



*The Simple High Available  
Linux File Server*

## **Virtuelle Cold-Standby Server mit Linux**

**Schlomo Schapiro**  
Principal Consultant  
Leitung Virtualisierung & Open Source

28.05.2008

# Agenda

- **Hintergrundinformationen**
- **The Simple High Available Linux File Server**
- **Vorteile: Virtuelle Cold-Standby Server**

# Kundensituation

- Landesvermessung und Geobasisinformation Brandenburg
  - Katasterdaten des Landes Brandenburg
  - Amtliche Karten, Luftbilder usw.
  - Sammelt, speichert und verarbeitet Geoinformationsdaten
  - In Potsdam und Frankfurt / Oder
  - Heavily IT dependant
  - HP ist Hauptlieferant
    - 3 EVA (~40 TB)
    - > 100 Server



# Die Problemsituation

- Viele Systeme speichern Daten auf vielen Fileserver und lokal
- Viele komplexe Datenbeziehungen mit viel FTP und Kopien zwischen UNIX/Linux Servern und Windows Systemen
- Z.T. sehr große Datenvolumina, Performanceprobleme
- Lokale Benutzeraccounts auf vielen Systemen (UNIX/Linux)
  
- Lösung: Storage Konsolidierung
- Laufendes Projekt

# Ziele der Storage Konsolidierung

- **Einheitliches Storage für UNIX/Linux und Windows**
- **Zentralisiertes Storage**
- **Hochverfügbarkeit & Notfallwiederherstellung**
- **Gute Unterstützung der Protokolle**
- **CIFS:**  
**Windows ACL, Access Based Enumeration, DFS Replication, AD Integration**
- **NFS:**  
**Version 2 und 3, Posix ACL, AD Integration mit RFC2307**
- **FTP, RSYNC, SCP**

# Entscheidungsprozess & Kriterien

## Wirtschaftliche und technische Argumente

### Wichtige Kriterien

- Investitionen und Betriebskosten für 3 Jahre (bis 2010)

- Verfügbarkeit und Zuverlässigkeit (Lösungen ohne eigenes Storage benutzen die vorhandene EVA 8xxx)

- „Qualität“ der CIFS und NFS Implementieren, Interoperabilität

- Einfache Nutzung und Verwaltung

- Effizientes Backup mit CommVault Galaxy, präferiert LAN-free

### Optionale Kriterien

- Asynchrone Protokolle unterstützt (RSYNC, FTP ...)

- Integrierte Replikation zu einem unabhängigen 2. Storage

- Zukünftige Storage Strategie stärken

# Ergebnis: Linux File Server

## ■ Positiv

- Sehr redundant
- NFS und CIFS ACL Integration
- Beste und weiteste Protokollunterstützung
- Backup (CommVault Galaxy) mit LAN-free Agent
- Integrierte Replikation der Daten auf 2. Storage
- Keine Implikation für Storage Strategie
- Linux Wissen kann genutzt werden

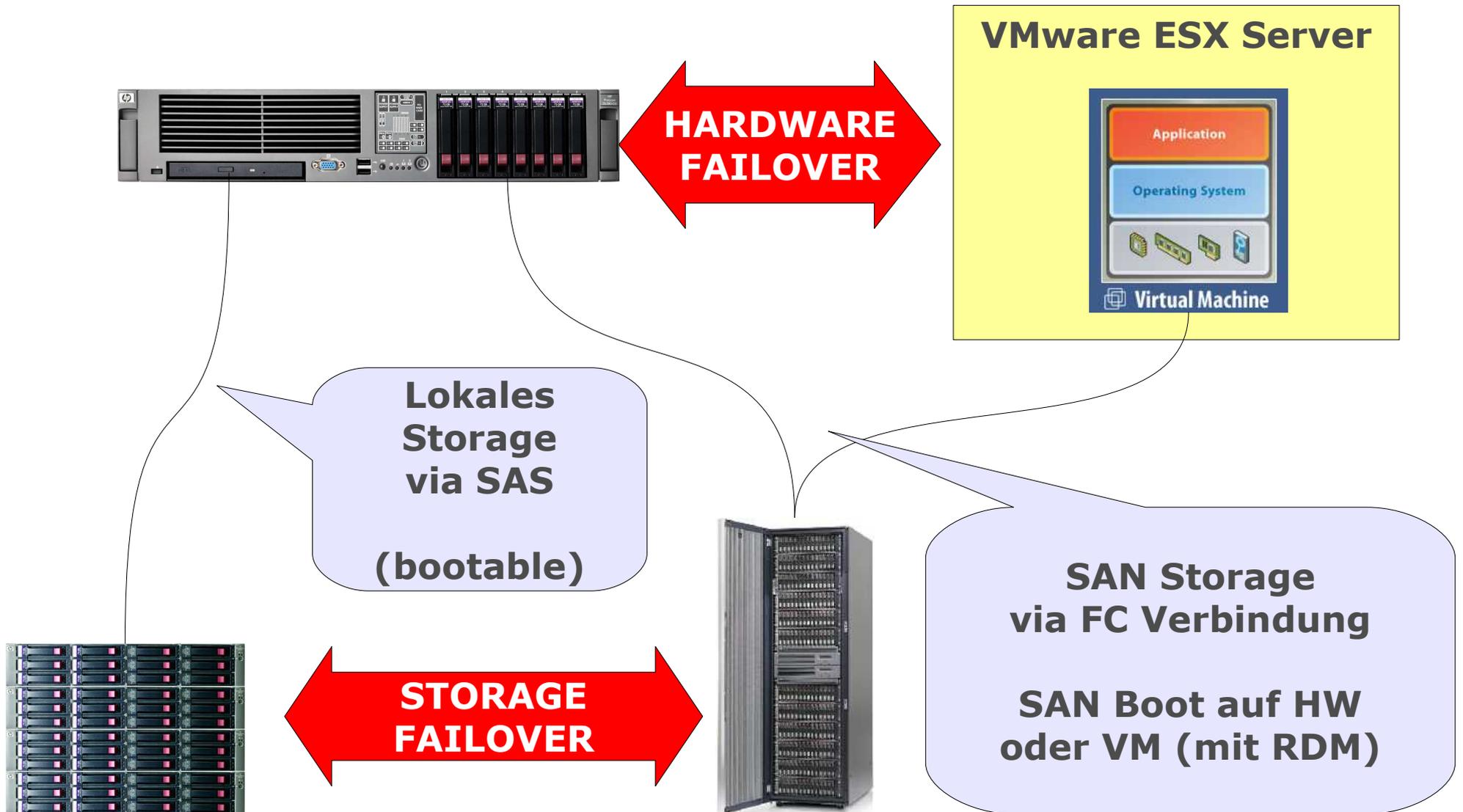
## ■ Negativ

- Unvollständige CIFS-Unterstützung (Samba)
- Verwaltung mit Shell Kommandos

# Agenda

- Hintergrundinformationen
- The Simple High Available Linux File Server
- Vorteile: Virtuelle Cold-Standby Server

# The Simple High Available Linux File Server



# Komponenten

## Hardware (alles von HP)

- DL 380 G5

- EVA 8000 (30TB)

- 4x MSA60 (36TB)

- mehrere DL Server als ESX Server (für das Rechenzentrum)

## Software

- SuSE Enterprise Linux 10 SP1

- Samba, NFS

- LVM, LVM Snapshots (smbsnap)

- rsync

- Virtuelle Maschine auf VMware ESX 3 (normalerweise aus)

# Aufgaben – SAN Boot

- **SAN Boot mit Multipathing (DM-MPIO)**
  - **Möglich mit SLES10SP1**
  - **dm-multipath bereits Teil der initrd**
  - **Installation in der VM**
  - **Siehe auch Artikel in i'X 04/2008 p. 142**
  - **Dual Boot Hardware und Virtuelle Maschine (Treiber ...)**
  - **Den versehentlichen Boot der VM verhindern (per ISO Image)**

# Aufgaben – Lokales Storage

## ■ Lokales Storage verwalten

- Automatisches Cloning des Produktionssystems auf das lokale Storage
- Modifikationen des Boot Prozesses vom lokalen Storage (RAID-1)
- Lokales Storage an Stelle des SAN Storage (ro) mounten
- Nächtliches rsync aller Daten aus dem SAN zum lokalen Storage
- Vorbereitung des lokalen Systems, als Produktivsystem genutzt zu werden (mit manueller Rückkopie)

# Vorteile

- **2 Ebenen der Redundanz**
  - **Hardware und VM laufen mit dem selben System und denselben Daten – Failover ohne Datenverlust**
  - **SAN Storage auf lokales Storage repliziert**
- **Wiederherstellung ~ 5 Min für HW oder Storage Probleme**
- **Sofortiges Disaster Recovery – auch mit vielen TB Daten**
- **Sehr einfache Lösung – keine komplexe Clusterkonfiguration**
- **Failover: HW neu booten (Storage) oder VM boot (Hardware)**
- **Administrator behält alles im Griff**
- **Sehr bezahlbare Lösung – keine Extrakosten für Hochverfügbarkeit**

# System & Samba Setup

## ┃ Lokales Storage

- ┃ GPT (>2TB)

- ┃ System auf RAID-1 (MD)

- ┃ LILO (GPT, MD)

## ┃ Alles andere via LVM

## ┃ rsync script mit Sicherheitsprüfungen

## ┃ AD Integration via RFC2307

## ┃ Volume Shadow Copy

## ┃ Map BUILTIN Accounts

```
passdb backend = tdbsam
smb ports = 445
disable netbios = Yes
name resolve order = wins
inherit acls = Yes
hide unreadable = Yes
idmap backend = ad
idmap uid = 100-20000000
idmap gid = 100-20000000
winbind enum users = Yes
winbind enum groups = Yes
winbind use default domain = Yes
winbind nss info = rfc2307
use sendfile = yes
```

# Performance Tuning

## ■ Benchmarks (1GBit):

■ 125 MB/s (NFS)

■ 100 MB/s (CIFS)  
(>2 streams)

```
net.core.rmem_max = 16777216
net.core.wmem_max = 16777216
net.ipv4.tcp_rmem = 4096 87380 16777216
net.ipv4.tcp_wmem = 4096 65536 16777216
net.ipv4.tcp_no_metrics_save = 1
net.ipv4.tcp_moderate_rcvbuf = 1
net.core.netdev_max_backlog = 2500
```

■ Nutzen XFS (auf SLES gut unterstützt)

■ sysctl.conf (auch auf dem client)

■ USE\_KERNEL\_NFSD\_NUMBER="16"

■ Bonding für redundante Netzwerkanbindung

■ Jumbo Frames hatten keinen meßbaren Effekt auf den Durchsatz, CPU Last um ca. 50% reduziert

# Ausblick

- **Automatisiertes Failover**
  - Heartbeat in der `initrd` vor dem Mounten
  - Storage und Netzwerk überwachen – schwierige Entscheidung
- **MD oder LVM Mirror zwischen SAN und lokalem Storage (ist aber kein Disaster Recovery !) als Alternative zu rsync**
  - Onlinespiegelung
  - Für Datenbanken usw.
- **Multipathing SAN Boot mit RHEL/CentOS, Ubuntu ...**

# Agenda

- Hintergrundinformationen
- The Simple High Available Linux File Server
- Virtuelle Cold-Standby Server

# Cold Standby

- Günstigste und einfachste Form der Hochverfügbarkeit
- Klassischerweise problematisch wegen
  - doppelte Hardwarekosten
  - Datenreplikation vom aktiven zum passiven (=stromlosen) System
- Daher keine Lösung für alle Server im Rechenzentrum

# Virtuelle Cold-Standby Server

- Ist ein Betriebskonzept, das die Vorteile der Virtualisierung auch für klassische Hardwareserver nutzt.
- Cold-Standby als günstiges Ausfallkonzept für alle Systeme
- Nebeneffekt:
  - Nachweis der Virtualisierbarkeit eines Systems
  - Bei Ausfall der Hardware – Produktivnutzung als VM
- Das ist primär eine Idee und eine Denkweise
- Schickt mir Emails mit Details Eurer Implementierung

# Fragen & Antworten

Mehr Open Source Software ([schapiro.org/schlomo/projects](http://schapiro.org/schlomo/projects))



Halle  
7.2b  
Stand  
102

- Relax & Recover (Linux Disaster Recovery)
- RSYNC BACKUP MADE EASY  
(Backup Software mit Hardlinks)
- OpenVPN Gateway Builder  
(Linux Router selber bauen mit zentraler Verwaltung)
- easyVCB (VMware VI3 Backup, w.i.p.)



**Schlomo Schapiro**  
Principal Consultant  
Leitung Virtualisierung und Open Source

[sschapiro@probusiness.de](mailto:sschapiro@probusiness.de)  
+49 160 97846168

**probusiness Berlin AG**  
Potsdamer Platz 11  
D-10785 Berlin

[berlin@probusiness.de](mailto:berlin@probusiness.de)  
+49 30 259378 0