

The Unbreakable Database System

**Real Application Cluster
auf Sun Cluster 3.0**

**Unterführung, 11.2002
M. Beeck, M. Kühn**

- **Comparisson HA**
 - HA Ziele, DataGuard, HA Oracle, RAC
- **Sun Cluster 3.0 Key Features**
 - Availability, Manageability, Maintenance
- **OPS+SC 2.2 / RAC+SC3.0**
 - Cache Fusion, TAF, Load Balancing
- **Konfiguration**
 - „best Practice“
- **Sun + Oracle zertifizierte Komplettlösungen**
- **Projekt: WireCard**
- **Live Demo**

Comparisson HA: HA Ziele

- Reduzieren bzw. Vermeiden von Ausfallzeiten
 - Software- , Hardware- , Human Error, Disaster
- Daten und Anwendungen für Benutzer verfügbar halten
- Erhöhung des Applikationsdurchsatzes durch horizontale Skalierung
- Erhöhte Verfügbarkeit auch während Wartung oder Upgrade

Comparisson HA: DataGuard

- Vorteile

- Schutz gegen Human Errors
- Standorte können beliebig sein (Disaster Recovery)

- Bis zu 10 Standby Server
- No Data Loss im Guaranteed

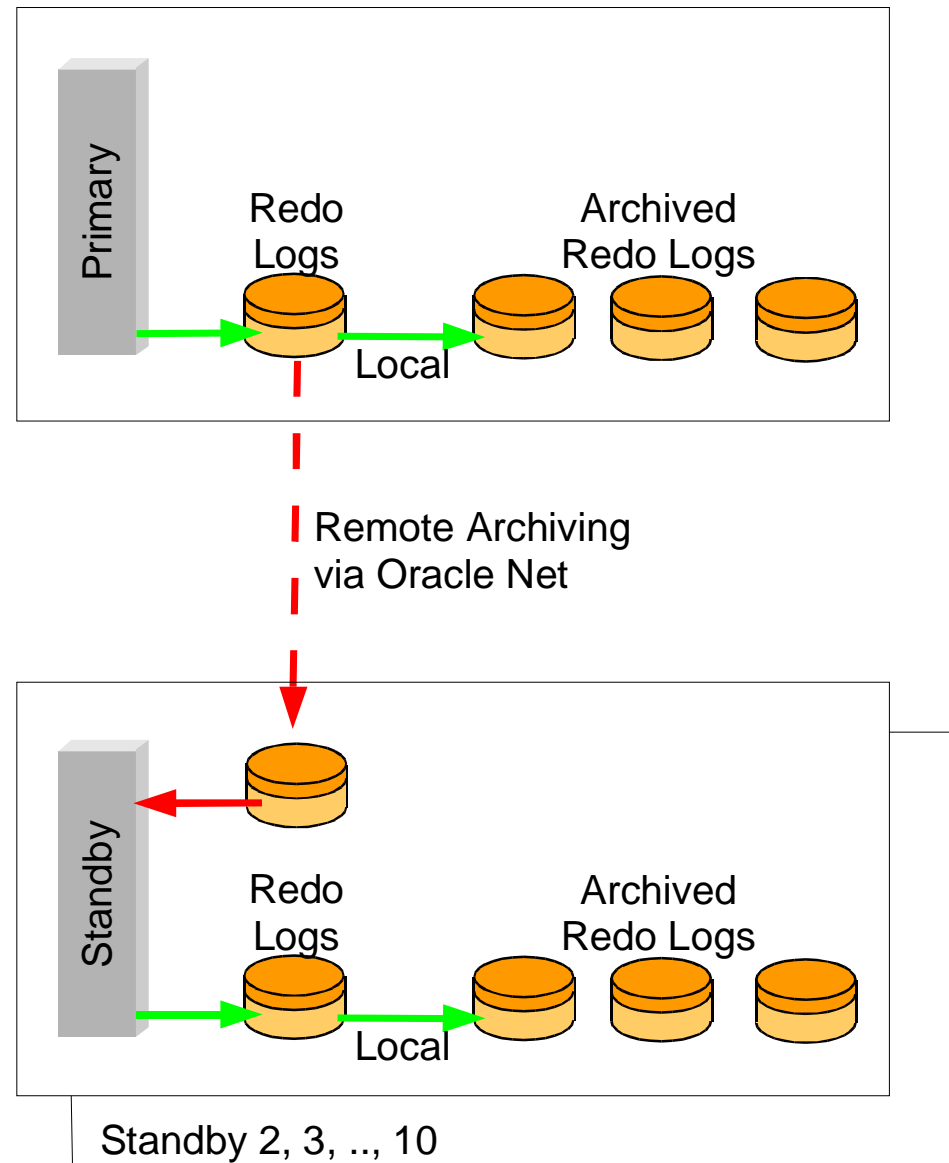
Protect Mode

- Gap Detection / Gap Resolution

- Nachteile

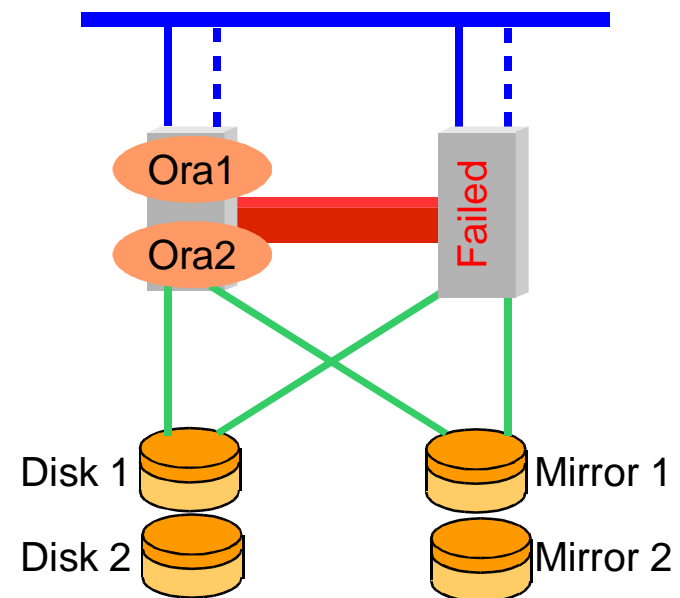
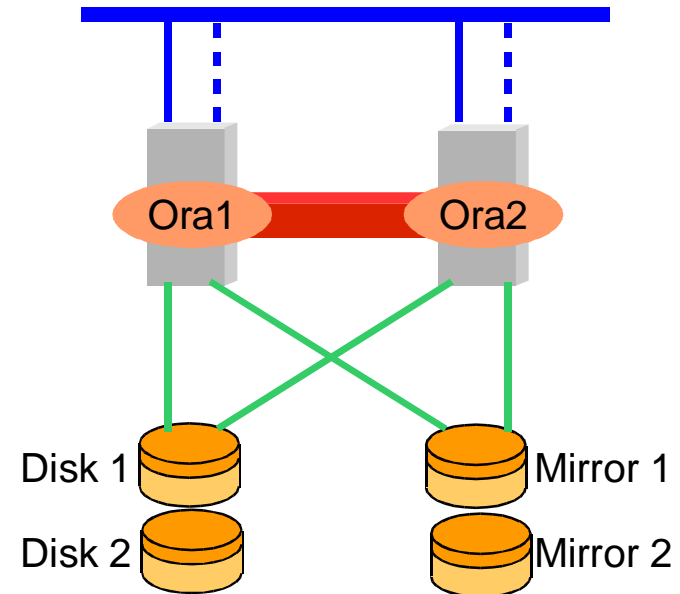
- Kein automatisches Failover
- Hoher administrativer Aufwand
- Gracefull Database Failover ist zeitintensiv

- Datenverlust, wenn nicht im Guaranteed Protect Mode
- Je höher der Datenschutz, desto schlechter die Performance



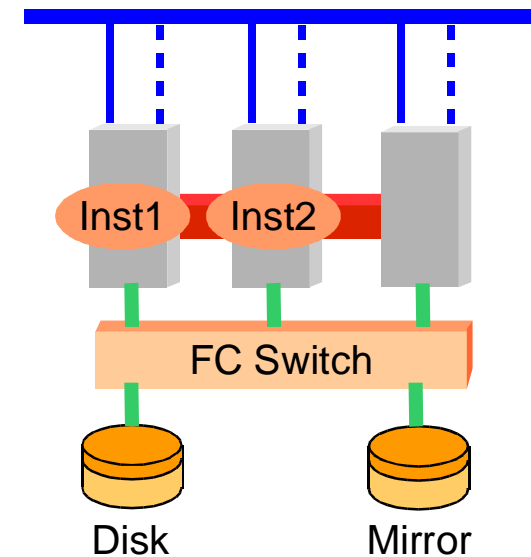
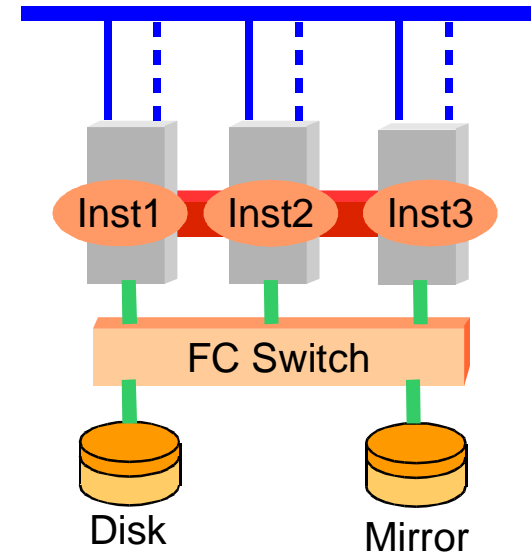
Comparisson HA: HA Oracle

- Vorteile
 - automatische Übernahme im Fehlerfall
 - schnelle Fehlererkennung
 - kein Datenverlust
 - Aktiv / Aktiv Konfiguration
 - geringer administrativer Aufwand
- Nachteile
 - keine horizontale Skalierung
 - Im Fehlerfall müssen sich alle User wieder verbinden
 - Standort maximal 10km getrennt (Campus Cluster)



Comparisson HA: Oracle 9i RAC

- Vorteile
 - Skaliert über alle Knoten
 - Hohe Performance durch Cache Fusion
 - Im Fehlerfall sind nur wenige Benutzer (hier 33%) betroffen
 - Automatische Benutzerübernahme
 - Sehr schnelle Übernahme (<1min)
 - Applikationen werden fortgeführt (TAF)
- Nachteile
 - Hohe Kosten
 - Standort maximal 10km getrennt



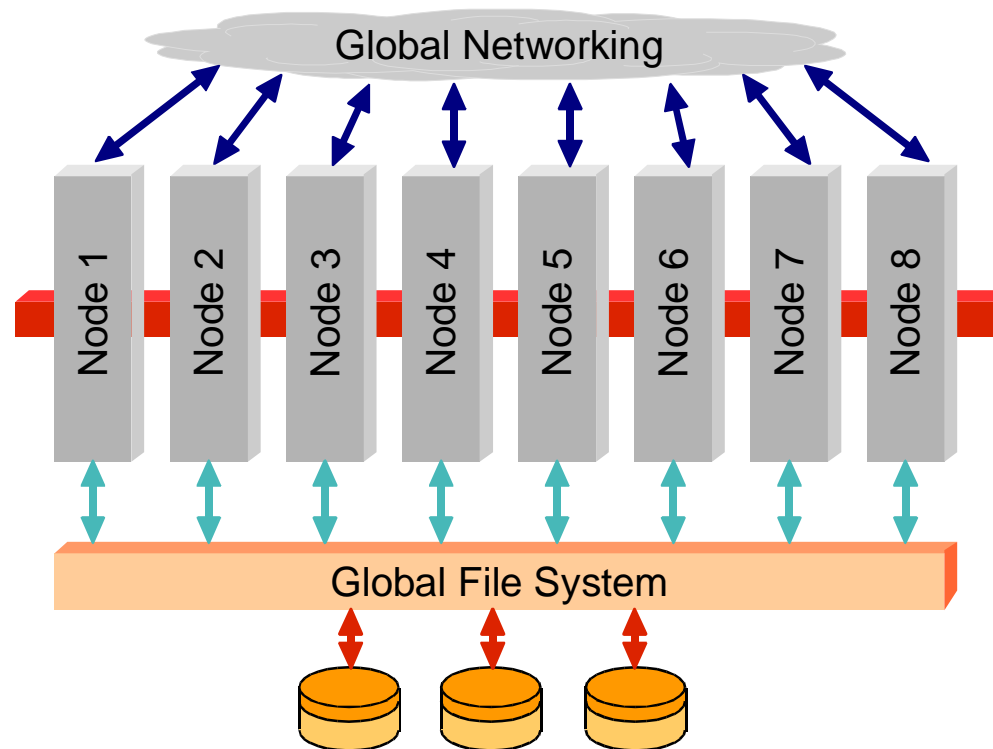
Comparisson HA: Bewertungstabelle

<u>Criteria</u>	Data Guard	HA Oracle	RAC
Performance	Low	Normal	High
Availability	Normal	High	Very high
Failover	8-10min	2-3min	< 1min
Data Loss	High	Low	Very low
Manageability	Complex	Easy	Easy
Cost	Low	Normal	High

- **Comparisson HA**
 - HA Ziele, DataGuard, HA Oracle, RAC
- **Sun Cluster 3.0 Key Features**
 - Availability, Manageability, Maintenance
- **OPS+SC 2.2 / RAC+SC3.0**
 - Cache Fusion, TAF, Load Balancing
- **Konfiguration**
 - „best Practice“
- **Sun + Oracle zertifizierte Komplettlösungen**
- **Projekt: WireCard**
- **Live Demo**

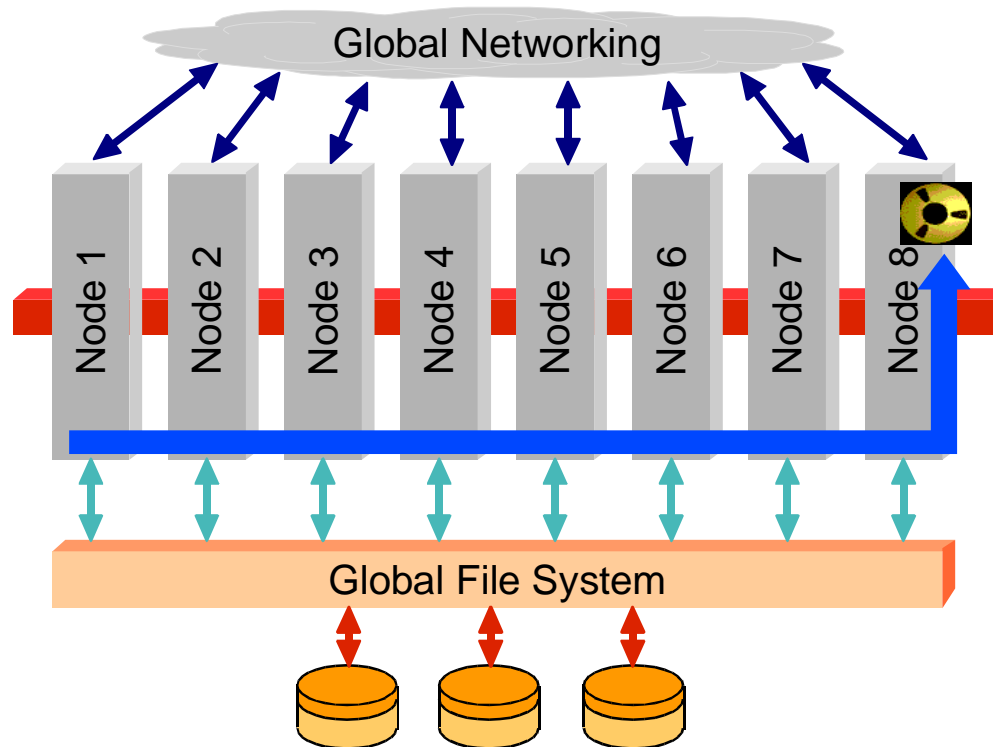
Sun Cluster 3.0: Overview

- Integriert in Solaris 8 Kernel
- Bis zu 8 Knoten
- Failover und Scalable Data Services
- clusterweite Filesysteme, Devices und Interfaces
- Builtin Load Balancing



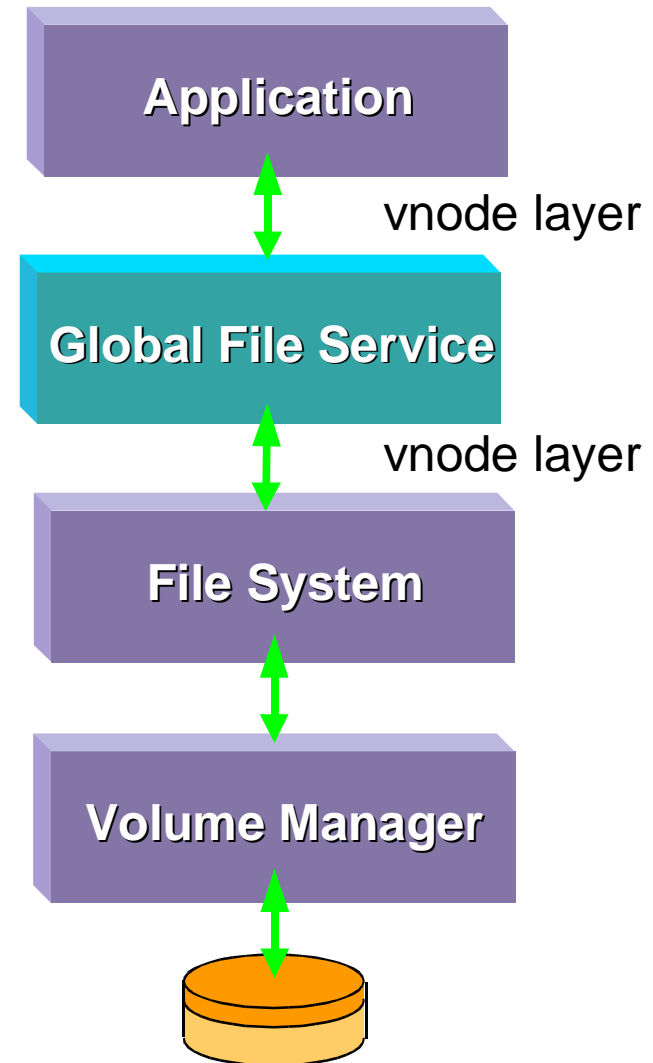
Sun Cluster 3.0: Global Devices

- Global Devices sind überall verfügbar
- Zugriff erfolgt über Global Namespace
 - überall gleiche Pfade: /dev/global
- dual-hosted Devices sind für Applikationen kontinuierlich verfügbar, selbst wenn der Primary Path ausfällt



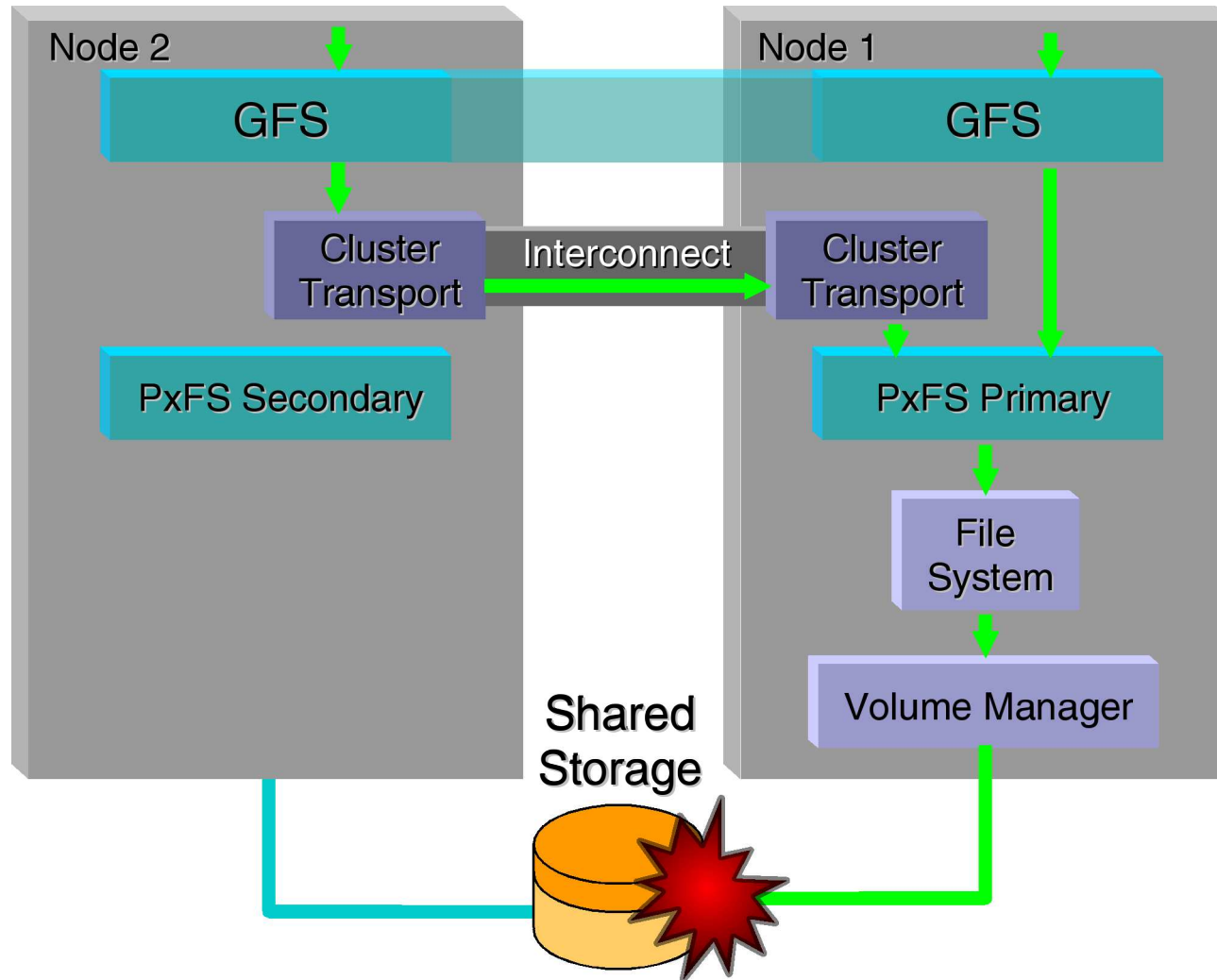
Sun Cluster 3.0: Global File Service

- **Kein neues FS**
- Cluster File System
 - hochverfügbar, distributed, cache-coherent
- Kernel-basierte Client/Server Architektur
 - PxFs Mechanism basiert auf *vnode* interface
- unabhängig vom FS Type und Volume Manager
- Failover/Switchover transparent zur Application und zum User
- Global Mount von einem Knoten für alle Knoten
 - mount -o global
 - /etc/vfstab



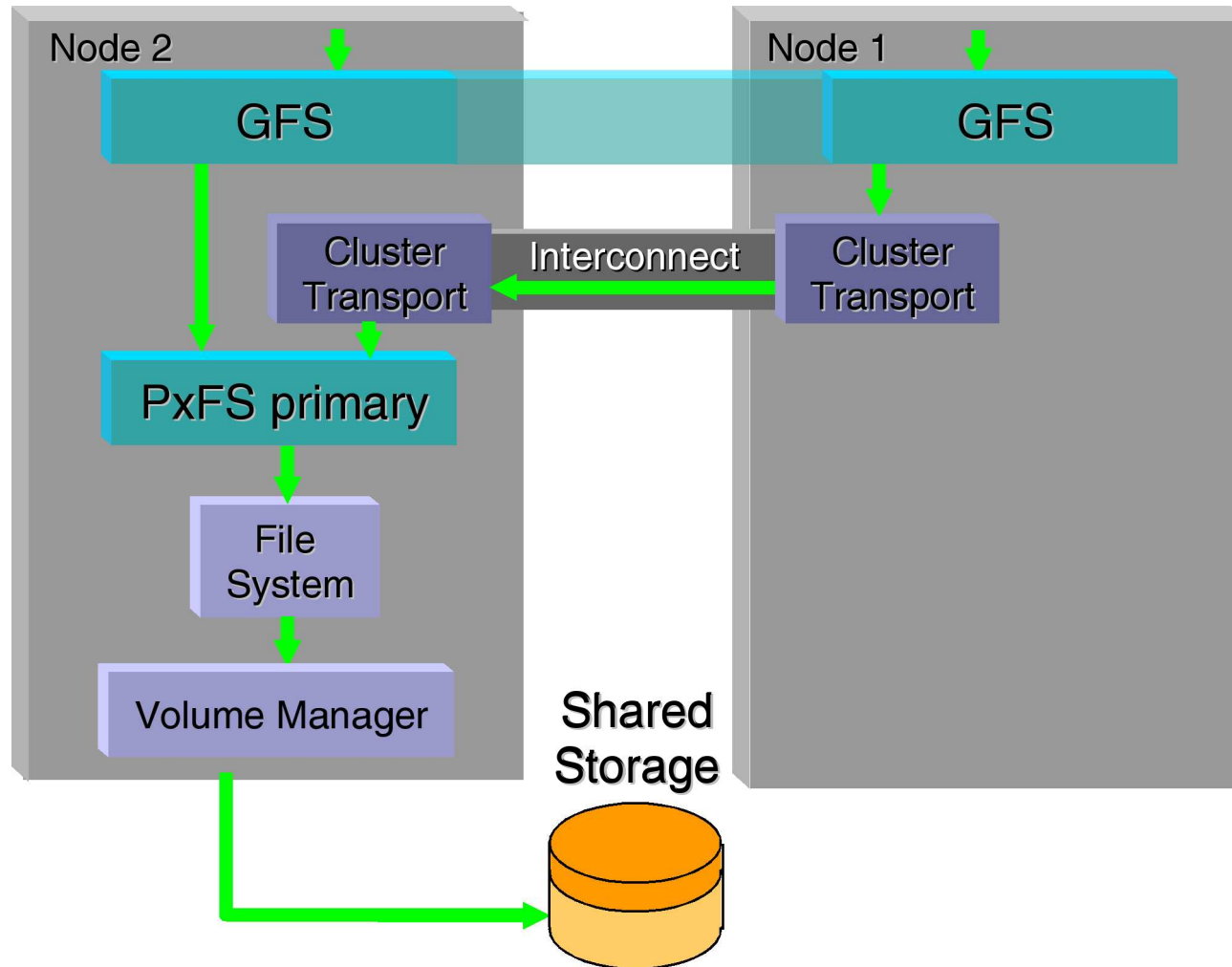
Sun Cluster 3.0: Global File Service

- Wie das Cluster File System funktioniert !!!



Sun Cluster 3.0: Global File Service

- Wie das Cluster File System funktioniert !!!



Sun Cluster 3.0: Vorteile für RAC

- GFS für Oracle Binaries und Config Files auf allen Knoten permanent verfügbar
- Vereinfachte Administration
- Bootet automatisch im Cluster Mode
- Update und Patches können im Non-Cluster-Mode Knoten für Knoten installiert werden, ohne den gesamten Cluster zu stoppen
- Dynamic Reconfiguration

Aber: Es sind max. 4 Knoten mit RAC auf Sun Cluster zertifiziert

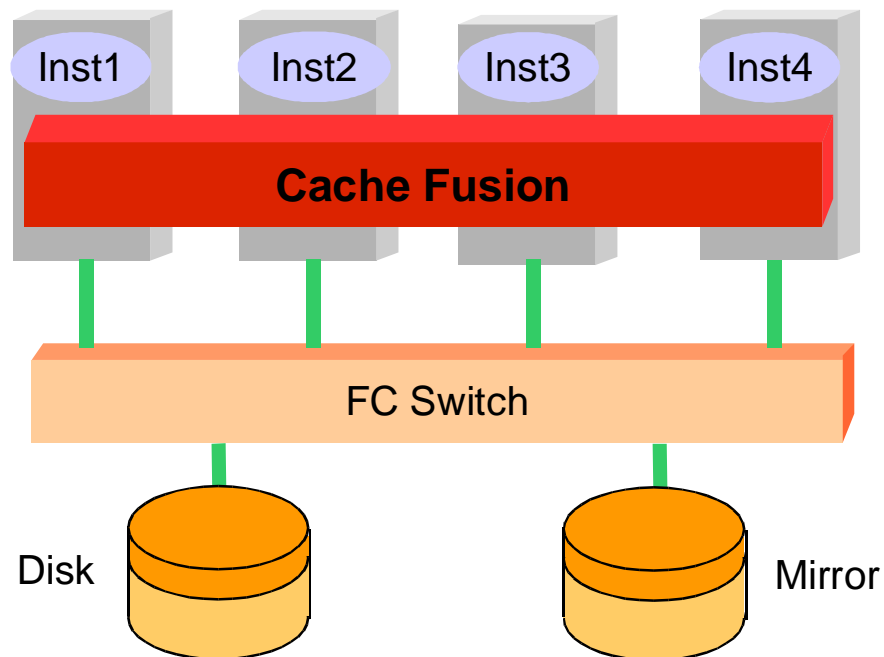
- **Comparisson HA**
 - HA Ziele, DataGuard, HA Oracle, RAC
- **Sun Cluster 3.0 Key Features**
 - Availability, Manageability, Maintenance
- **OPS+SC 2.2 / RAC+SC3.0**
 - Cache Fusion, TAF, Load Balancing
- **Konfiguration**
 - „best Practice“
- **Sun + Oracle zertifizierte Komplettlösungen**
- **Projekt: WireCard**
- **Live Demo**

OPS+SC 2.0 / RAC+SC 3.0: SC 2.2 -> SC 3.0

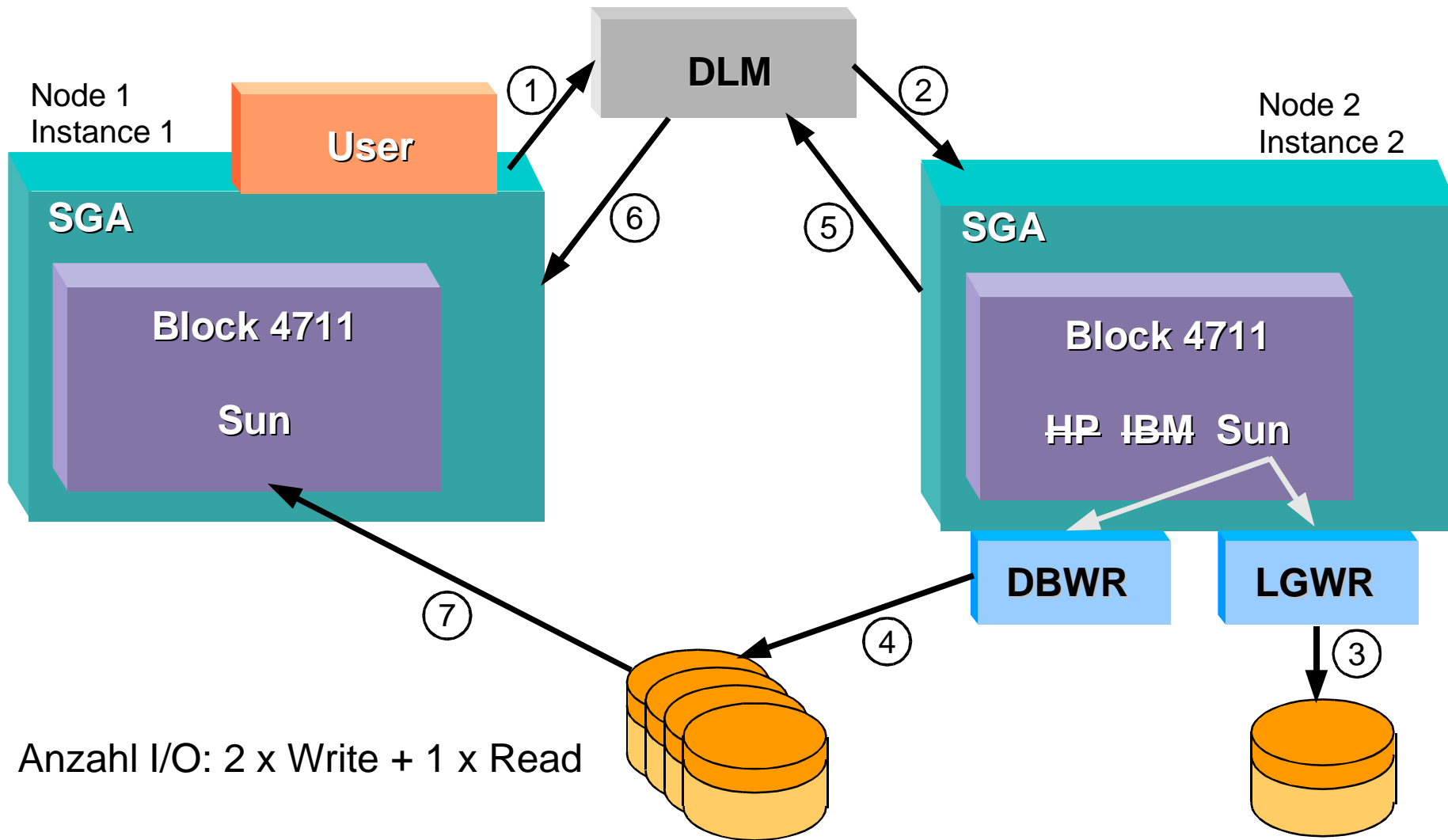
- Global Devices
 - kontinuierliche Verfügbarkeit für File System Services
 - kontinuierlicher Verfügbarkeit für Network Services
- Interconnects
 - bis zu 6 Interconnects
 - Load Balancing
- Booting in Cluster Mode
 - Cluster Framework in Kernel
 - Fast Failure Detection + Recovery
- Resource Group
 - kein „logischer Host“
 - *Manage the Service not the Server* ⇒ Service Components entkoppelt

OPS+SC 2.0 / RAC+SC 3.0: Cache Fusion

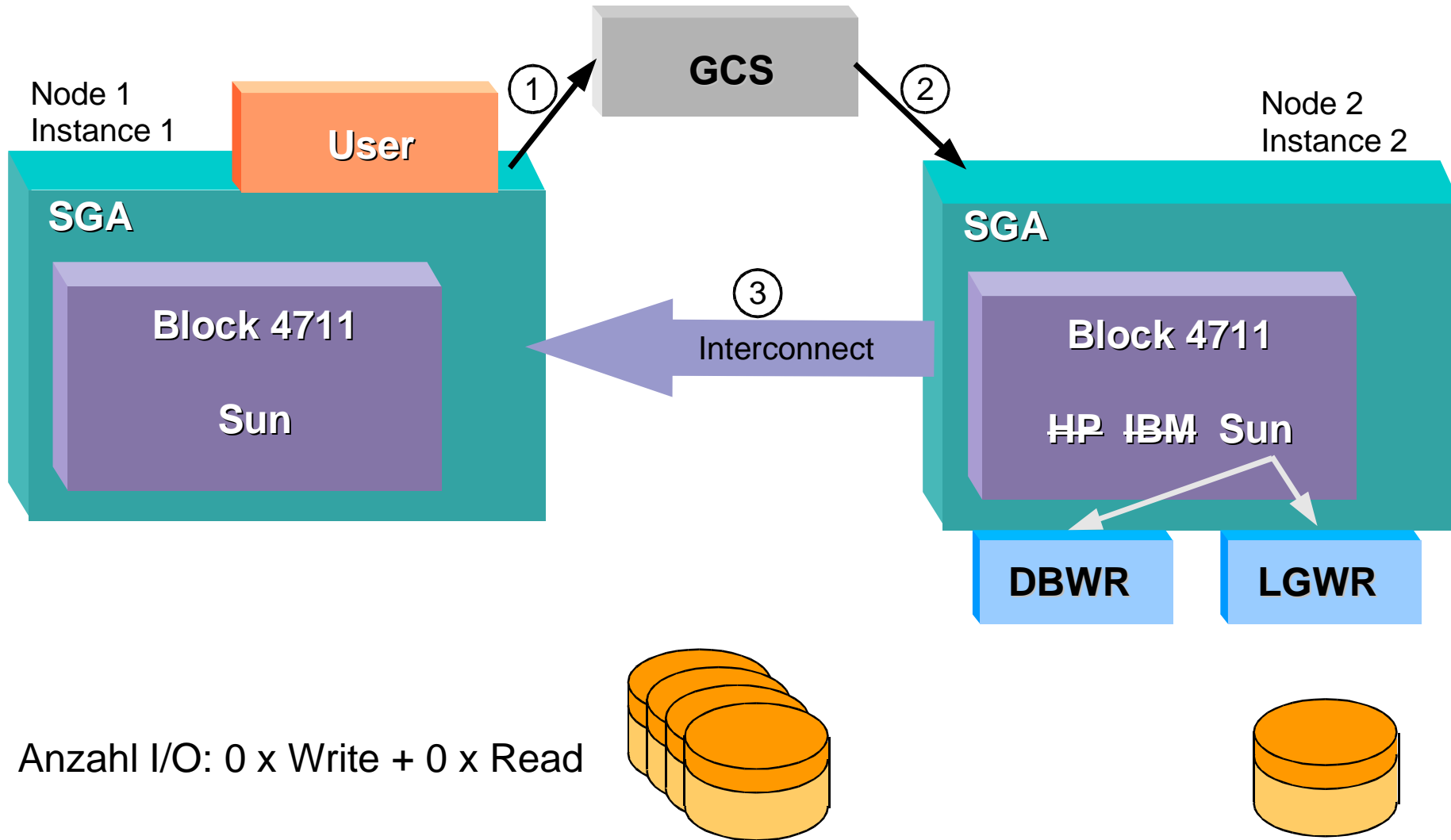
- Optimierung des *Cache Coherency* Protokolls
 - Reduzierung der Anzahl Messages
 - Vermeidung von I/O
- Eindruck einer global SGA (Shared Global Area)



OPS+SC 2.0 / RAC+SC 3.0: OPS Block Ping

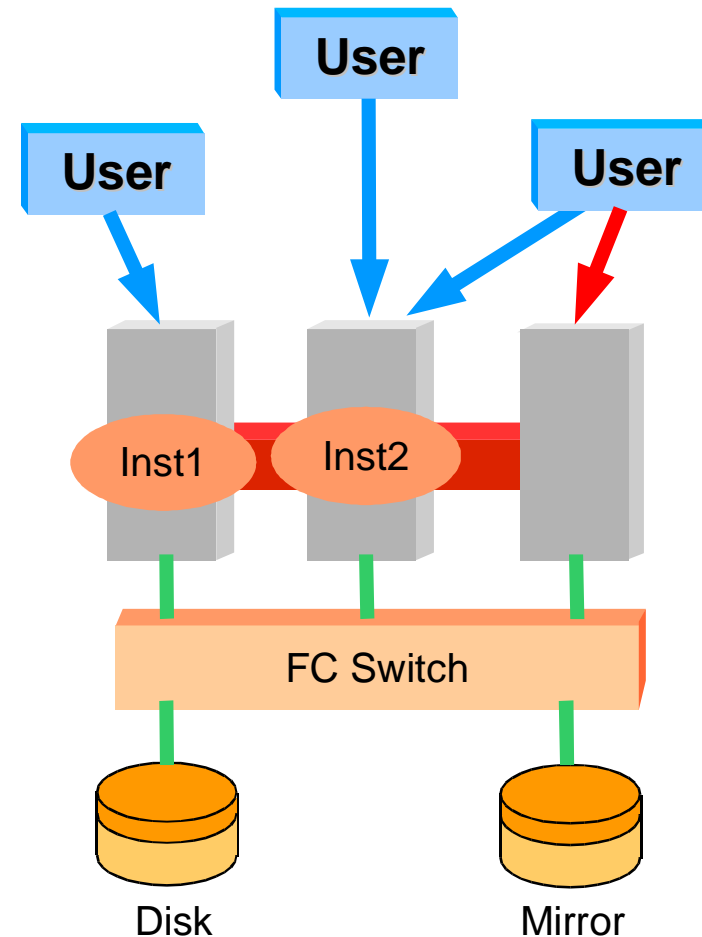


OPS+SC 2.0 / RAC+SC 3.0: RAC Block Shipping

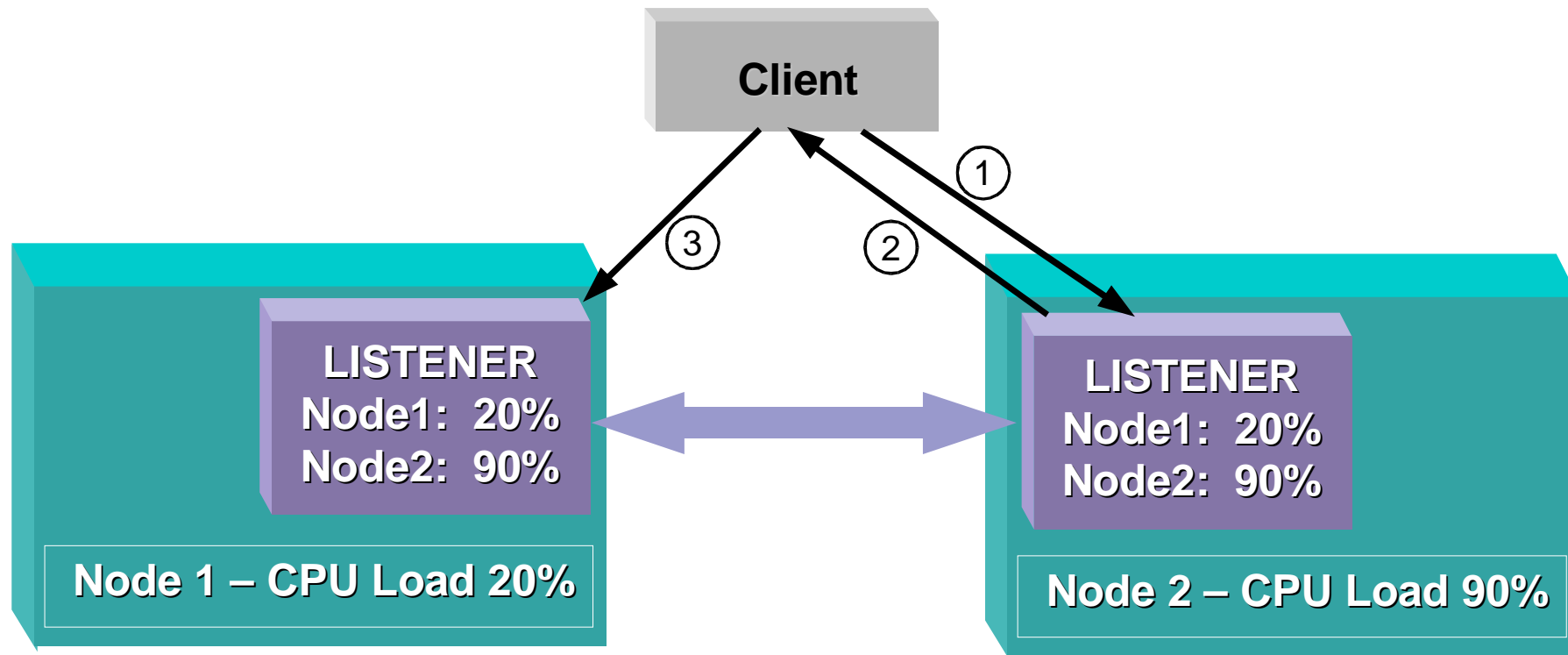


OPS+SC 2.0 / RAC+SC 3.0: Total Application Failover (TAF)

- Applikationen und Benutzer werden bei Systemausfall automatisch und transparent mit der nächsten laufenden Instanz verbunden
- Es sind „nur“ wenige Benutzer betroffen. Hier sind es 1/3 aller Benutzer
- Übernahme der Benutzer liegt bei weniger als 1 Minute
- Client pre-connect möglich, um große Anzahl an Benutzer schneller zu migrieren



OPS+SC 2.0 / RAC+SC 3.0: Connection Load Balancing



- Listener tauschen sich alle 30 Sekunden über CPU Load aus

Shared Server:

1. kleinster Node Load
2. kleinster Instance Load
3. kleinster Dispatcher Load

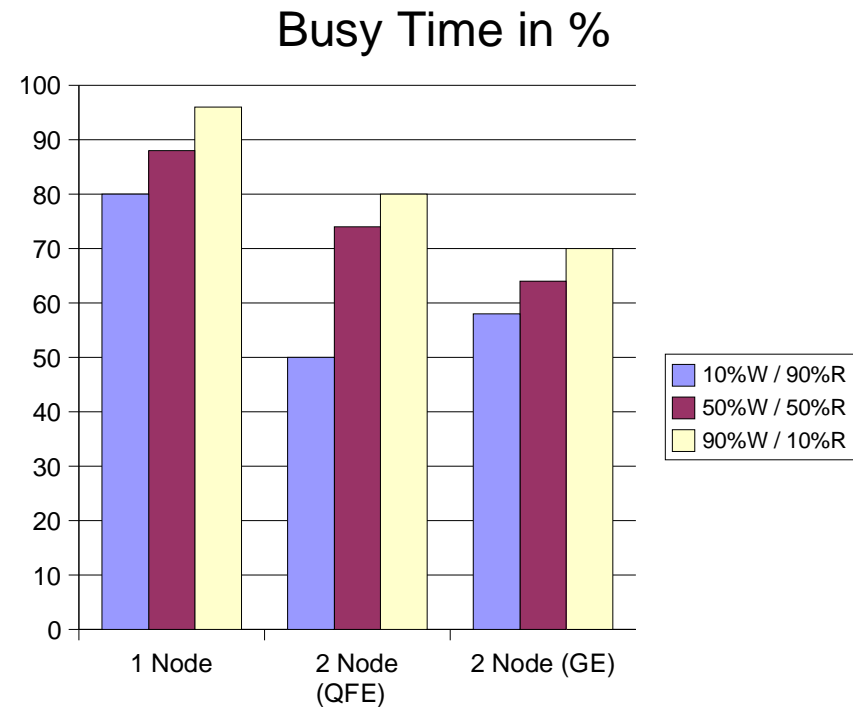
Dedicated Server:

1. kleinster Node Load
2. kleinster Instance Load

- **Comparisson HA**
 - HA Ziele, DataGuard, HA Oracle, RAC
- **Sun Cluster 3.0 Key Features**
 - Availability, Manageability, Maintenance
- **OPS+SC 2.2 / RAC+SC3.0**
 - Cache Fusion, TAF, Load Balancing
- **Konfiguration**
 - „best Practice“
- **Sun + Oracle zertifizierte Komplettlösungen**
- **Projekt: WireCard**
- **Live Demo**

Konfiguration : "best practise" - Skalierung

- Availability durch horizontale Skalierung
- Performance durch vertikale Skalierung
- Viele kleine Server produzieren Overhead
 - Systemload aller Server steigt
 - Erfahrung: 1 Server 80%, 2 Server 50%
 - > 20% Overhead bei Lastverteilung
- Empfehlung:
 - ++ wenige große Server
 - viele kleine Server



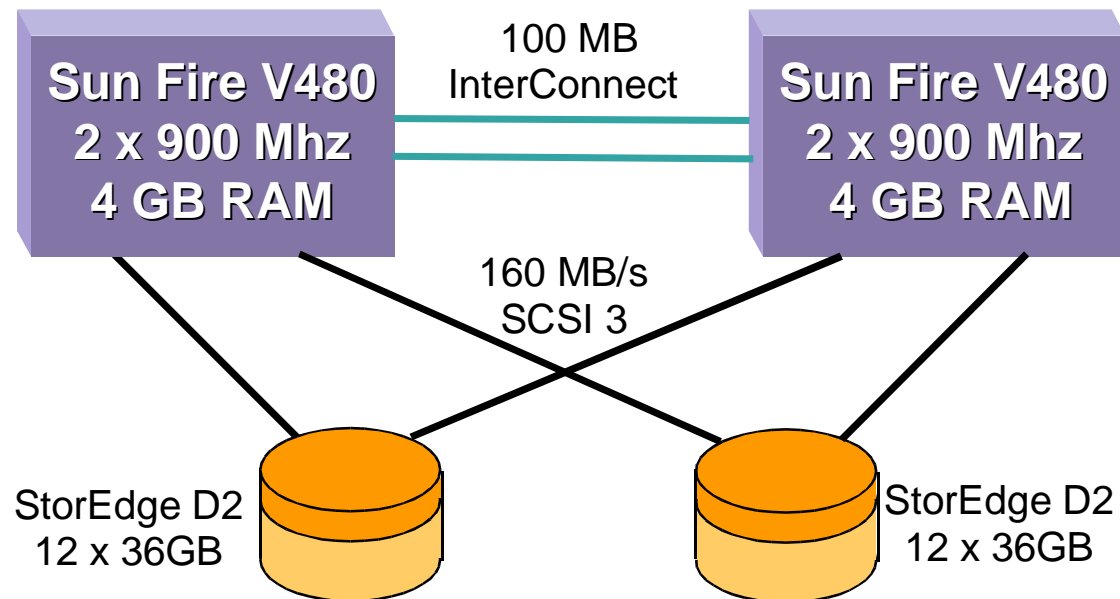
Konfiguration : “best practise“ - Transparent Application Failover

- Scenario 1
 - shutdown abort von Instance 1
 - DB Connect geht automatisch und transparent zum nächsten Knoten
 - Failover < 1min
- Scenario 2
 - Power off von Knoten 2
 - DB Connect bzw. Client bekommt nichts vom Ausfall mit
 - TCP/IP Client Timeout Parameter laufen ab, erst jetzt ist der Ausfall bekannt
 - Failover findet nach 11min statt
- Lösung zu Scenario 2
 - Threshold 1: tcp_ip_abort_interval (default: 480000 ms)
 - Threshold 2: tcp_ip_abort_cinterval (default: 180000 ms)

- **Comparisson HA**
 - HA Ziele, DataGuard, HA Oracle, RAC
- **Sun Cluster 3.0 Key Features**
 - Availability, Manageability, Maintenance
- **OPS+SC 2.2 / RAC+SC3.0**
 - Cache Fusion, TAF, Load Balancing
- **Konfiguration**
 - „best Practice“
- **Sun + Oracle zertifizierte Komplettlösungen**
- **Projekt: WireCard**
- **Live Demo**

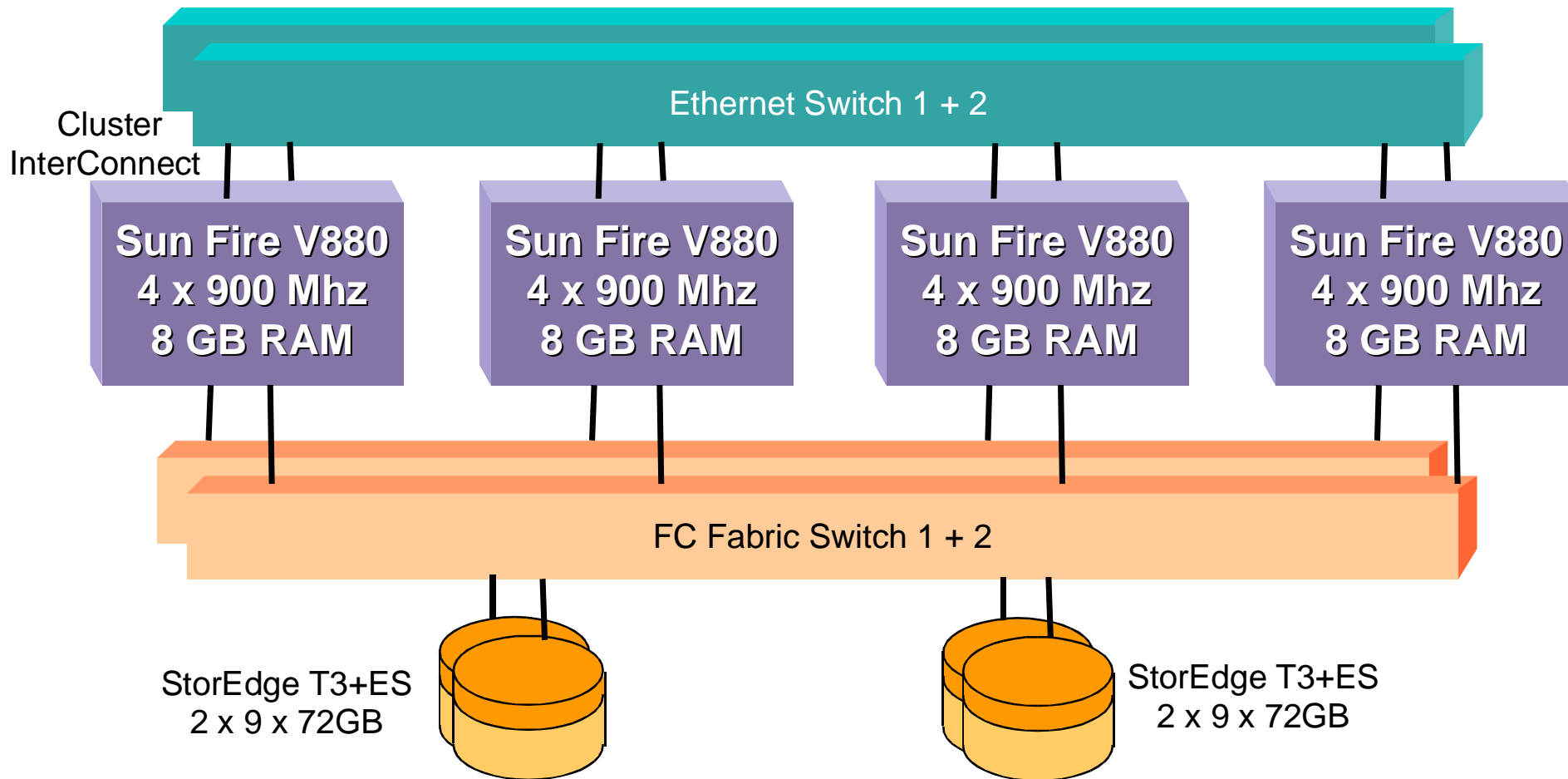
Sun + Oracle zertifizierte Komplettlösungen: 2 Node Cluster

- Lizenzierung
 - 2 x Sun Cluster
 - 2 x VxVM
 - 2 x CVM
 - 2 x Oracle Enterprise
 - 2 x Oracle RAC



Sun + Oracle zertifizierte Komplettlösungen: 4 Node Cluster

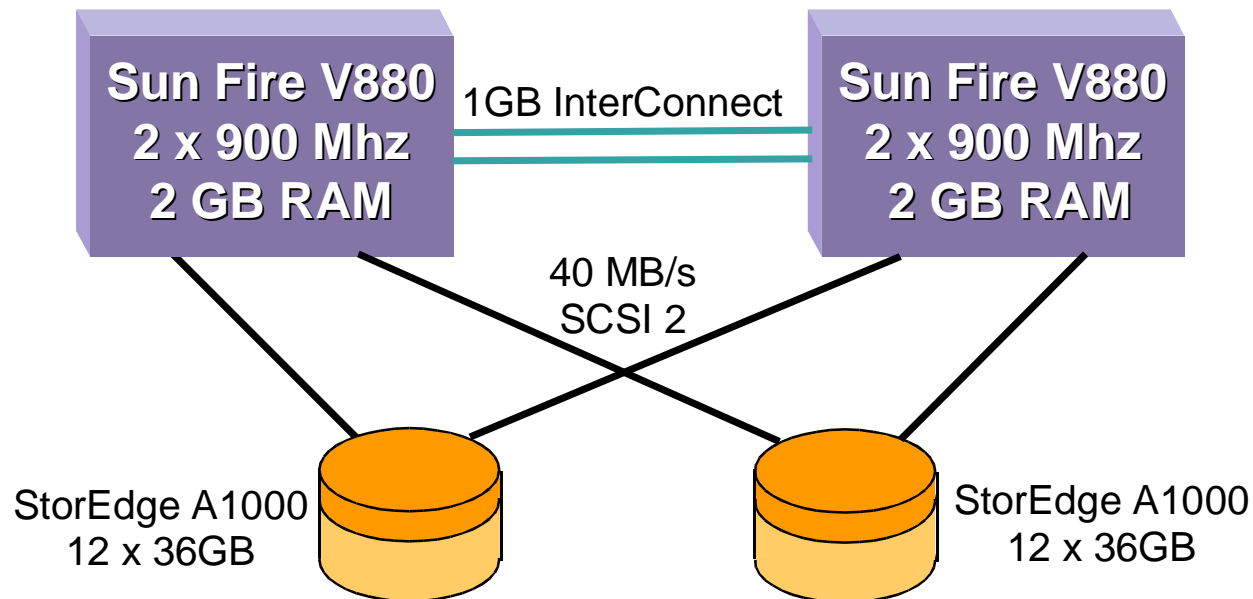
- Lizenzierung
 - 4 x Sun Cluster, 4 x VxVM, 4 x CVM
 - 4 x Oracle Enterprise, 4 x Oracle RAC



- **Comparisson HA**
 - HA Ziele, DataGuard, HA Oracle, RAC
- **Sun Cluster 3.0 Key Features**
 - Availability, Manageability, Maintenance
- **OPS+SC 2.2 / RAC+SC3.0**
 - Cache Fusion, TAF, Load Balancing
- **Konfiguration**
 - „best Practice“
- **Sun + Oracle zertifizierte Komplettlösungen**
- **Projekt: WireCard**
- **Live Demo**

Projekt: WireCard : 2 Node Cluster

- Warum 2 Node RAC ?
 - „Wir können uns keinen Datenverlust erlauben. Die Ausfallzeit muß so klein wie möglich sein !“
- Warum Sun Fire V880 Server ?
 - „Die Wachstumsrate der User war eine Unbekannte. Vertikale Skalierung war die richtige und günstigere Lösung zum Auffangen der Userlast.“
- Warum als Disk Array A1000er ?
 - „Wir wollten zu Beginn die vorhandene HW nutzen.“



- **Comparisson HA**
 - HA Ziele, DataGuard, HA Oracle, RAC
- **Sun Cluster 3.0 Key Features**
 - Availability, Manageability, Maintenance
- **OPS+SC 2.2 / RAC+SC3.0**
 - Cache Fusion, TAF, Load Balancing
- **Konfiguration**
 - „best Practice“
- **Sun + Oracle zertifizierte Komplettlösungen**
- **Projekt: WireCard**
- **Live Demo**

Live Demo: Oracle 9i RAC

- Stand: Sun – best Systeme