

Einführungsaufwand von Filesystemen für virtualisierte parallele Datenbanken

best Systeme GmbH, Unterföhring

Wolfgang Stief

stief@best.de

Dipl.-Ing. (FH)
Systemingenieur Unix

2004-07-08
GIMS Zugspitze

Agenda

Verteilte Dateisysteme im SAN

Virtualisierung von Plattenkapazität

Anwendung im richtigen Leben?

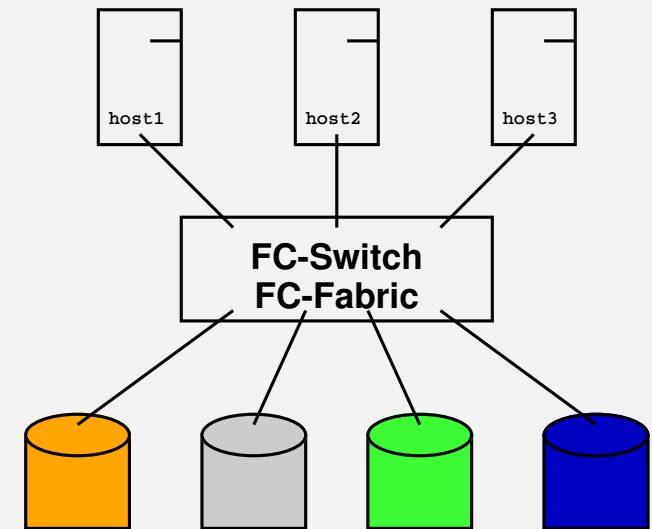
Einführungsaufwand?

Motivation

- große bis sehr große Datenbanken
- mehr Daten in kürzerer Zeit
- Gridcomputing, *horizontal scaling*
- möglichst geringer Administrationsaufwand bei maximaler Leistung
- brachliegende Disk-Kapazitäten

Von SCSI zu Fiber Channel

- SCSI hat Bus-Topologie, also potentiell Zugriffskonflikte möglich.
- SCSI ist nicht vorgesehen für mehrere Hosts an einem Bus.
- SCSI ist nicht vorgesehen für große Entfernungen ($>25\text{m}$).
- SAN erlaubt Vernetzung von Hosts mit Storage.
- Zeitgleicher Zugriff von Hosts auf gleiches Storage wird prinzipiell möglich.
- SAN hat Switch-Topologie, dadurch strukturierte Verkabelung und kurze Latenzzeiten möglich.



Einschränkung von SANs?

Problem:

Bisherige Filesysteme sind *exklusiv* (NTFS, ufs, ext2, ext3, ReiserFS, VxFS, JFS etc.)

⇒ kein konkurrierender Read/Write Zugriff auf gleiche Datenbasis möglich.

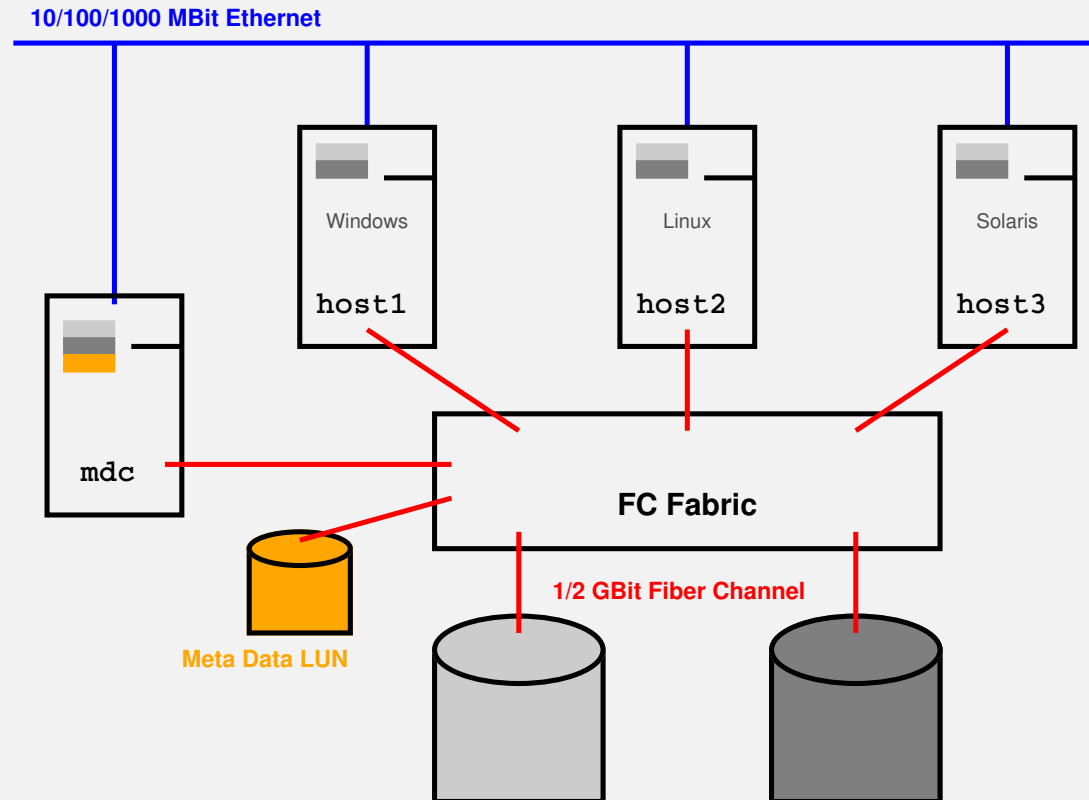
⇒ Shared Data muss durch (langsames) Netz.

Lösung:

Einsatz von verteilten Filesystemen mit *Multiple Writer* Funktion.

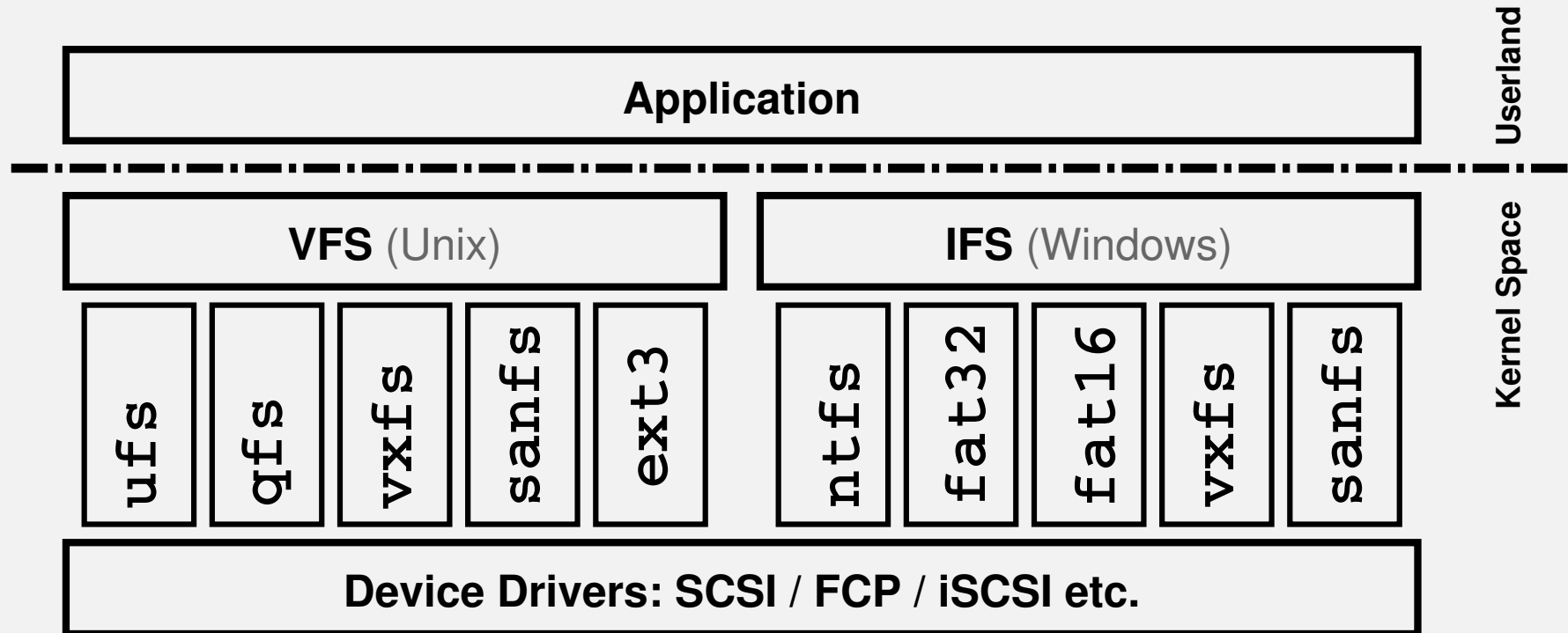
Beispiele: Sun QFS (homogen, Solaris), IBM GPFS (Linux + AIX), RedHat GFS (Linux, GPL), Oracle OCFS (Linux, GPL), IBM SAN-FS (heterogen), ADIC StorNext Filesystem (heterogen)

Wie funktioniert ein verteiltes Dateisystem?



- Request über TCP/IP an Metadata Controller (MDC).
- Read/Write über FCP + SAN direkt auf Storage.

Verteilte Filesysteme – Implementierung



Verteilte Filesysteme – Pro und Contra

Pro

- + breitbandiger Zugriff von **allen** angeschlossenen Hosts auf das Storage
- + einheitliche Sicht auf Daten
- + Trennung von Daten und Metadaten \Rightarrow schnelles Filesystem

Contra

- erhöhter Aufwand für Infrastruktur
- kein *natives* Filesystem \Rightarrow womöglich Probleme mit Betriebssystemwerkzeugen (dump, fsck u. ä.)

Agenda

Verteilte Dateisysteme im SAN

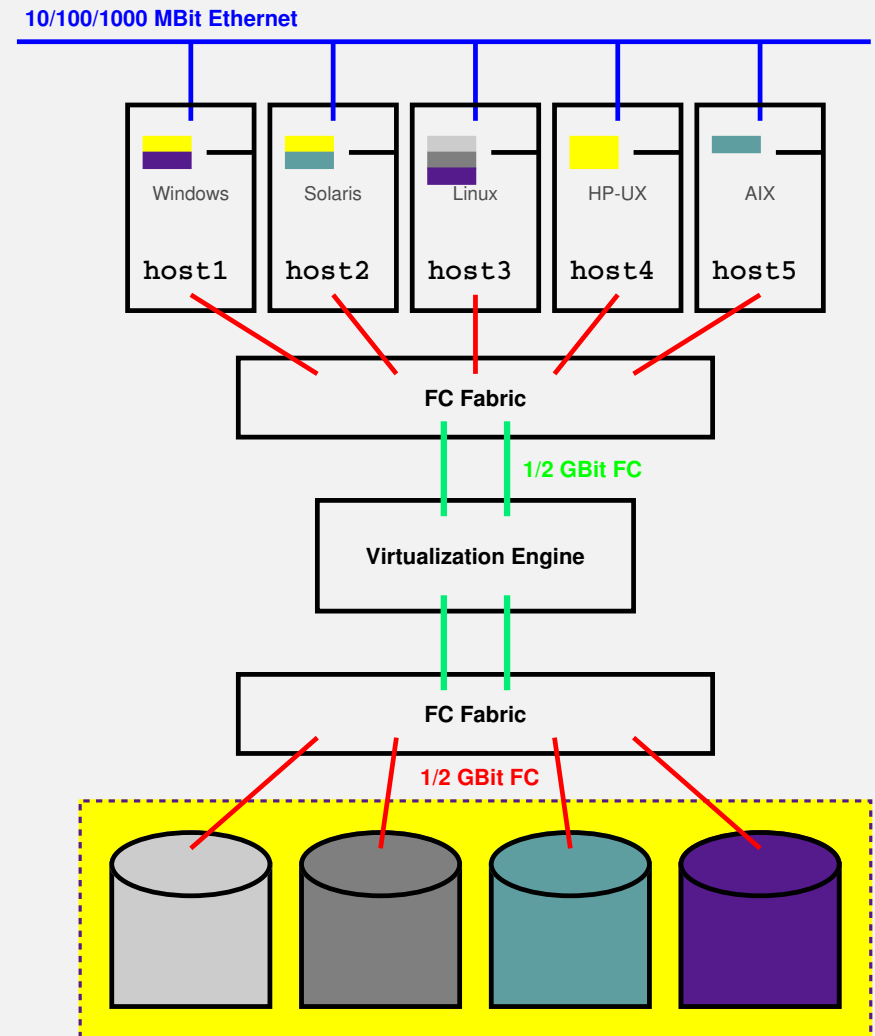
Virtualisierung von Plattenkapazität

Anwendung im richtigen Leben?

Einführungsaufwand?

Storage Virtualisierung – Prinzip

- Plattensubsysteme werden vollständig von Hosts entkoppelt.
- Kein direkter Zugriff von Host auf physikalische Disk.
- *Virtualization Engine* kann in Software oder Hardware realisiert sein.



Storage Virtualisierung – Features

- Storage Pooling („*Da rotieren doch auch noch ein paar Gigabytes unbenutzt rum!*“)
- Vergabe von Policies („*Welche Daten kommen in welchen Topf?*“)
- Volume Management über Rechner- und Storage-Grenzen hinweg
- zentrale – möglicherweise einheitliche – Verwaltung heterogener Storage-Systeme (Sun, HDS, IBM, HP, EMC²...)
- LAN-less Backup u. Restore
- Block Level Snapshots und Remote Mirrors, auch wenn Disk-Subsysteme sowas nicht originär bieten

Storage Virtualisierung – Pro und Contra

Pro

- + einfachere Administration
- + verbesserte Nutzung der Ressourcen
- + höhere Durchsätze durch RAID-Striping
- + LAN-less od. sogar Server-less Backup

Contra

- erhöhter Administrationsaufwand
- wenig Einfluß auf Datenlokation
- *langsamstes Kettenglied* kann Gesamtperformance ausbremsen
- hohe Initialkosten

Agenda

Verteilte Dateisysteme im SAN

Virtualisierung von Plattenkapazität

Anwendung im richtigen Leben?

Einführungsaufwand?

Storage Virtualisierung – Wer macht sowas?

Oracle: ASM – Automatic Storage Management

- Verwaltet Disks in verteilter Umgebung
- Vergrößern der Speicherkapazität durch einfaches Zufügen von physikalischen Disks
- automatisches Loadbalancing – Daten werden nach Zufügen von Disks im Hintergrund neu verteilt
- kümmert sich selbständig um Mirroring
- Bestandteil von Oracle 10g
- für Mainframer ist das natürlich ein **sehr** alter Hut :-)

Agenda

Verteilte Dateisysteme im SAN

Virtualisierung von Plattenkapazität

Anwendung im richtigen Leben?

Einführungsaufwand?

Einführungsaufwand?

- Umdenken erforderlich
- wesentlicher Eingriff in bestehende Infrastruktur
- u. u. relativ hohe Initialkosten
- Zeitaufwand zur Datenmigration
- ggf. mehrere Zwischenschritte notwendig, wenn alte Plattensysteme zukünftig in Virtualisierung einbezogen werden sollen

Danke für die Aufmerksamkeit.

Fragen?