

Lower Layer Technologien in Grids

best Systeme GmbH, Unterföhring

Wolfgang Stief

stief@best.de

Dipl.-Ing. (FH)

Senior Systemingenieur Unix

2005-05-23

Open Systems Day Fall '05

Compute Grid vs. Compute Grid

Compute Grid vs. Data Grid

Grid Interconnects

Distributed Filesystems

Data Access at Block Level

Shared Filesystems

Datenverteilung mal ganz anders

Compute Grid vs. Compute Grid

Nicht überall, wo *Grid* draufsteht, ist auch *Grid* drin:

- Oracle 10g ist **kein** Grid.
- Cluster im HPC sind **kein** Grid. Bestenfalls sind sie Bestandteil eines Grids.
- Charakteristisch: Nutzung geographisch mehr oder minder stark verteilter Ressourcen über standardisierte Protokolle.
- Häufig zitiertes Beispiel: Strom aus der Steckdose.

Agenda

Compute Grid vs. Compute Grid

Compute Grid vs. Data Grid

Grid Interconnects

Distributed Filesystems

Data Access at Block Level

Shared Filesystems

Datenverteilung mal ganz anders

Compute Grid vs. Data Grid

Grid bedeutet nichts anderes als Zugriff auf in einem Netz **verteilte** Ressourcen:

- **Compute Grid:** Verteilte Rechenleistung, u. U. auch heterogen, **kein** Single System Image.
- **Data Grid:** Verteilte Festplatten bzw. verteilte Daten.

Grid sagt noch nichts aus über Protokolle, Betriebssysteme, Anwendungen, zu lösende Probleme.

+ Bessere Ausnutzung von Ressourcen.

+ Verteilte Administration.

– Nicht für jedes Problem geeignet.

– Overhead für Kommunikation erforderlich.

– Verteilte Administration.

Compute Grid vs. Compute Grid

Compute Grid vs. Data Grid

Grid Interconnects

Distributed Filesystems

Data Access at Block Level

Shared Filesystems

Datenverteilung mal ganz anders

Grid Interconnects

- Myrinet, Infiniband, Scalable Coherent Interface (SCI) etc. sind Cluster Interconnects bei HPC-/HA-Clustern.
⇒ **Keine** gültigen Grid Interconnects.
- Auf Layer 2: Ethernet, Token Ring, FDDI, ATM etc.
- Auf Layer 3: IPv4, IPv6

⇒ **Standards!**

Grid Interconnects (cont'd)

Am gebräuchlichsten ist sicherlich Ethernet:

- Aktuell bis 10 GBit/s verfügbar.
- Praxis zeigt: 80% - 90% Nutzbandbreite bei GBE.
- 10 GBit/s derzeit noch sehr teuer:
 - ⇒ Verwendung im Netzwerk-Core und teilweise für WAN-Strecken.
 - ⇒ Noch kaum Serveranbindung, obwohl HW verfügbar.
- Trunking ist möglich und standardisiert (*Link Aggregation*, IEEE 802.3ad, seit 2000).
- Weitstrecken über Glasfaser möglich, Technik existiert.
- ATM aufgrund QoS und geringerer Latenzzeiten im WAN besser geeignet.

Compute Grid vs. Compute Grid

Compute Grid vs. Data Grid

Grid Interconnects

Distributed Filesystems

Data Access at Block Level

Shared Filesystems

Datenverteilung mal ganz anders

Distributed Filesystems

Network File System (NFS), Andrew File System (AFS), Distributed File System (DFS), Coda, InterMezzo etc.

- + vergleichsweise einfach administrierbar (common knowledge)
- + standardisiert und dokumentiert
- + Support in allen gebräuchlichen Betriebssystemen
- + Erweiterungen zu Authentifizierung verfügbar (Kerberos u. ä.)
- + Zugriffsschutz mittels Access Control Lists (ACL)
- viel Overhead \Rightarrow schlägt auf Performance \Rightarrow vergleichsweise langsam
- Namespace (Mountpoints) müssen ggf. innerhalb des Grids synchronisiert werden \Rightarrow erhöhter administrativer Aufwand.

Compute Grid vs. Compute Grid

Compute Grid vs. Data Grid

Grid Interconnects

Distributed Filesystems

Data Access at Block Level

Shared Filesystems

Datenverteilung mal ganz anders

Data Access at Block Level

- Ziel: performanter Zugriff auf Daten
 - Derzeit aktuelle Protokolle: SCSI, FCP, iSCSI, FCIP, iFCP.
- + auf Datentransport optimiert
- + schnell, weil wenig Overhead
- *shared access* ist nicht trivial

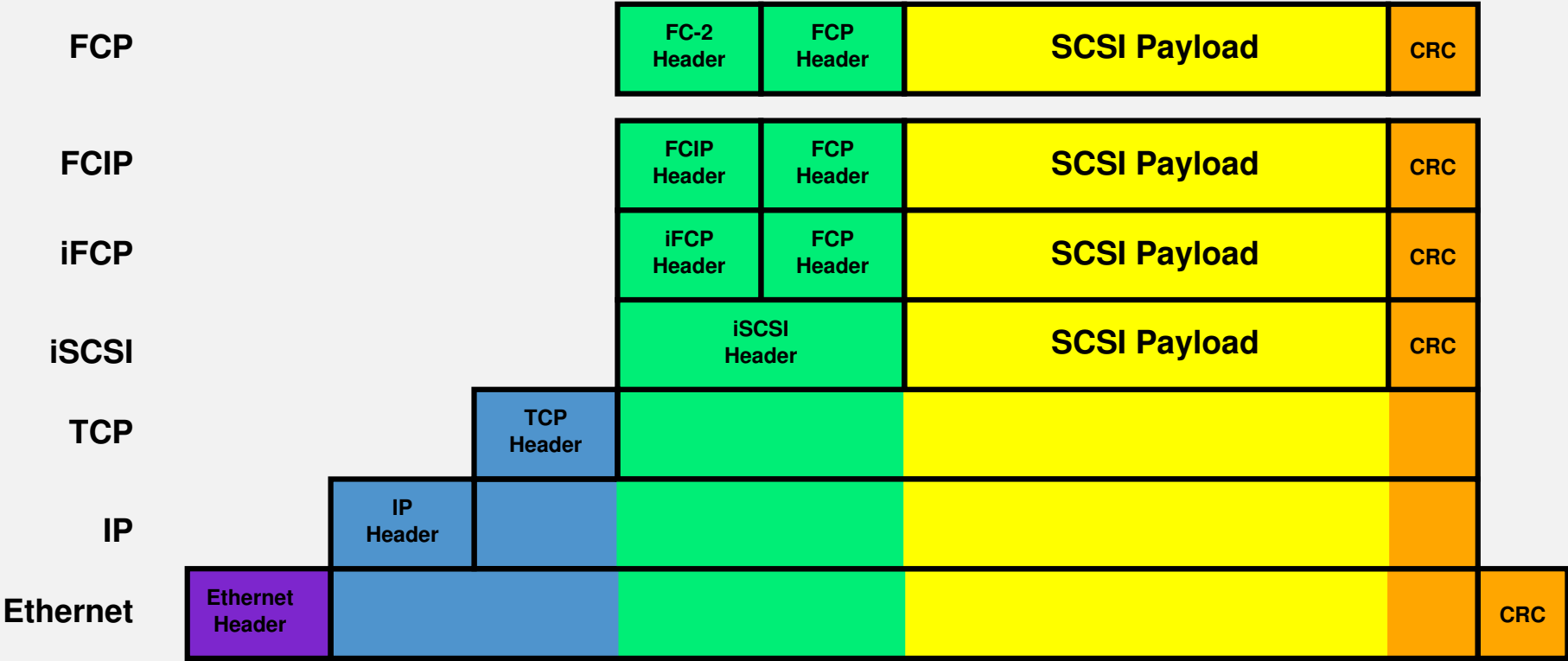
Data Access at Block Level – SCSI und FCP

- SCSI als Bus, limitierte Länge und limitierte Anzahl der Geräte.
 - FCP hat Features ähnlich wie Netzwerk (Switching, Zoning, WWN-Security).
- + Protokolle sind ausgereift und standardisiert.
- + Support in jedem handelsüblichen Betriebssystem on board.
- + Bandbreitenerhöhung durch *Load Balancing* möglich ⇒ muss von Hardware und Treiber unterstützt sein.
- Kein *concurrent write access* möglich.
- Nicht bzw. nur begrenzt „routbar“
⇒ nicht WAN-tauglich

Data Access at Block Level – iSCSI, FCIP und iFCP

- zu Grunde liegendes Protokoll hier wie oben: SCSI
 - verpackt in TCP/IP \Rightarrow routing-fähig
-
- + Kombinierbar mit IPsec/TLS, damit verschlüsselter Kanal.
 - + Zugriff auf Storage über WAN wird möglich.
 - + Einfache Bandbreitenerhöhung durch Link Aggregation (IEEE 802.3ad).
 - + Layer 2 austauschbar.
 - + Nur **ein** Layer 2 Medium innerhalb des Grids realisierbar.
-
- Hohe Latenz \Rightarrow IP kennt kein durchgängiges QoS.
 - Auch hier kein *concurrent write access*.
 - Anfällig gegen allerlei bekannte Security-Probleme aus der TCP/IP-Welt (DoS, Spoofing, Man in the Middle. . .)

Data Access at Block Level – Protokoll-Overhead



Compute Grid vs. Compute Grid

Compute Grid vs. Data Grid

Grid Interconnects

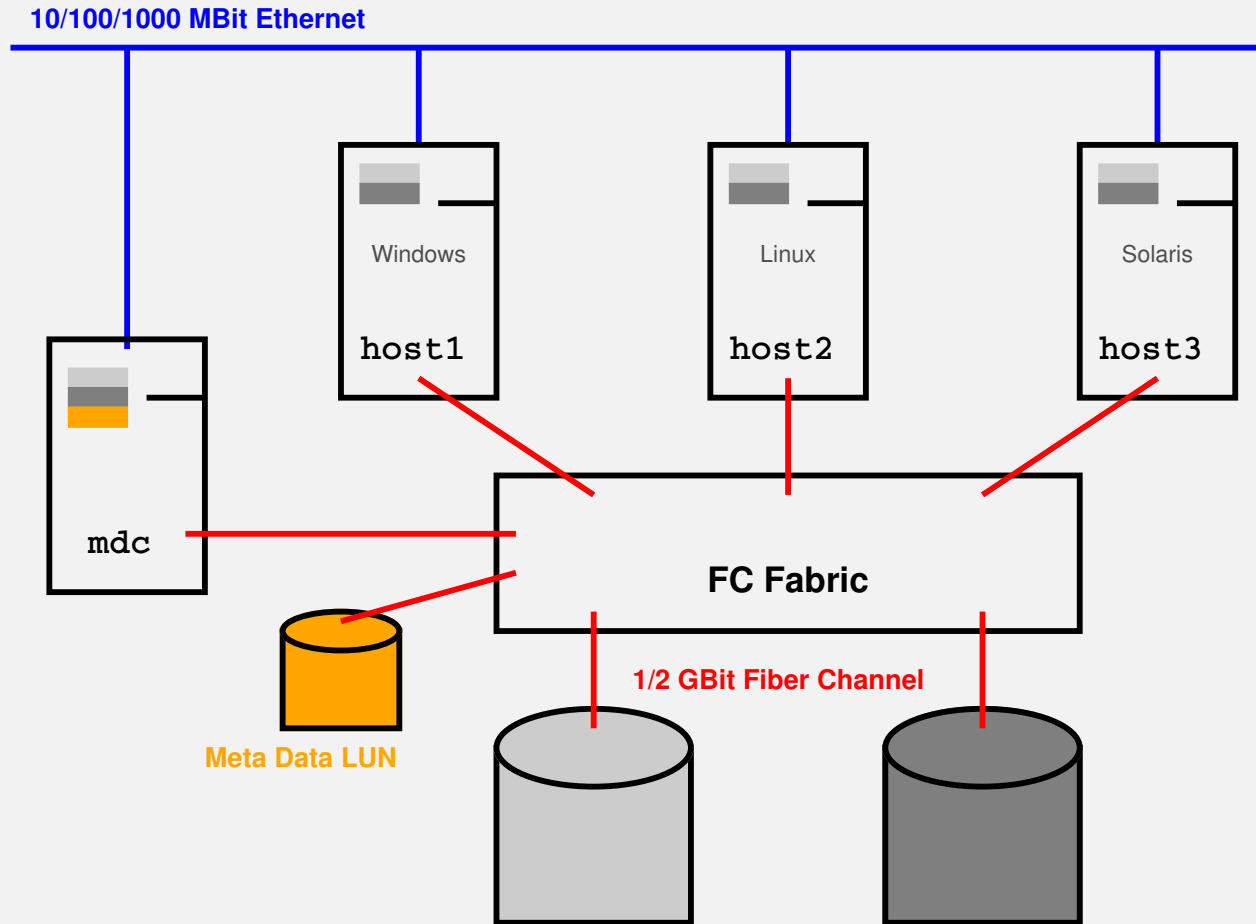
Distributed Filesystems

Data Access at Block Level

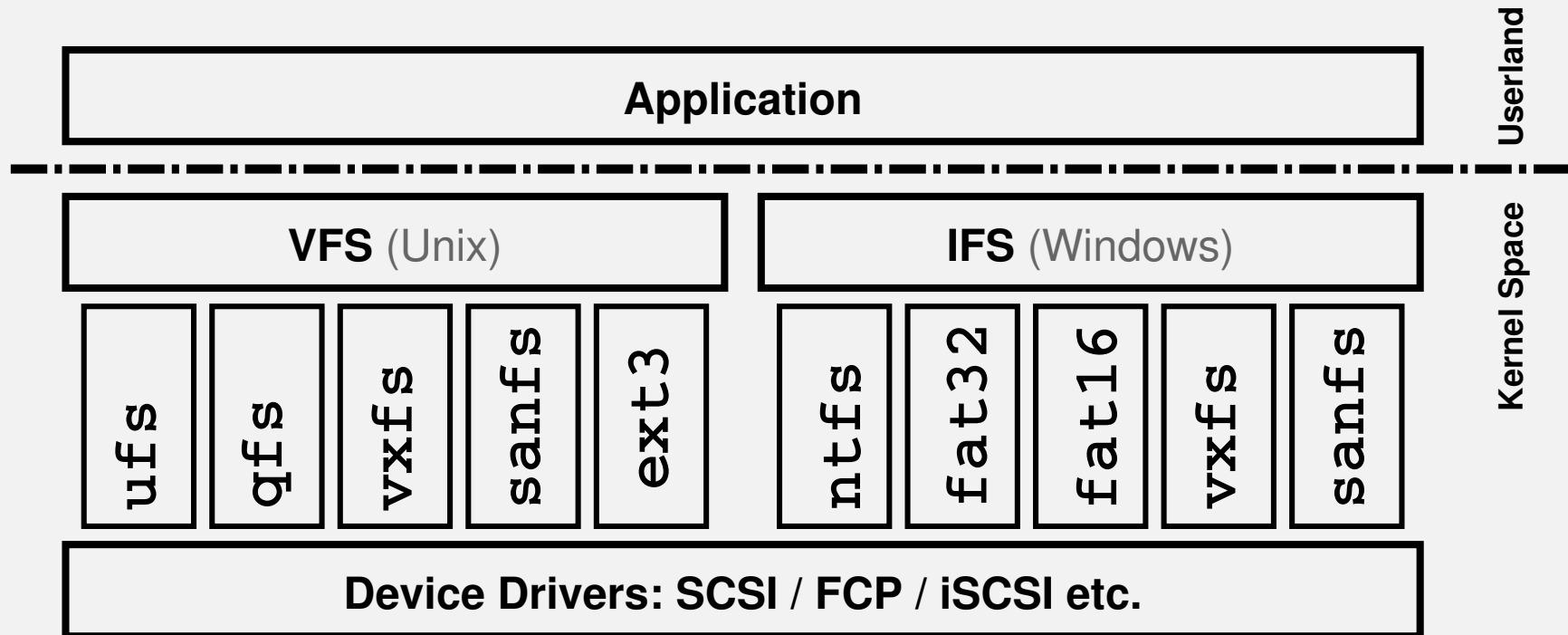
Shared Filesystems

Datenverteilung mal ganz anders

Shared Filesystems – Grundprinzip



Shared Filesystems – Implementierung



Shared Filesystems – Produkte

ADIC StorNext Filesystem, Sun QFS, Sun PlexFS, IBM Tivoli SANergy, IBM GPFS, Oracle Cluster Filesystem, RedHat GFS...

- + Multiple Writer Funktionalität
- + Snapshots, Striping, Mirroring etc. im Filesystem möglich.
- + Heterogenes Umfeld möglich.
- + *Anflanschen* von Zusatzfeatures
 - ⇒ Hierarchical Storage Management
- großer Einführungsaufwand
- komplexe Administration
- u. U. hohe Initialkosten (Lizenzen, Infrastruktur)

Compute Grid vs. Compute Grid

Compute Grid vs. Data Grid

Grid Interconnects

Distributed Filesystems

Data Access at Block Level

Shared Filesystems

Datenverteilung mal ganz anders

Datenverteilung mal ganz anders

ftp, scp, sftp, rsync

- + scriptingfähig
 - + einfaches Handling
 - + für quasi jedes Betriebssystem zu haben
 - + heterogenes Umfeld sehr einfach möglich
- keine Echtzeit

Quellen und weiterführende Infos

- Ulf Troppens, Rainer Erkens
Speichernetze
dpunkt.verlag, Heidelberg, 1. Auflage, 2003
- **Grid-Computing**
Artikel in der deutschen Wikipedia, Version v. 2005-11-09, 17:02h
<http://de.wikipedia.org/wiki/Grid>
- **Grid computing**
Artikel in der englischen Wikipedia, Version v. 2005-11-18, 19:22h
http://en.wikipedia.org/wiki/Grid_computing
- **The DataGrid Project**
<http://eu-datagrid.web.cern.ch/eu-datagrid/>
- Viktors Berstis
Fundamentals of Grid Computing
IBM Redbooks Paper, 2002
- Srikumar Venugopal, Rajkumar Buyya, Ramamohanarao Kotagiri
A Taxonomy of Global Data Grids
University of Melbourne, Department of Computer Science and Software Engineering, Grid Computing and Distributed Systems Laboratory, April 2005

Danke für Ihre Aufmerksamkeit.

Fragen?

Wolfgang Stief – stief@best.de
November 2005