

Snapshot-Backups mit Dirvish & Rsync

Robert Sander

clt2012@gurubert.de

Herausforderung

- Datensicherung soll sein
 - konsistent
 - komplett
 - kostengünstig
 - kompakt
 - komfortabel
 - schnell wiederherstellbar

Lösung

- Dirvish
 - LVM Snapshots als Quelle für konsistenten Zustand
 - benutzt rsync mit --link-dest für hardlinks
 - geschrieben von JW Schultz († 27.3.2004)
- Ähnliche Tools:
 - rdiff-backup
 - rsnapshot

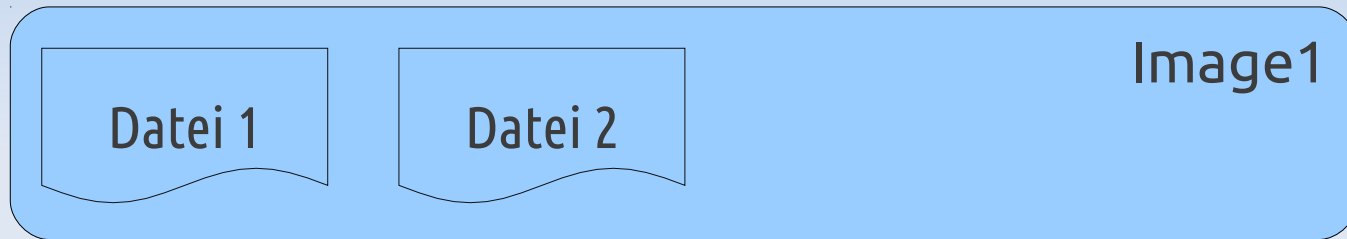
Platzsparen mit Hardlinks

Datei 1

Datei 2

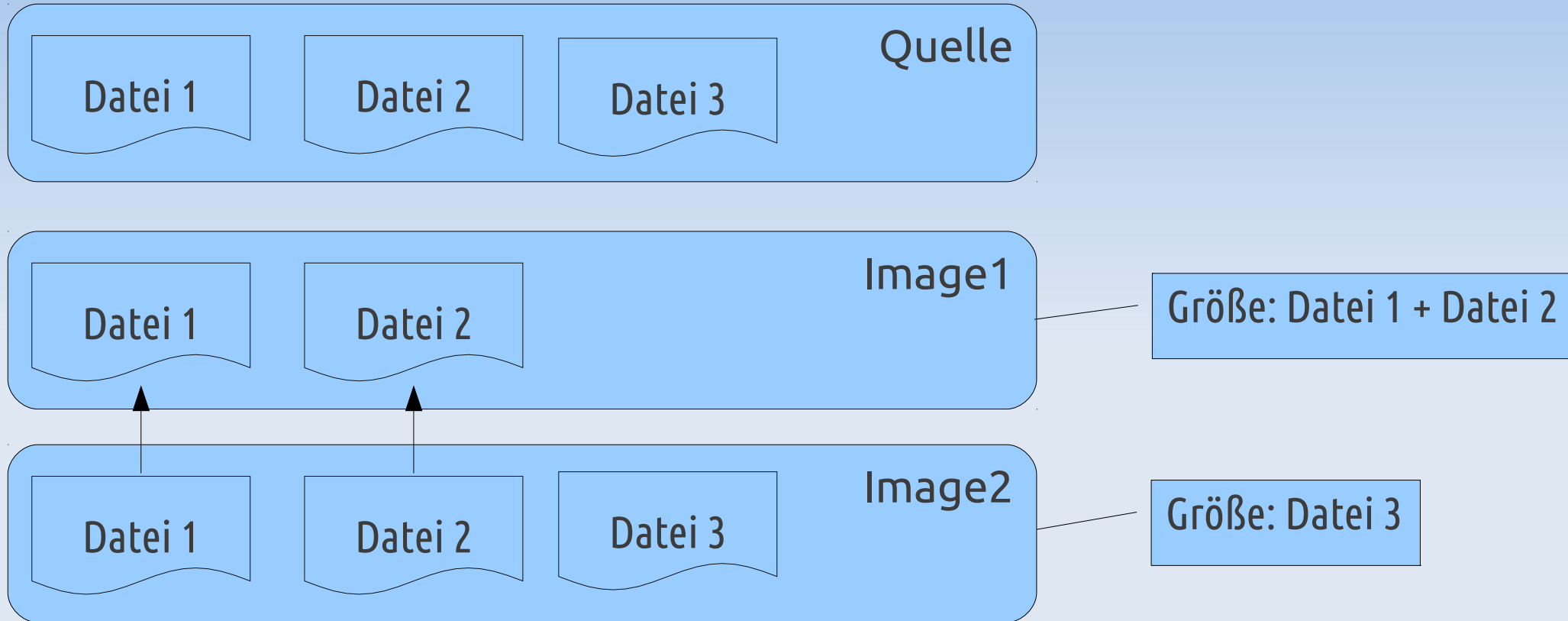
Quelle

Platzsparen mit Hardlinks

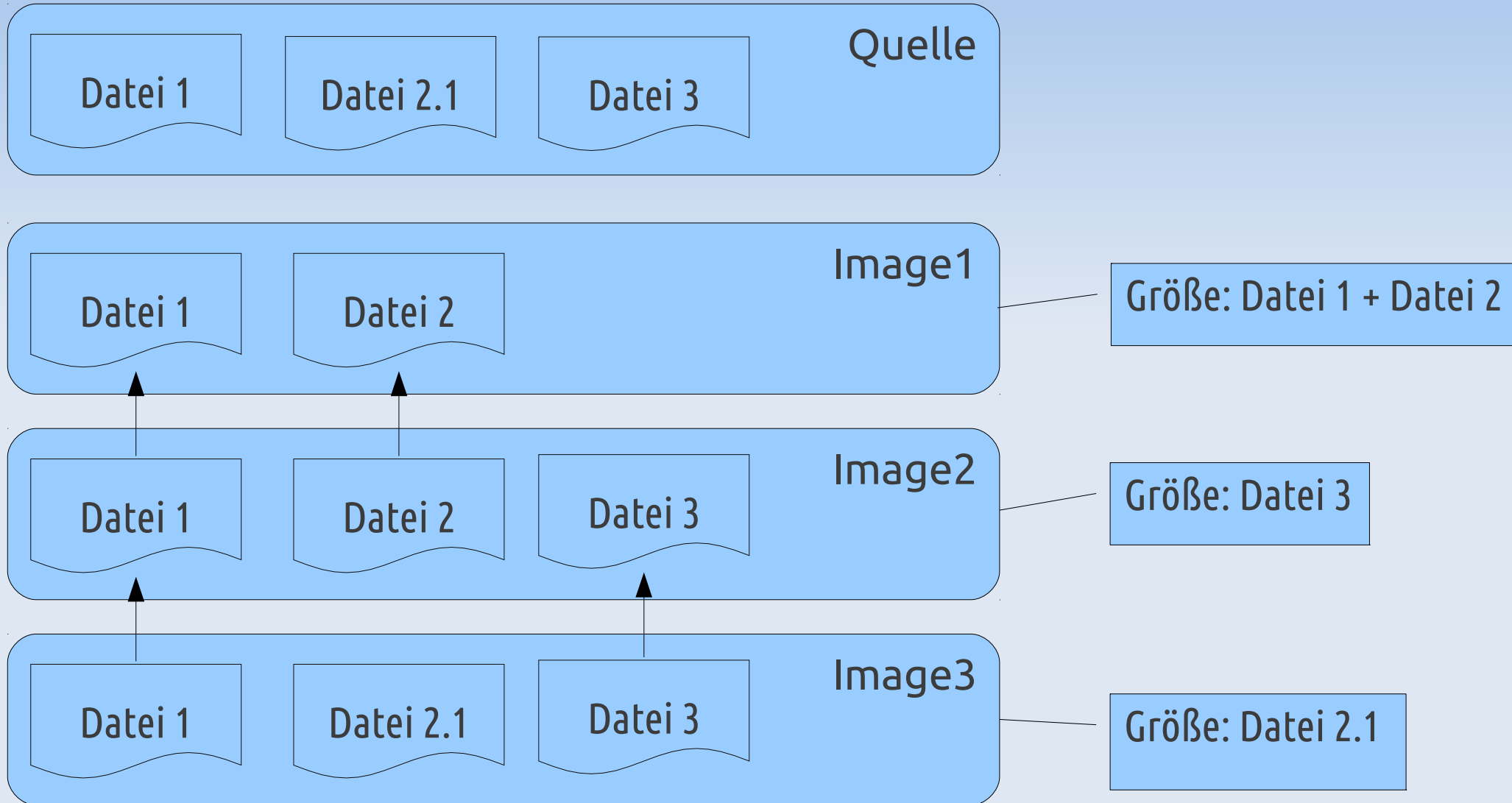


Größe: Datei 1 + Datei 2

Platzsparen mit Hardlinks



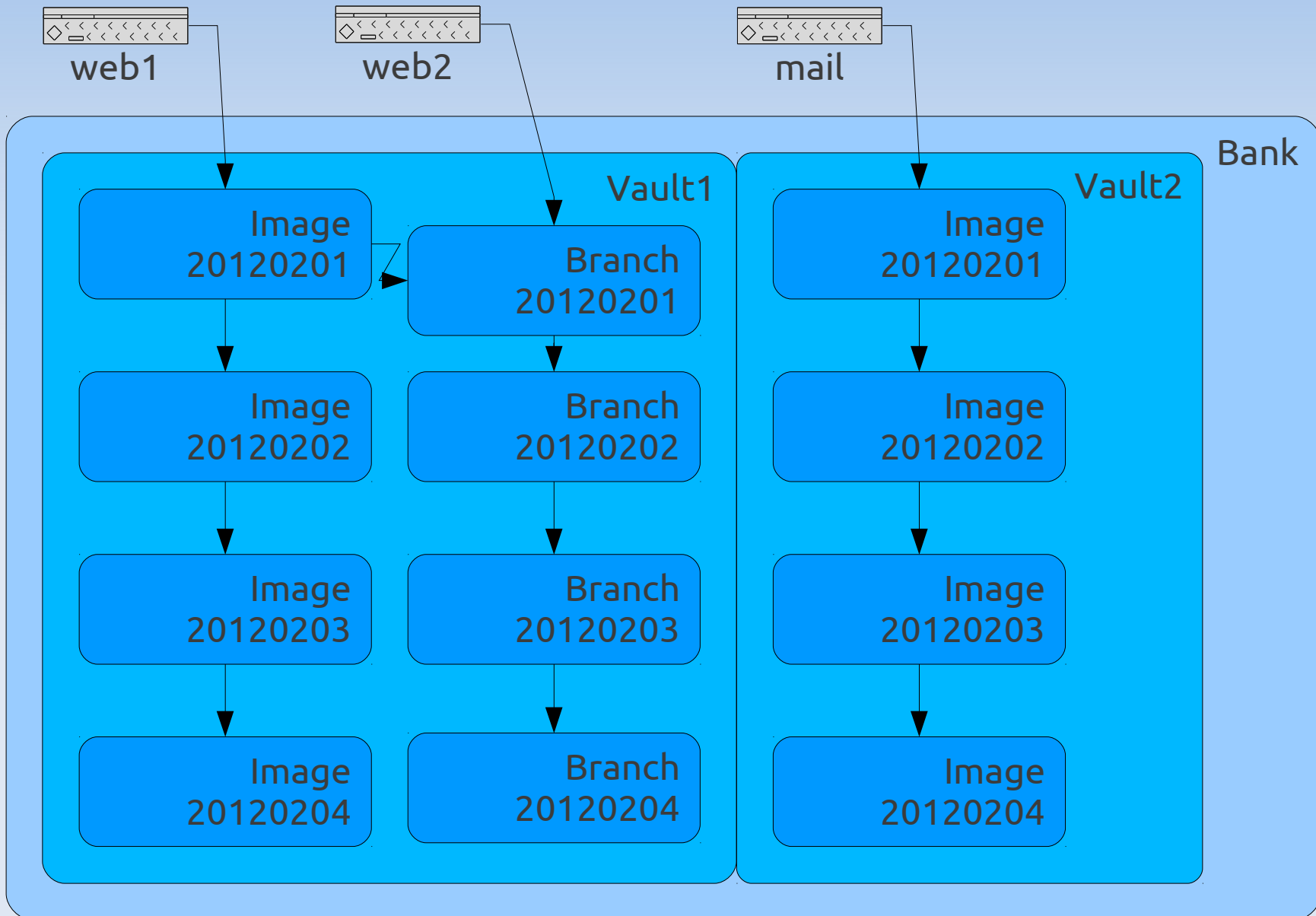
Platzsparen mit Hardlinks



Dirvish Übersicht

- Bank: Verzeichnis mit Vaults
- Vault: Enthält die Images einer Quelle
 - auf eigenem Filesystem
- Image: ein "SaveSet" , bestehend aus:
 - index
 - log
 - summary
 - tree
- Branch: enthält die Images einer fast identischen Quelle

Dirvish Übersicht



Dirvish Konfiguration

- `/etc/dirvish/master.conf`
 - Globale Optionen inkl. Bank(s)
- `/etc/dirvish/client.conf`
 - Optionen für einzelne Clients
- `bank/vault/dirvish/default.conf`
 - Quellenangabe, Skripte, Exclusionen
- `bank/vault/dirvish/branch.conf`
 - Das gleiche für den jeweiligen Branch

`bank: /path/to/bank`

`index: gzip`

`log: gzip`

`image-default: %Y%m%d`

`Runall:`

`web:web1`

`web:web2`

`mail`

`expire-default: +6 weeks`

`expire-rule:`

`* * 1-7 * 1 +10 years`

Dirvish Konfiguration

- `/etc/dirvish/master.conf`
 - Globale Optionen inkl. Bank(s)
- `/etc/dirvish/client.conf`
 - Optionen für einzelne Clients
- `bank/vault/dirvish/default.conf`
 - Quellenangabe, Skripte, Exclusionen
- `bank/vault/dirvish/branch.conf`
 - Das gleiche für den jeweiligen Branch

- `/etc/dirvish/mail.conf:`
`speed-limit: 3`
`zxfer: yes`

Dirvish Konfiguration

- `/etc/dirvish/master.conf`
 - Globale Optionen inkl. Bank(s)
- `/etc/dirvish/client.conf`
 - Optionen für einzelne Clients
- `bank/vault/dirvish/default.conf`
 - Quellenangabe, Skripte, Exclusionen
- `bank/vault/dirvish/branch.conf`
 - Das gleiche für den jeweiligen Branch

- `/path/to/bank/mail/dirvish/default.conf:`
`client: mail`
`tree: /`
`xdev: 0`
`pre-client: /sbin/pre.sh`
`post-client: /sbin/post.sh`
`exclude:`
 - `lost+found/`
 - `/tmp`
 - `/var/tmp/`

Dirvish Konfiguration

- `/etc/dirvish/master.conf`
 - Globale Optionen inkl. Bank(s)
- `/etc/dirvish/client.conf`
 - Optionen für einzelne Clients
- `bank/vault/dirvish/default.conf`
 - Quellenangabe, Skripte, Exclusionen

- `bank/vault/dirvish/branch.conf`
 - Das gleiche für den jeweiligen Branch

- `/path/to/bank/web/dirvish/web1.conf:`

`client: web1`

`tree: /var/www`

- `/path/to/bank/web/dirvish/web2.conf:`

`client: web2`

`tree: /var/www`

Dirvish Konfiguration

- tree
 - Quellverzeichnis
- image-temp: *current*
 - Das aktuelle Image hat immer den gleichen Namen
- client: *servername*
 - kopiert von lokaler oder NFS-Quelle
 - ohne ssh als Transport
- exclude
 - Wenn Teile der Quelle nicht gesichert werden sollen
 - analog zu rsync --exclude
- Weitere rsync-Parameter
 - speed-limit, xdev, zxfcr, rsh

Branches

- Wenn mehrere Quellen identisch starten
- Master in der default.conf
- Weitere Quellen in eigenen branch.conf, init dann mit --link-dest auf Master
- ➔ Erstes Backup schon mit Hardlinks
- /etc/dirvish/master.conf:
 - vault1
 - vault1:branch
 - vault2

Scripting

- Pre und Post auf Server und Client möglich
 - `$DIRVISH_{SERVER,CLIENT,SRC,DEST,IMAGE,EXCLUDE}`
- Pre: Erzeugen von LVM Snapshots, mounten
 - Aus einem Skript, welches `$DIRVISH_SRC` auswertet
 - z.B. `DIRVISH_SRC=/mnt/lvm/dirvish/VG01/LV01`
 - `$LV=`basename $DIRVISH_SRC`; $VG analog`
 - `lvcreate -s -L 20G -n dirvish_$LV /dev/$VG/$LV`
 - `mount -o ro /dev/$VG/dirvish_$LV $DIRVISH_SRC`
- Post: Unmounten, löschen von LVM Snapshots
 - `umount /dev/$VG/dirvish_$LV`
 - `lvremove -f /dev/$VG/dirvish_$LV`

Vergleich

	dirvish	rdiff-backup	rsnapshot
Branches	Ja	Nein	Nein
Speichersparteknik	hardlinks	libsyntax erlaubt auch Änderungen in Dateien	hardlinks
Scripting	pre / post client / server DIRVISH-Variablen	Nicht eingebaut	pre client
History Expiration	1 Verzeichnis pro Image Expire-Rules	"Ewige" Aufzeichnung Manuelles Purging	Intervalldefinition (hourly, daily, weekly)
Automation	1 Cronjob	1 Cronjob pro Quelle	1 Cronjob pro Intervall
Extras	index.gz dirvish-locate full-text available	Fuse-FS für einzelne Images	

Links

- <http://dirvish.org>
- <http://wiki.dirvish.org>
- <http://rdiff-backup.nongnu.org>
- <http://rsnapshot.org>

**Wir suchen:
Admins, Consultants, Trainer!**

Wir bieten:
Spannende Projekte, Kundenlob, eigenständige
Arbeit, keine Überstunden, Teamarbeit

...und natürlich: Linux, Linux, Linux...

<http://www.heinlein-support.de/jobs>

Snapshot-Backups mit Dirvish & Rsync

Robert Sander
clt2012@gurubert.de

Herausforderung

- Datensicherung soll sein
 - konsistent
 - komplett
 - kostengünstig
 - kompakt
 - komfortabel
 - schnell wiederherstellbar

Lösung

- Dirvish
 - LVM Snapshots als Quelle für konsistenten Zustand
 - benutzt rsync mit --link-dest für hardlinks
 - geschrieben von JW Schultz († 27.3.2004)
- Ähnliche Tools:
 - rdiff-backup
 - rsnapshot

3 / 19

Dirvish ist eine kleine Sammlung von Perl-Skripten.

Platzsparen mit Hardlinks

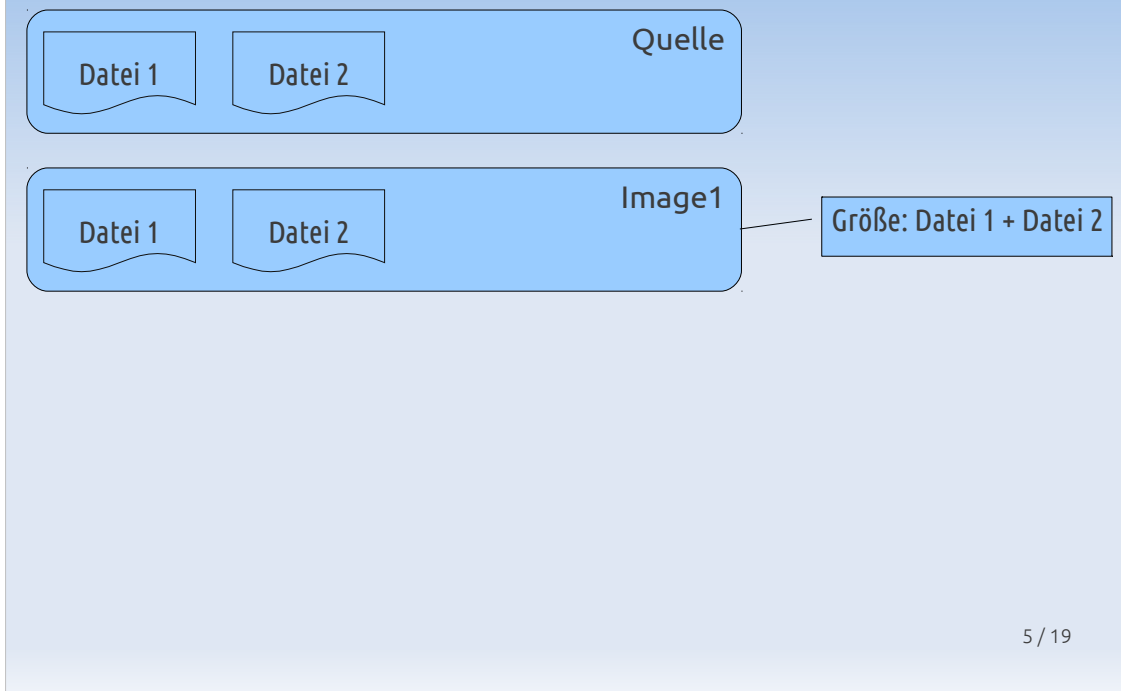


Hardlinks verweisen aus unterschiedlichen Verzeichnissen auf den selben Dateiinhalt

Wird ein Image gelöscht, ist die Datei in den anderen Images noch vorhanden

Bei eher kleiner durchschnittlicher Dateigröße oder Dateien, die sich inhaltlich nicht mehr ändern sehr effizient

Platzsparen mit Hardlinks

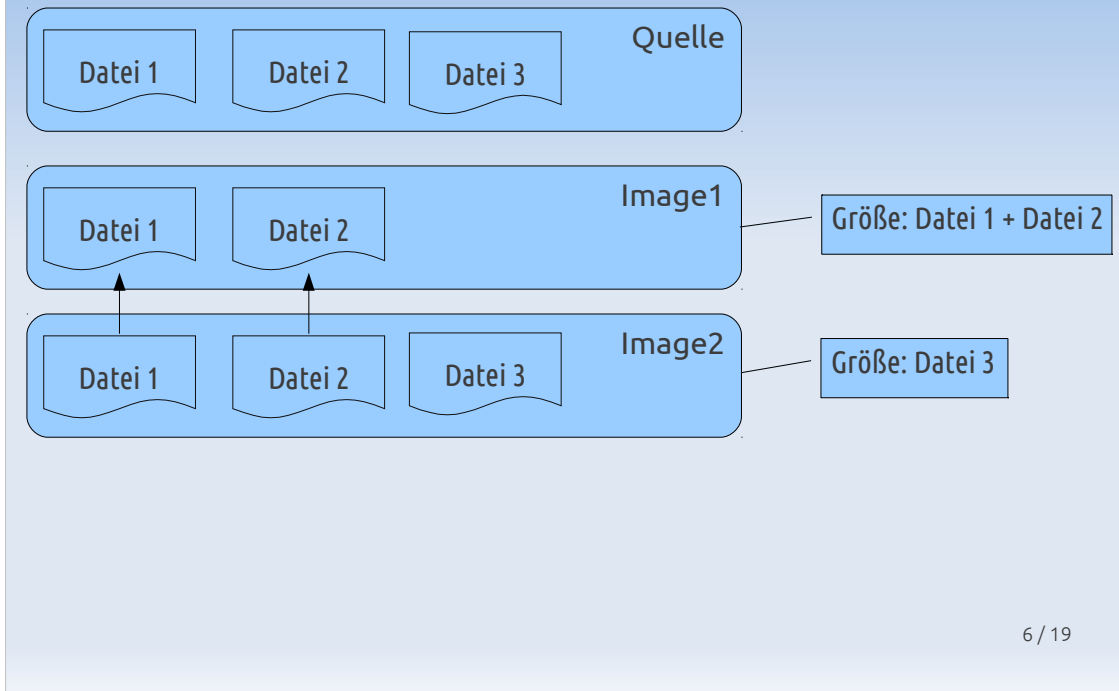


Hardlinks verweisen aus unterschiedlichen Verzeichnissen auf den selben Dateiinhalt

Wird ein Image gelöscht, ist die Datei in den anderen Images noch vorhanden

Bei eher kleiner durchschnittlicher Dateigröße oder Dateien, die sich inhaltlich nicht mehr ändern sehr effizient

Platzsparen mit Hardlinks

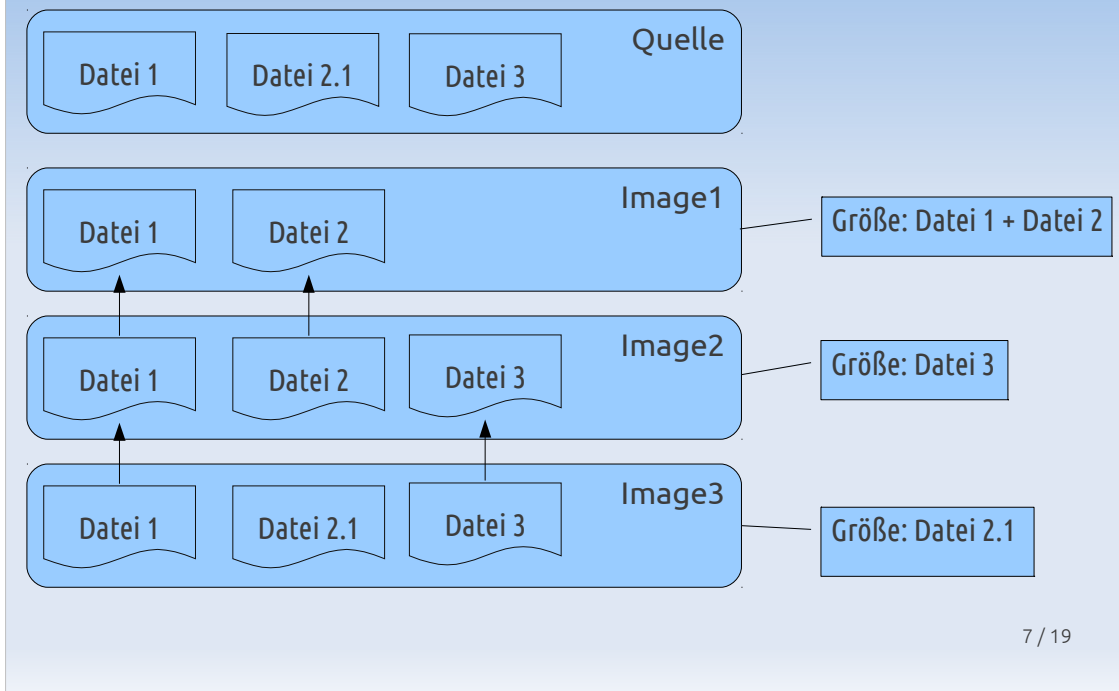


Hardlinks verweisen aus unterschiedlichen Verzeichnissen auf den selben Dateiinhalt

Wird ein Image gelöscht, ist die Datei in den anderen Images noch vorhanden

Bei eher kleiner durchschnittlicher Dateigröße oder Dateien, die sich inhaltlich nicht mehr ändern sehr effizient

Platzsparen mit Hardlinks



Hardlinks verweisen aus unterschiedlichen Verzeichnissen auf den selben Dateiinhalt

Wird ein Image gelöscht, ist die Datei in den anderen Images noch vorhanden

Bei eher kleiner durchschnittlicher Dateigröße oder Dateien, die sich inhaltlich nicht mehr ändern sehr effizient

Dirvish Übersicht

- Bank: Verzeichnis mit Vaults
- Vault: Enthält die Images einer Quelle
 - auf eigenem Filesystem
- Image: ein "Saveset" , bestehend aus:
 - index
 - log
 - summary
 - tree
- Branch: enthält die Images einer fast identischen Quelle

8 / 19

Es kann auch mehrere Banks geben

Ein Vault mit den Images sollte auf einem eigenen Dateisystem (eigenes LVM Volume) existieren, das vereinfacht das Management

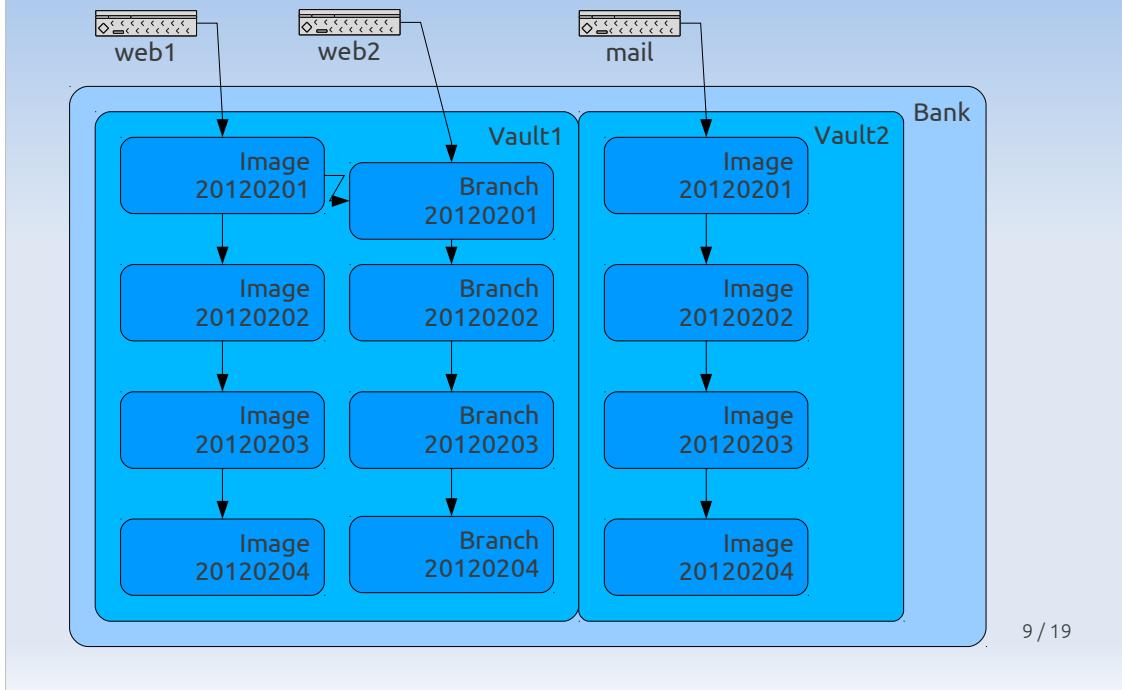
Ein eigenes Dateisystem ermöglicht die Trendanalyse (Änderungsrate, Füllstand) und ist bei einem Filesystemcheck schneller

Ein einzelnes Image enthält die Zusammenfassung (summary), die Logdatei und einen Index über alle im Image enthaltenen Dateien.

Index ermöglicht dirvish-locate die Suche nach Dateien im Vault

Ein Branch ist ein Image, das ein existierendes Image als --link-dest hat, d.h. die Quellen stimmen in großen Teilen überein

Dirvish Übersicht



Ein Vault enthält die Daten einer Quelle (bzw. mehrerer bei Benutzung von Branches)

Eine Bank enthält mehrere Vaults

Es können mehrere Banks existieren

Dirvish Konfiguration

- `/etc/dirvish/master.conf`
 - Globale Optionen inkl. Bank(s)
- `/etc/dirvish/client.conf`
 - Optionen für einzelne Clients
- `bank/vault/dirvish/default.conf`
 - Quellenangabe, Skripte, Exclusionen
- `bank/vault/dirvish/branch.conf`
 - Das gleiche für den jeweiligen Branch

```
bank: /path/to/bank
index: gzip
log: gzip
image-default: %Y%m%d
Runall:
    web:web1
    web:web2
    mail
expire-default: +6 weeks
expire-rule:
    * * 1-7 * 1 +10 years
```

10 / 19

Die Konfiguration verteilt sich auf mehrere Dateien, so daß flexibel globale Optionen gesetzt und Einstellungen pro Client und Vault gemacht werden können

Dirvish Konfiguration

- `/etc/dirvish/master.conf`
 - Globale Optionen inkl. Bank(s)
 - `/etc/dirvish/client.conf`
 - Optionen für einzelne Clients
 - `bank/vault/dirvish/default.conf`
 - Quellenangabe, Skripte, Exclusionen
 - `bank/vault/dirvish/branch.conf`
 - Das gleiche für den jeweiligen Branch
- `/etc/dirvish/mail.conf:`
`speed-limit: 3`
`zxfer: yes`

11 / 19

Die Konfiguration verteilt sich auf mehrere Dateien, so daß flexibel globale Optionen gesetzt und Einstellungen pro Client und Vault gemacht werden können

Dirvish Konfiguration

- `/etc/dirvish/master.conf`
 - Globale Optionen inkl. Bank(s)
- `/etc/dirvish/client.conf`
 - Optionen für einzelne Clients
- **`bank/vault/dirvish/default.conf`**
 - Quellenangabe, Skripte, Exclusionen
- `bank/vault/dirvish/branch.conf`
 - Das gleiche für den jeweiligen Branch

```
▪ /path/to/bank/mail/dirvish/default.conf:  
client: mail  
tree: /  
xdev: 0  
pre-client: /sbin/pre.sh  
post-client: /sbin/post.sh  
exclude:  
    lost+found/  
    /tmp  
    /var/tmp/
```

12 / 19

Die Konfiguration verteilt sich auf mehrere Dateien, so daß flexibel globale Optionen gesetzt und Einstellungen pro Client und Vault gemacht werden können

Dirvish Konfiguration

- `/etc/dirvish/master.conf`
 - Globale Optionen inkl. Bank(s)
 - `/etc/dirvish/client.conf`
 - Optionen für einzelne Clients
 - `bank/vault/dirvish/default.conf`
 - Quellenangabe, Skripte, Exclusionen
 - `bank/vault/dirvish/branch.conf`
 - Das gleiche für den jeweiligen Branch
- `/path/to/bank/web/dirvish/web1.conf:`
`client: web1`
`tree: /var/www`
- `/path/to/bank/web/dirvish/web2.conf:`
`client: web2`
`tree: /var/www`

13 / 19

Die Konfiguration verteilt sich auf mehrere Dateien, so daß flexibel globale Optionen gesetzt und Einstellungen pro Client und Vault gemacht werden können

Dirvish Konfiguration

- tree
 - Quellverzeichnis
- image-temp: *current*
 - Das aktuelle Image hat immer den gleichen Namen
- client: *servername*
 - kopiert von lokaler oder NFS-Quelle
 - ohne ssh als Transport
- exclude
 - Wenn Teile der Quelle nicht gesichert werden sollen
 - analog zu rsync --exclude
- Weitere rsync-Parameter
 - speed-limit, xdev, zxfcr, rsh

14 / 19

Image-Temp ist nützlich, wenn das Backup von Disk noch auf ein anderes Medium (z.B. Tape) übertragen werden soll. Dadurch gibt es immer das aktuelle Image an einem festen Pfad

Ist eine Quelle nur als NFS-Share verfügbar, wird Client auf den eigenen Hostnamen gesetzt, rsync läuft dann nicht über ssh

Branches

- Wenn mehrere Quellen identisch starten
- Master in der default.conf
- Weitere Quellen in eigenen branch.conf, init dann mit --link-dest auf Master
- ➔ Erstes Backup schon mit Hardlinks
- /etc/dirvish/master.conf:
 - vault1
 - vault1:branch
 - vault2

Scripting

- Pre und Post auf Server und Client möglich
 - `$DIRVISH_{SERVER,CLIENT,SRC,DEST,IMAGE,EXCLUDE}`
- Pre: Erzeugen von LVM Snapshots, mounten
 - Aus einem Skript, welches `$DIRVISH_SRC` auswertet
 - z.B. `DIRVISH_SRC=/mnt/lvm/dirvish/VG01/LV01`
 - `$LV=`basename $DIRVISH_SRC`; $VG analog`
 - `lvcreate -s -L 20G -n dirvish_$LV /dev/$VG/$LV`
 - `mount -o ro /dev/$VG/dirvish_$LV $DIRVISH_SRC`
- Post: Unmounten, löschen von LVM Snapshots
 - `umount /dev/$VG/dirvish_$LV`
 - `lvremove -f /dev/$VG/dirvish_$LV`

16 / 19

DIRVISH_SERVER: Hostname des Backupsystems
DIRVISH_CLIENT: Hostname des Clientsystem
DIRVISH_SRC: Pfad zur Quelle auf dem Client
DIRVISH_DEST: Pfad zum Image auf dem Server
DIRVISH_IMAGE: Vault:Image:Branch
DIRVISH_EXCLUDE: Liste der Exclude-Statements

Vergleich

	dirvish	rdiff-backup	rsnapshot
Branches	Ja	Nein	Nein
Speichersparteknik	hardlinks	libsyc erlaubt auch Änderungen in Dateien	hardlinks
Scripting	pre / post client / server DIRVISH-Variablen	Nicht eingebaut	pre client
History Expiration	1 Verzeichnis pro Image Expire-Rules	"Ewige" Aufzeichnung Manuelles Purging	Intervalldefinition (hourly, daily, weekly)
Automation	1 Cronjob	1 Cronjob pro Quelle	1 Cronjob pro Intervall
Extras	index.gz dirvish-locate full-text available	Fuse-FS für einzelne Images	

Links

- <http://dirvish.org>
- <http://wiki.dirvish.org>
- <http://rdiff-backup.nongnu.org>
- <http://rsnapshot.org>

**Wir suchen:
Admins, Consultants, Trainer!**

Wir bieten:
Spannende Projekte, Kundenlob, eigenständige
Arbeit, keine Überstunden, Teamarbeit

...und natürlich: Linux, Linux, Linux...

<http://www.heinlein-support.de/jobs>