

Ein XEN iSCSI HA-Cluster administrieren

Thomas Groß



Cluster ?

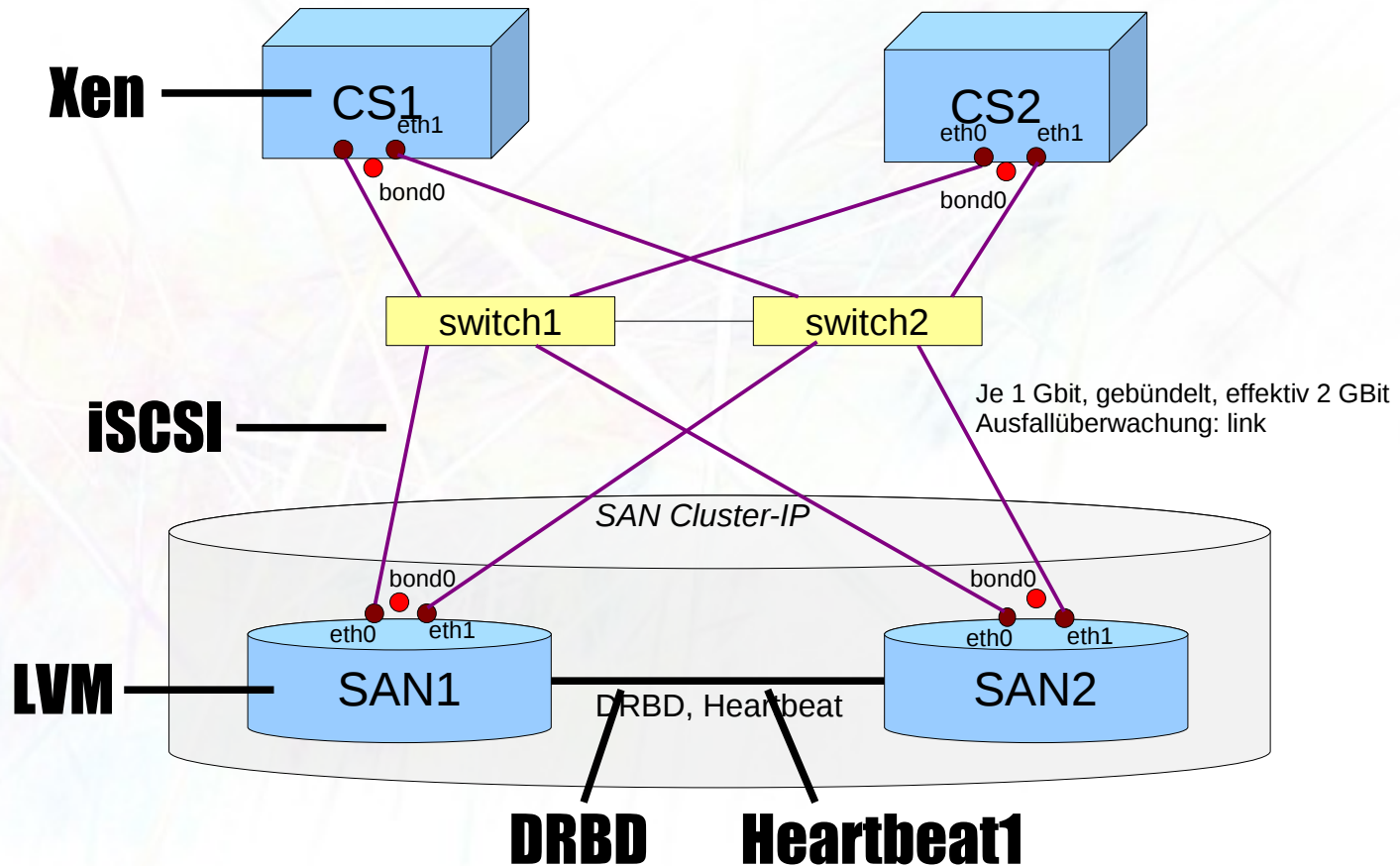
- zwei oder mehr Clusterserver mit **XEN**
 - ein gemeinsamer **iSCSI** Speicher mit **LVM**
(zwei Servern, die per **DRBD** gespiegelt und mit **Heartbeat1** synchronisiert werden)
 - Redundante Netzwerkverbindungen zwischen Clusterservern und Speicherservern (**bonding**)
 - alle Server mit **opensuse 11.1**
- ⇒ Aufbau mit Linux Standard-Werkzeugen ⇐

Ziele Virtualisierungscluster

- höhere Verfügbarkeit der Applikationen
- Serverkonsolidierung
- bessere Auslastung der Hardware
- flexible Erweiterung
- Energieeinsparung
- *leichtere Verwaltung ?*

Virtualisierungs-Cluster

Clusterverser und SAN gebündelt und redundant verbinden



Technologie: LVM

Logische Volumina statt Partitionen

- leicht zu vergrößern / verkleinern
- einfach mit Namen zu bezeichnen
- (kleine) Snapshots ermöglichen

Sichern eines LVM (einer virtuellen Maschine)

Clonen eines LVM → viele ähnliche vM

Disks → Raid → LVM → OS Installation

Technologie: iSCSI

- SCSI über TCP / Ethernet
 - + vorhandene Netzwerktechnik nutzen
 - Overhead
- Daten publizieren: Target (Paket **iscsitarget**)
iet-Daemon; */etc/ietd.conf*: Targets mit LUNs definieren
- Daten einbinden: Initiator (Paket **open-iscsi**)
nach „Login“ stehen die LUNs eines iSCSI-Targets
unter */dev/disk/by-path* als Blockdevice zur Verfügung

Virtualisierung: Xen

- frei verfügbar und einsetzbar (GPL/Opensource)
- Support in der Industrie
- Performance fast wie auf Hardware
- unmodifizierte Systeme (Windows) virtualisierbar
- Live-Migration von vM
- kann Hardware an vM exklusiv übergeben
- Subnetting / NAT

Virtualisierung: XEN

XEN 3.3

- SLES11
- openSUSE 11.1

XEN 3.4

- openSUSE 11.2

!! openSUSE liefert XEN zum aktuellen Kernel

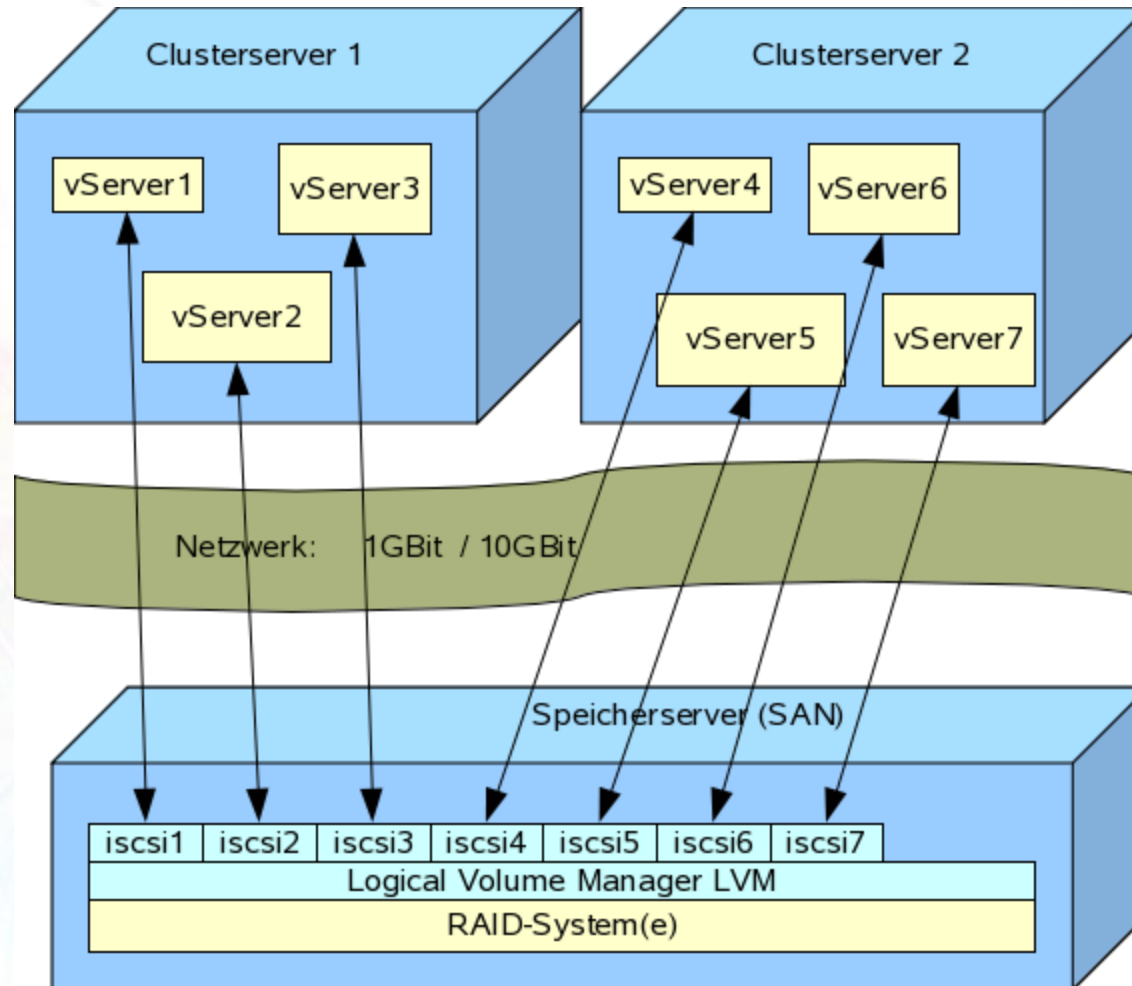
Clusterserver

- Clusterserver tragen die virtuellen Maschinen und damit die Verarbeitungsleistung
- Sie benötigen selbst nur Massenspeicher für das Grundsystem (XEN: dom0)
- Handhaben die iSCSI-Verbindung in der dom0

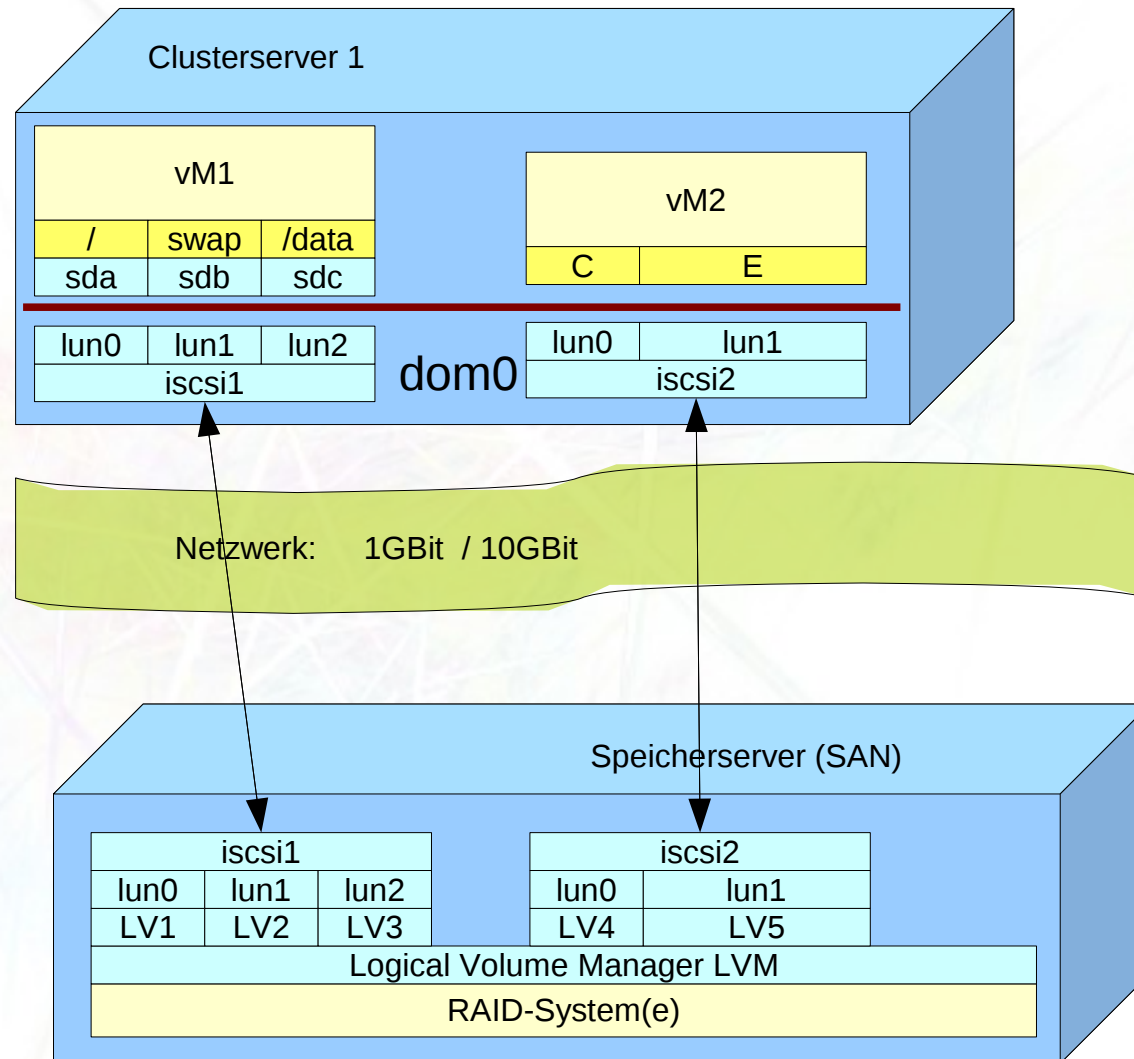
Virtuelle Maschinen

- arbeiten in den Clusterservern (verwenden deren Speicher und CPU)
- booten via iSCSI vom SAN („sehen“ nur SATA/IDE -Laufwerke)
- Umzug im laufenden Betrieb auf andere Clusterserver möglich (Migration)

Clusterstruktur



Clusterstruktur im Detail



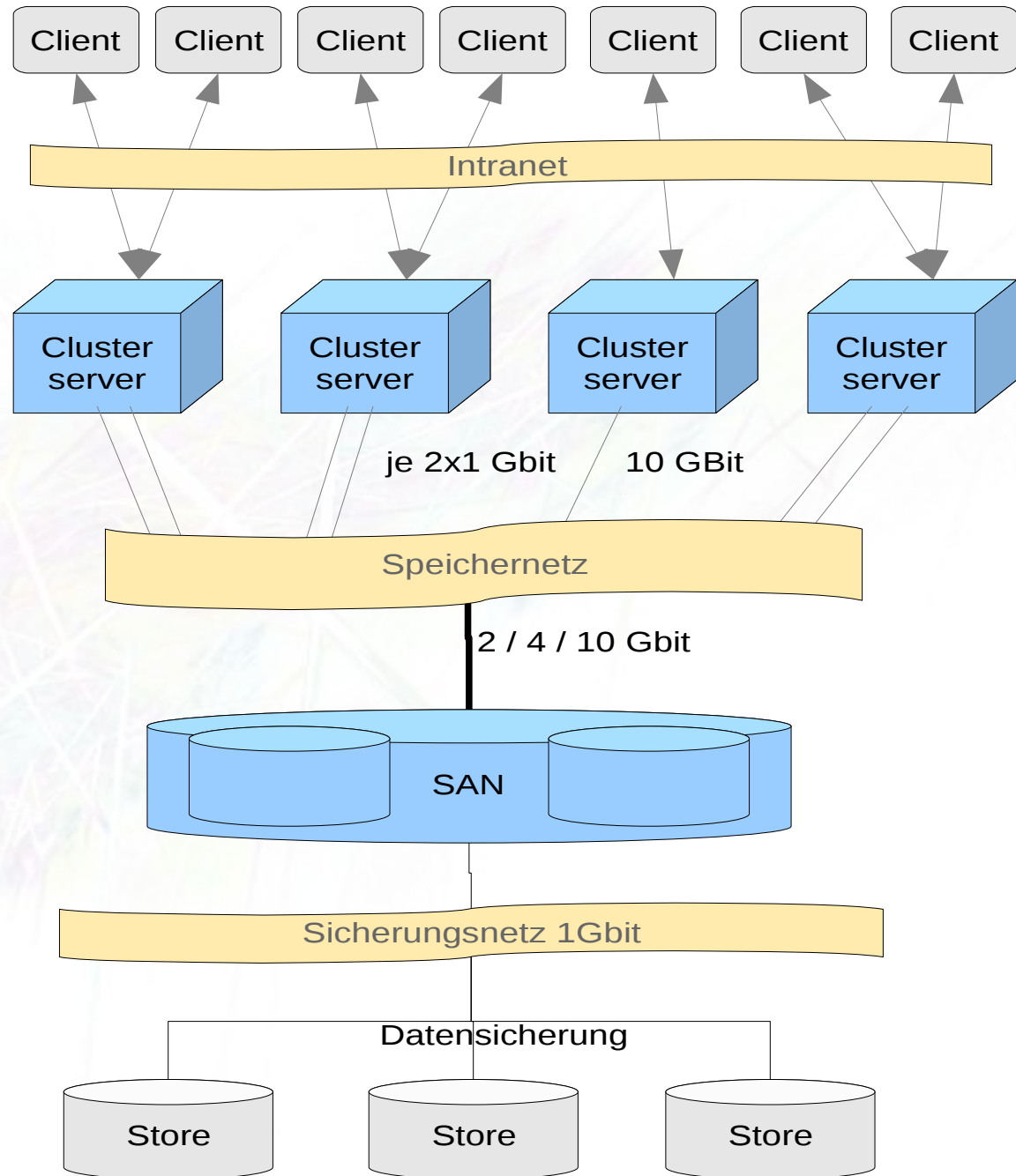
SAN

- Disks → RAID → DRBD → LVM → iSCSI
- *Logische Volumina* beherbergen vollständig die Betriebssystem-Installationen und ggf. zusätzliche Datenbereiche
- Mehrere LVs in einem iSCSI-Target bündeln:
ein iSCSI-Target == eine vM

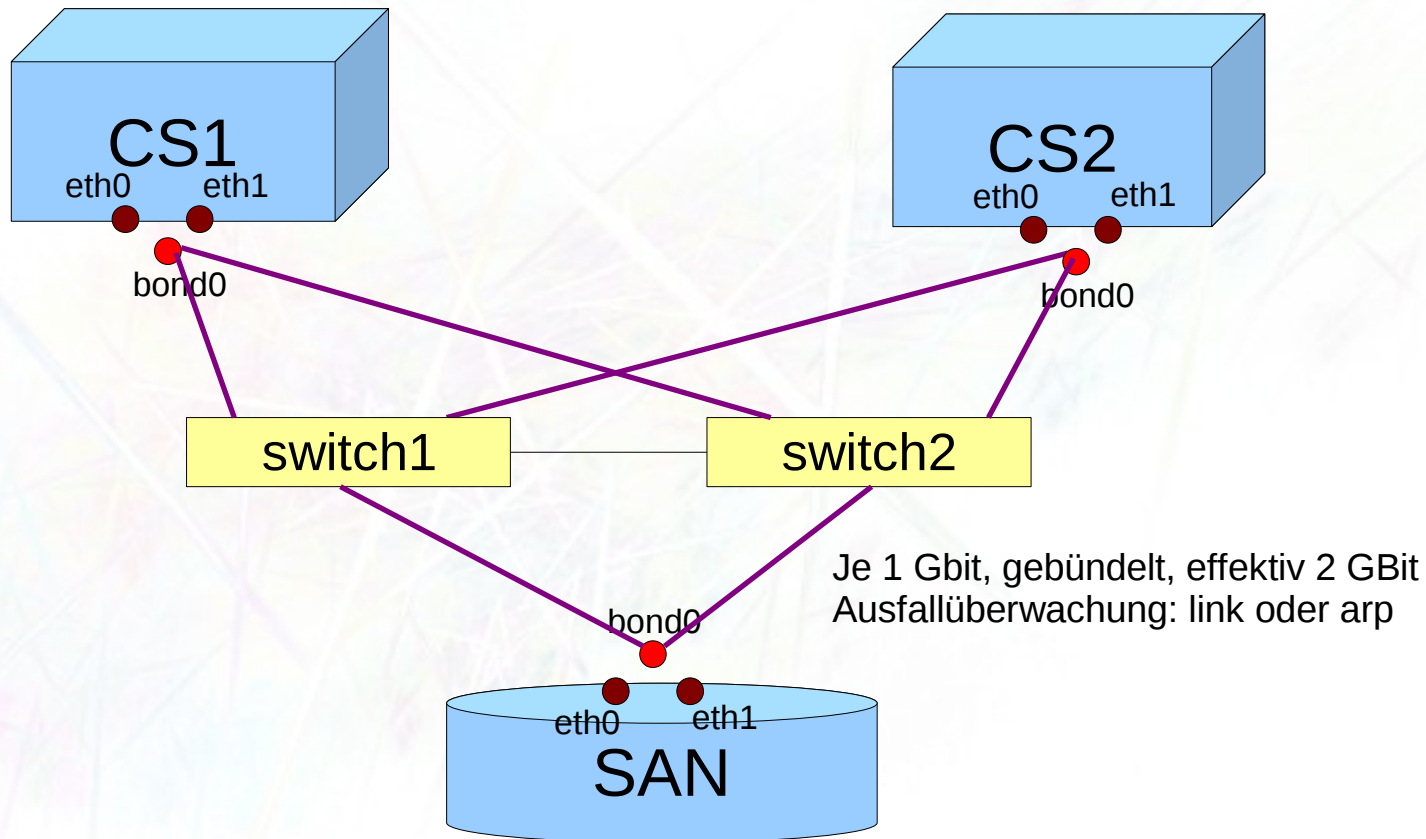
SAN: DRBD + Heartbeat!

- Echtzeit-Datenabgleich über ein separates Netz
aktiver SAN-Server → passiver SAN-Server
- Überwachung der SAN-Server gegenseitig
- Bildung der **SAN Cluster-IP** und Zuweisung auf den aktiven (DRBD) SAN-Server
- Umschaltung innerhalb weniger Sekunden

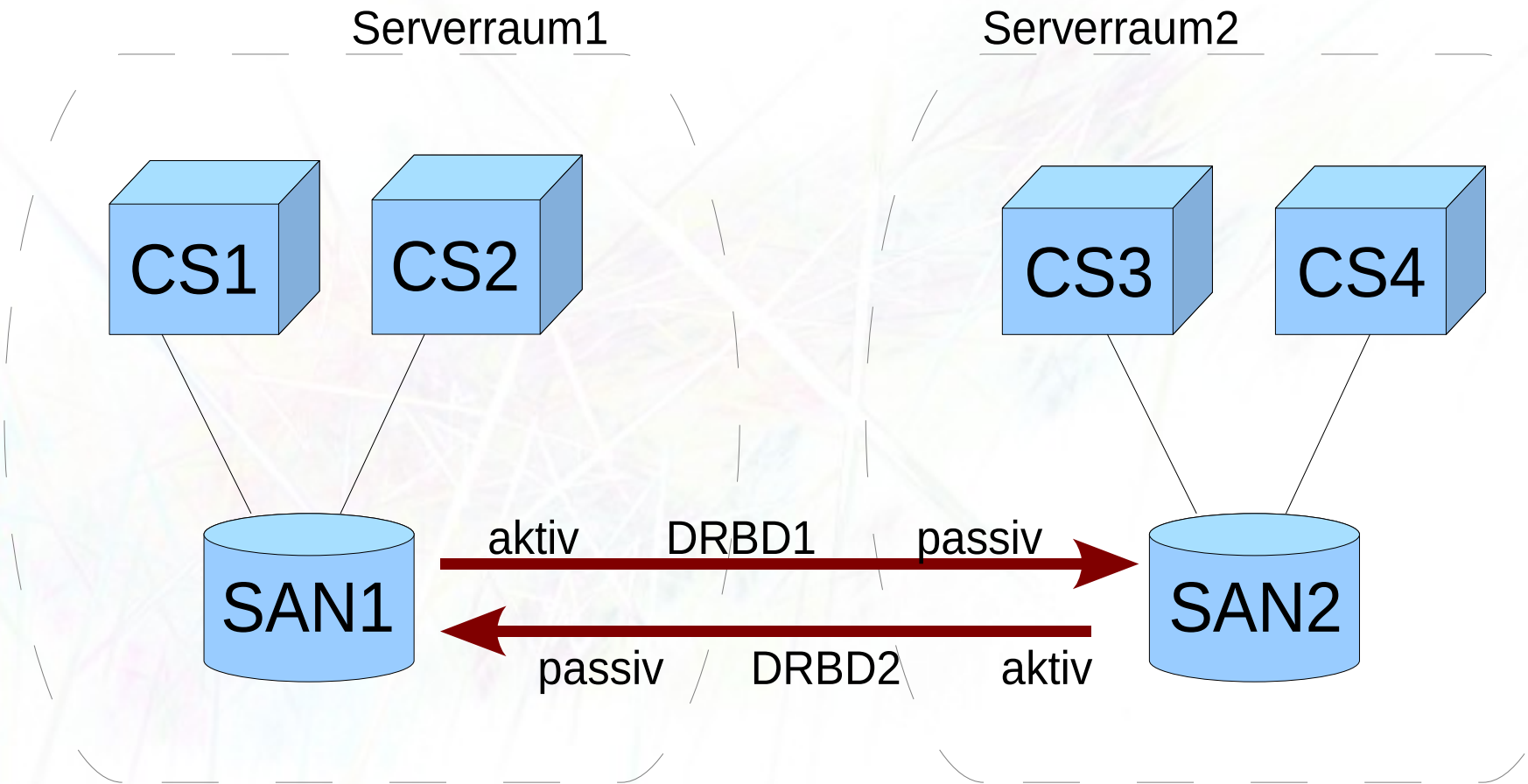
Vier Ebenen Architektur



Ausfallsicherheit Speichernetz



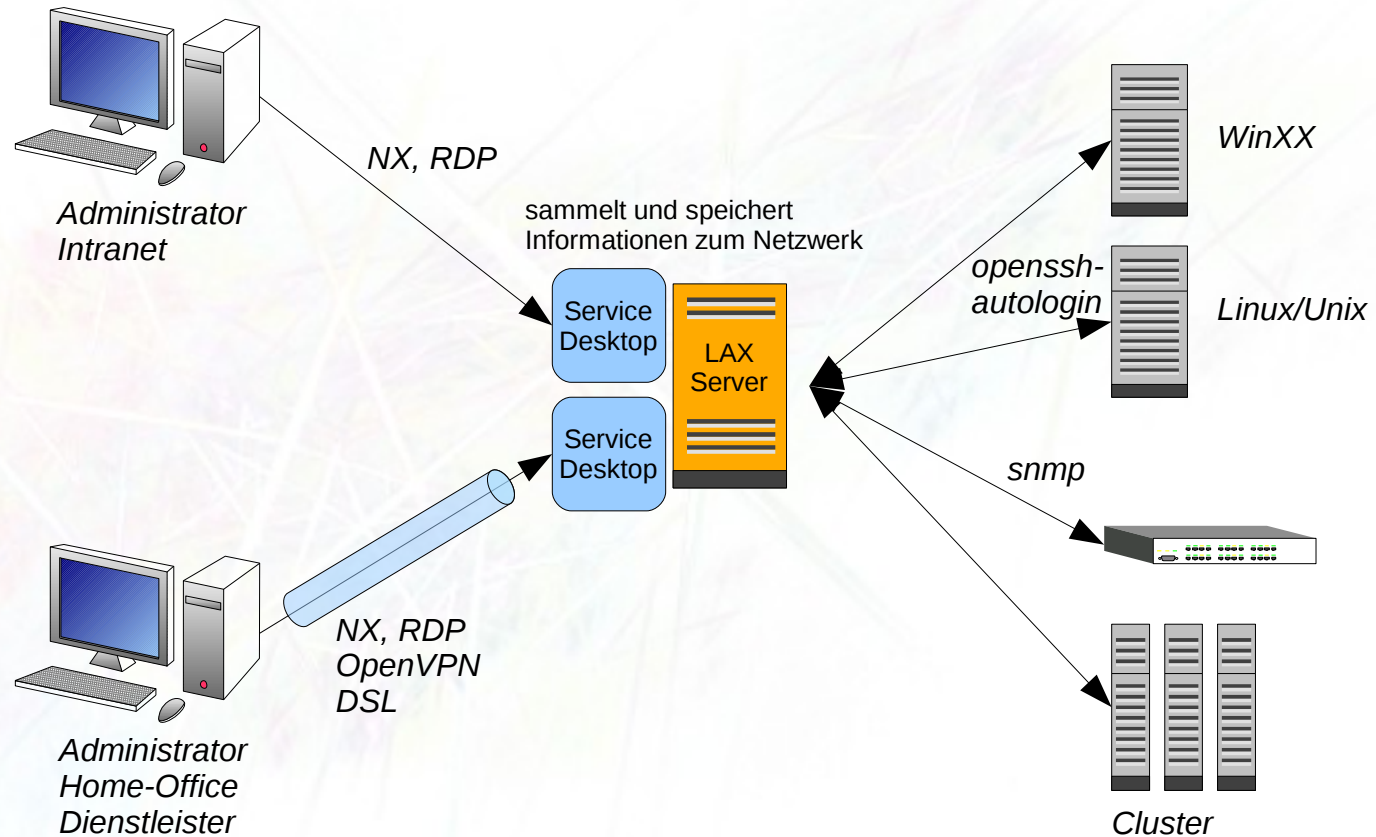
Ausfallsicherheit Cluster



Verwaltung: LAX Admin Framework

- Scriptsammlung für typische Admin-Aufgaben
- Admin-Server als Ausgangspunkt der Operationen und Sammelpunkt der Informationen
- openssh-Autologin Channels zu den Ziel-Hosts
- KDE-basierter Admin-Desktop
- Visualisierung, Monitoring
- Einfache Admin-GUIs zur Server-Steuerung
 - X.509 Zertifikate-Manager + OpenVPN Server
 - Verschiedene Serverdienste
 - Virtualisierungscluster

Administration via LAX Server



LAX Desktop

The screenshot displays the LAX Desktop interface, which is a remote control desktop environment. The interface is organized into several sections:

- Host:** Located on the left side, it lists system components: openSUSE, hermes, mailscaan.mailscaan, and alarmzeiten.vym.
- Controller:** Located in the center, it lists system components: ttpc, hoth, snoopy, devel, and winxp2. Each component has a status indicator (green or red) and a list of services.
- Service:** Located at the bottom, it lists services: laxmonitor, PostgresDB, laxtad, Tunnel80, apache-ttpc, Monitoring-Fehler, offene_Alarme, openvpn-Server, firewall-DL, intraweb, apache-devel, and winxp2-sshd.
- Gruppe:** Located on the right side, it lists system components: 99, mail, and herkules. Each component has a status indicator and a list of services.
- System Information:** Located in the top right, it shows "Anz. Prozesse Anzahl: 186" and "KDE.ORG".
- Calendar:** Located in the bottom right, it shows a date "2009-02-19 13:19" and a phone number "Tel: 0987654321".
- Taskbar:** Located at the bottom, it shows the LAX logo, the text "Remote Control Desktop", and several icons (LAX, doc, vat).

The interface is designed to provide a comprehensive overview of the system's status and services, allowing for remote control and management.

LAX Clustermanager

- Bash-Skripte rund um LVM, iSCSI, DRBD und Xen zur Vereinfachung der Administration
- einfache GUIs
- 3 Typen von Cluster-Instanzen:
 - einzelner Virtualisierungsserver
 - Active-Active-Cluster (DRBD)
 - iSCSI SAN Cluster

LAX Clustermanager

Namenskonventionen Hostname / LV / iSCSI

Bsp: LVs für Host **deb1**:

deb1-root system -wi-ao 10,00G

deb1-swap system -wi-ao 256,00M

iSCSI:

Target iqn.san1.klasse.kurs:**deb1**

Lun 1 Path=/dev/system/**deb1-swap**,Type=fileio

Lun 0 Path=/dev/system/**deb1-root**,Type=fileio

LAX Clustermanager

OS-Installation aus Vorlagen

Kopie + Änderung zur gebrauchsfertigen VM
Scripte zur Anpassung der Netzwerkanbindung für
openSUSE, Debian und WinXP (sysprep)

```
vx-install -i 192.168.30.223 -r 512M -f  
10G -s 1G -c 1 -d "Mein Testerver"  
cl1 os10.3-base testsrv1
```

/
Cluster

/
Vorlage

/
Name vM

LAX Clustermanager

Betrieb

`vx-start cl1 testserv1`

`vx-active | -inactive cl1`

`vx-stop | -del | -kill cl1 testserv1`

`vx-move cl1 testserv1 cs2`

Cluster

vM

Ziel-Clusterserver

LAX Clustermanager

Sicherung / Disaster Recovery

- vM im mittels LVM Snapshots sichern
 - ! laufenden Betrieb: Vorsicht bei Datenbanken!
- Sicherung per openssh auf Server als .tb2

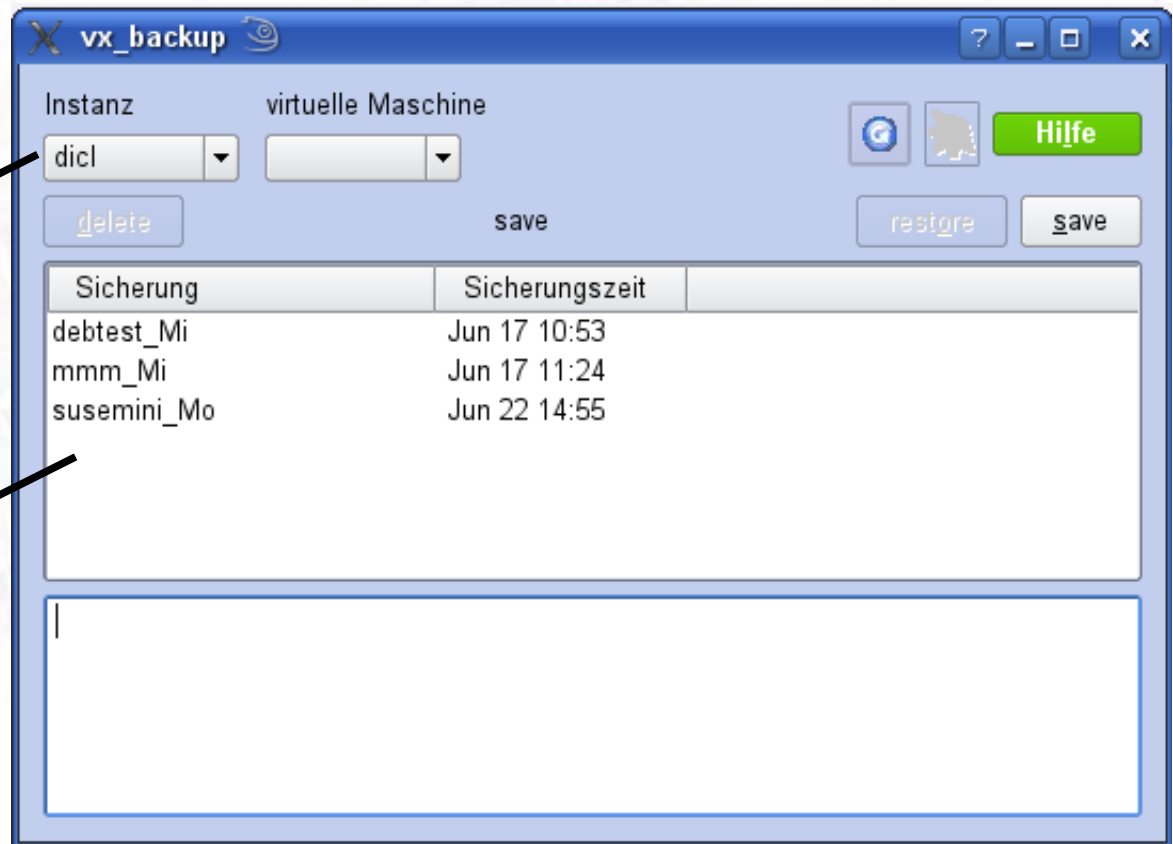
`vx-backup-add | -del | -list | -restore cl1 testsrv1`

- Vollsicherung inkl. Org-Daten, z.B. Größe der LV's
- Win-Maschinen per *dd* und *ntfsclone* sichern

LAX Clustermanager

Sicherung / Disaster Recovery

Cluster-Instanz



The screenshot shows the 'vx_backup' application window. At the top, there are two dropdown menus: 'Instanz' (set to 'dicl') and 'virtuelle Maschine'. To the right are icons for help and a green 'Hilfe' button. Below these are buttons for 'delete', 'save', 'restore', and 'save'. The main area contains a table with two columns: 'Sicherung' and 'Sicherungszeit'. The table lists three backup entries: 'debttest_Mi' (Jun 17 10:53), 'mmm_Mi' (Jun 17 11:24), and 'susemini_Mo' (Jun 22 14:55). Below the table is a large empty text area.

Sicherung	Sicherungszeit
debttest_Mi	Jun 17 10:53
mmm_Mi	Jun 17 11:24
susemini_Mo	Jun 22 14:55

Liste der Sicherungen

LAX Clustermanager GUI

The screenshot displays the LAX Clustermanager GUI. At the top left, the window title is "laxClustermanager". The main interface is divided into several sections:

- Cluster-Baum (Doppelklick):** A tree view showing the cluster structure. The root is "dicl", which contains sub-clusters "cs1" and "cs2". Another cluster "intra" is shown with a sub-cluster "herkules".
- Eigenschaften Clusterserver:** A panel showing server properties for the selected cluster. It includes fields for "freier Speicher (MB)", "CPU", "Eigenschaften", and "Disk:". The "Disk:" field shows "sancl: 226.26G" and "42.13G".
- virtuelle Maschinen:** A section with various control buttons: a grey circle, "Stop", "Kill", another grey circle, "Move", "Defs", "New", another grey circle, "Con", "SSH", "RDP", and "VNC".
- AKTIVE virtuelle Maschinen:** A table listing active virtual machines. The status "RUHENDE" is shown in a grey button on the right. The table has columns for Maschine, Cserver, Speicher, vCPU, Status, Zeit, Prio, Disk, MAC, IP-Adresse, Template, and Info.

Maschine	Cserver	Speicher	vCPU	Status	Zeit	Prio	Disk	MAC	IP-Adresse	Template	Info
clamxp2	cs1	512	1	-b----	34719.5	3	4G	00:16:3E:3F:99:66	192.168.30.144	wxp	Windows
debtest	cs1	256	1	-b----	74.0	1	4G	00:16:3E:3D:3E:D9	192.168.30.142	debian	Linux
newjoomla	cs1	512	1	-b----	363.1	2	4G	00:16:3E:31:EC:7A	192.168.30.157	joomla	Linux
t1	cs2	512	1	-b----	266.2	3	6G	00:16:3E:6F:FD:43	192.168.30.158	os11.1-base	Linux
win3	cs2	512	1	-b----	32155.4	3	4G	00:16:3E:DD:B7:7D	192.168.30.140	wxp	Windows
winxpc12	cs1	512	1	-b----	29093.6	3	4G	00:16:3E:99:39:31	192.168.30.143	wxp	Windows

LAX Monitoring + Alarmierung

Monitoring-Daemon, min: 1 Minute

Portcheck, ping, Init-Script, Nagios-Script

Alarmscripte zu Ressource zuordnen:

- Dienst oder vM neu starten
- Cluster reorganisieren
- Mail / SMS / PopUp-Fenster

Alarmer und Monitoring in Postgres-DB

Cluster reorganisieren

Ausfall eines Clusterservers durch Monitoring erkannt



Alarmscript *clusterreorg* gestartet



Neuplanen der Clusterbelegung (Prio, Speicher)



Stoppen, Verschieben, Starten von VM

Vielen Dank!

www.teegee.de

www.teegee.de/download/lax.tgz