

Hochverfügbarkeitslösungen mit MySQL

**Lenz Grimmer
Community Relations Manager
MySQL GmbH**

2. November 2006

Überblick

- MySQL: Überblick
- Begriffe und Konzepte für Hochverfügbarkeit
- Anforderungen an Hochverfügbarkeit
- Hochverfügbarkeitslösungen mit MySQL
 - Replikation
 - MySQL Cluster
- Weitere Informationen sowie Fragen & Antworten

Firmengeschichte

- Entwicklung des MySQL-Servers begann 1994 (TCX DataKonsult AB)
- MySQL AB wurde 1995 von Michael “Monty” Widenius, David Axmark and Allan Larsson gegründet
- Server-Entwicklung orientierte sich an den Anforderungen im produktiven Einsatz: wenig Features, aber schnell und stabil
- Häufige Releases mit überschaubaren Änderungen
- Einfach zu installieren und verwenden (15-Minuten-Regel)



Zahlen und Fakten

- Mehr als 320 Mitarbeiter in mehr als 25 Ländern (Top 5: USA, Schweden, Deutschland, Russland und Irland)
- Über 10 Standorte weltweit (Uppsala, Cupertino, München)
- Ca. 2/3 of aller Mitarbeiter arbeiten zuhause
- Kommunikation via E-Mail, VoIP und IRC
- Aktiengesellschaft in privater Hand, nicht börsennotiert und schuldenfrei
- Finanziell unterstützt von verschiedenen Venture Capital-Firmen (Benchmark, Index, Intel, Red Hat, SAP, etc.)
- <http://www.mysql.com/>

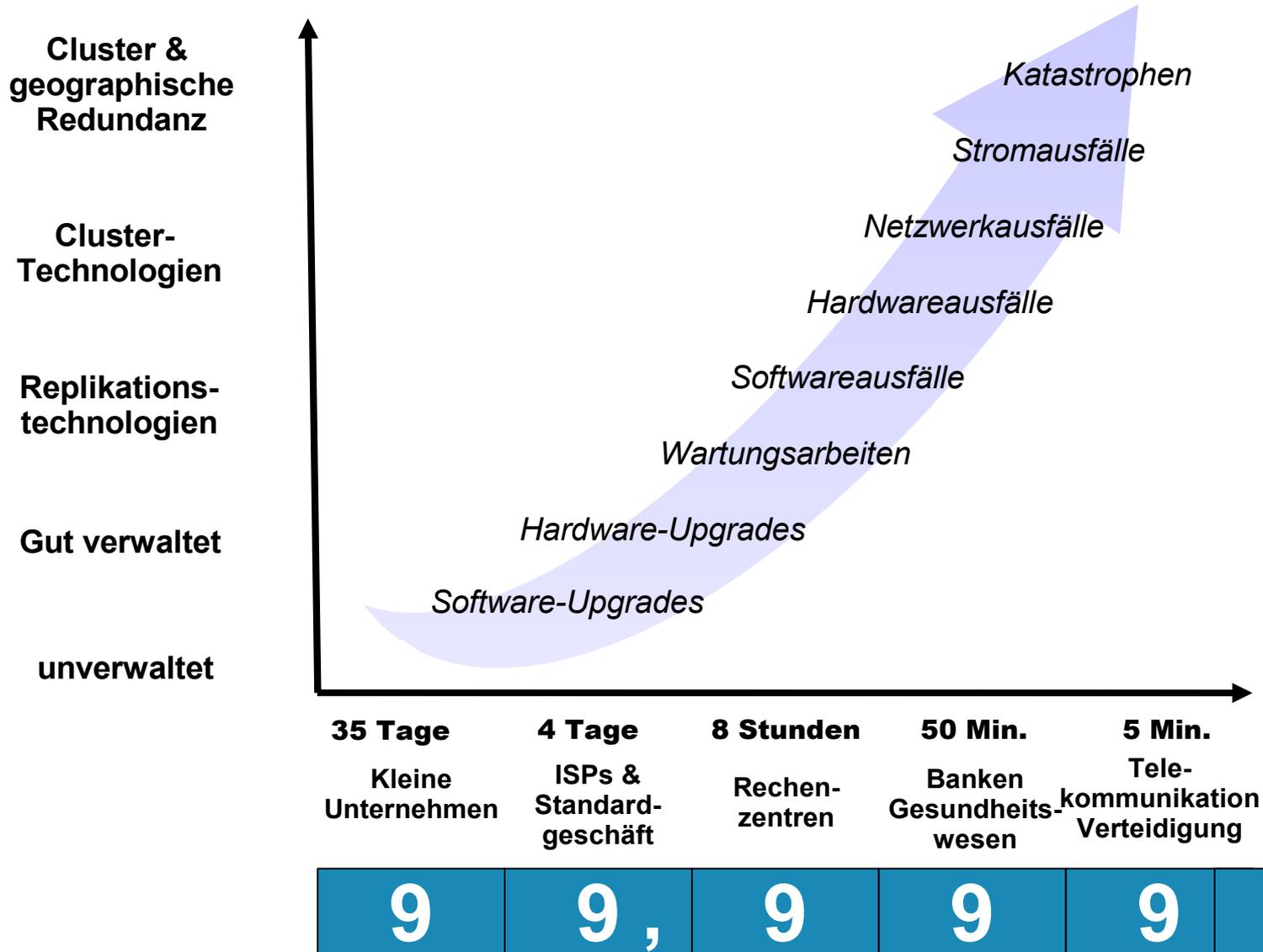
Geschäftsmodell

- Hauptprodukt: MySQL Server
 - Bestandteil des sog. “LAMP-Stack”
 - Über 8 Mio. aktive Installationen
- Einnahmequellen
 - Service-Angebote rund um die MySQL Produkte (Support, Schulung, Zertifikation, Consulting)
 - Subscriptions-Services wie MySQL Network (Updates, Support, Knowledge-Base)
 - Kommerzielle Lizenzen von MySQL für OEM Kunden
 - Franchising von MySQL-Produkten und Services unter dem MySQL-Warenzeichen an Partner
- Duales Lizenzmodell bietet kommerzielle Lizenzen als Alternative zur GPL (“Quid pro quo”)
- MySQL ist Inhaber der Marke (Produktnamen and Logos)

Was ist Hochverfügbarkeit?

- Hochverfügbarkeit bezieht sich auf die Verfügbarkeit von Ressourcen in einem Computersystem auch bei Ausfall einzelner Komponenten
- Zwei Hauptkategorien
 - Kundenspezifische Lösungen/redundante Soft- und Hardware
 - Software-Lösungen aufbauend auf COTS-Komponenten
- Ständige Verfügbarkeit
 - Ausfallfreier Betrieb
 - Keine Betriebsstörung selbst bei Systemausfällen
 - Sehr hoher Verfügbarkeitsgrad (oft unnötig)

Die 5 Neuner zur Verfügbarkeit



Andere Hochverfügbarkeitskonzepte/Begriffe

- Fehlertoleranz
 - Aufrechterhaltung der Verfügbarkeit von Ressourcen trotz Störungen/Ausfällen
 - Erreichung durch Redundanz von Komponenten
- Single Point of Failure (SPOF)
 - Hard- oder Softwarekomponente, deren Ausfall einen Komplettausfall des Systems nach sich zieht
 - Komponente, der typischerweise ein Gegenstück/ redundanter Ersatz fehlt
- Failover
 - Ungeplanter Wechsel eines Dienstes vom Primär-Server auf ein Standby-System im Fehlerfall
 - Überwachung z.B. mittels Heartbeats

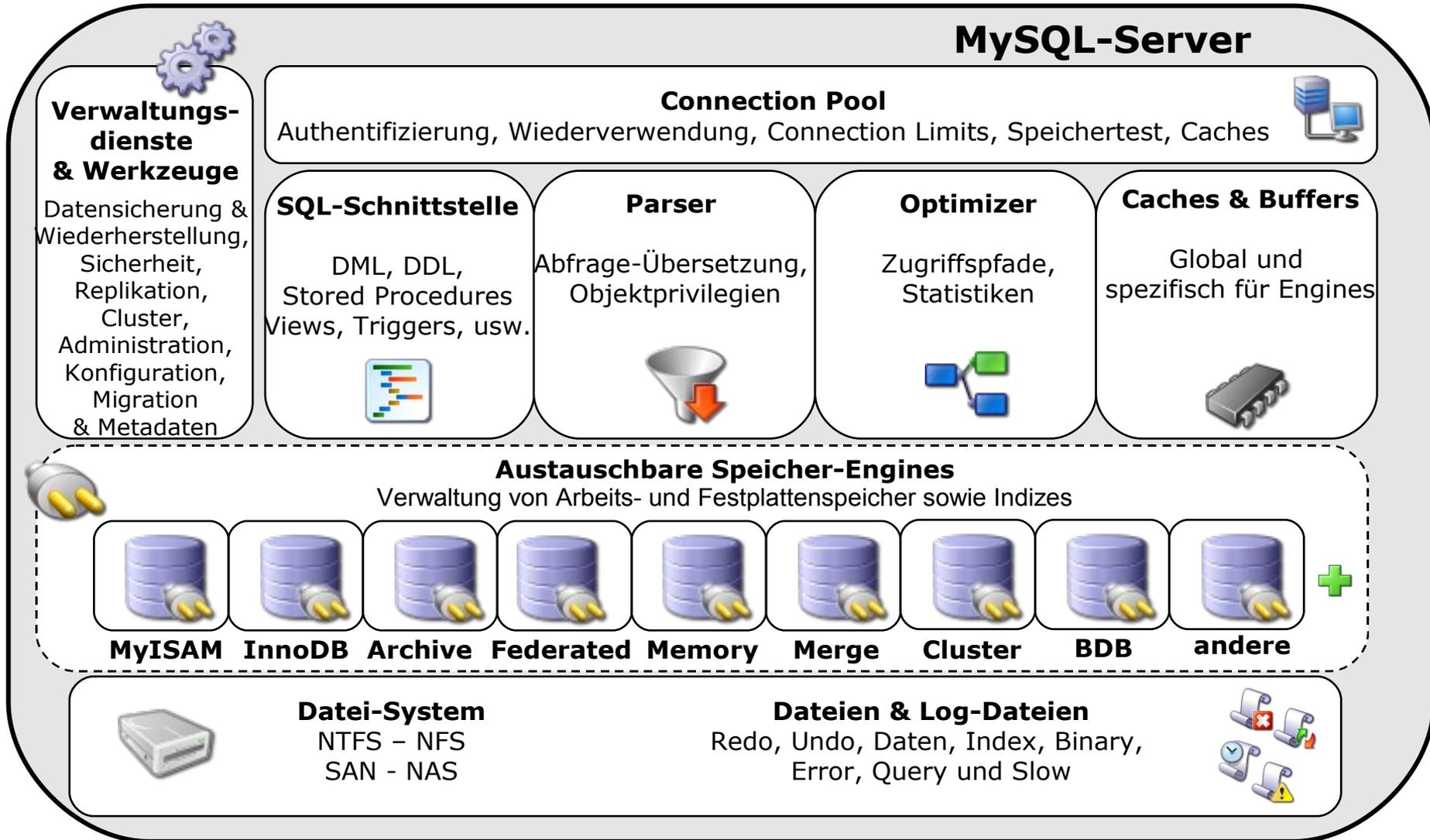
Überlegungen und Anforderungen zu Hochverfügbarkeit

- Service Level Agreements (SLAs)
- Budgets & Zeitrahmen für Fertigstellung
- Anforderungen an Failover
- Fault Tolerance – Single Point of Failures
- Open Source, proprietäre & gemischte Technologie-Plattformen
- Sicherheit
- Hardware-Anforderungen
- Überlegungen zum Storage
 - RAID
 - SAN
 - NAS
- Betriebssystemunterstützung
- Geographische Redundanz
- Welchen Einfluß hat ein Hard- oder Software-Fehler auf den Betrieb?

Hochverfügbarkeitslösungen mit MySQL

- MySQL-Replikation
- Linux Heartbeat & MySQL-Replikation
- Linux Heartbeat, DRBD & MySQL
- Lastverteilung & MySQL
- MySQL mit Shared-Storage & Clustering-Agenten
- MySQL Cluster
- MySQL Cluster & MySQL-Replikation
- Andere Hochverfügbarkeitslösungen für MySQL

MySQL Server Architektur



Was ist Replikation?

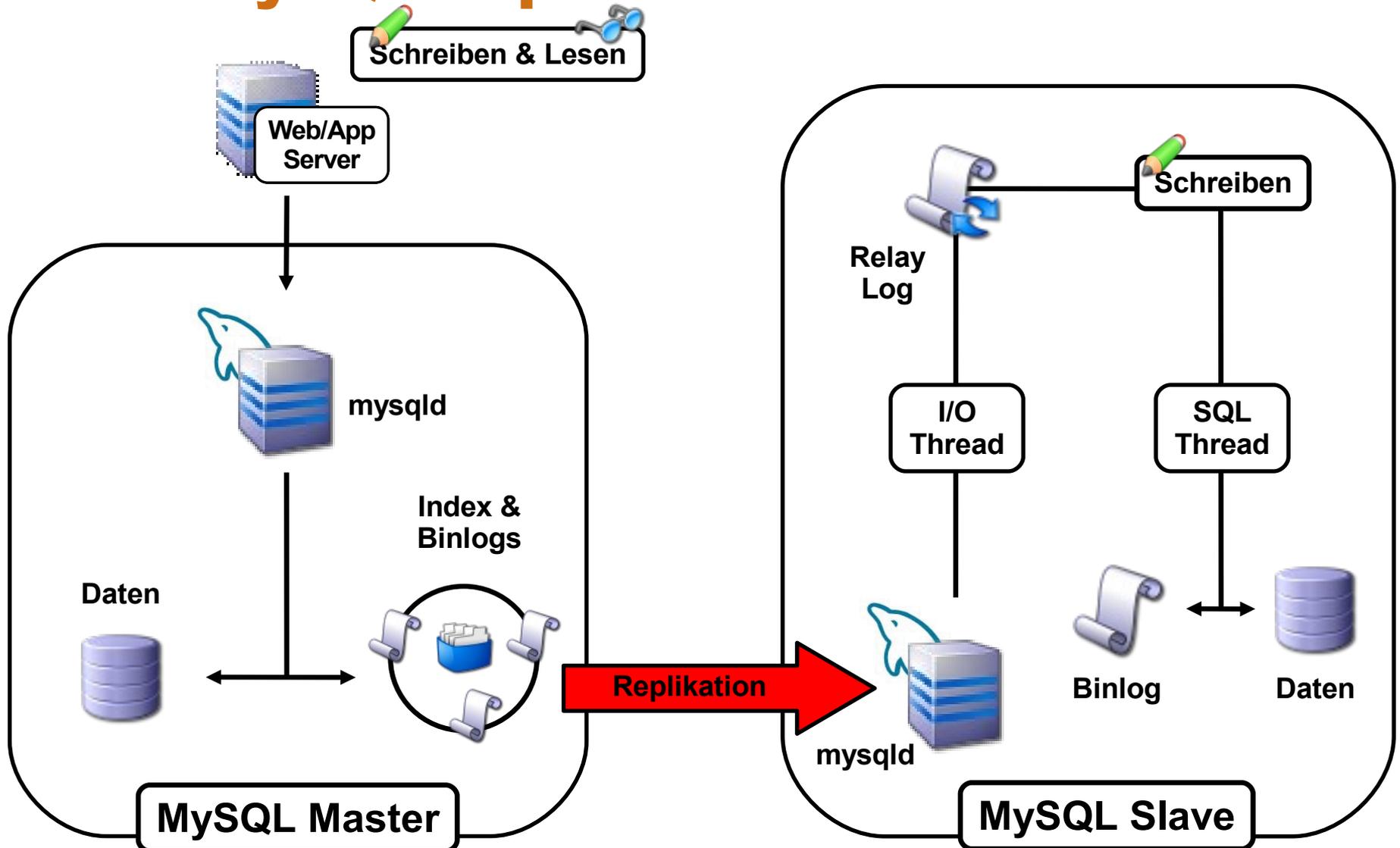
- Vervielfältigung von Daten auf mehrere Standorte
- Nutzung zusätzlicher Hard- und Software für:
 - Verfügbarkeit
 - Absicherung vor Datenverlust
 - Zuverlässigkeit
 - Fehlertoleranz
 - Performanz
- Asynchron vs. synchron
- Latenz
- Dateibasierte Replikation (z.B. rsync) vs. Block-Level-Replikation (z.B. DRBD)

MySQL-Replikation

- Asynchron
 - befehlsbasiert
- Lesen vs. Schreiben
 - Antwortzeiten
 - Lastverteilung
 - Horizontale Skalierung (Scale out)
- Verwaltung
 - Datensicherung, Upgrades, Testen
- Ermöglicht geographische Redundanz
 - Aufrechterhaltung der Betriebsbereitschaft & Wiederherstellung nach Ausfällen
- Multi-Source vs. Multi-Master
- Manueller Failover

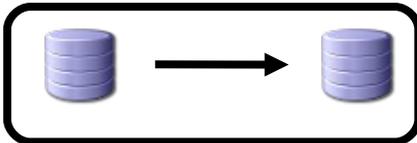
MySQL-Replikation im Überblick

Schreiben & Lesen

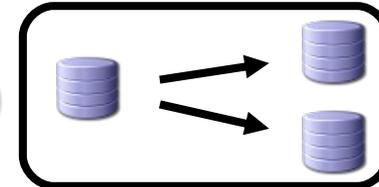


Topologien zur MySQL-Replikation

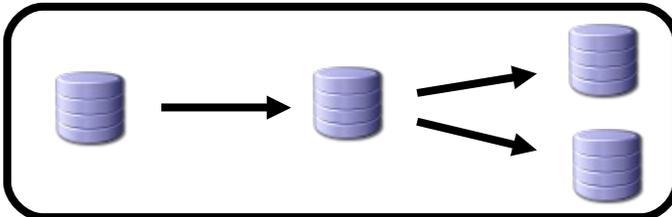
Master > Slave



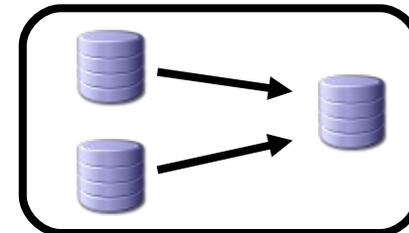
Master > Slaves



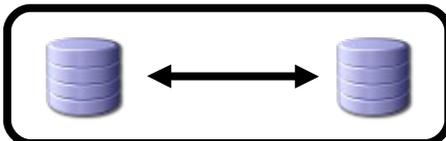
Master > Slave > Slaves



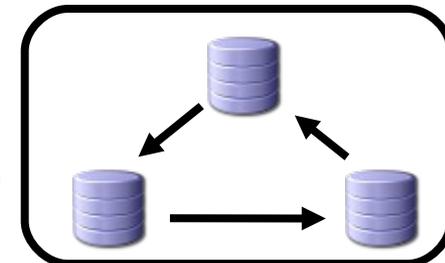
Masters > Slave (Multi-Source)



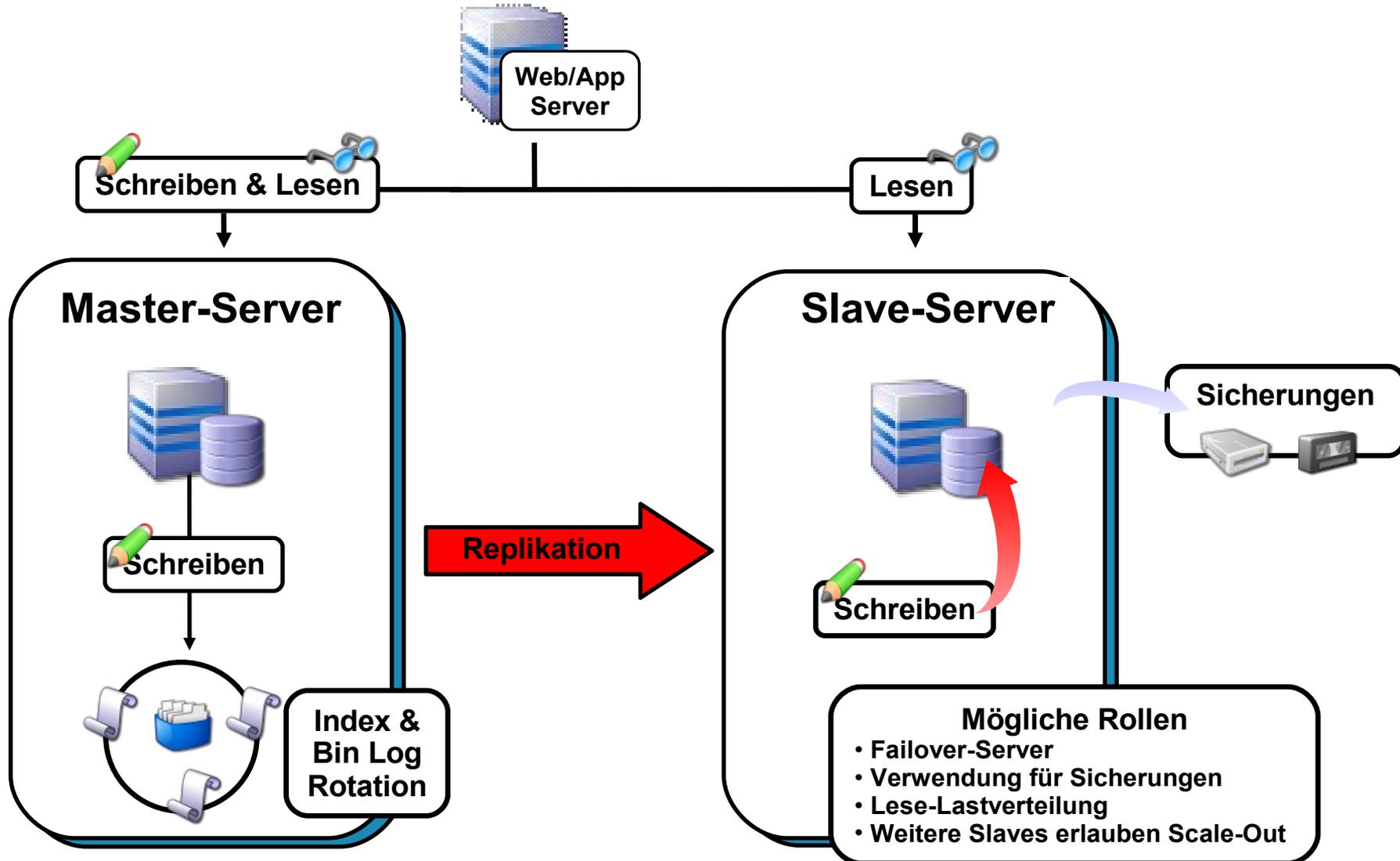
Master < > Master (Multi-Master)



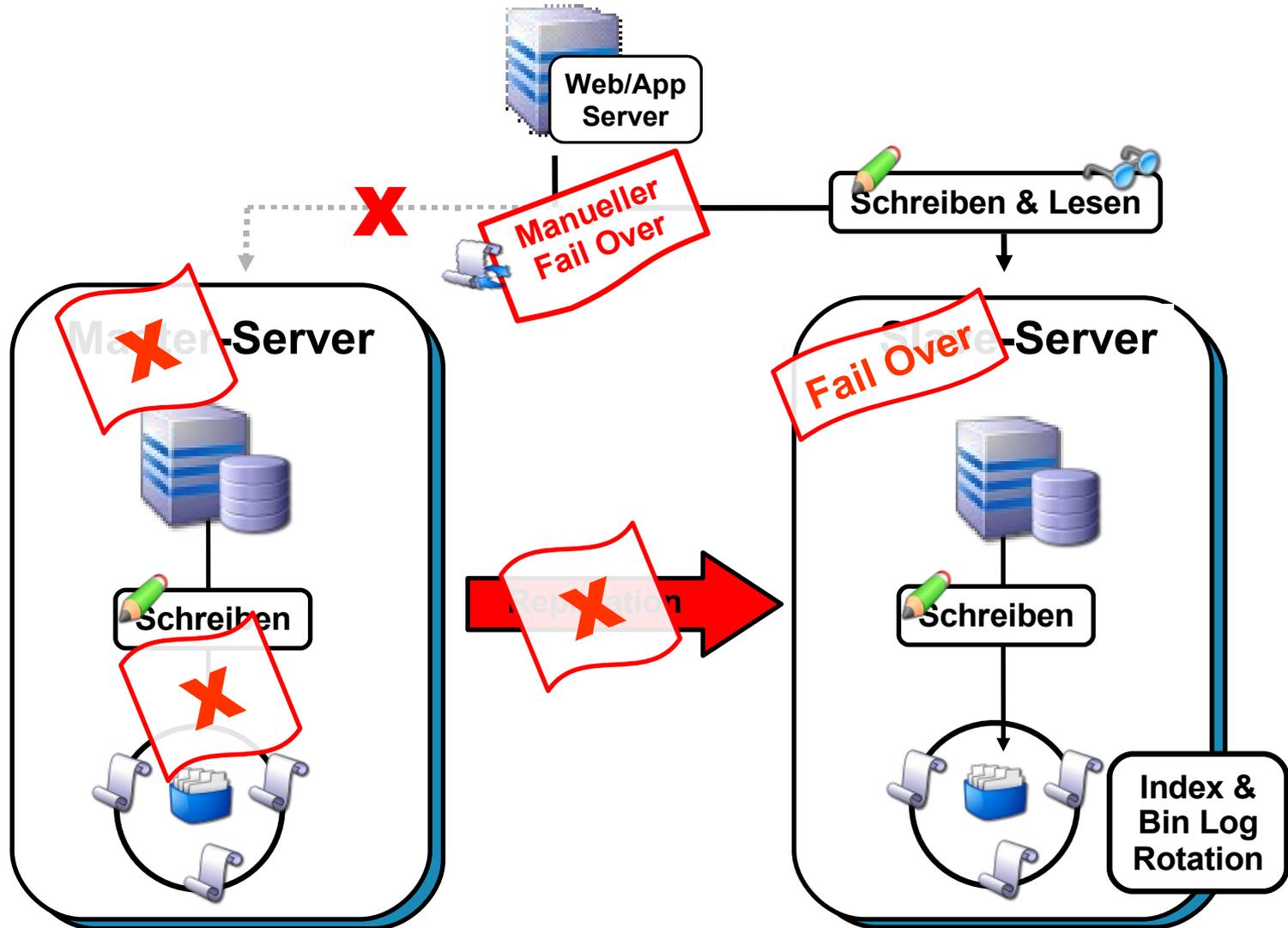
Ring (Multi-Master)



MySQL-Replikation (Scale-Out)



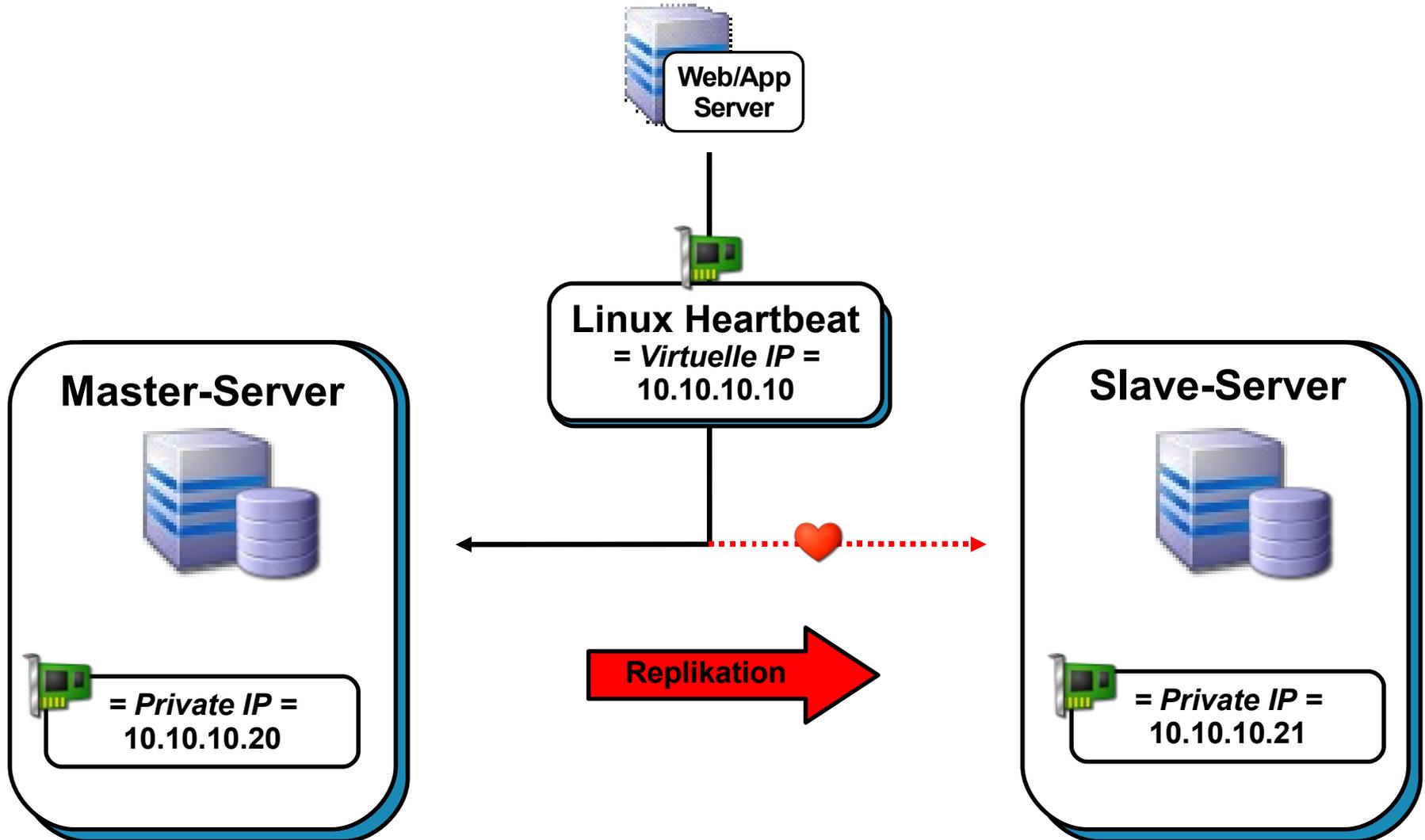
MySQL Replikation (Failover)



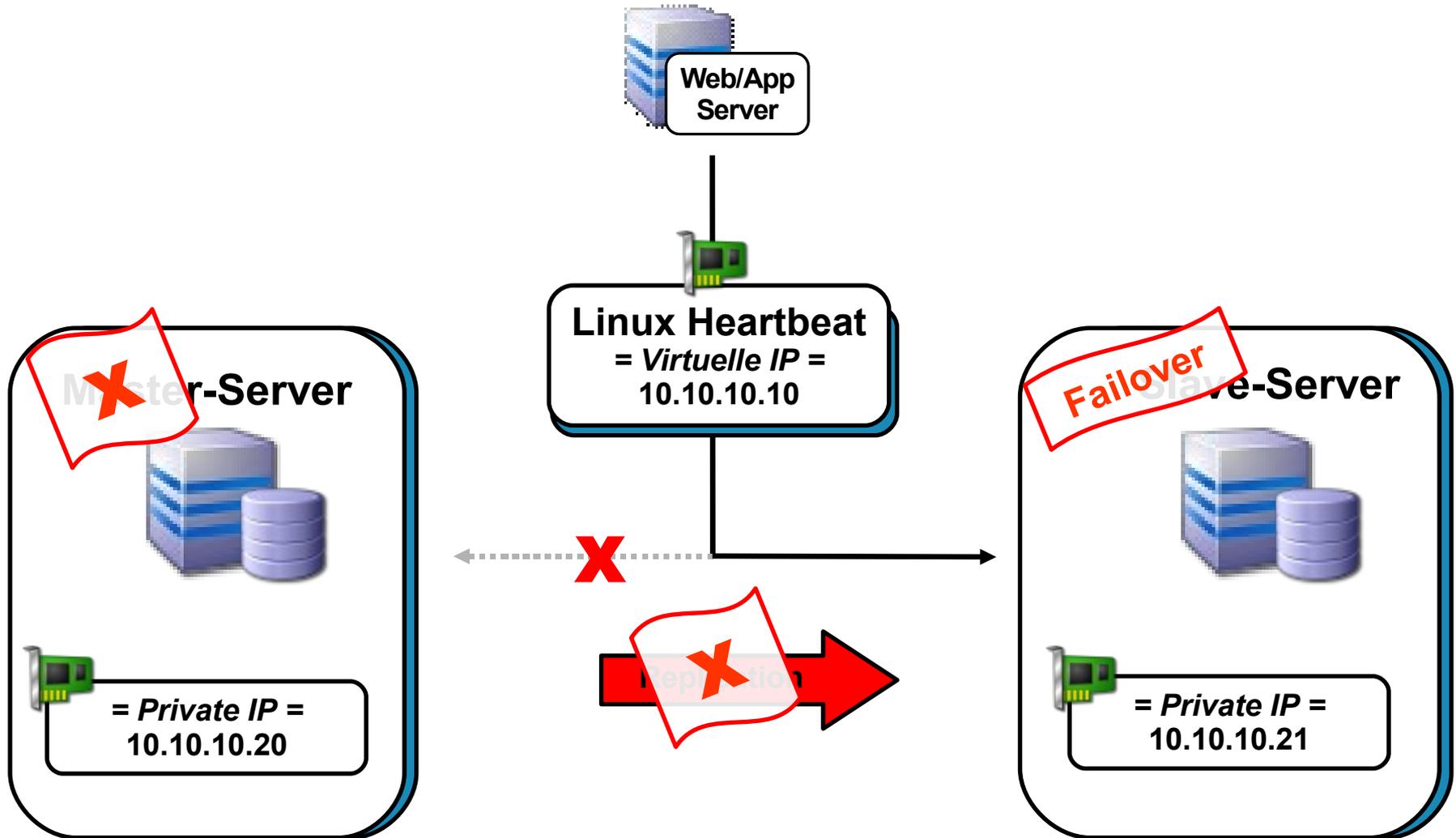
Linux Heartbeat & MySQL-Replikation

- Heartbeat-Protokoll
 - Mitteilungen werden von Knoten zu Knoten geschickt
 - Fehlgeschlagener Heartbeat startet einen Failover-Prozeß
- Verwaltung einer virtuellen IP
 - Übernahme der IP Adresse
- Merkmale
 - Einfach einzurichten, geringe Komplexität
 - Open Source
 - Kostengünstig
 - Automatische virtuelle IP-Verwaltung
 - Asynchrone Datenreplikation
 - Manuelles oder automatisches (bevorzugt) Failover
 - Wiederherstellung und mehrfache Ausfälle nicht trivial

Linux Heartbeat & MySQL Replikation



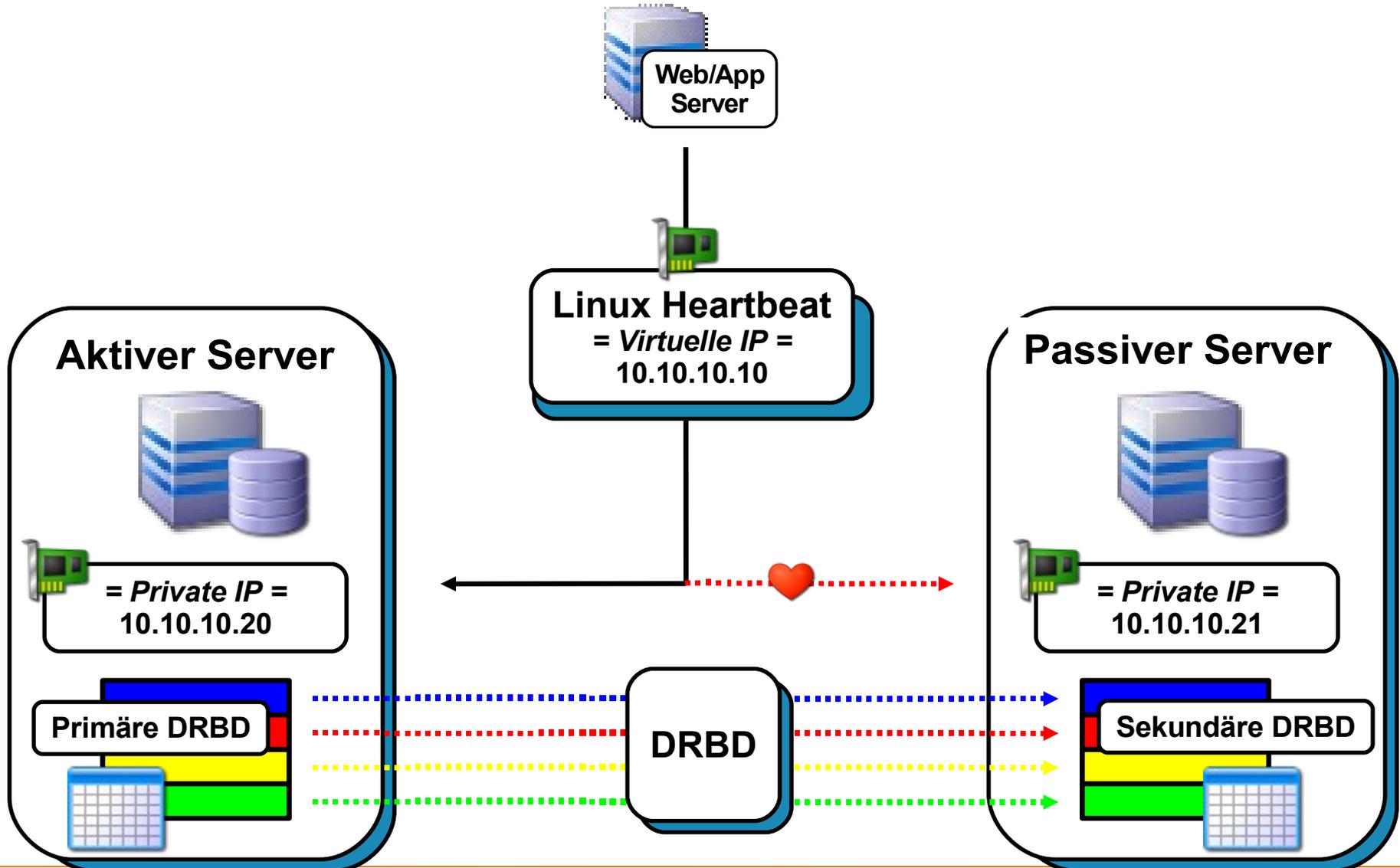
Linux Heartbeat & MySQL Replikation (Failover)



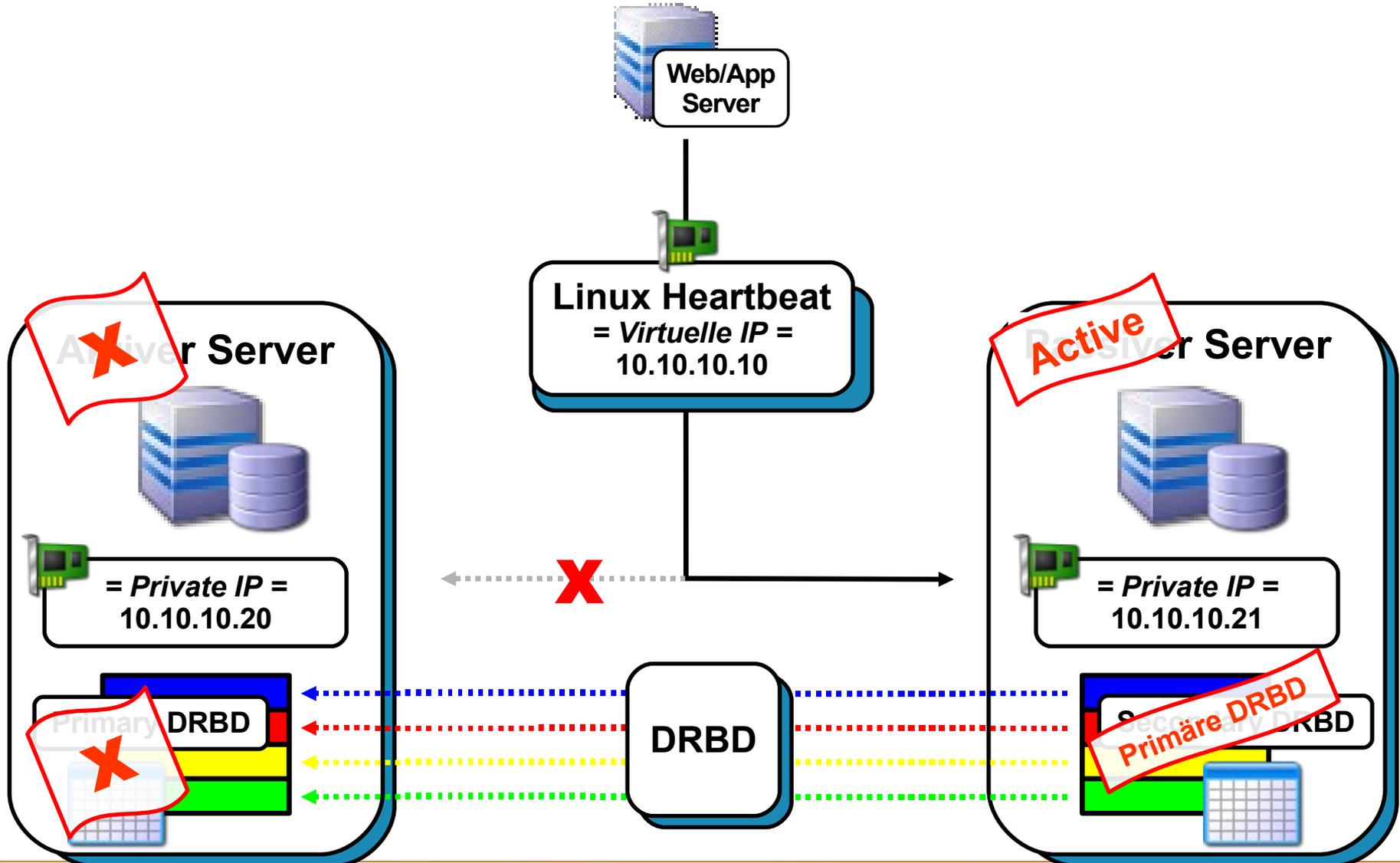
Linux Heartbeat, Block-Replikation & MySQL

- Distributed Replicated Block Device (DRBD)
 - Läuft über Standard IP-Netzwerke
 - Verteilte Datenhaltung auf Blockebene
 - “RAID 1 via TCP/IP”
- Synchron
- Merkmale
 - Installation und Konfiguration sind komplexer
 - Keine speziellen Netzwerkkomponenten (außer Heartbeat)
 - Ausgezeichnete Performanz (Datenblöcke vs. Zeilen)
 - Vereinfacht in vielen Fällen die Wiederherstellung
 - Linux Heartbeat verwaltet Failover und virtuelle IPs

Linux Heartbeat, Block-Replikation & MySQL



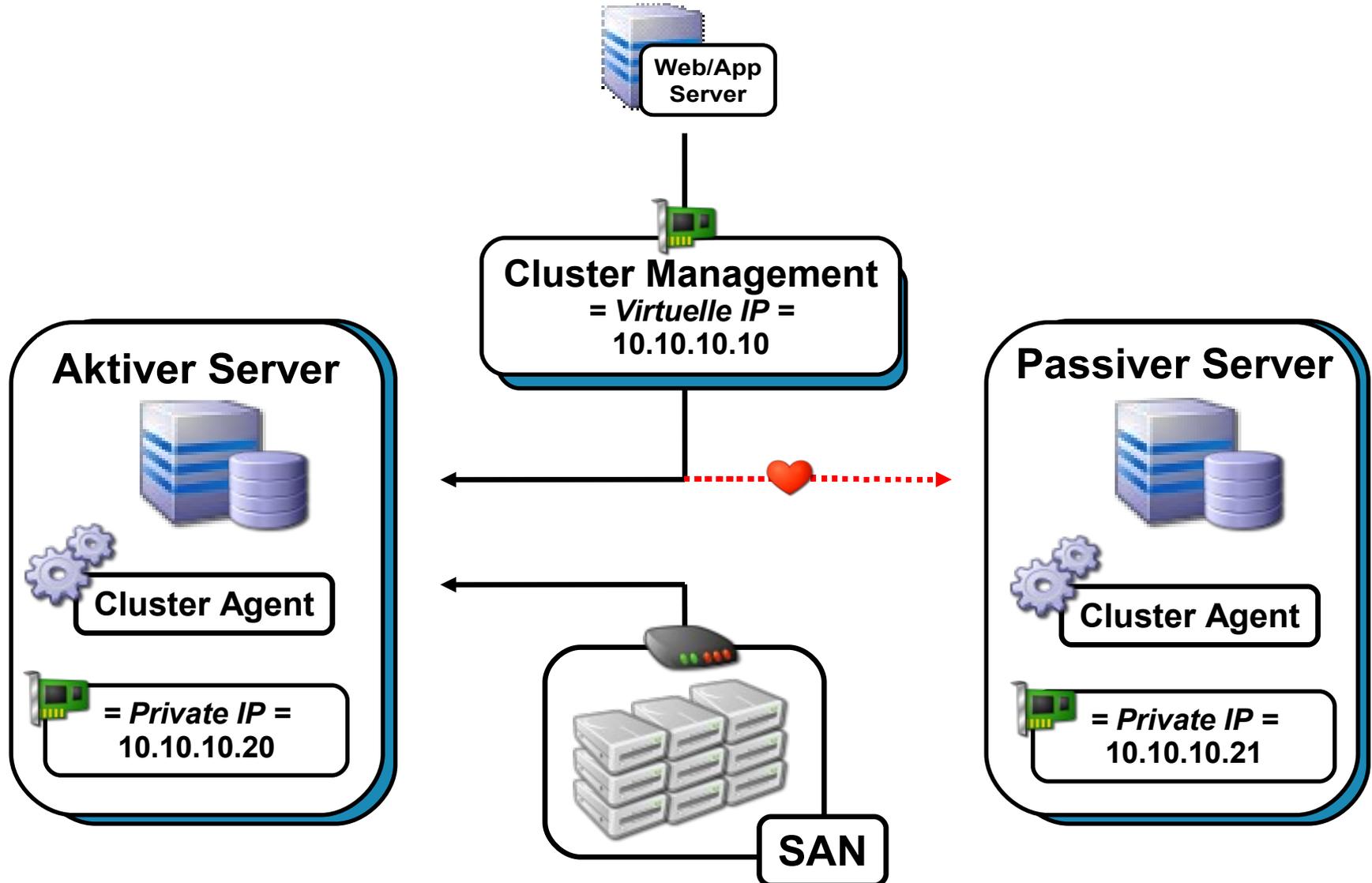
Linux Heartbeat, Block-Replikation & MySQL



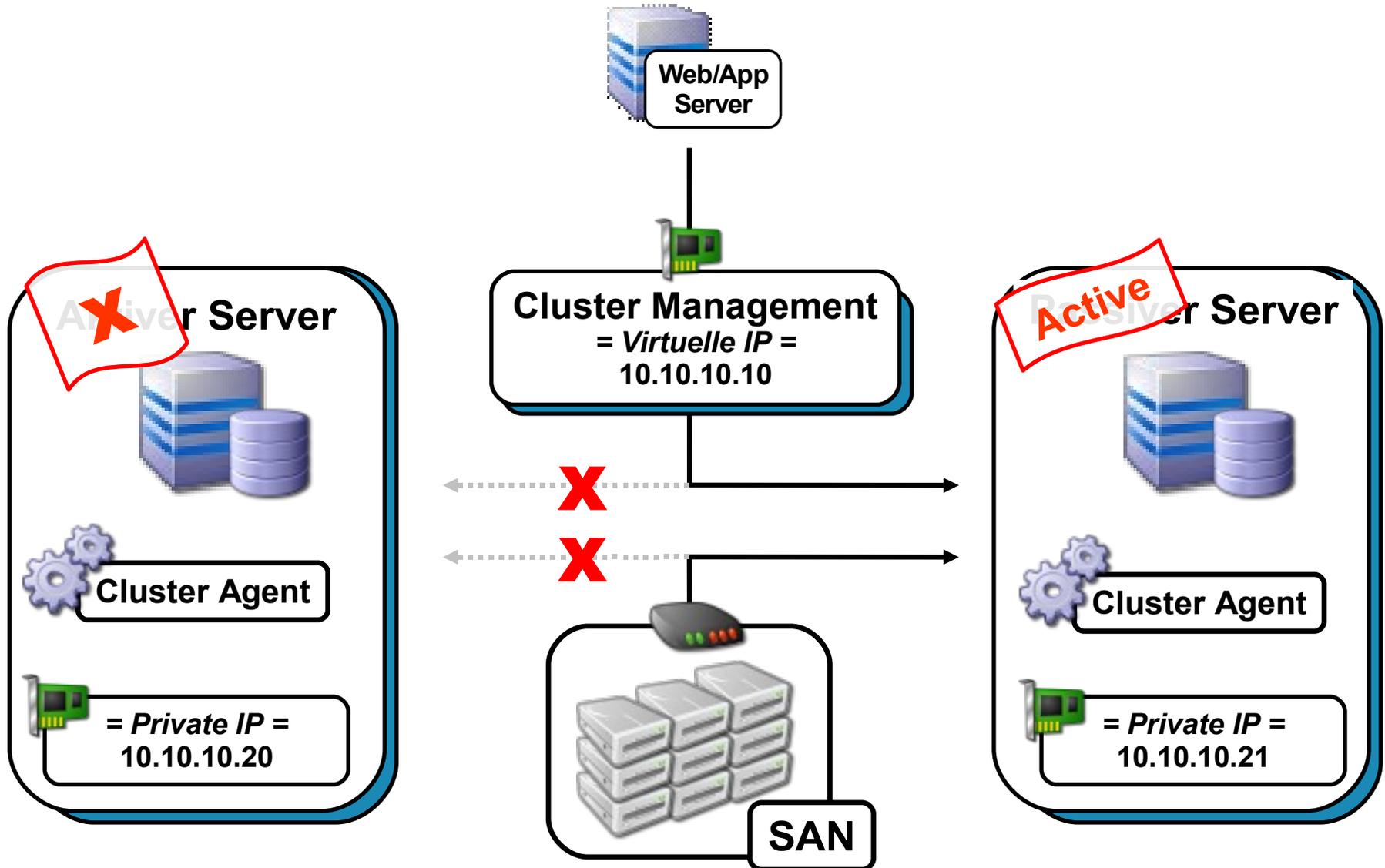
MySQL mit Shared Storage & Cluster-Agenten

- Konfiguration ähnlich Aktiv/Passiv
- Mehrere Instanzen können nicht gleichzeitig auf die selben Daten-Dateien zugreifen
- Automatische Verwaltung
 - Virtuelle IPs
 - Failover
 - Umhängen von Dateisystemen
- Merkmale
 - teuer (Speicher, Hardware, Software)
 - Ungenutzte Ressourcen
 - Lange Ausfallzeiten
 - Anfänglich sehr komplex
 - Viele Möglichkeiten und bewährte Anbieter

MySQL mit Shared Storage & Cluster-Agenten



MySQL mit Shared Storage & Cluster-Agenten



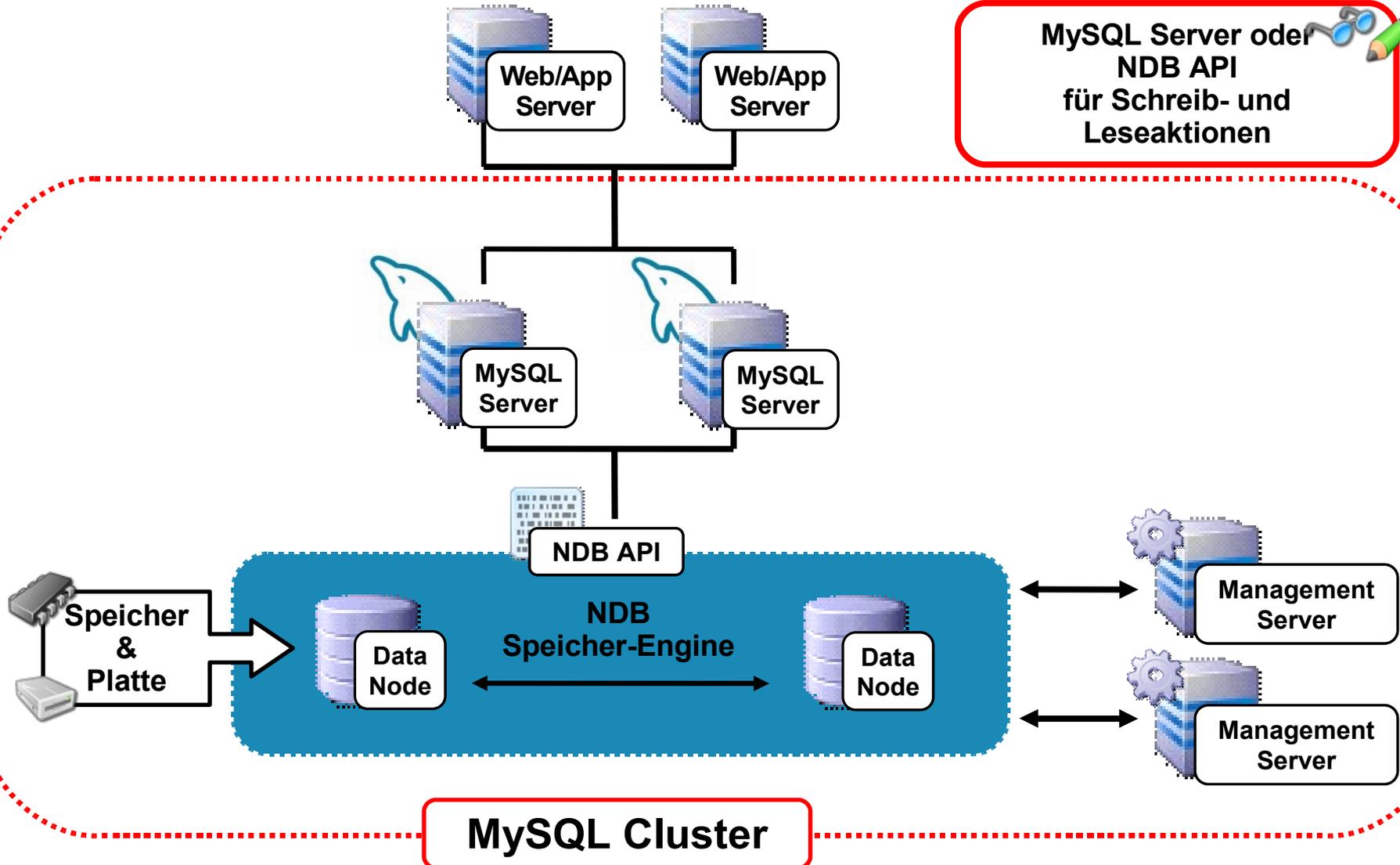
MySQL Cluster

- Shared-Nothing
- Synchron (2-Phasen-Commit)
- Schnelles, automatisches Failover
- Hohe Performanz
- Hoher Transaktionsdurchsatz
- Keine spezifischen Anforderungen an die Komponenten
- In-Memory & Unterstützung von plattenbasierten Daten (5.1)
- Heartbeat-Protokoll

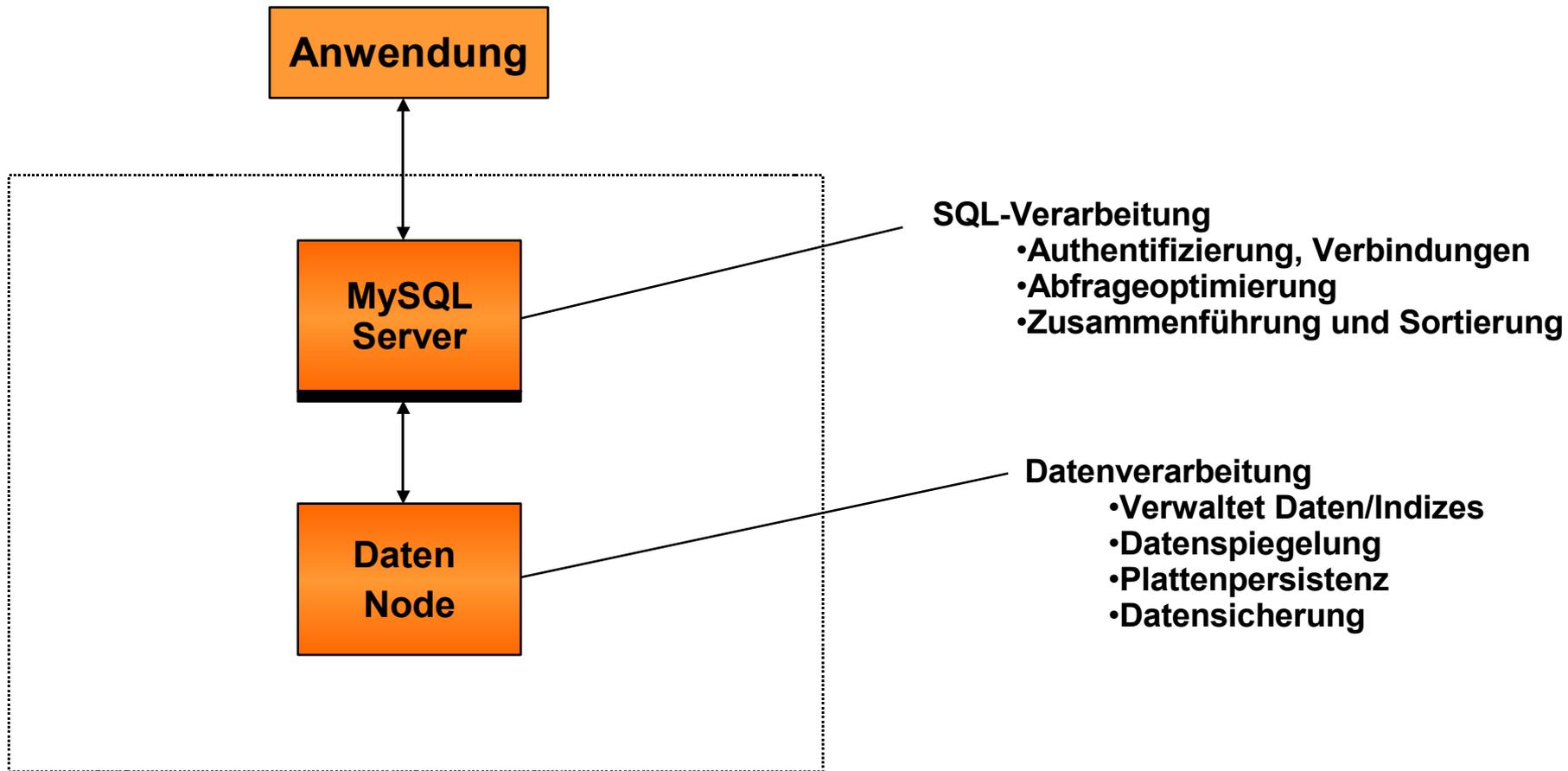


Architektur von MySQL Cluster

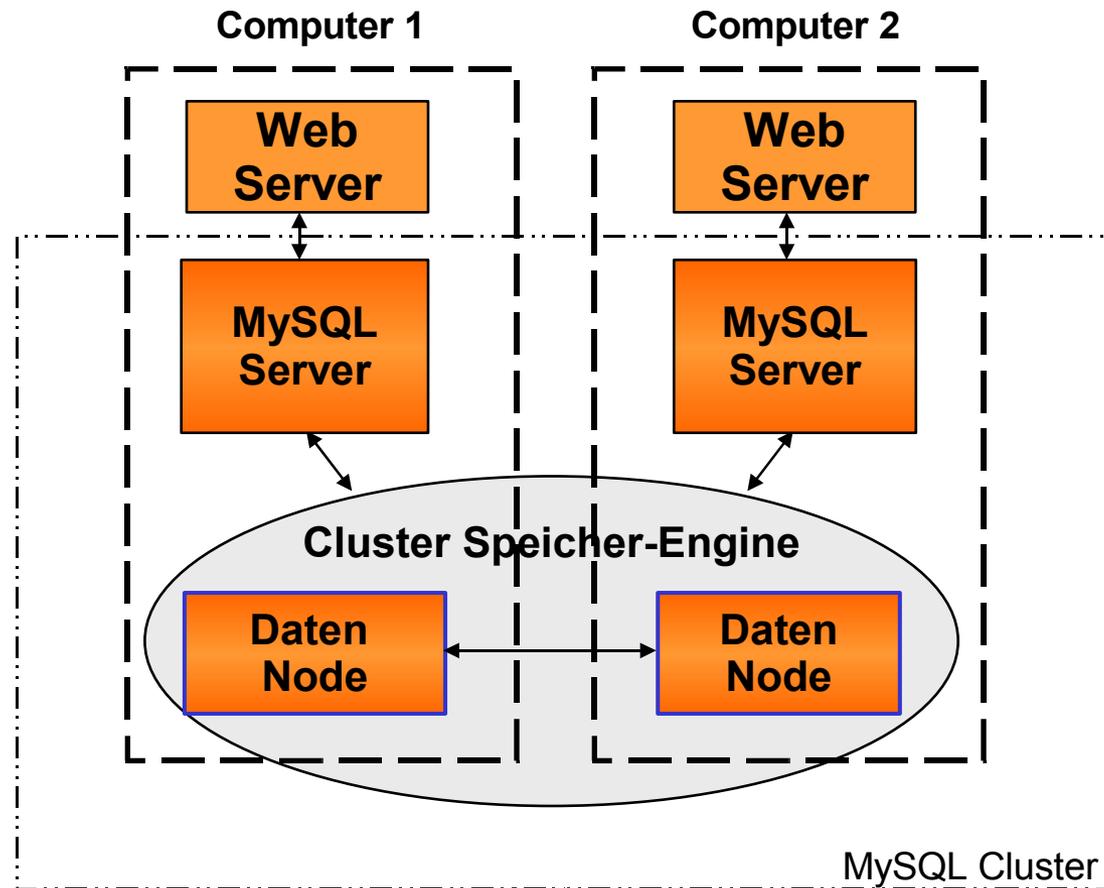
MySQL Server oder NDB API für Schreib- und Leseaktionen



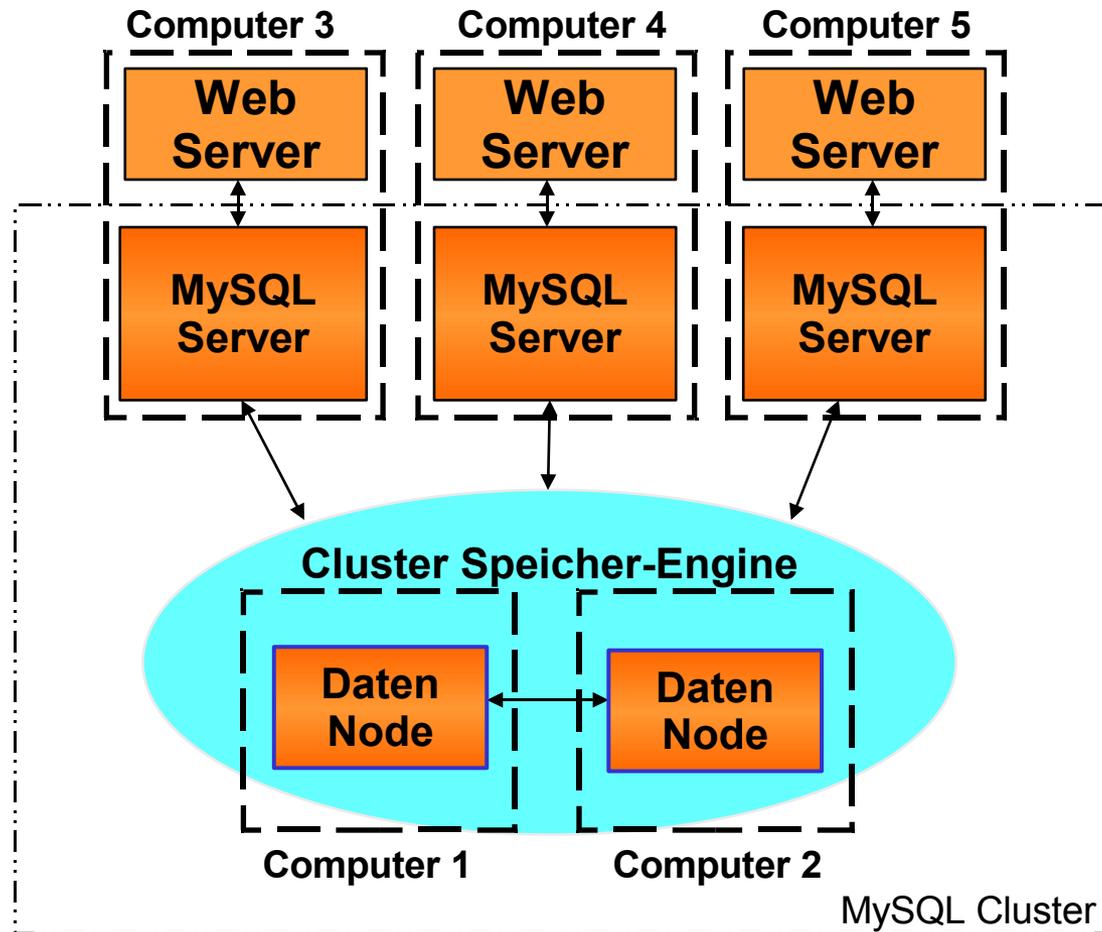
Getrennte SQL- und Datenschicht



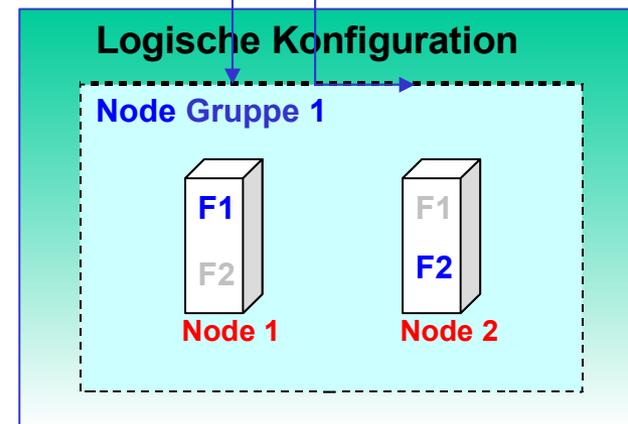
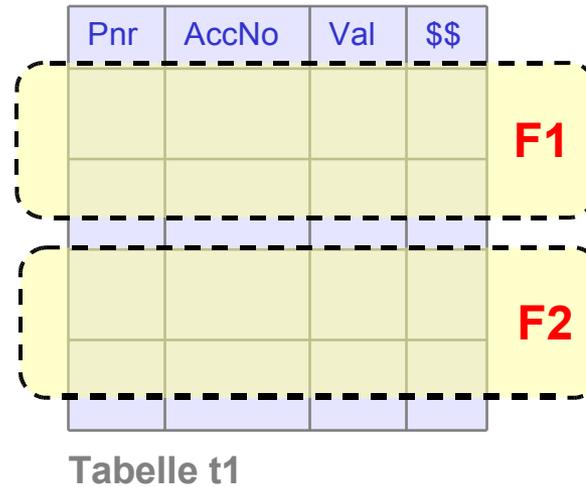
Minimale Konfiguration mit Redundanz



Konfiguration für Hochperformanz



Datenverteilung auf 2 Nodes



2 Kopien der Daten

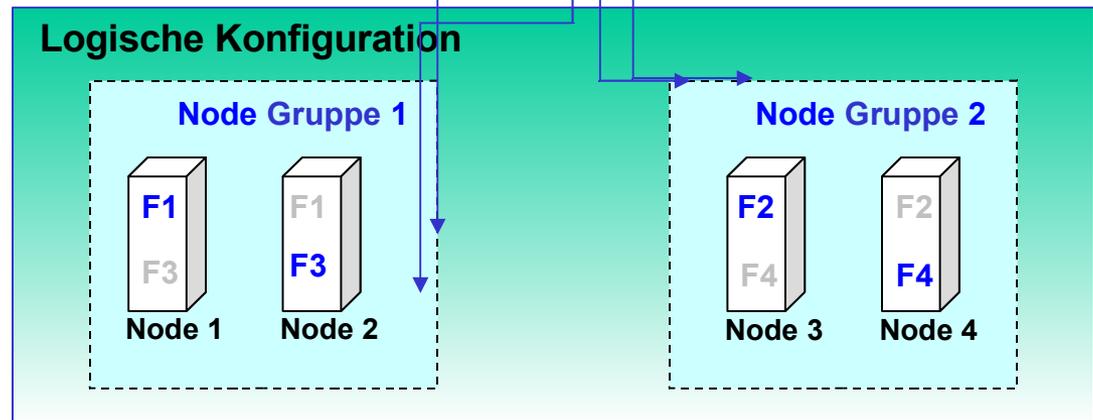
Fx – primäre Replik

Fx – sekundäre Replik

Datenverteilung auf 4 Nodes

Pnr	AccNo	Val	\$\$	
				F1
				F2
				F3
				F4

Tabelle 1

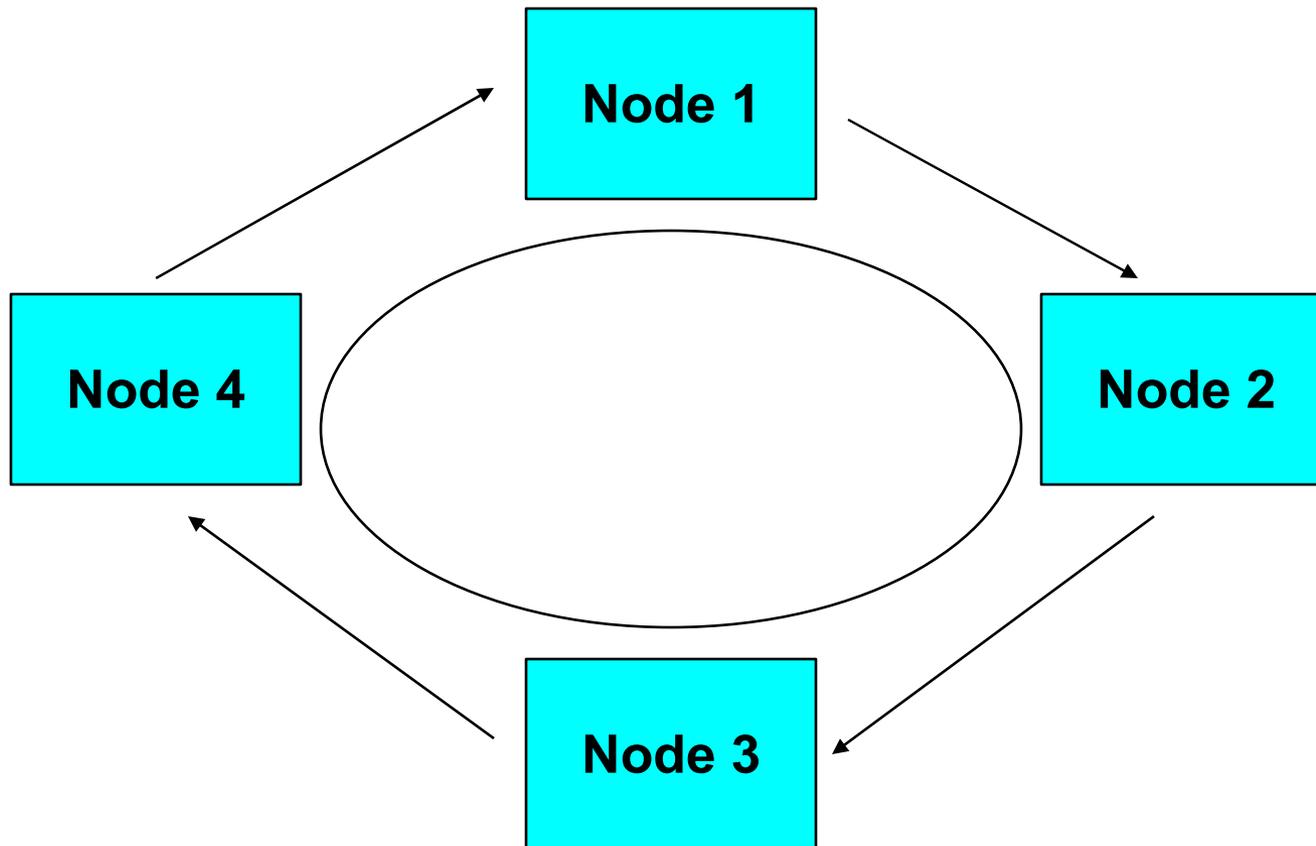


2 Kopien der Daten

F_x – primäre Replik

F_x – sekundäre Replik

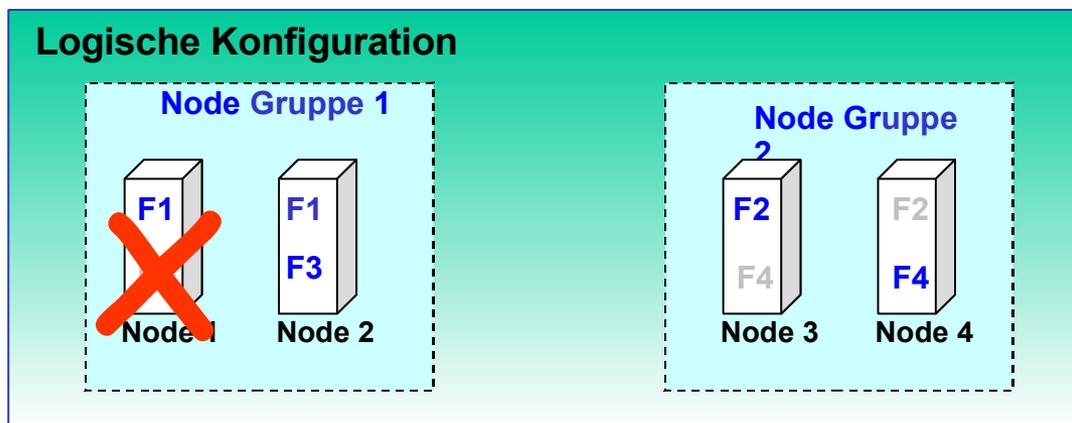
Fehlererkennung: Heartbeats, verlorene Verbindungen



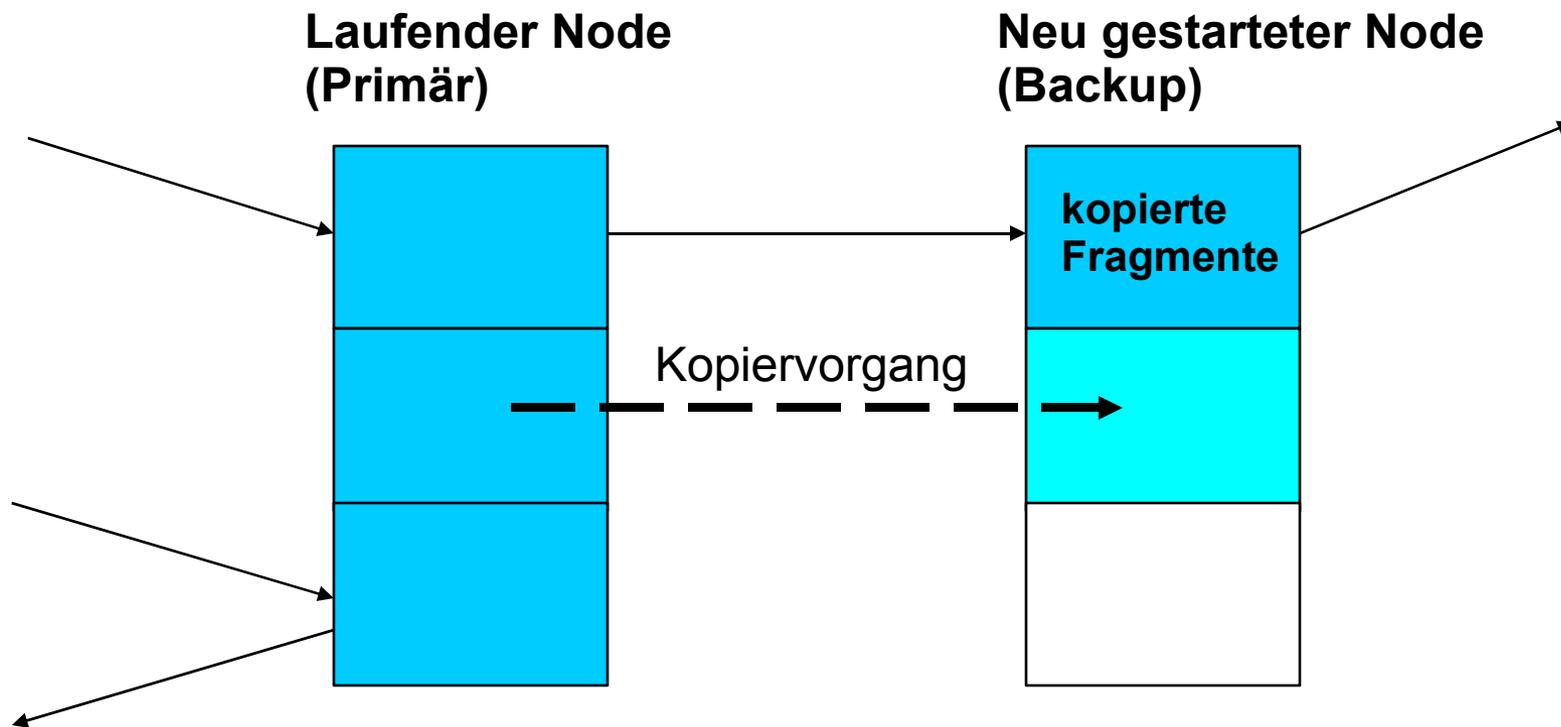
Ausfall eines Nodes

Was passiert wenn Node 1 ausfällt?

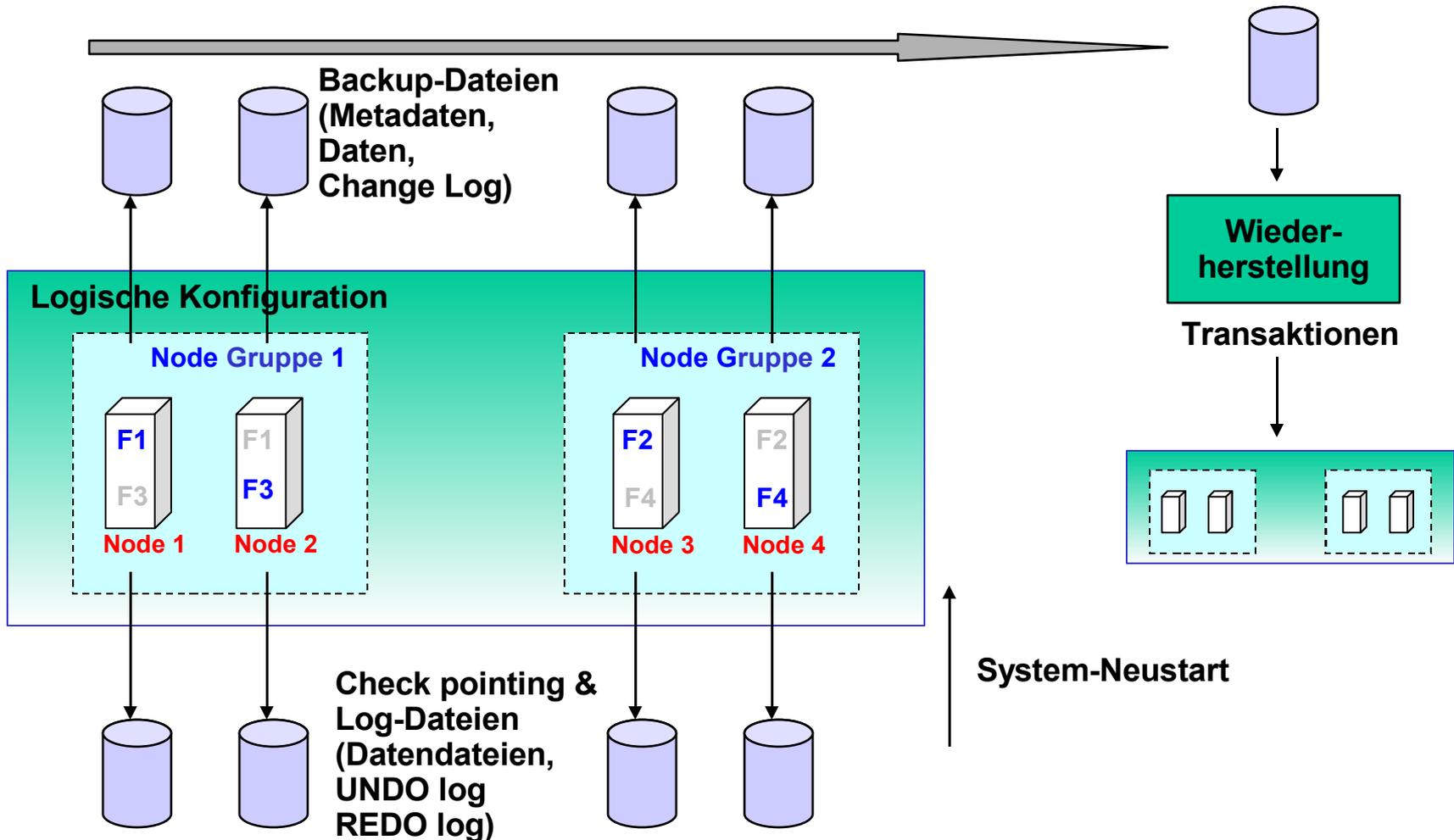
- Erkennung des Ausfalls des Nodes. Primäres Fragment F1 wird von Node 2 übernommen
- Automatischer Wiederherstellung von Node 1 nach Neustart
 - Node 1 stellt F1 und F3 von Node 2 wieder her und tritt dem Cluster wieder bei



Wiederherstellung eines Nodes



System-Wiederherstellung



Verfügbarkeitsfunktionen – Zusammenfassung (1/2)

Zuverlässigkeit betreffend Ausfall eines Nodes

- **Datenredundanz** mit synchroner Replikation
- **Schnelles nahtloses Umschalten**, da Daten schon repliziert wurden
- **Ausfall mehrerer Nodes** wird ohne Systemstop wiederhergestellt
- **Network Partitioning Protokoll**, um Datenbankpartitionierung zu vermeiden

Zuverlässigkeit betreffend Systemausfall

- **Protokollierung mit Check Points**
- **Datenbankwiederherstellung** auf einen transaktionskonsistenten Stand nach einem Systemausfall

Zuverlässigkeit betreffend Anwendungsfehlern

- **Online-Backup & -Restore**

Verfügbarkeitsfunktionen – Zusammenfassung (2/2)

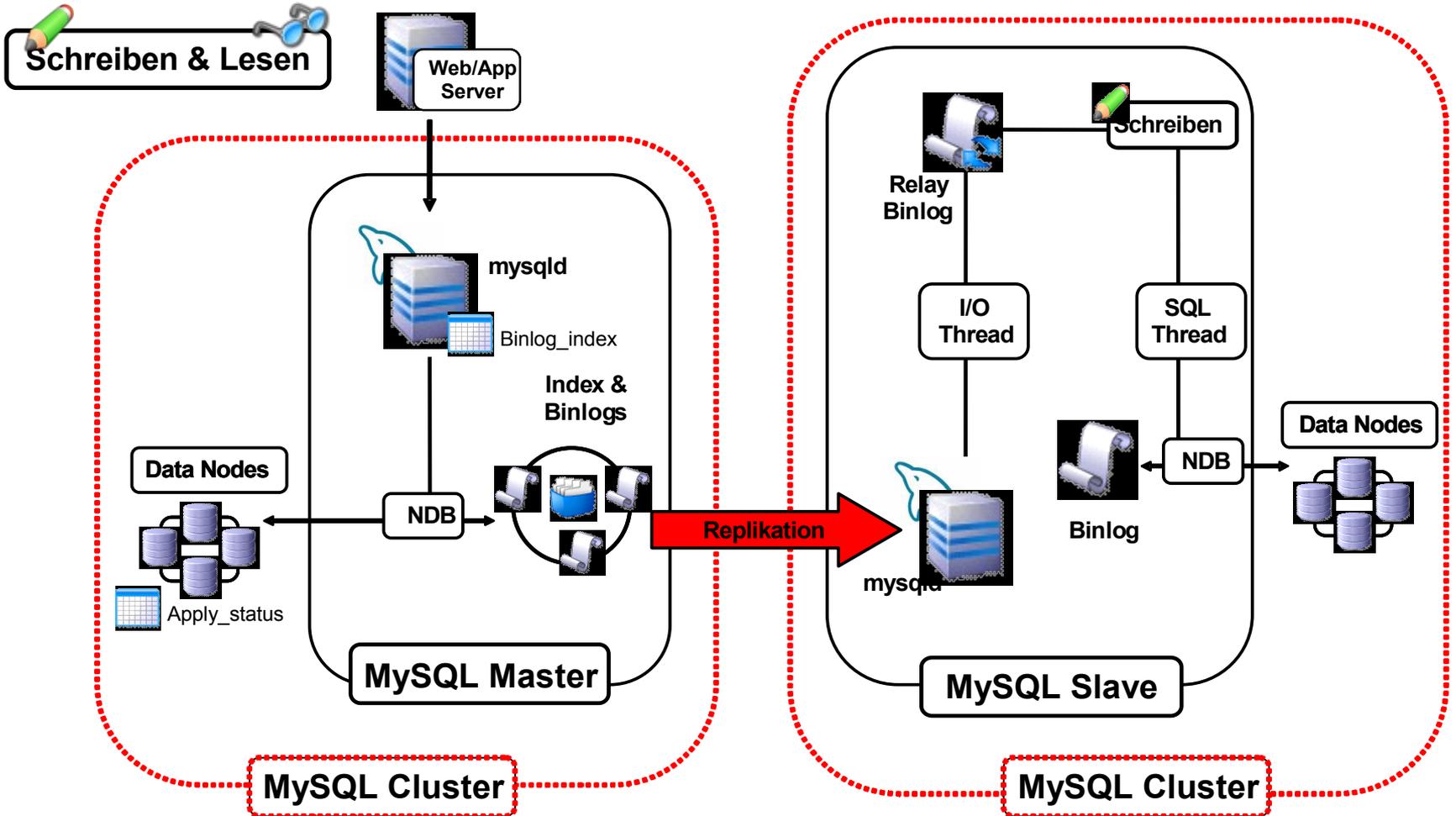
Node-Architektur, um Single-Point-of-Failure auszuschließen

- **MySQL-Server** sind mit allen Daten Nodes verbunden
- **Data Node** mit replizierten Daten auf verschiedenen Servern
- **Management Server Node** kann ohne Einfluß auf die Datenbank ausfallen

Disaster Recovery

- **Geografische Daten-Redundanz** mit MySQL-Replikation zwischen entfernten Systemen

MySQL Cluster mit Replikation



Globale Redundanz: Business Continuity & Disaster Recovery

Andere Hochverfügbarkeitslösungen für MySQL

- HA/Performance Networking für MySQL Cluster
 - Dolphin SCI Interconnect
- Betriebssystem Clustering
 - Sun Cluster Manager (Sun Solaris)
 - Microsoft Windows Clustering Services
 - Red Hat Cluster Suite
- Hochverfügbarkeits-Middleware
 - Continuent m/Cluster
- Backup Software
 - Zmanda

<http://solutions.mysql.com/>



Ausblick: MySQL 5.1

- Replikation
 - Zeilenbasierte Replikation
 - Verbesserte Replikation zwischen MySQL Cluster Instanzen
- MySQL Cluster
 - Unterstützung von festplattenbasierter Speicherung nicht indizierter Tabellenspalten
 - Online ADD INDEX und DROP INDEX für MySQL Cluster
 - Verbesserter Cluster-Backup

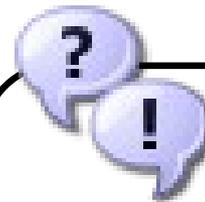
Beratungs- und Supportleistungen von MySQL

- MySQL Cluster Starthilfe
- MySQL Hochverfügbarkeitslösungen
- Starthilfe für horizontale Skalierung, Hochverfügbarkeit und Replikation von MySQL
- MySQL Architektur & Design
<http://www.mysql.de/consulting/>
- 24 x 7 Support für produktive Systeme
<http://www.mysql.de/support/>

Weitere Informationen

- MySQL-Schulungen
- MySQL Enterprise
- White Paper
- Anwenderberichte
- Pressemitteilungen, Veranstaltungen
- Webseminare
- Webseminar-Mitschnitte

Lenz Grimmer
Community Relations
lenz@mysql.com



**Fragen
&
Antworten**