Herzlich Willkommen

zum Workshop

Vergleich AFS und GFS
Konfiguration eines GFS Clusters

Referent: Dipl. Informatiker Mark Grimme

ATIX GmbH

E-mail: grimme@atix.de

ATIX

Inhalt theoretischer Teil

- ATIX GmbH
  - Philosophie
  - Leistungen
- Grundlagen Clustering
- Cluster Filesysteme
  - GFS
- Shared Root Cluster
  - Shared Storage Cluster
  - Diskless Shared Root Cluster
  - Aufbau
  - Implementierung
  - Anwendungen
- Beispiel

ATIX

Inhalt praktischer Teil (1)

- RedHat Cluster Suite
  - Cluster Volume Manager
  - CMAN
  - Fencing
  - Lock Manager
    - GULM
    - DLM
  - Resource Manager
  - Cluster Configuration
    - cluster.conf
    - ccsd

ATIX

Inhalt praktischer Teil (2)

- Installation und Konfiguration
  - RPM Pakete (rpm, up2date, yum)
  - Netzwerkkonfiguration
- Cluster Layer
  - Erstellen der Cluster Konfiguration
    - GUI
    - Manuell
  - Starten des Cluster Layers
- Storage Layer
  - Cluster Volume Manager
- GFS
- Resource Manager

ATIX

ATIX GmbH

ATIX

Die ATIX Philosophie

ATIX

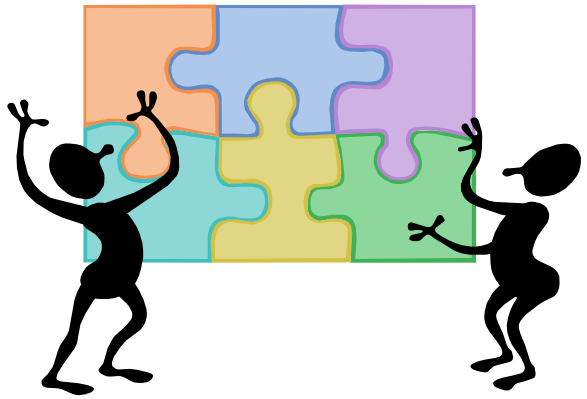
Advanced Technology for Individual Success

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

ATIX

ATIX Geschäftsfelder

- Beratung
  - Linux im Rechenzentrum (Cluster-Lösungen, HA)
  - Speichernetzwerke
  - Verfügbarkeitsanalyse / Katastrophenvorsorge
- Services
  - CC - Proof of Concept
  - Projektbegleitung / Integration
  - Workshops / Schulungen
- Produkte
  - com.oonics NAS-Serie (NASBox, NASHead, NASCluster, Grayhead)
  - com.oonics Cluster-Suite (Multi-Node Active/Active)
  - com.oonics Parallel Enterprise Plattform



ATIX

ATIX Partner

- Einige wenige ausgewählte strategische Partnerschaften
- Partner:
  - Hewlett Packard
    - Hardware für Speicherlösungen und IA Server
    - HP Linux Elite Program
  - Red Hat
    - Software für Cluster und Speicherlösungen
    - ATIX ist der einzige Red Hat Advanced Partner for Clustering und Storage/GFS in Europa
  - EMIC
    - Software für MySQL und Apache Cluster
  - Zeta (HP Enterprise Distributor)

ATIX

ATIX Referenzen

- Messe München, Dutschland
  - Web Server Infrastruktur
- Messe Leipzig, Deutschland
  - Active/Active NAS-Cluster mit ADS
- IP-Tech AG, Schweiz
  - Rechenzentrums-Infrastuktur für die verschiedensten Services
- T-Systems, Deutschland
  - Rechenzentrums-Infrastruktur für Software-Entwicklung
- Holländische Bank, Niederlande
  - Rechenzentrums-Infrastruktur für ein online Versicherungsportal
- Deutsche Großbank, Deutschland
  - Infrastruktur für eine Streaming Lösung für bis zu 4.000 Broker
- Internationaler Pharmakonzern, Niederlassungen weltweit
  - Beratung hinsichtlich Katastrophenvorsorge der IT Infrastruktur

ATIX

Zusammenfassung

- 10 Jahre Erfahrung im Bereich Clustering und Storage
- Spezialist für optimal skalierbare und hochverfügbare Enterprise-Infrastrukturen
  - Entwicklung und Implementierung maßgeschneiderter Storage-Lösungen basierend auf SAN/NAS
  - Umsetzung hochskalierbarer IT-Infrastrukturen für Unternehmensanwendungen vor allem auf Linux-Basis
- Wichtige strategische Partnerschaften mit Marktführern
- com.oonics Serie - das maßgeschneiderte Rundum-Paket für Enterprise-Infrastrukturen

ATIX

Grundlagen Clustering für IT Plattformen

ATIX

Clustertechnologien

„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

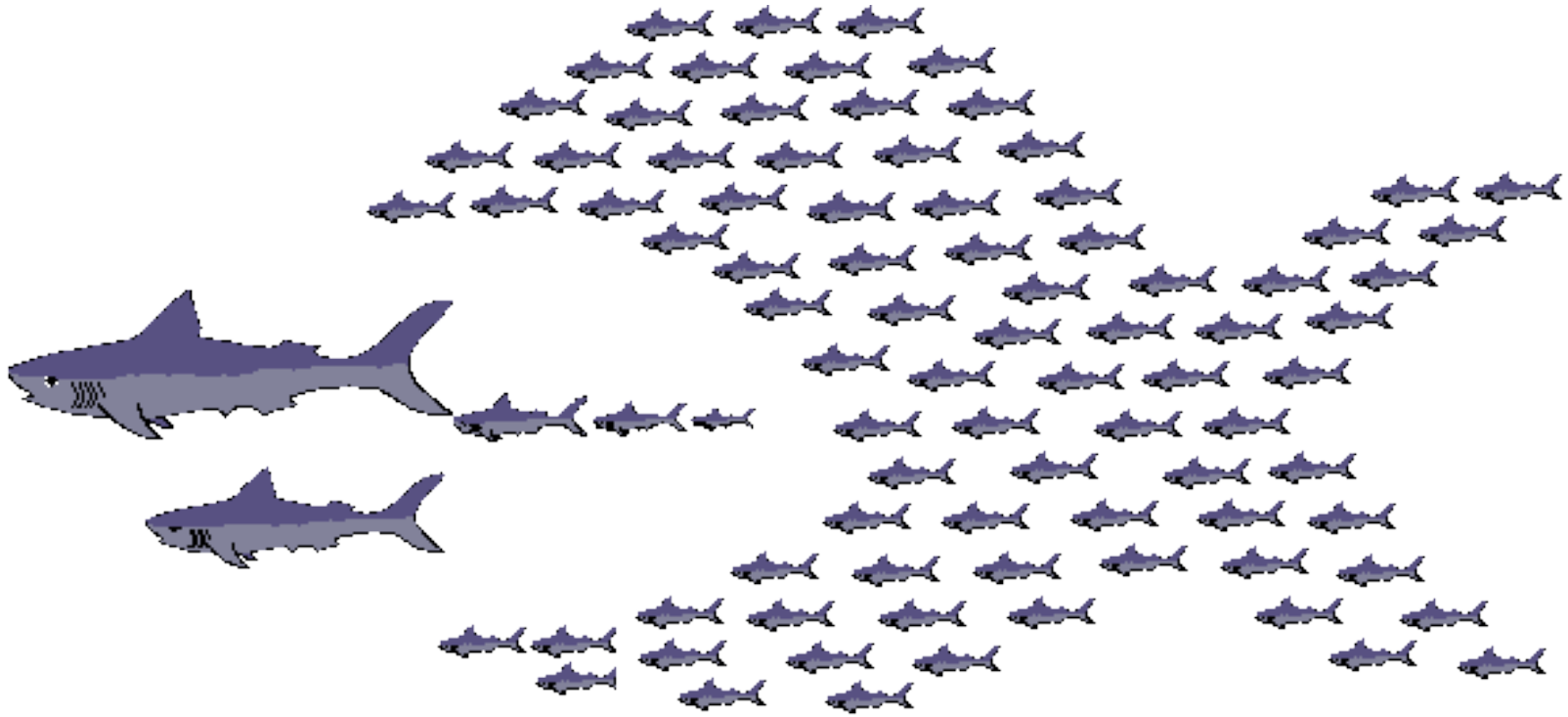
ATIX

Was ist ein Cluster?

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

ATIX

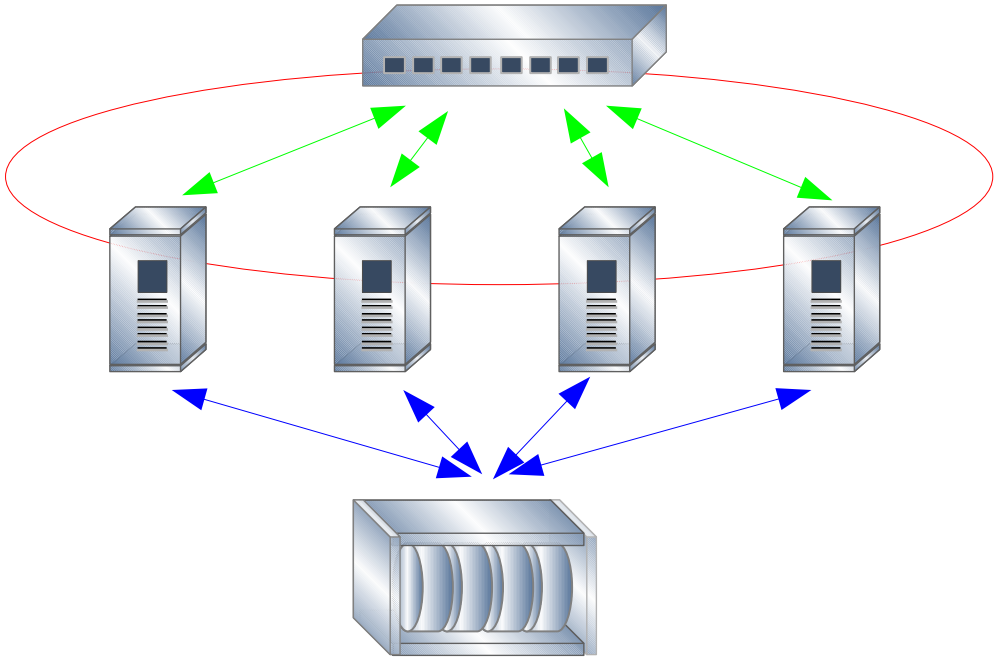
Rechencluster



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

ATIX

Network Loadbalancing Cluster



Loadbalancer

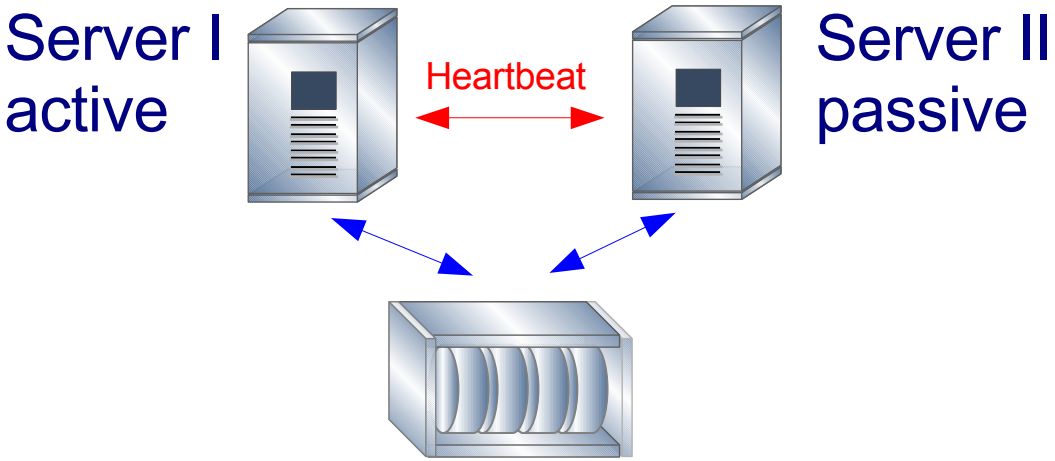
Heartbeat Netzwerk

Shared Storage

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

ATIX

HA Cluster Active/Passive

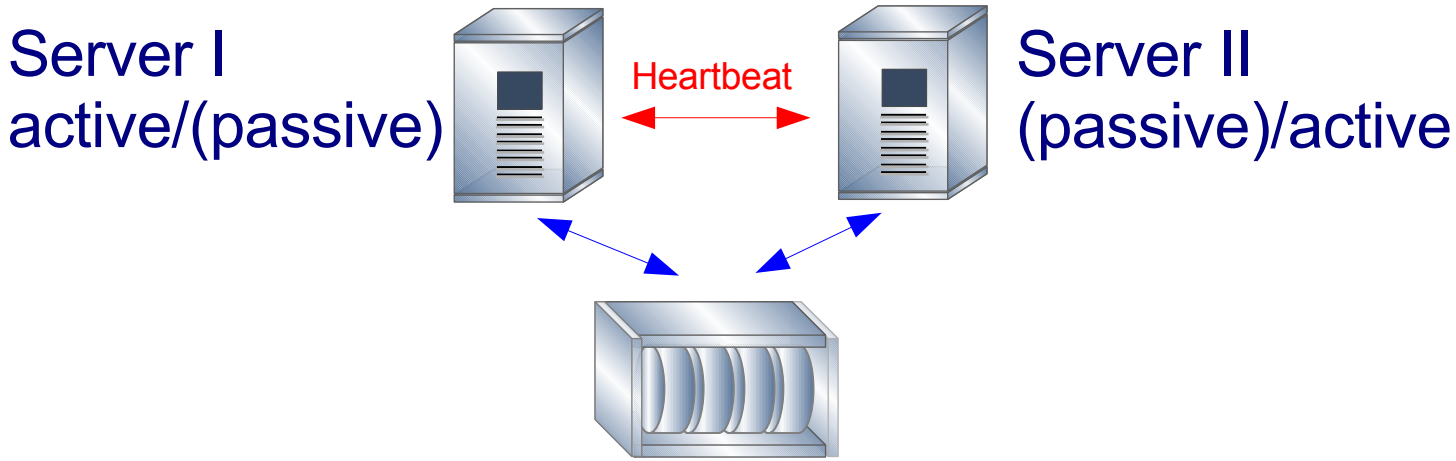


Konzept:

- 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

ATIX

HA Cluster Active/Active

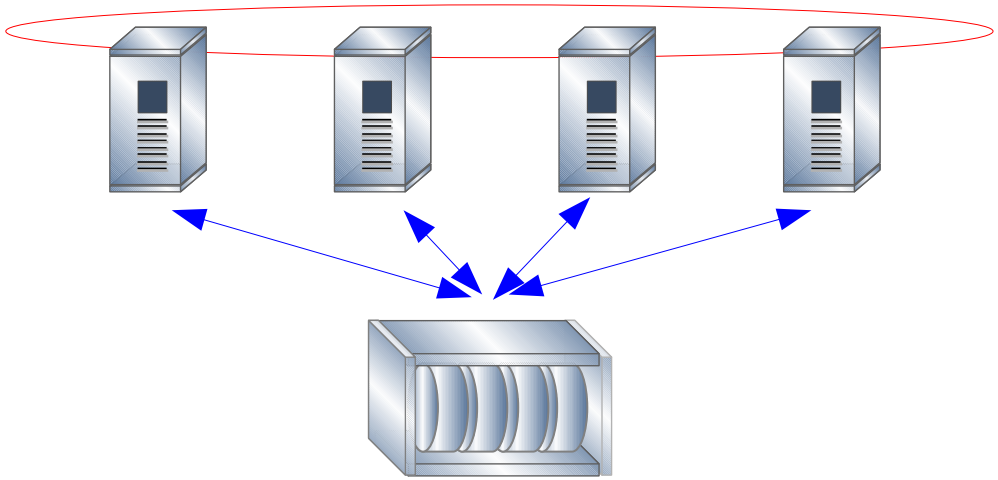


Konzept:

- Auf jeden Knoten laufen unterschiedliche Dienste
- Dadurch ist jeder Knoten aktiv und passiv zugleich
- Im Fehlerfall gibt es aber Performanceeinbussen
- An sonsten gilt das gleiche wie für Active/Passive Cluster

ATIX

HA Cluster N+1



Heartbeat Netzwerk

Shared Storage

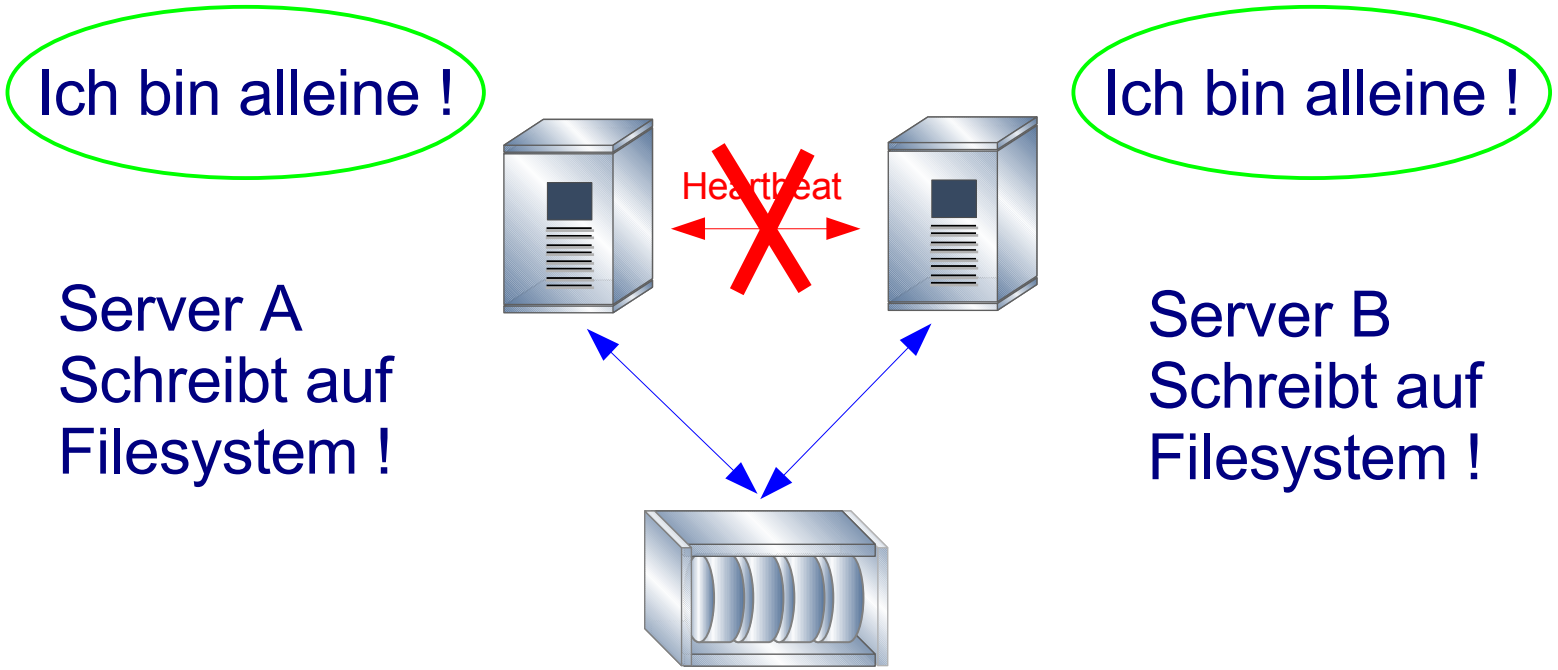
Konzept:

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

ATIX

HA Cluster: Split Brain Problem

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



Datenkorruption ?!?

HA Cluster: Split Brain Problem

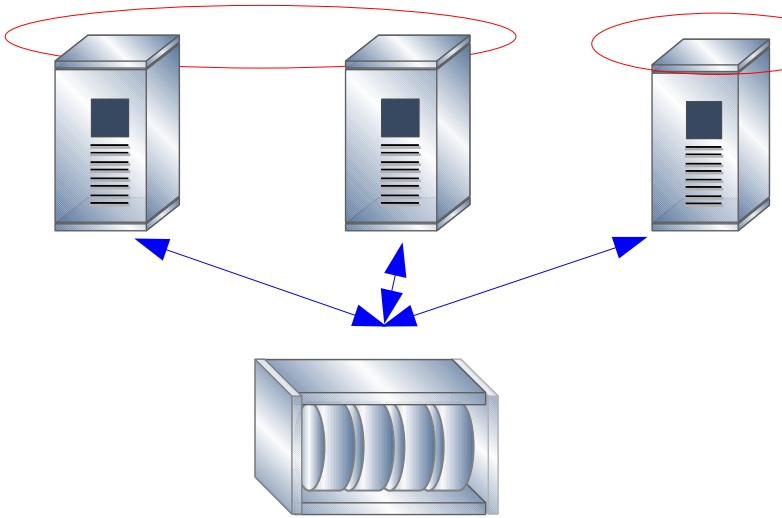
- Lösung1: SCSI reserve/release
  - SCSI Device kann nur einmal gemounted sein.
  - z.B. Windows Cluster
- Lösung2: Network Tie Breaker
- Lösung3: Quorum Requirement
  - Activer Teil muss mehr als die Hälfte der „votes“ bekommen.
  - Voraussetzung: Ungerade Anzahl an Votes
    - Unterschiedliche Anzahl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

ATIX

HA Cluster: Quorum Requirement

- Bsp: 3 Knoten, Quorum Requirement

We've got 2/3 votes
-> Quorum



I've got 1/3 votes
dying ...
reboot/fencing

ATIX

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)

ATIX

Zusammenfassung

- Clustering
  - Aufbau hochverfügbarer Enterprise Infrastrukturen
  - Hochskalierbare Parallel Cluster (Loadbalancing, HA)
  - High Performance Computing
  - Commodity Hardware
- ATIX
  - Spezialist für optimal skalierbare und hochverfügbare Enterprise-Infrastrukturen auf Linux Basis
  - Wichtige strategische Partnerschaften mit Marktführern, wie Red Hat und HP
  - com.oonics Serie - Enterprise-Infrastrukturen

ATIX

Cluster Filesysteme

- GFS -

- Diskless Shared Root Cluster

ATIX

GFS History (1)

- 1995: GFS Project an der University of Minnesota
  - Matthew O‘ Keefe
  - Shared Cluster Filesystem für enorme Supercomputing Datenmengen
- 1997: Gründung der Firma Sistina
  - Matthew O‘ Keefe
  - GFS Version 4.11 unter GPL
- 2001: GFS Version 4.2 unter Sistina Public License
  - Kein Source Code mehr verfügbar
  - Lizenzen kostenpflichtig
  - OpenGFS auf Basis GFS 4.11
- 2004: RedHat übernimmt Sistina
  - GFS Version 6.0 unter GPL
- 2005: GFS 6.1
  - Integration in RedHat Cluster Suite

ATIX

GFS History (2)

- GFS 2
  - Locking: DLOCK
    - SCSI Device Locks (Storage System)
  - Erste PoC Installationen
- GFS 4.11
  - Locking: DMEP
    - SCSI Memory export (Storage System oder memexpd)
  - Produktive Installationen
- GFS 5.1
  - Locking: DMEP (HA), GULM Server
- GFS 5.2/6.0
  - Locking: GULM Server (HA)
- GFS 6.1
  - Locking: GULM Server (HA), DLM

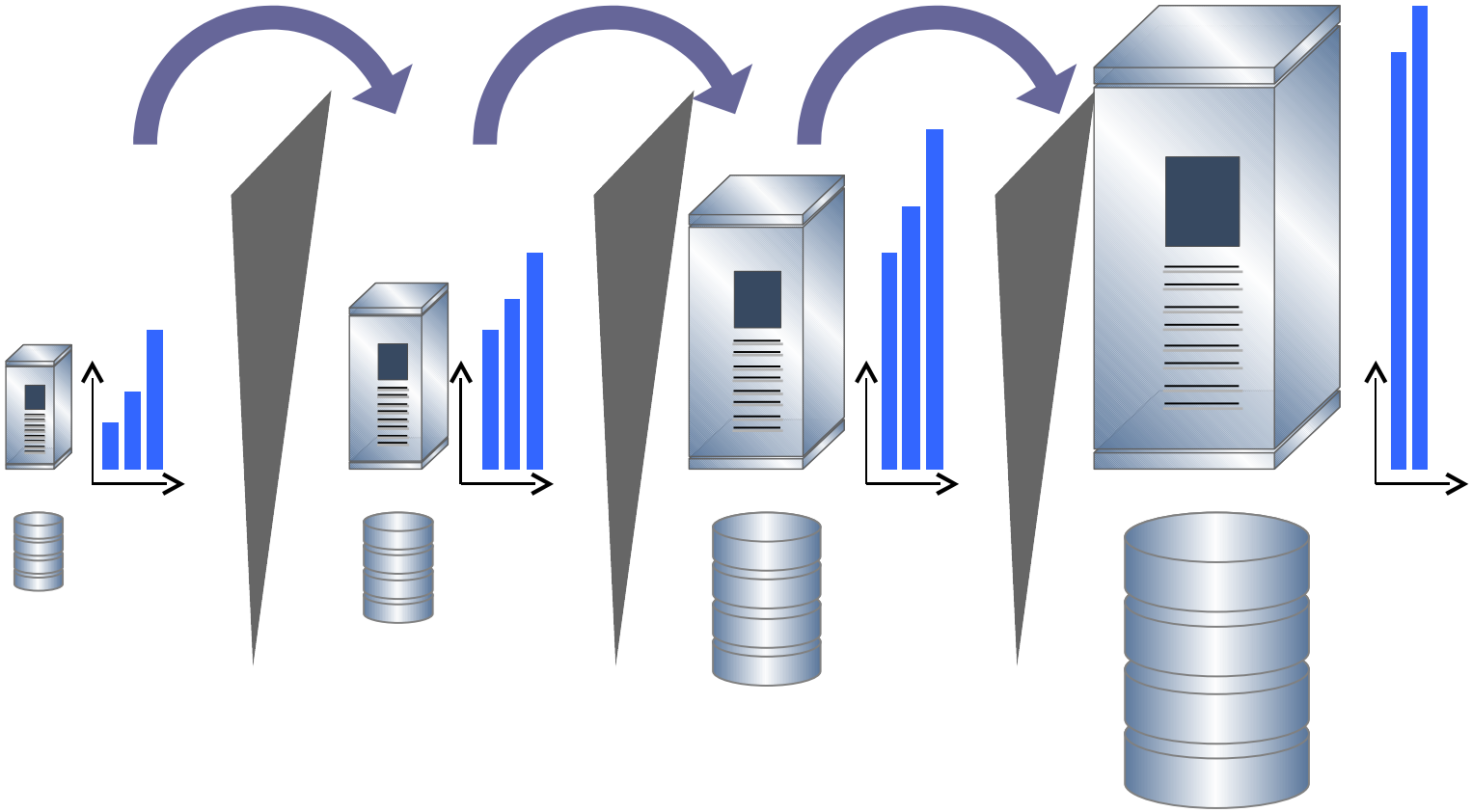
ATIX

Herausforderungen für IT Dienste

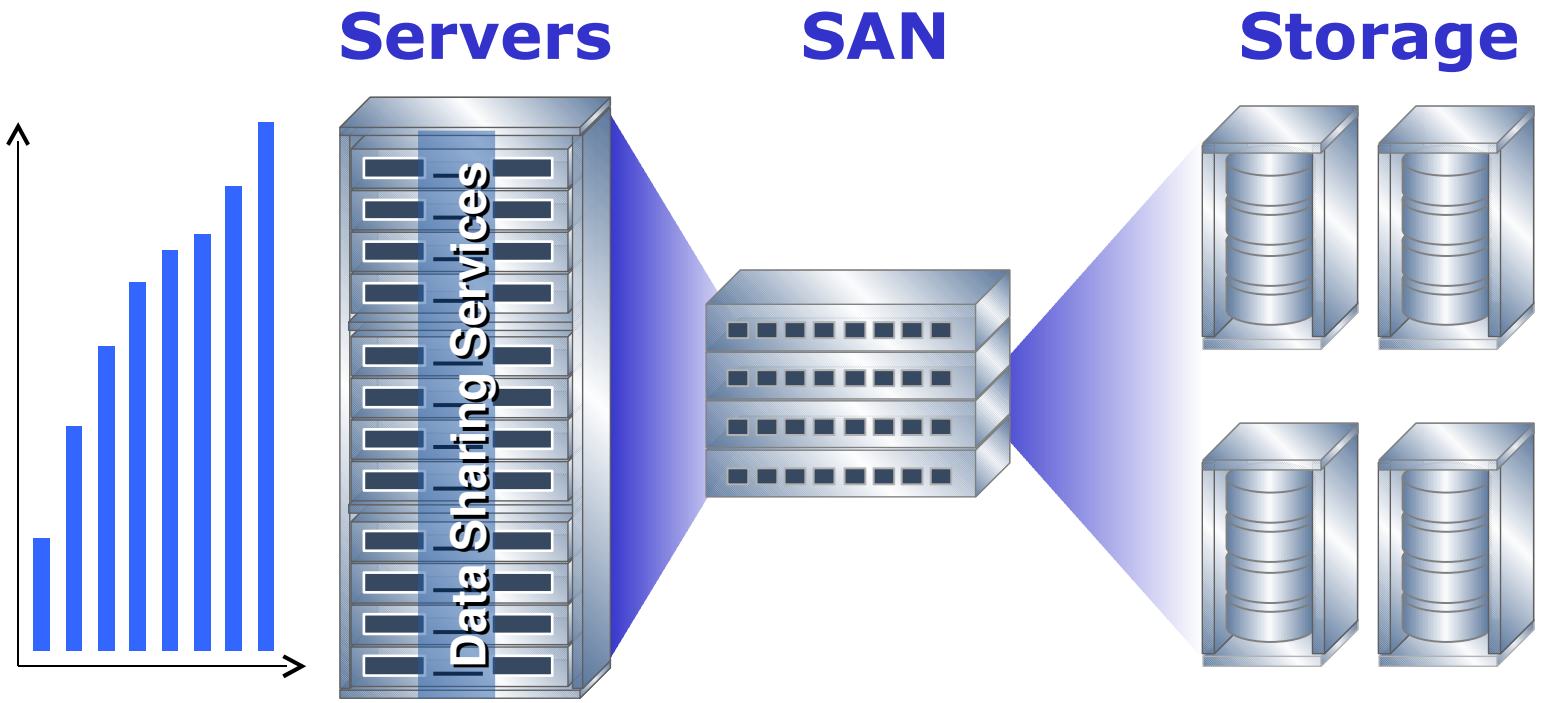
- Verfügbarkeit (24x7)
  - Kein Single Point of Failure
  - Minimum MTTR
  - Desaster Toleranz, Desaster Recovery
- Skalierbarkeit
  - Investitionsschutz
  - „Pay as you grow“ - optimale Leistung
  - Performance / Kapazität
- Optimales Preis – Leistungsverhältnis
- Einfaches Management
- Automatische Überwachung

ATIX

Skalierbarkeit



Skalierbarkeit Storage-Cluster

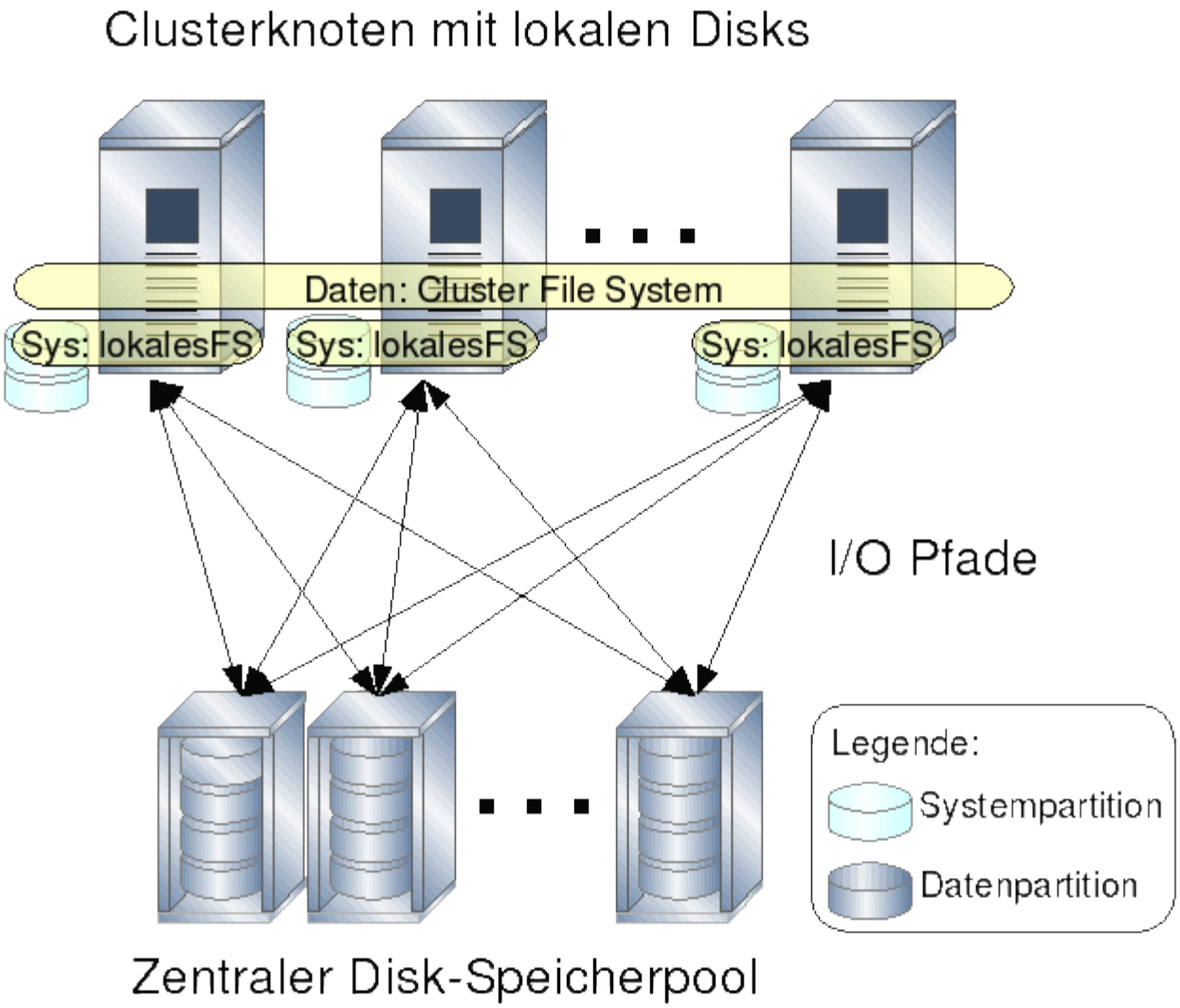


SAN + Linux + Data Sharing = Incremental Computing

- Incrementally and independently add compute, I/O and storage capacity
- Avoid architectural or application changes
- Lower cost of deployment and management

ATIX

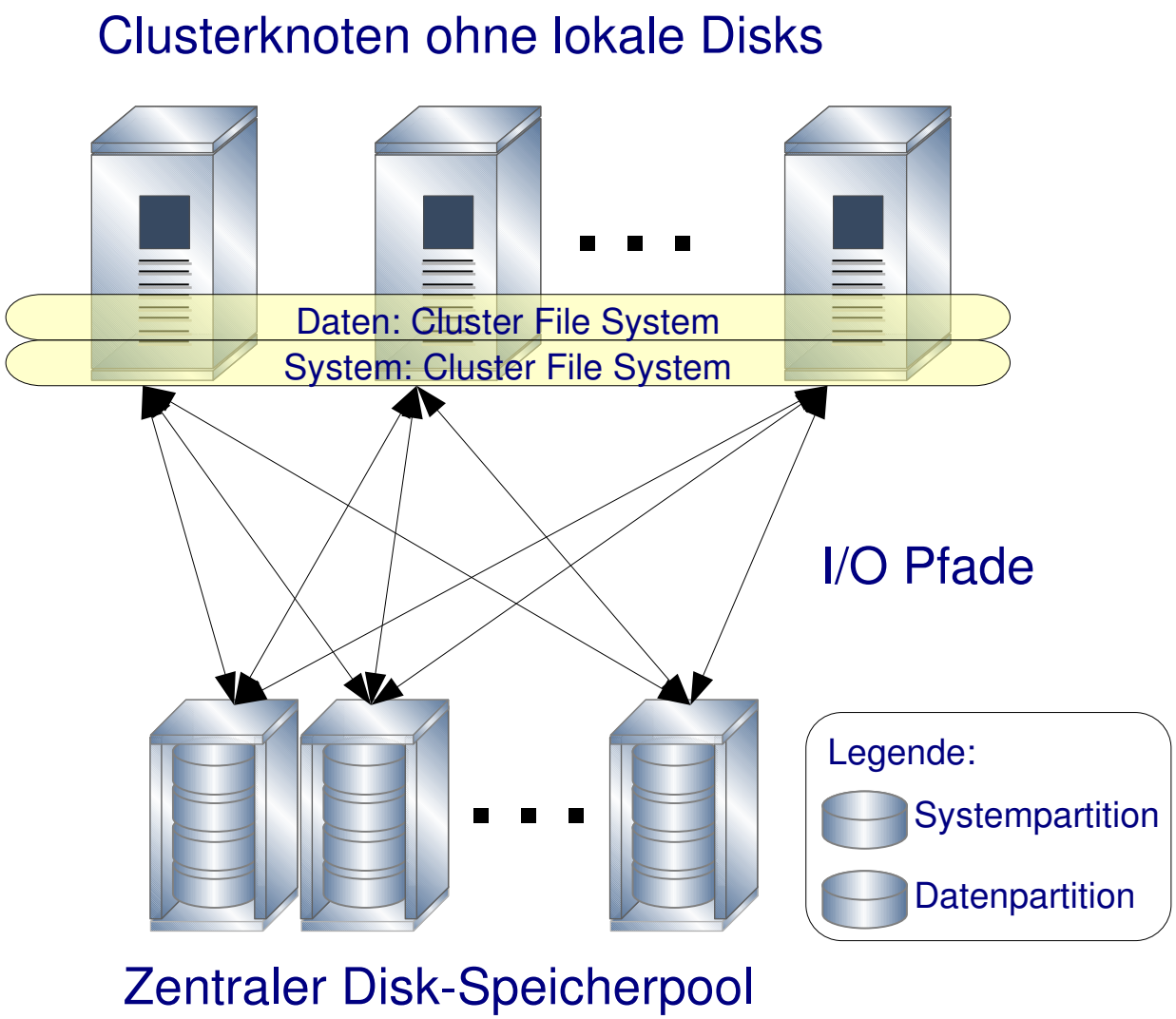
Shared Storage Cluster



- Peer to Peer Kommunikation
- Cluster Filesystem
- Server Cluster
  - Active/Active
- Storage Cluster
  - Storage Pooling
  - Volume Manager
- Hochleistungs Speichernetzwerk
- Management ??

ATIX

Diskless Shared Root Cluster



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

ATIX

Diskless Shared Root

- Standard Linux OS (RedHat, SuSe)
- Booten vom SAN
- Cluster Filesystem – Global Filesystem
- Shared Root im initial Ramdisk (initrd)
- Erweiterte Bootinformation durch Cluster Archiv
- Context Dependent Path Names
  - Symbolik - Tools
- Erweiterungen
  - „Application“-Serverless Backup
  - Desastertoleranz, Desaster Recovery
  - Cluster Clone, Snapshot

ATIX

Global Filesystem (GFS)

- Entwicklung seit 1995
  - University of Minnesota – Sistina – RedHat
  - Version 6
- Symetrisches Cluster Filesystem
- Vollwertiges POSIX Dateisystem
- Direct IO (Datenbanken)
- HA Locking Server (GULM)
- Distributed Lock Manager (DLM)
- Online Vergrößerbar
- Context Dependent Path Names
- Weiteres: ACLs, Quotas, Multipath

Context Dependent Path Names

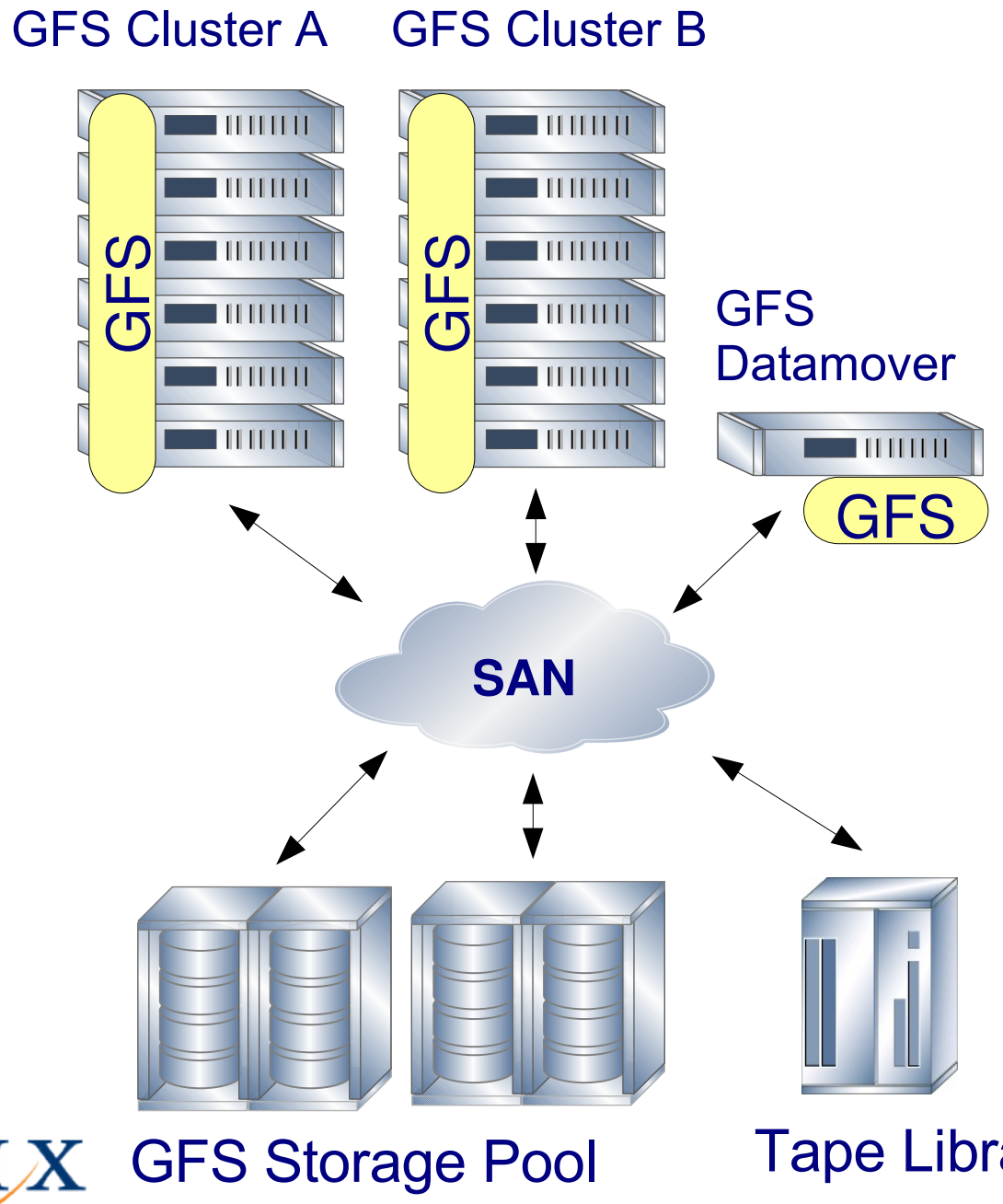
- Hostabhängige Dateien, Verzeichnisse
  - z.B. /tmp, /var/run, etc
- Verschiedene Abhängigkeiten
  - @hostname 'uname -n'
  - @mach 'uname -m'
  - @os 'uname -s'
  - und @sys, @uid, @gui
- Beispiel
  - /cluster/cdsl/{node1, node2, node3, default}
  - /cdsl.local -> /cluster/cdsl/@hostname
  - /var -> /cdsl.local/var

ATIX

Shared Root Werkzeuge

- com_create_hostdependent_file
  - macht eine Datei zu einer hostabhängigen Datei
- com_create_shared_file
  - macht aus einer hostabhängigen Datei eine „gesharte“
- com_create_cdsl
  - erzeugt automatisch einen cdls-Baum mittels CCA
- mkinitrd
  - erzeugt ein Shared Root initrd aus einer Standarddistribution
- ccs_read
  - Anpassungen an das RedHat Tool um mittels regulären Ausdrücken zu suchen.

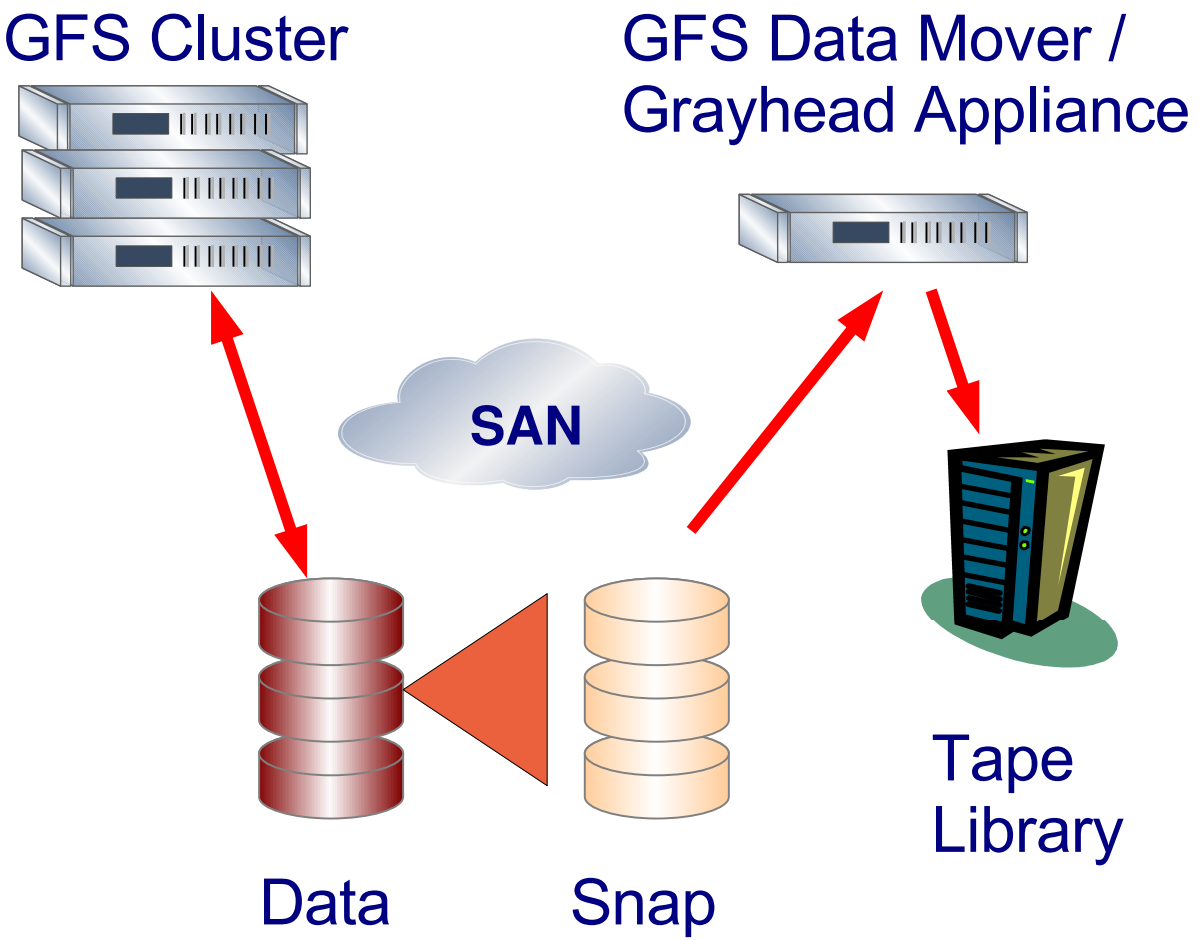
Aufbau GFS Shared Root



- SAN
- Storage Pool
- Datamover
- Shared Root

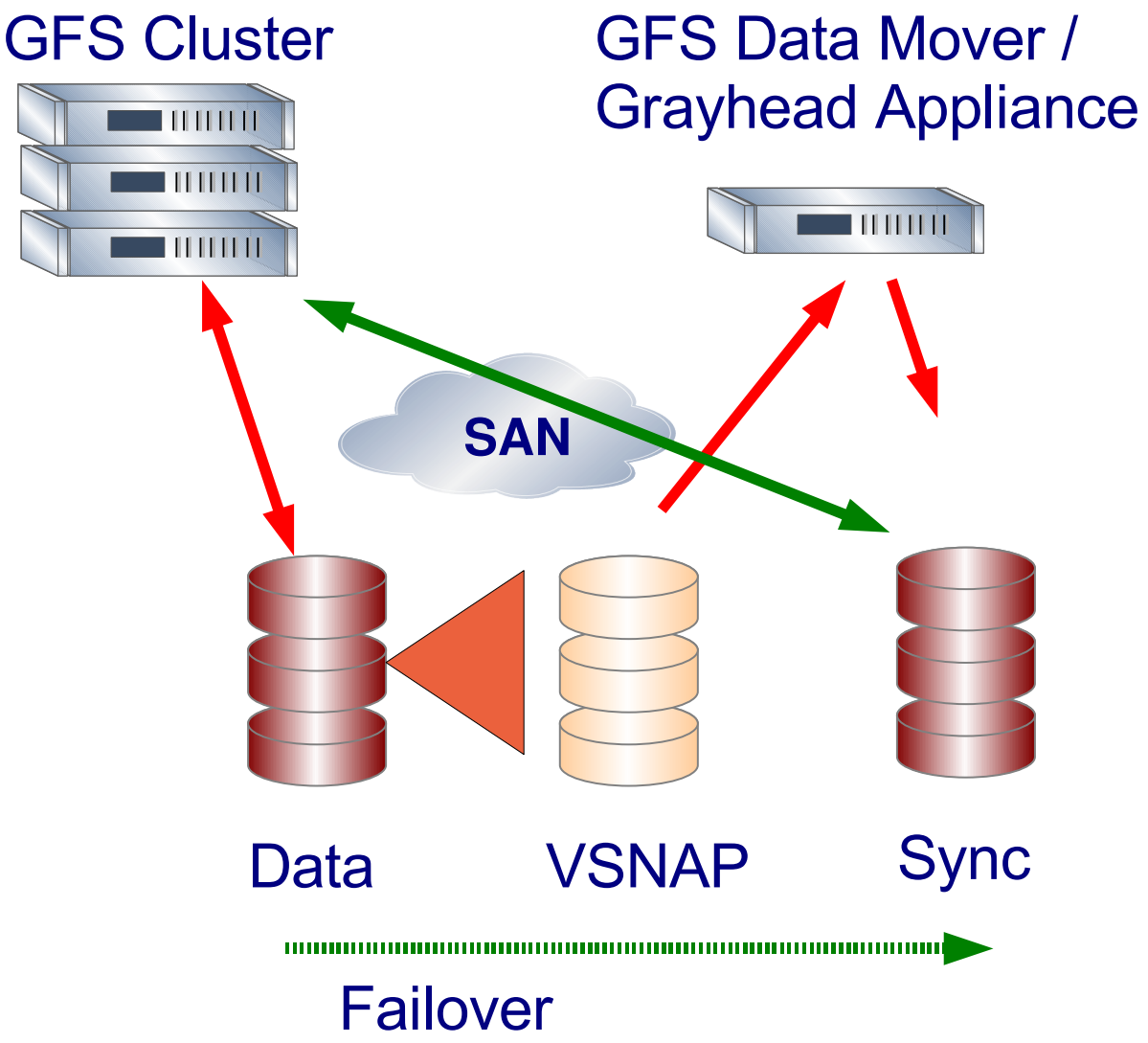
ATIX

Application-Server-less Backup



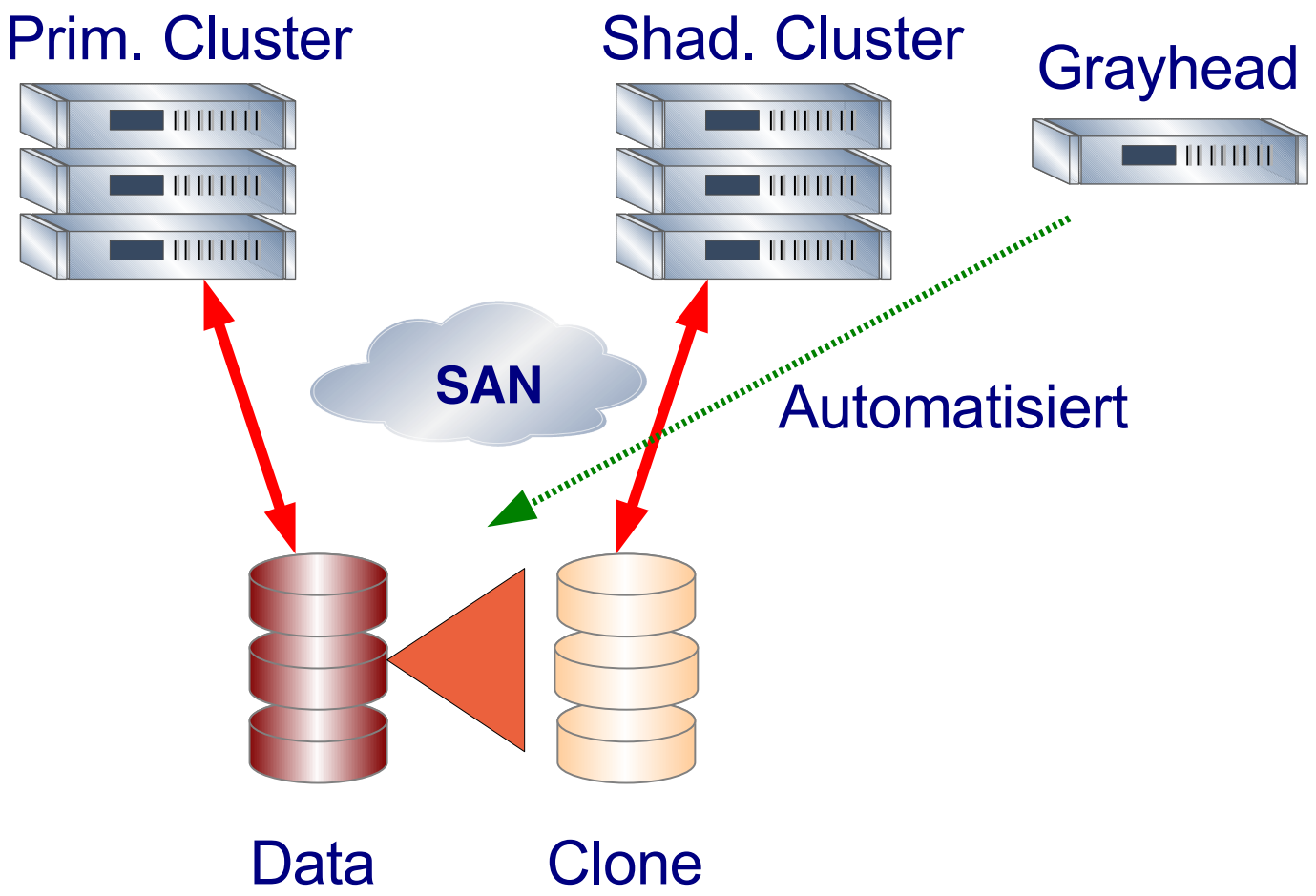
ATIX

Desaster Recovery Option



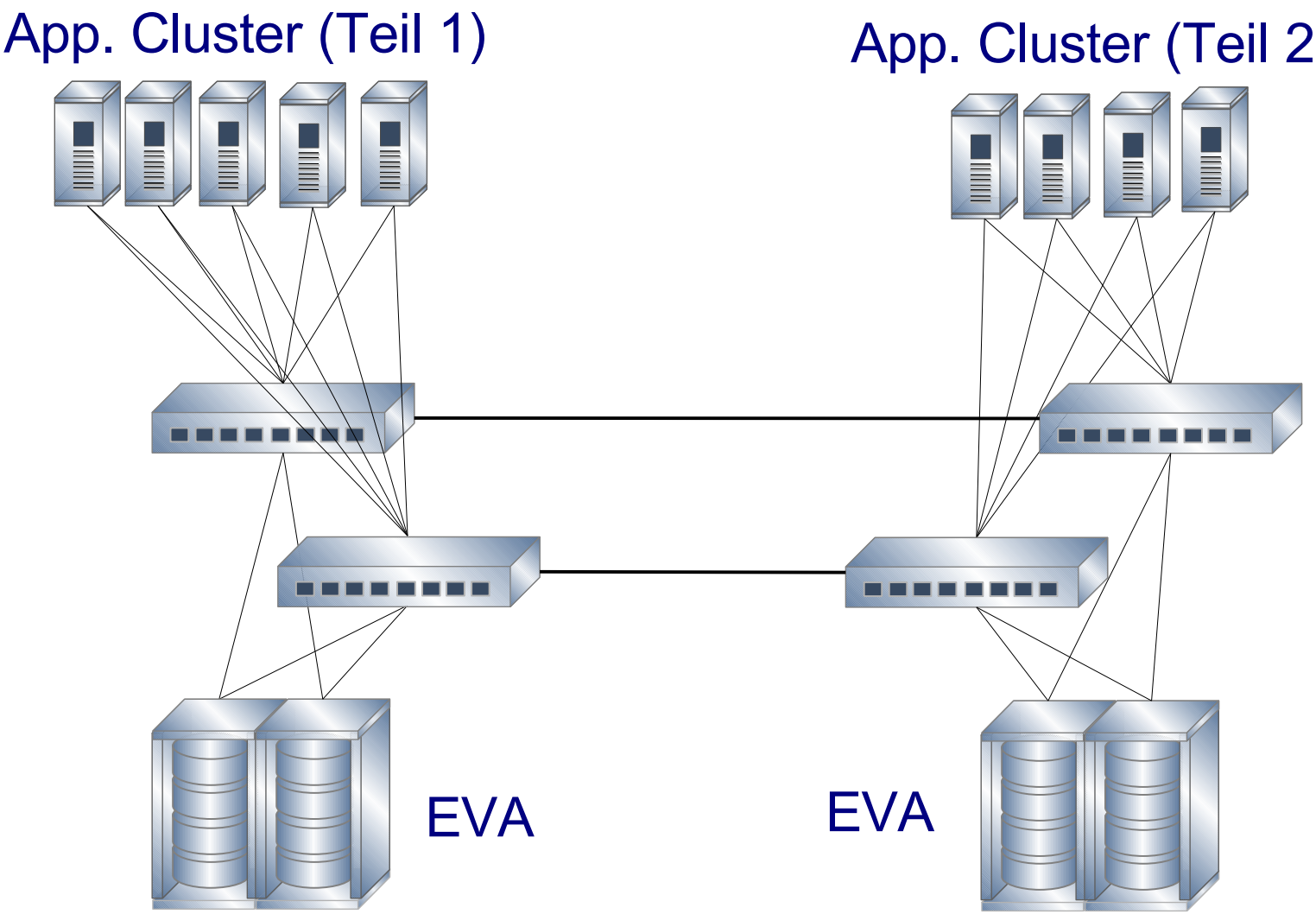
ATIX

Testsystem / DataWarehouse



ATIX

Data Replication



Shared Root Anwendungen

- Datenbanken
  - Oracle RAC, MySQL (1/N)
- Mail-Server
  - qmail, sendmail
- Web-Application-Server
  - Tomcat IBM Websphere
- Web-Server (apache)
- VMWare-Server
- File-Server
  - com.oonics NAS-Cluster

ATIX

Fazit Shared Root Cluster

- Hochskalierbare und hochverfügbare Cluster lassen sich mit standard Linux-Technologien und notwendigen Anpassungen verwirklichen
- Cluster-Architektur
  - Blades/diskless Server
  - Zentraler, virtualisierter und steuerbarer Storage
  - Cluster FS als Datenzugriffsschnittstelle
  - Shared Root für diskless Cluster
- Betrieb und Überwachung mit standard Opensource Software

ATIX

GFS / Cluster Suite

- Installation
- Administration

ATIX

RedHat Cluster Suite rpms (1)

- rgmanager — Manages cluster services and resources
- system-config-cluster — Cluster configuration and monitoring GUI
- ccsd — Contains the cluster configuration services daemon (ccsd) and associated files
- magma — Contains an interface library for cluster lock management
- magma-plugins — Contains plugins for the magma library
- cman — Contains the Cluster Manager (CMAN), which is used for managing cluster membership, messaging, and notification
- cman-kernel — Contains required CMAN kernel modules

ATIX

RedHat Cluster Suite rpms (2)

- dlm — Contains distributed lock management (DLM) library
- dlm-kernel — Contains required DLM kernel modules
- fence — The cluster I/O fencing system that allows cluster nodes to connect to a variety of network power switches, fibre channel switches, and integrated power management interfaces
- iddev — Contains libraries used to identify the file system (or volume manager) in which a device is formatted
- lvm2-cluster — Cluster extensions for the logical volume manager

ATIX

RedHat GFS 6.1 rpms

- GFS — The Red Hat GFS module
- GFS-kernel — The Red Hat GFS kernel module
- gnbd — The GFS Network Block Device module
  - Optional only when using shared network storage devices
- gnbd-kernel — Kernel module for the GFS Network Block Device
  - Optional only when using shared network storage devices
- GFS-kernheaders — GFS kernel header files
- gnbd-kernheaders — gnbd kernel header files
  - Optional only when using shared network storage devices

GFS Requirements

- 2 or 3 and more cluster nodes (quorum requirement)
- Shared storage device accessible from all nodes
  - Fibre Channel Storage
  - GNBD
  - VMWare shared disks (for test scenarios)
- 2 independent network devices
  - One for cluster communication
  - One for data
- Fencing Device for each node
  - fence_manual for test scenarios

ATIX

Network Requirements

- Make sure that each node can “see” all other nodes in the cluster
  - Make sure that the required network ports are accessible
    - Disable all iptables first
    - When you are sure your cluster configuration is running, you can configure the iptables again.
- Make sure that all node names can be resolved
  - Ping all nodes

```
[root@gfs-node-1 ~]# ping -c 1 gfs-node-2
PING gfs-node-2 (192.168.3.91) 56(84) bytes of data.
64 bytes from gfs-node-2 (192.168.3.91): icmp_seq=0 ttl=64 time=9.48 ms

[root@gfs-node-1 ~]# ping -c 1 gfs-node-3
PING gfs-node-3 (192.168.3.92) 56(84) bytes of data.
64 bytes from gfs-node-3 (192.168.3.92): icmp_seq=0 ttl=64 time=2.60 ms
```

ATIX

Package Installation

- Using the RedHat Network
  - # up2date --installall --channel rhel-i386-as-4-cluster
  - # up2date --installall --channel rhel-i386-as-4-gfs-6.1
- Manual rpm installation
  - Check installed kernel version
    - # uname -r
      2.6.9-11.ELsmp
  - Choose the rpms that match your kernel version
  - Copy all needed rpms into an empty directory e.g. /tmp/RPMS
  - # rpm -ivh /tmp/RPMS/*

ATIX

Cluster Tools

- ccs_tool (8)
  - The tool used to make online updates of CCS config files
- ccs_test (8)
  - The diagnostic tool for a running Cluster Configuration System
- cman_tool (8)
  - Cluster Management Tool
- fence_tool (8)
  - A program to join and leave the fence domain
- clustat (8)
  - Cluster Status Utility
- clusvcadm(8)
  - Cluster User Service Administration Utility
- See man pages for more information

Cluster Configuration

- Cluster configuration is written in XML format to /etc/cluster/cluster.conf
- The cluster configuration file has to be deployed to every node in the cluster
- The ccsd is responsible for
  - Locally serving the cluster configuration to other services (fenced, cman)
  - Online updates and deployment of configuration changes
  - The latest revision of the cluster.conf wins
    <cluster config_version="3" name="my_cluster">
- cluster.conf can be created and modified
  - Manually (for complex configurations)
  - system-config-cluster GUI (recommended)
- Updates via
  - ccs_tool update; cman_tool version -r <new_version>
  - system-config-cluster GUI

ATIX

Cluster Basic Configuration Steps

- # ssh -X root@<node>
- # system-config-cluster
- Create new configuration
- Define cluster name
- Create fencing devices
- Create cluster nodes
- Add fencing devices to cluster nodes
- Save
- Deploy /etc/cluster/cluster.conf
  - # scp -r /etc/cluster root@<node_name>:/etc/
- See RedHat's Cluster Configuration Manual for detailed GUI instructions

ATIX

Start Cluster Services

- Do for all nodes in cluster:
- Load ccsd
  # ccsd
- Load Distributed Locking Manager
  # modprobe dlm
- Start cluster manager
  # cman_tool join -w
- Check cluster status
  # cman_tool nodes
- Start fencing layer
  # fence_tool join -w -c
- Start clvmd
  # clvmd

ATIX

Create Storage Configuration

- Check for shared scsi disks
  # cat /proc/scsi/scsi
  # cat /proc/partitions
  # fdisk -l
- Create lvm2 physical volume
  # pvcreate <device_name>
- Create lvm2 volume group
  # vgcreate <vg_name> <pv_name>
- Create lvm2 logical volume
  # lvcreate -n <lv_name> -L <size> <vg_name>
- Check lvs on all nodes
  # lvscan

ATIX

Create GFS

- # man gfs_mkfs: show gfs_mkfs manual
- Some gfs_mkfs options
  - -h: print all options
  - -t <LockTableName>: defines name of cluster locking table. Has to be clustername:fsname
  - -j <journals>: defines the number of journal spaces (Every node needs its own journal space)
  - -p <lockproto>: defines the locking protocol (For DLM use lock_dlm)
- Example:

```
# gfs_mkfs -t mycluster:httplogs -j 4 -p lock_dlm
/dev/vg_gfs/lv_httplogs
```

  - This will create a gfs with logs for 4 nodes, dlm locking on the lv /dev/vg_gfs/lv_httplogs

Mount GFS

- Mount options (see man gfs_mount, man mount)
  - acl Enables POSIX Access Control List acl(5) support within GFS.
- Mount GFS manually
- Example

```
# mount-t gfs /dev/vg_gfs/lv_httplogs /mnt/gfs
```

- Mount GFS automatically
  - Edit /etc/fstab
  - Add line

    /dev/vg_gfs/lv_hhtplogs /mnt/gfs gfs defaults 0 0
  - GFS will be mounted from startscript
    /etc/init.d/gfs start
  - GFS will be umounted from startscript
    /etc/init.d/gfs stop

ATIX

GFS Tools

- gfs_mount
  - Mounting a GFS file system
- gfs_fsck
  - The GFS file system checker
- gfs_grow
  - Growing a GFS file system
- gfs_jadd
  - Adding a journal to a GFS file system
- gfs_mkfs
  - Make a GFS file system
- gfs_quota
  - Manipulate GFS disk quotas
- gfs_tool
  - Tool to manipulate a GFS file system

ATIX

Context Dependent Path Names

- Create symbolic links with context dependent variables
  - # ln -s /mnt/gfs/@<variable>/bin /usr/local/bin
- Variables:
  - @hostname: value of uname -n
  - @mach: value of uname -m
  - @os: value of uname -s
  - @sys: value of @mach_@os
  - @uid: value of id -u
  - @gid: value of id -g

ATIX

GFS Journaling

- GFS supports both data and metadata journaling
- Use gfs_tool to enable data journaling
  - setflag Flag File1 File2 ...
    - Set an attribute flag on a file. There are four currently supported flags. They are jdata, directio, inherit_jdata, and inherit_directio.
    - The jdata flag causes all the data written to a regular file to be journaled. The jdata flag can only be set (or cleared) if the file is zero length.
    - The inherit_jdata flag is set on a directory. It causes all new regular files created in that directory automatically inherit the jdata flag. The inherit_jdata is also inherited by any new subdirectories created in that directory.

GFS Direct IO

- GFS supports Direct IO
- Use gfs_tool to enable Direct IO
  - setflag Flag File1 File2 ...
    - Set an attribute flag on a file. There are four currently supported flags. They are jdata, directio, inherit_jdata, and inherit_directio.
    - The directio flag causes all I/O to a regular file to be Direct I/O, even if the O_DIRECT flag isn't used on the open() command.
    - The inherit_directio flag is set on a directory. It causes all new regular files created in that directory automatically inherit the directio flag. The inherit_directio is also inherited by any new subdirectories created in that directory.

GFS Quota Support

- GFS uses its gfs_quota command to manage quotas
- Two quota settings are available for each user ID (UID) or group ID (GID): a hard limit and a warn limit.
- Setting Quota hard limit
  - gfs_quota limit -u <User> -l <Size> -f <MountPoint>
  - gfs_quota limit -g <Group> -l <Size> -f <MountPoint>
- Setting Quota soft limit
  - gfs_quota warn -u <User> -l <Size> -f <MountPoint>
  - gfs_quota warn -g <Group> -l <Size> -f <MountPoint>
- Use gfs_quota get or list to see Quotas
- Synchronizing Quotas
  - Manually: gfs_quota sync -f MountPoint
  - Automatically every 60 sec or value set by
    - gfs_tool settune <MountPoint> quota_quantum <Seconds>

ATIX

Grow GFS

- GFS data space can be extended online
  - Use lvresize to extend the logical volume
  - Use gfs_grow <mountpoint> to grow gfs data space
- GFS journal space can be extended online
  - Usefull if you want to add new servers to your cluster
  - Use lvresize to extend the logical volume
  - Use gfs_jadd -j <num> <mountpoint> to add new journals to the GFS

ATIX

GFS Documentation

- Open-Shared Root Site
  "mini HowTo Build a shared root cluster"
  - http://sourceforge.net/docman/display_doc.php?docid=31675&group_id=156679#id2849103
- M. Hlawatschek, M. Grimme
  "Data sharing with a GFS storage cluster"
  - http://www.redhat.com/magazine/006apr05/features/gfs/
- RedHat Cluster Suite
  - http://www.redhat.com/docs/manuals/csgfs/browse/rh-cs-en/
- RedHat GFS
  - http://www.redhat.com/docs/manuals/csgfs/browse/rh-gfs-en

Noch Fragen?

Vielen Dank für Ihre Aufmerksamkeit!

Viel Spass !!

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

ATIX

HA Cluster: Stateless und Statefull Services

- Problem des transparenten Failovers
- Theoretisch muss der Dienst in genau dem gleichen Zustand auf dem 2. Knoten wieder laufen
- Dienste, die stateless sind speichern im Memory keine clientabhängigen Zustände ab
  - => Transparenter Failover ist kein Problem
- Dienste, die Zustände im Memory haben, müssen die Möglichkeit bieten diese persistent zu machen

ATIX

HA Cluster: Stateless und Statefull

- Statefull Services:
  - Perfektes Beispiel: DBMS
  - Lösung: Write-Logs, Redo-Logs etc.
  - Aber auch Dienste wie NFS/CIFS können Statefull sein
  - Es ist in diesem Fall immer notwendig, dass solche Dienste auch von der HA-Software unterstützt werden
    - Meist geschieht das über spezielle Resource Types

ATIX

Nagios

General
- Home
- Documentation

Monitoring
- Tactical Overview
- Service Detail
- Host Detail
- Status Overview
- Status Summary
- Status Grid
- Status Map
- 3-D Status Map
- Service Problems
- Host Problems
- Network Outages
- Comments
- Downtime
- Process Info
- Performance Info
- Scheduling Queue

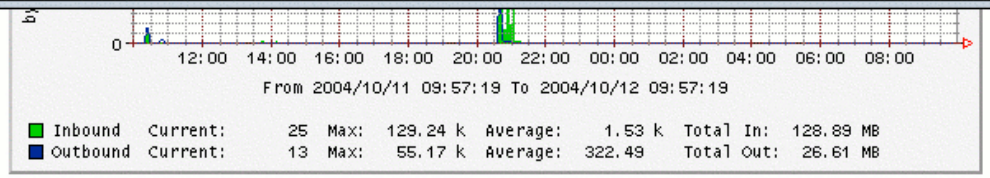
Reporting
- Trends
- Availability
- Alert Histogram
- Alert History
- Alert Summary
- Notifications
- Event Log

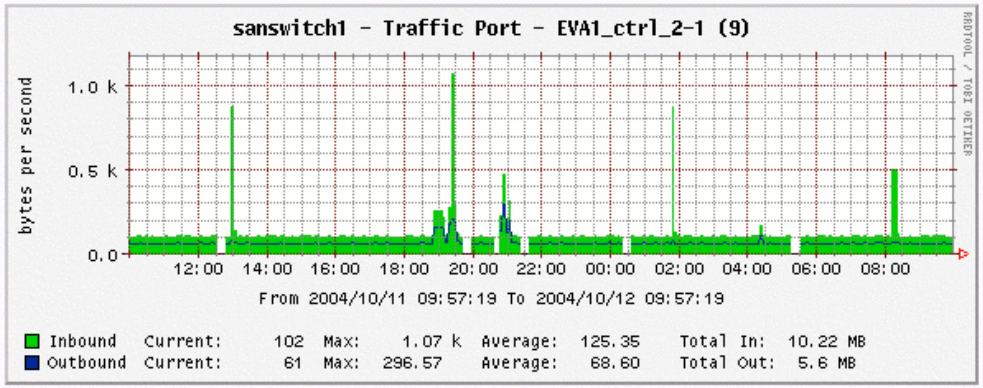
Configuration
- View Config

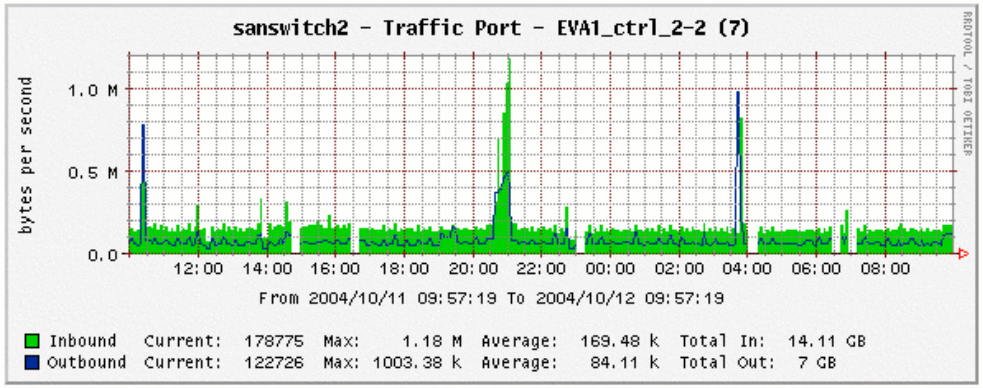
Cacti
- Internal

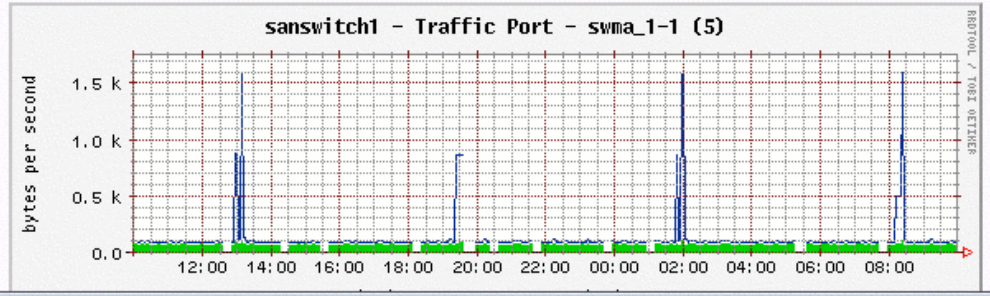
Host	Service	Status	Last Check	Duration	Attempt	Status Information
blade1-1	Check for warning load array of 4-3-3 and critical 5-4-4	OK	26-08-2004 18:26:45	0d 17h 45m 14s	1/3	Load : 0.04 0.04 0.00 : OK
	Checks for the given severity of the ucd-snmp monitored Disk /	OK	26-08-2004 18:27:27	16d 7h 58m 9s	1/3	Mountpoint with 51% available space (7208400/3543192/3665208) is [OK]!
	Checks the severity of monitored Disk /mnt/local/archiv	OK	26-08-2004 18:28:07	2d 6h 38m 45s	1/3	Mountpoint with 36% available space (20635700/12492956/7094508) is [OK]!
	Com-backup	OK	26-08-2004 00:56:37	2d 7h 25m 3s	1/1	Backup Job kollersv2 finished successfully ! Backup Job kollersv1 finished successfully !
	NFSserver	OK	26-08-2004 18:29:30	2d 6h 18m 51s	1/3	NFS server version 2 running OK!
	TRAPCPQDriveArray	PENDING	N/A	0d 0h 59m 12s+	0/1	Service check scheduled for Thu Aug 26 18:35:12 2004
	TRAPCPQFibreChannel	PENDING	N/A	0d 0h 59m 12s+	0/1	Service check scheduled for Thu Aug 26 18:30:54 2004
	TRAPCPQHealth	PENDING	N/A	0d 0h 59m 12s+	0/1	Service check scheduled for Thu Aug 26 18:31:34 2004
	TRAPCPQNic	PENDING	N/A	0d 0h 59m 12s+	0/1	Service check scheduled for Thu Aug 26 18:31:50 2004
	TRAPCPQRack	PENDING	N/A	0d 0h 59m 12s+	0/1	Service check scheduled for Thu Aug 26 18:32:31 2004
	TRAPCPQThreshhold	PENDING	N/A	0d 0h 59m 12s+	0/1	Service check scheduled for Thu Aug 26 18:33:13 2004
	TRAPCpqHost	PENDING	N/A	0d 0h 59m 12s+	0/1	Service check scheduled for Thu Aug 26 18:33:55 2004
	TRAPCpqSInfo	PENDING	N/A	0d 0h 59m 12s+	0/1	Service check scheduled for Thu Aug 26 18:34:35 2004
	TRAPDEFAULT	OK	10-08-2004 15:49:04	16d 4h 32m 4s	1/1	checked
	TRAPRMON	PENDING	N/A	0d 0h 59m 12s+	0/1	Service check scheduled for Thu Aug 26 18:30:58 2004
	cacti@mysql	OK	26-08-2004 18:26:40	2d 3h 10m 45s	1/3	Uptime: 18812 Threads: 1 Questions: 30682 Slow queries: 0 Opens: 13 Flush tables: 1 Open tables: 7 Queries per second avg: 1.631
	monitoring webserver	OK	26-08-2004 18:26:55	2d 6h 16m 5s	1/3	HTTP ok: HTTP/1.1 200 OK - 0.054 second response time
	ping	OK	26-08-2004 18:27:36	2d 6h 14m 43s	1/3	PING OK - Packet loss = 0%, RTA = 0.02 ms
blade1-2	Check for warning load array of 4-3-3 and critical 5-4-4	OK	26-08-2004 18:28:18	0d 4h 33m 32s	1/3	Load : 0.00 0.00 0.00 : OK
	Checks for the given severity of the ucd-snmp monitored Disk /	OK	26-08-2004 18:28:59	0d 4h 35m 42s	1/3	Mountpoint with 68% available space (7208368/2336604/4871764) is [OK]!
	Checks the severity of ucd-snmp monitored Disk /mnt/pool/data	OK	26-08-2004 18:29:41	0d 4h 32m 2s	1/3	Mountpoint with 0% available space (628956848/628953708/3140) is [OK]!
	PING	OK	26-08-2004 18:30:22	0d 18h 10m 44s	1/3	PING OK - Packet loss = 0%, RTA = 0.12 ms
	TRAPCPQDriveArray	PENDING	N/A	0d 0h 59m 12s+	0/1	Service check scheduled for Thu Aug 26 18:31:03 2004
	TRAPCPQFibreChannel	PENDING	N/A	0d 0h 59m 12s+	0/1	Service check scheduled for Thu Aug 26 18:31:45 2004
	TRAPCPQHealth	PENDING	N/A	0d 0h 59m 12s+	0/1	Service check scheduled for Thu Aug 26 18:32:00 2004

Comments
Downtime
Process Info
Performance Info
Scheduling Queue
Reporting
Trends
Availability
Alert Histogram
Alert History
Alert Summary
Notifications
Event Log
Configuration
View Config
Cacti
Internal
Host: Localhost
Host: Sanworks Management Appliance
SAN-Portstats
Host: sanswitch1
Host: sanswitch2
Port Traffic Blades
Port Errors Blades
Port Traffic EVA SWMA3
Port Errors EVA SWMA3
koller









Betrieb
„Prozesse und Sensoren...“

- Betrieb
  - „Welche Prozesse fallen wann, wo an und sind wie gelöst?“
    - Updates Software (Kernel, etc), HW (Firmware, etc)
    - Analyse von Schwellwertüberschreitungen und Reaktion
    - Hinzufügen neuer Server, Storage, Kundenprojekte etc.
    - Disaster Recovery
    - Backup/Restore
  - Dokumentation der Prozesse
  - Dokumentation aller Einstellungen
- SOPs

ATIX

Überwachung und Betrieb
„Prozesse und Sensoren...“

- Überwachung
  - Standardprotokoll SNMP (Unterstützung aller HW/SW Komponenten)
  - Definition von Sensoren
    - aktiv/passiv, eventgesteuert/polling
    - langfristig/kurzfristig
  - Opensource Software zur Überwachung
    - Nagios kurzfristig, aktiv/passiv
    - Cacti langfristig/passiv
  - Integration in eine gemeinsame Oberfläche

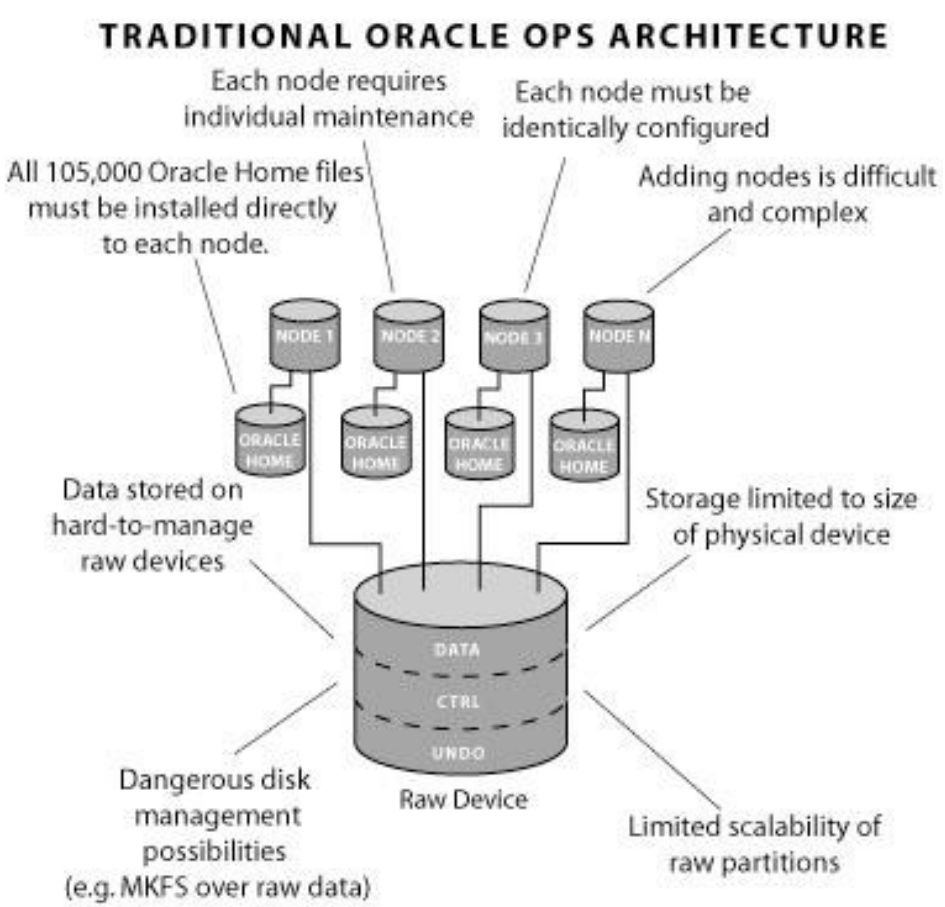
ATIX

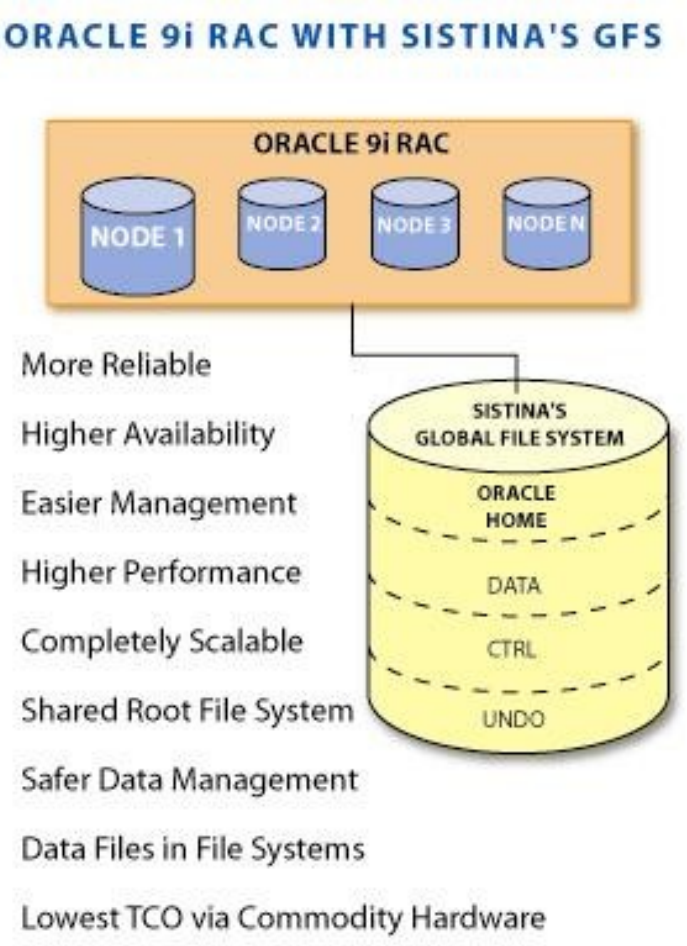
com.oonics Cluster-Suite

- Automatisiertes Management
  - Storage-Management (EVA Snap, Clone)
  - Einfügen neuer Server (Skalierbarkeit)
  - Automatisierte Konfiguration via CCA (z.B. Rollenverteilung, IP)
- Überwachung
  - Optische Anzeige des Systemzustands
  - Alerting (email, SNMP traps, pager)
  - Trend Analyse
- Availability / Data protection
  - Desaster Recovery Option
  - Application-Server-less Backup

ATIX

Oracle: Night and Day





ATIX