



ADVANCED TECHNOLOGY FOR INDIVIDUAL SUCCESS



Herzlich willkommen zum Workshop: Konfiguration eines GFS-Clusters



Dipl.-Ing. (FH)

Reiner Rottmann
rottmann@atix.de
7 March 2007

Agenda

- **Vorstellung** der ATIX GmbH
 - Was kann ATIX für Sie tun?
 - Geschäftsfelder
- **Theorie** GFS Cluster
- **Praktische Umsetzung** in VmWare



ADVANCED TECHNOLOGY FOR INDIVIDUAL SUCCESS

„Wir helfen Ihnen, Ihre Daten sicher und
effizient zu speichern“

Was kann ATIX für Sie tun?

- **Beratung** im Linux, **Storage** und **Cluster** Umfeld
- **Konzepterstellung** und Machbarkeits-**Studien**
- **Unterstützung** bei anspruchsvollen **IT Projekten**
- **Support** von Red Hat **Storage** und **Cluster** Lösungen
- **Know-How Transfer** und **Workshops**
- Entwicklung von **Diskless Shared Root Cluster** Lösungen



SPEICHERNETZWERK LÖSUNGEN

Ein neues Zuhause
für kritische Windows
und UNIX Daten

SAN

NAS

Shared
Root



SAP UND ERP LÖSUNGEN

Eine neue Heimat
für hochverfügbare
Tru64 SAP und ERP
Systeme



CONSULTING

Para-Virtualization Application-Virtualization

Singlesystem Management



High Availability

Oracle

Hochverfügbar

NAS

Perfekte Integration

SAP

WEB

Optimal skalierbar

Simple Management

Hoher Investitionsschutz

Linux

Kosteneffizienz



DATENBANK LÖSUNGEN

Eine zuverlässige
Umgebung für
Linux Datenbanken

ORACLE
DATABASE

PostgreSQL



WEB SERVICES LÖSUNGEN

Eine Enterprise
IT-Plattform für
anspruchsvolle
Internet-Dienst-
leistungen



SUPPORT UND MONITORING

Cluster Technologien

„There are three ways to do anything faster

In a computer

1. Work harder

+ Processor speed

2. Work smarter

+ Algorithms

3. Get help“

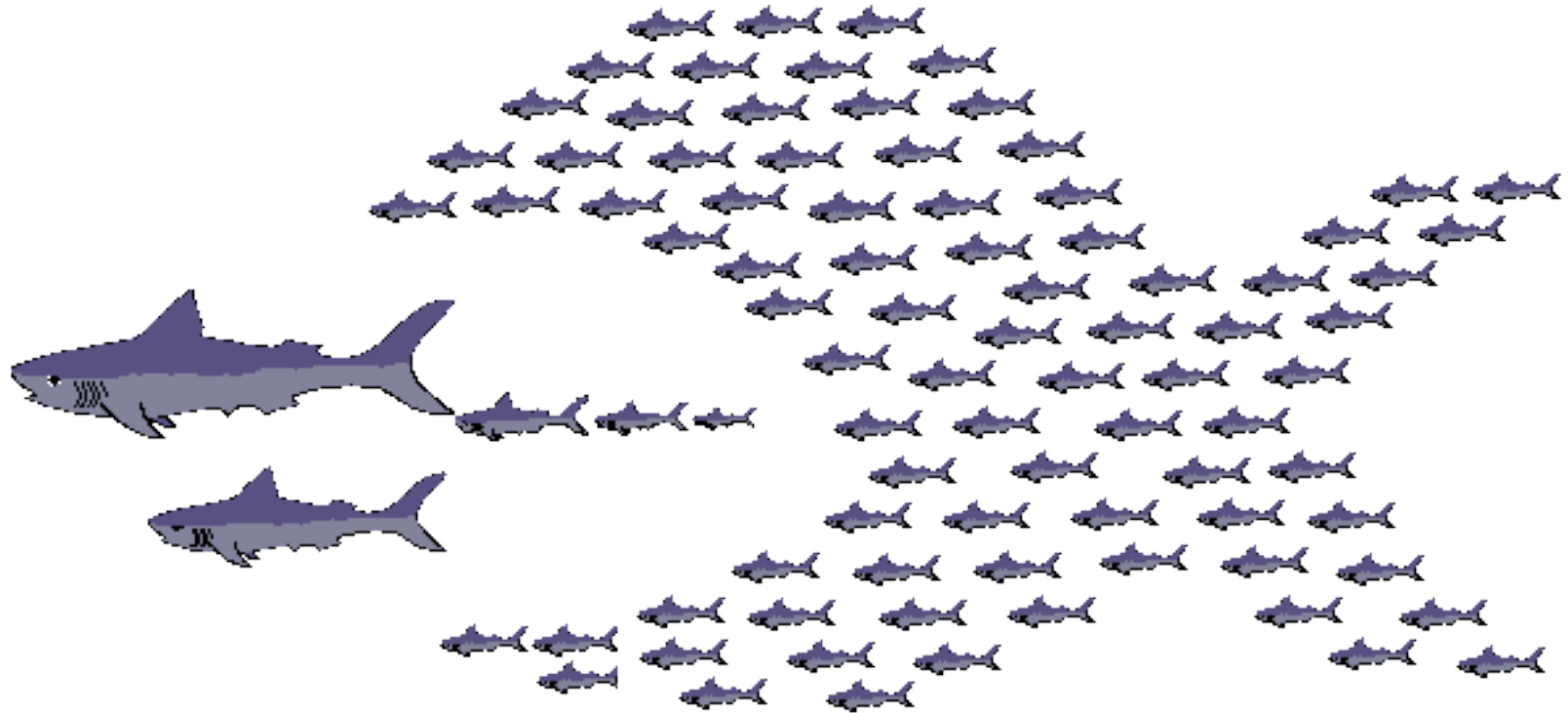
+ Parallel processing

* In search of clusters, G.F. Pfister, 2nd Edition

Was ist ein Cluster?

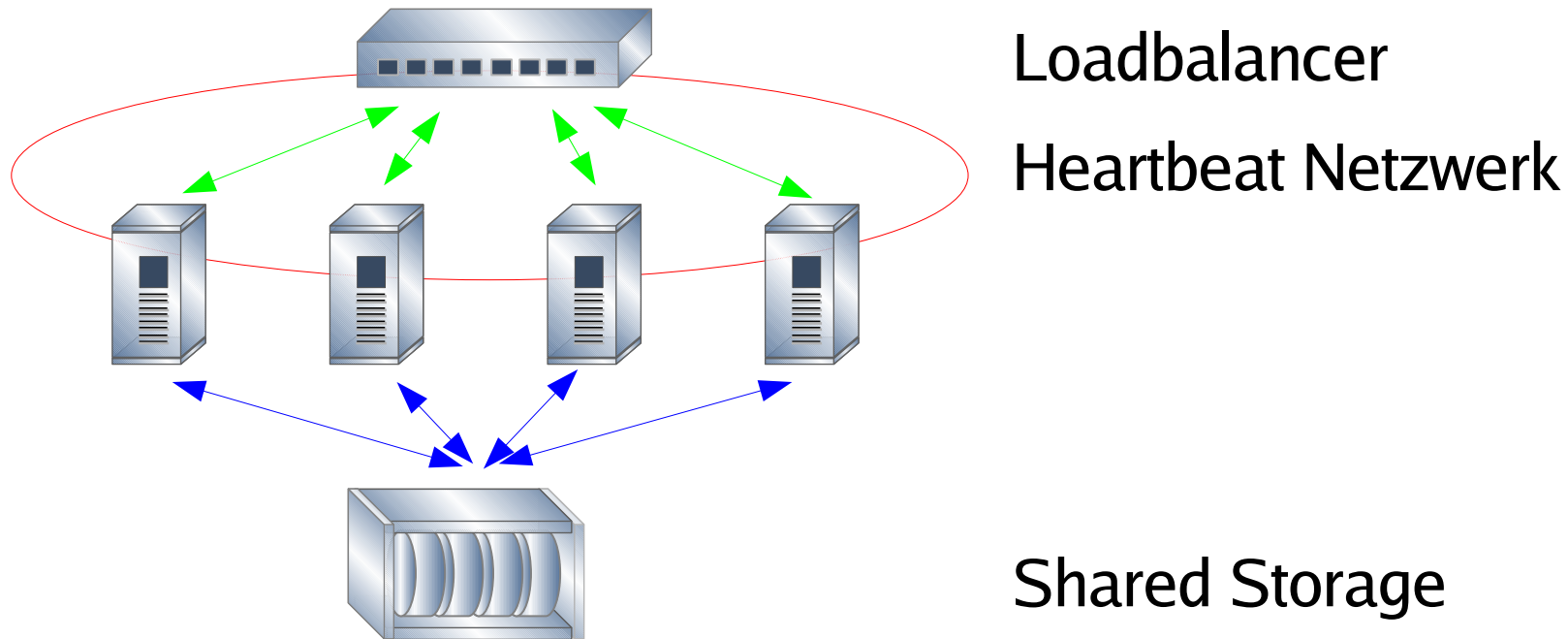
- Paralleles oder verteiltes Rechensystem
- Zusammenschluss von eigenständigen Computern
- Nach aussen eine Ressource
- Schnelle Netzwerkanbindung
- Lockere Verbindung der Einheiten

Rechencluster



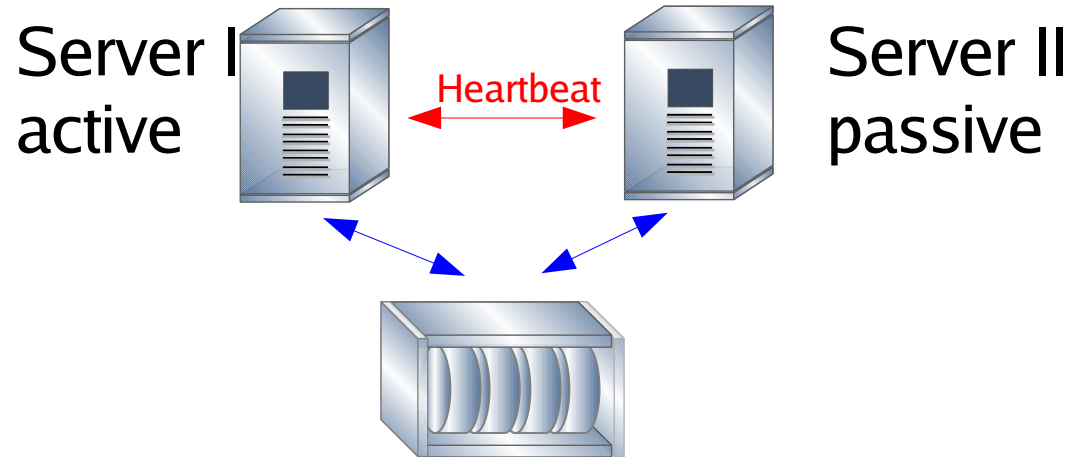
- Idee: Viele kleine Einheiten erledigen die Aufgabe einer großen
- Einsatz: Parallel Processing, „Number Crunching“ etc.

Network Loadbalancing Cluster



- Eine virtuelle IP Adresse auf dem Director/LB
- Verschiedene Arten des Loadbalancing
 - NAT/Tunneling/Direct-Routing
 - Scheduling Algorithmus (RR, LWR, LCR, ..)
- Dynamische Konfiguration (Heartbeating)

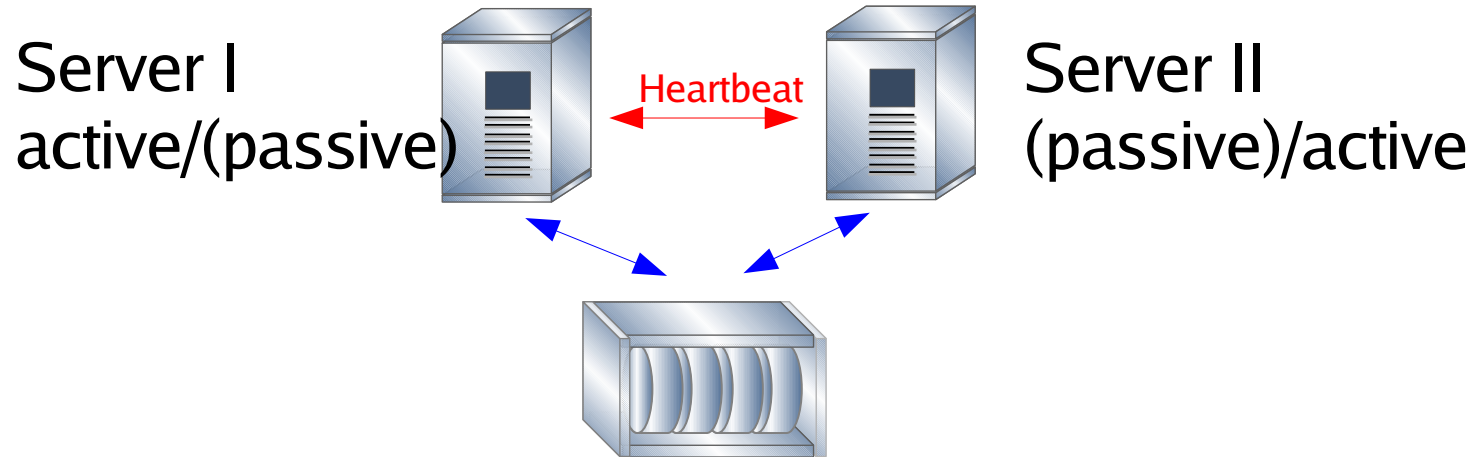
HA-Cluster Active/Passive



Konzept:

- Der passive Knoten wartet auf den Ausfall des aktiven
- Ein oder mehrere Dienste laufen auf dem aktiven Knoten
- Eine Software sorgt für die Überwachung und den Failover
- Voraussetzung sind „gleiche“ Daten
 - Shared Storage (SAN/Parallel-SCSI)
 - Datenreplikation
 - Separate Filesysteme

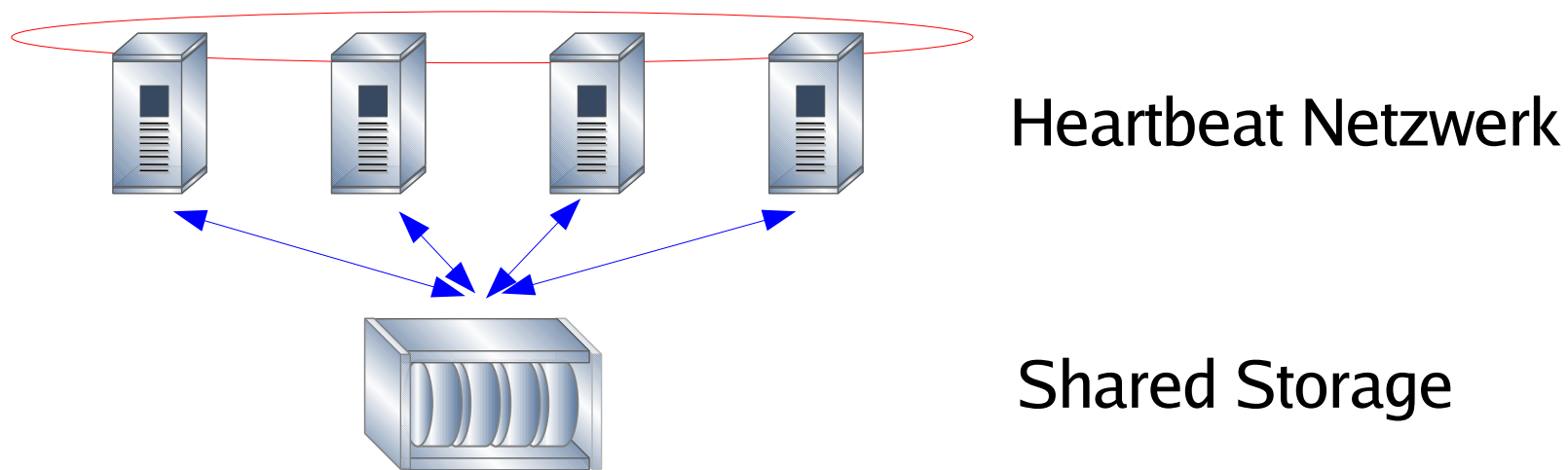
HA-Cluster Active/Active



Konzept:

- Auf jedem Knoten laufen unterschiedliche Dienste
- Jeder Knoten aktiv und passiv zugleich
- Im Fehlerfall gibt es aber Performanceeinbussen
- Ansonsten gilt das gleiche wie für Active/Passive Cluster

HA-Cluster N+1

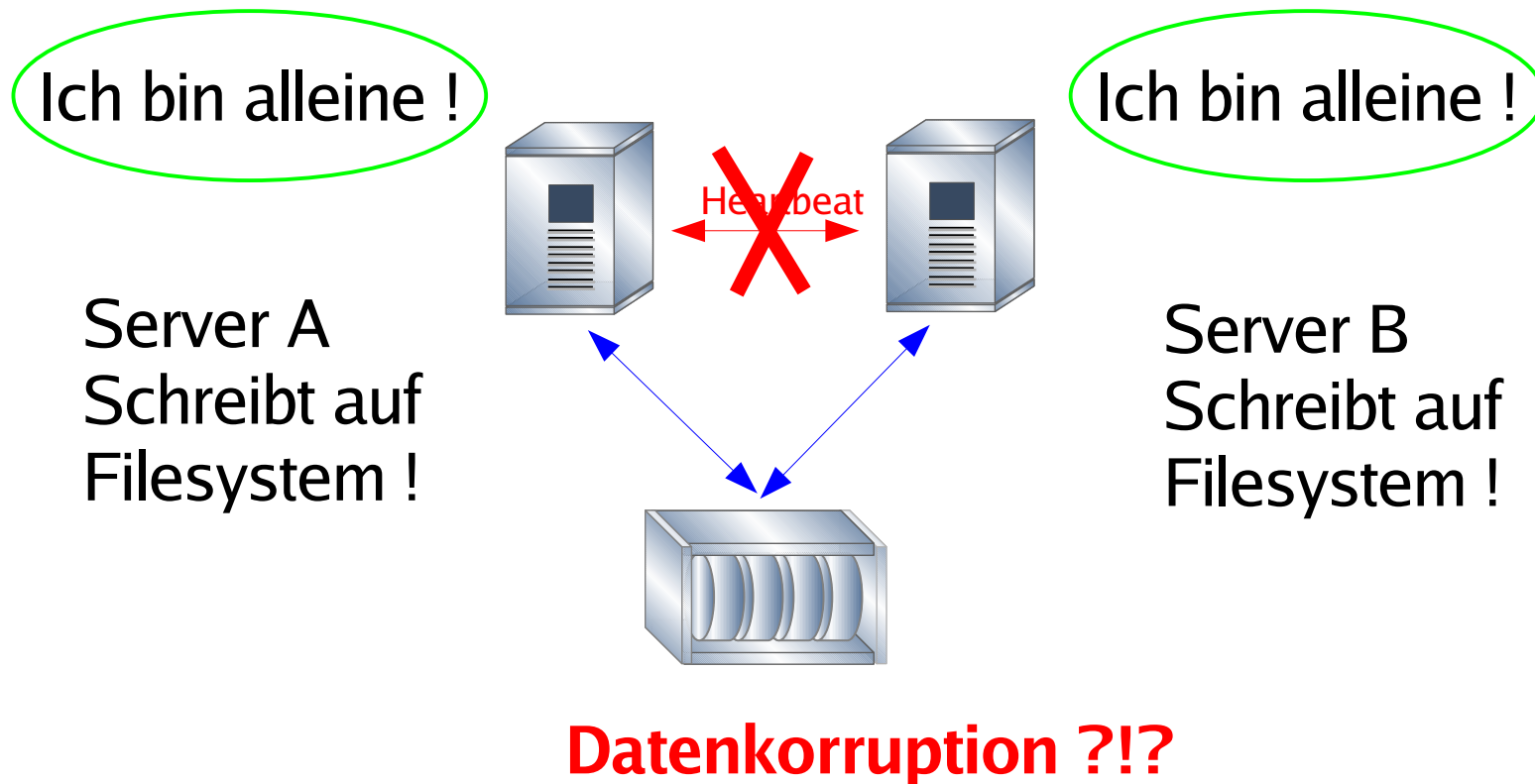


Konzept:

- Es werden um einen mehr Server verwendet als notwendig (N+1)
- Der Cluster kann den Ausfall eines Knoten ohne Performanceeinbussen hinnehmen
- N+2, N+3 für grössere Ns auch möglich
- Typischerweise unterstützt HA-Software maximal ca. 16 Server
- Dieses Konzept kann auch für NLB Cluster benutzt werden

HA-Cluster: Split Brain Problem

- Frage: „Was passiert, wenn sich der Cluster teilt?“

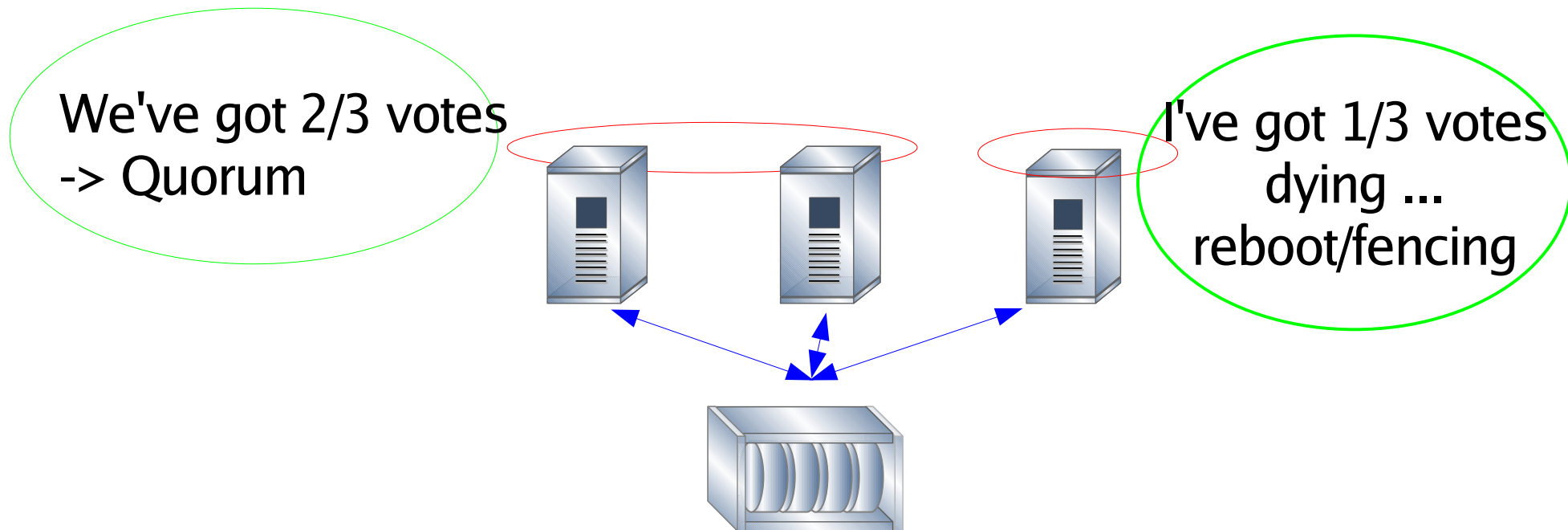


Lösung für Split Brain:

- Lösung1: SCSI reserve/release
 - SCSI Device kann nur einmal gemounted sein.
 - z.B. Windows Cluster
- Lösung2: Network Tie Breaker
- Lösung3: Quorum Requirement
 - Activer Teil muss mehr als die Hälfte der „votes“ bekommen.
 - Voraussetzung: Ungerade Anzahl an Votes
 - Unterschiedliche Anzahl pro Server
 - Quorum Device hat N-1 votes

HA-Cluster Quorum Requirement

- Bsp: 3 Knoten, Quorum Requirement

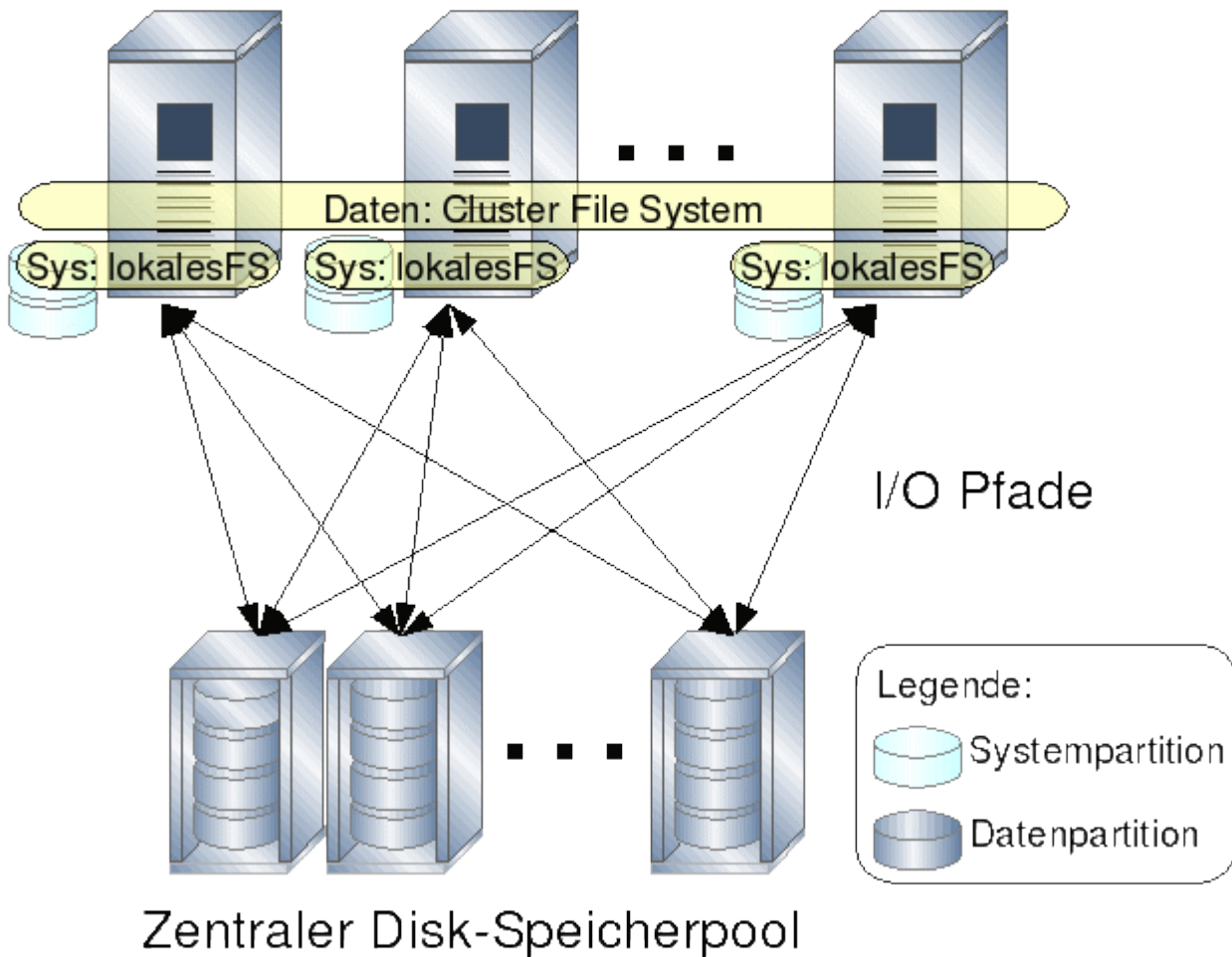


HA-Cluster I/O Fencing

- Split Brain: Alle Server, die nicht im Cluster sind, werden „gefenced“
- Fehlverhalten: Jeder Knoten, der ein Fehlverhalten zeigt, wird „gefenced“
- Sicherstellung der Datenkonsistenz
- Fencing über:
 - Stonith (Shoot the other node in the head)
 - z.B. Network power switch, ILO
 - Zoning („versperren“ des I/O Pfades)
 - Failfast (Server „killed“ sich selbst)

Shared Storage Cluster

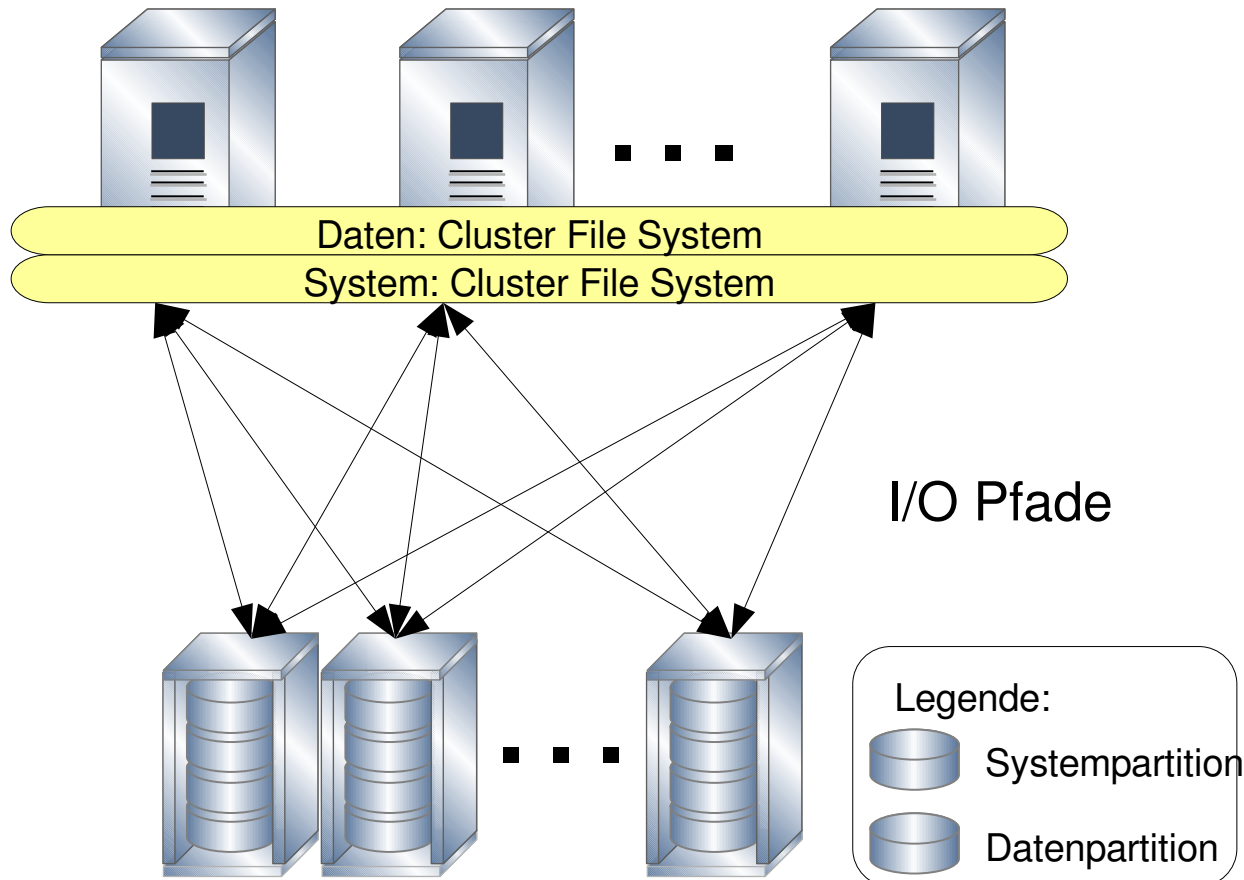
Clusterknoten mit lokalen Disks



- Cluster Filesystem
- Server Cluster
 - Active/Active
 - OS auf lokalen Platten
- Storage Cluster
 - Storage Pooling
 - Volume Manager
- Hochleistungs
Speichernetzwerk
- Relativ hoher
Managementaufwand
- Relativ hoher
Monitoringaufwand

Diskless Shared Root Cluster

Clusterknoten ohne lokale Disks



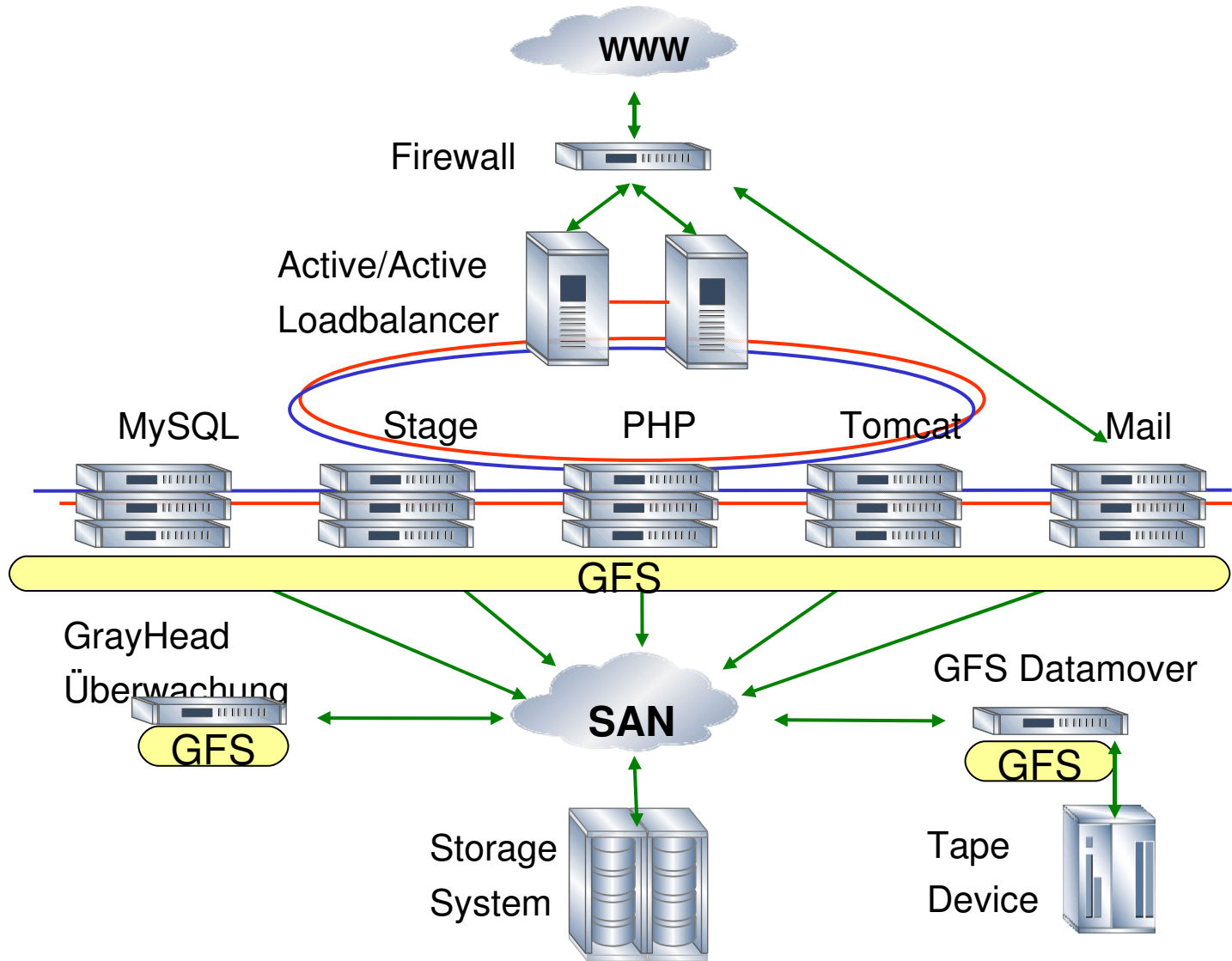
Zentraler Disk-Speicherpool

- Shared Root Partition (/)
- Clusterfilesystem
 - Daten + System
- SSI auf FS Ebene
 - Einfaches Management
- Skalierbarkeit
- Performance
- Storage Cluster
- **Monitoring!!**

Beispiel: Messe München International (MMI)

- Die Messe München GmbH betreibt 73 Internetauftritte
- Ø Anzahl der Zugriffe pro Monat > 3 Mio Visits
- Die Zugriffszahlen steigen seit dem Jahr 2003 kontinuierlich um den Faktor 4-5
- Hohe Verfügbarkeit komplexer Dienste muss gewährleistet sein
- Aussteller- und Besucherdatenbanken, Shopsysteme, Organizer
- Besucherregistrierung
- Umsetzung in einem sehr eng gefassten Zeitfenster innerhalb der „messefreien“ Zeiten
- Skalierbarkeit ohne großen Aufwand

MMI: Architektur



MMI: Vorteile der Architektur

- Flexibler Aufbau
- Bessere Skalierbarkeit der Systeme
 - Zusätzliche Server können bei Bedarf durch „boot from SAN“ innerhalb kürzester Zeit in die Serverlandschaft integriert werden
- Hohe Ausfallsicherheit
 - Mehr virtuelle Server
 - Bessere Netzwerkinfrastruktur
 - Integrierte Hochverfügbarkeits-Funktionalität
- Optimale Ausnutzung bereits bestehender Ressourcen
 - Hardware kann flexibel in den Cluster integriert werden
- Einfache Administration
 - Single Point of Control durch Single System Image

MMI: Business Values

- Initialkosten für die Softwarebeschaffung konnten minimiert werden
- Reduzierung der Betriebskosten auf **35 %**
- Geringe Folgekosten bei Versionsupdates
- Optimaler Investitionsschutz
- Bedarfsgerechte Investitionen - “pay as you grow”

Praktischer Teil



Ihre Fragen?



Atix GmbH
Einsteinstraße 10
87516 Unterschleißheim
www.atix.de info@atix.de



ADVANCED TECHNOLOGY FOR INDIVIDUAL SUCCESS