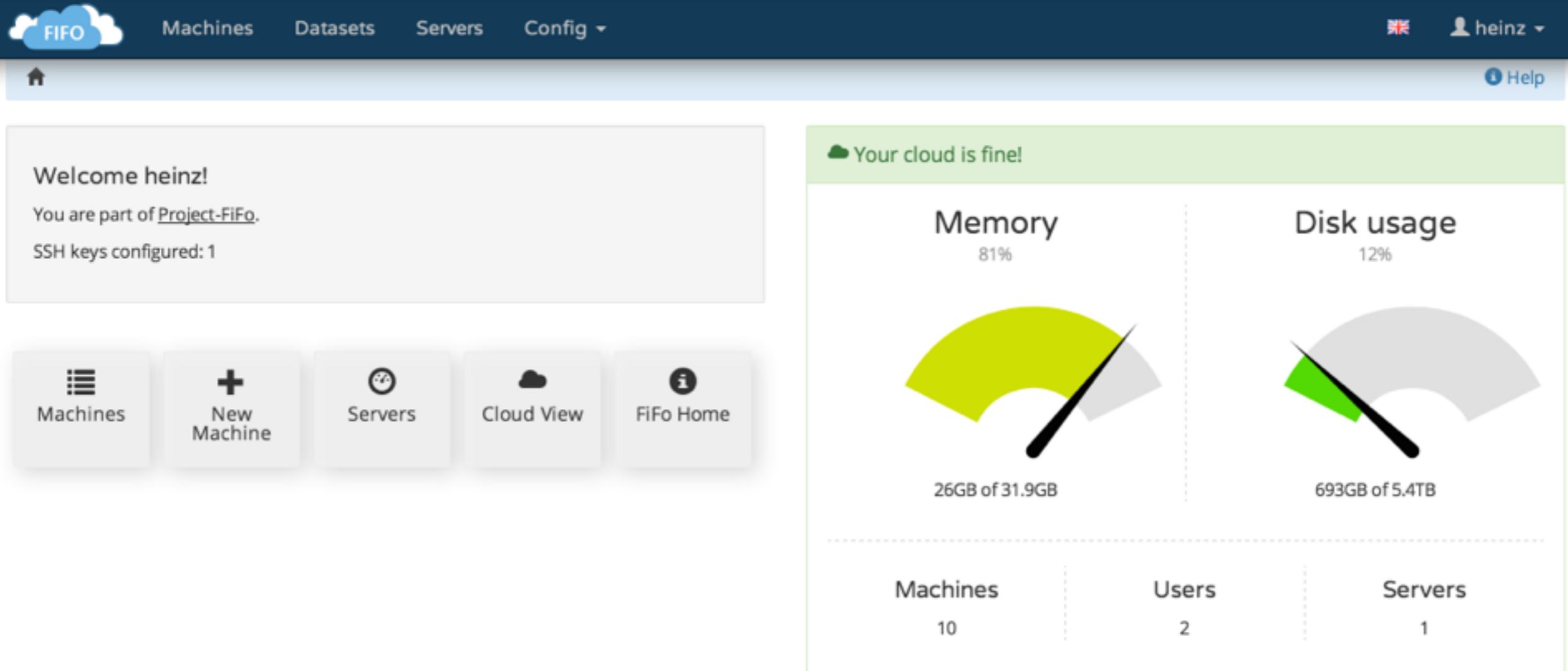
# Projekt FiFo

- <http://project-fifo.net>
- Open Source Cloud Management für SmartOS
- Entwicklung in Erlang
- Development Version
- mDNS Service Discovery
- VNC Support via noVNC

# Projekt FiFo - Architektur

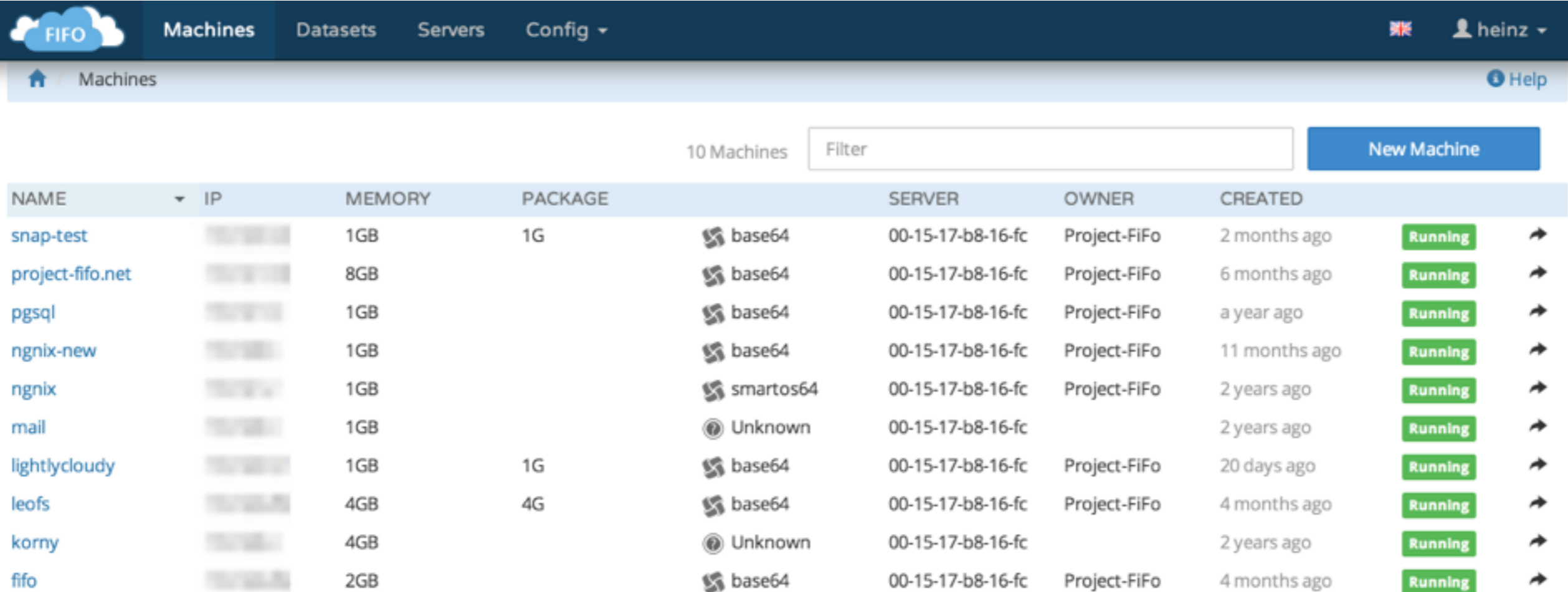


# Projekt FiFo - Status





# Projekt FiFo - VMs



The screenshot shows a cloud management interface with a dark blue header. The header contains the 'FIFO' logo, navigation tabs for 'Machines', 'Datasets', 'Servers', and 'Config', a user profile for 'heinz', and a 'Help' link. Below the header, a breadcrumb trail shows 'Machines'. The main content area displays a table of 10 VMs. Above the table, there is a '10 Machines' label, a search filter box, and a 'New Machine' button. The table columns are: NAME, IP, MEMORY, PACKAGE, SERVER, OWNER, CREATED, and a status column with 'Running' labels and external links.

NAME	IP	MEMORY	PACKAGE	SERVER	OWNER	CREATED	Status	
snap-test		1GB	1G	base64	00-15-17-b8-16-fc	Project-FiFo	2 months ago	Running
project-fifo.net		8GB		base64	00-15-17-b8-16-fc	Project-FiFo	6 months ago	Running
pgsql		1GB		base64	00-15-17-b8-16-fc	Project-FiFo	a year ago	Running
nginx-new		1GB		base64	00-15-17-b8-16-fc	Project-FiFo	11 months ago	Running
nginx		1GB		smartos64	00-15-17-b8-16-fc	Project-FiFo	2 years ago	Running
mail		1GB		Unknown	00-15-17-b8-16-fc		2 years ago	Running
lightlycloudy		1GB	1G	base64	00-15-17-b8-16-fc	Project-FiFo	20 days ago	Running
leofs		4GB	4G	base64	00-15-17-b8-16-fc	Project-FiFo	4 months ago	Running
korny		4GB		Unknown	00-15-17-b8-16-fc		2 years ago	Running
fifo		2GB		base64	00-15-17-b8-16-fc	Project-FiFo	4 months ago	Running

# Projekt FiFo - VM details

The screenshot shows the OpenStack VM details page for a machine named 'fifo'. The interface includes a top navigation bar with 'Machines', 'Datasets', 'Servers', and 'Config' menus. The user 'heinz' is logged in. The machine 'fifo' is in a 'running' state. The 'Details' tab is active, showing four main sections: Machine, Package, Dataset, and Network.

**Machine **fifo****

Created	4 months ago
Type	zone
Server	00-15-17-b8-16-fc
Owner	Project-FiFo
List Color	Color

**Package **custom****

CPU	Share: 64
	Cap: 100
Ram	2GB
Disk	50GB
IO Prio	128

**Dataset **base64 13.2.1****

UUID: 17c98640-1fdb-11e3-bf51-3708ce78e75a

Info: A 64-bit SmartOS image with just essential packages installed. Ideal for users who are comfortable with setting up their own environment and tools.

**Network **admin****

NIC	
IP	
Netmask	
Gateway	
Mac	





# Global Zone build environment



# Warum?

- Eigene Software-Pakete in der Globalen Zone
  - munin-node
  - postfix statt sendmail
- Anpassungen von SMF / Start-Scripte
- Eigene Erweiterungen

# Build zone

- SmartOS multiarch Zone 13.3.0

```
# imgadm import a1d74530-4212-11e3-8a71-a7247697c8f2
```

- min. 32 GB RAM für die Zone
- ca. 20GB Speicherplatz



# Build zone

- Zone über JSON Datei erstellen

```
{  
  "brand": "joyent",  
  "max_physical_memory": 32768,  
  "tmpfs": 8192,  
  "fs_allowed": "ufs,pcfs,tmpfs",  
  "image_uuid": "a1d74530-4212-11e3-8a71-a7247697c8f2",  
  "quota": 15,  
}
```

# The source

- Aktueller Quellcode aus dem Git Repository

```
build-zone # pkgin in scmgit  
build-zone # git clone https://github.com/joyent/smartos-live  
build-zone # cd smartos-live
```

```
build-zone # cp sample.configure.smartos configure.smartos  
build-zone # ./configure
```

- Nicht genug Leistung (RAM, CPUs)?

```
gsed -e '/^PARALLEL/{s/-j[0-9]*/-j8/}' -i.bak projects/illumos-extra/Makefile.defs
```

# The source

- Backe backe SmartOS

```
build-zone # gmake world && gmake live
```

- Packaging

```
build-zone # pkgin in cdrtools pbzip2
```

```
build-zone # export LC_ALL=C
```

```
build-zone # tools/build_iso
```

```
build-zone # tools/build_usb
```

# Links

- SkyLime /opt/custom Scripts
  - <https://github.com/skylime/smartos-config>
- Joyent Wiki
  - <http://wiki.smartos.org/>
  - <https://github.com/joyent/smartos-live>
- Blog von Jonathan Perkin (Joyent Entwickler / pkgsrc)
  - <http://www.perkin.org.uk/>
- SmartOS mit IPv6 und AMD Support
  - <http://imgapi.uqcloud.net/builds>
- Community Datasets
  - <http://datasets.at>

# Community

- IRC ([irc.freenode.net](http://irc.freenode.net))
  - #smartos
  - #illumos
  - #joyent
  - #project-fifo
- Mailinglist
  - <http://smartos.org/smartos-mailing-list/>
  - <https://groups.google.com/forum/?fromgroups#!forum/project-fifo>





**Fragen?**