



Sun Cluster Überblick für Solaris OS

Sun Microsystems, Inc.
4150 Network Circle
Santa Clara, CA 95054
U.S.A.

Teilenr.: 819-0154
September 2004, Version A

Copyright 2004 Sun Microsystems, Inc. 4150 Network Circle, Santa Clara, CA 95054 U.S.A. Alle Rechte vorbehalten.

Dieses Produkt und die Dokumentation sind urheberrechtlich geschützt und werden unter Lizenzen vertrieben, durch die die Verwendung, das Kopieren, Verteilen und Dekompilieren eingeschränkt werden. Ohne vorherige schriftliche Genehmigung durch Sun und gegebenenfalls seiner Lizenzgeber darf kein Teil dieses Produkts oder Dokuments in irgendeiner Form reproduziert werden. Die Software anderer Hersteller, einschließlich der Schriftentechnologie, ist urheberrechtlich geschützt und von Lieferanten von Sun lizenziert.

Teile des Produkts können aus Berkeley BSD-Systemen stammen, die von der University of California lizenziert sind. UNIX ist eine eingetragene Marke in den Vereinigten Staaten und anderen Ländern und wird ausschließlich durch die X/Open Company Ltd. lizenziert.

Sun, Sun Microsystems, das Sun-Logo, docs.sun.com, AnswerBook, AnswerBook2, und Solaris sind in den USA und anderen Ländern Marken von Sun Microsystems Inc. Sämtliche SPARC-Marken werden unter Lizenz verwendet und sind Marken oder eingetragene Marken von SPARC International Inc. in den Vereinigten Staaten und anderen Ländern. Produkte mit der SPARC-Marke basieren auf einer von Sun Microsystems Inc. entwickelten Architektur.

Die grafischen Benutzeroberflächen von OPEN LOOK und Sun™ wurden von Sun Microsystems Inc. für seine Benutzer und Lizenznehmer entwickelt. Sun erkennt dabei die von Xerox Corporation geleistete Forschungs- und Entwicklungsarbeit auf dem Gebiet der visuellen oder grafischen Benutzeroberflächen für die Computerindustrie an. Sun ist Inhaber einer einfachen Lizenz von Xerox für die Xerox Graphical User Interface. Diese Lizenz gilt auch für Lizenznehmer von SUN, die mit den OPEN LOOK-Spezifikationen übereinstimmende grafische Benutzerschnittstellen implementieren und die schriftlichen Lizenzvereinbarungen einhalten.

U.S. Government Rights – Commercial software. Government users are subject to the Sun Microsystems, Inc. standard license agreement and applicable provisions of the FAR and its supplements.

DIE DOKUMENTATION WIRD "AS IS" BEREITGESTELLT, UND JEDLICHE AUSDRÜCKLICHE ODER IMPLIZITE BEDINGUNGEN, DARSTELLUNGEN UND HAFTUNG, EINSCHLIESSLICH JEDLICHER STILLSCHWEIGENDER HAFTUNG FÜR MARKTFÄHIGKEIT, EIGNUNG FÜR EINEN BESTIMMTEN ZWECK ODER NICHTÜBERTRETUNG WERDEN IM GESETZLICH ZULÄSSIGEN RAHMEN AUSDRÜCKLICH AUSGESCHLOSSEN.

Copyright 2004 Sun Microsystems, Inc. 4150 Network Circle, Santa Clara, CA 95054 U.S.A. Tous droits réservés.

Ce produit ou document est protégé par un copyright et distribué avec des licences qui en restreignent l'utilisation, la copie, la distribution, et la décompilation. Aucune partie de ce produit ou document ne peut être reproduite sous aucune forme, par quelque moyen que ce soit, sans l'autorisation préalable et écrite de Sun et de ses bailleurs de licence, s'il y en a. Le logiciel détenu par des tiers, et qui comprend la technologie relative aux polices de caractères, est protégé par un copyright et licencié par des fournisseurs de Sun.

Des parties de ce produit pourront être dérivées du système Berkeley BSD licenciés par l'Université de Californie. UNIX est une marque déposée aux Etats-Unis et dans d'autres pays et licenciée exclusivement par X/Open Company, Ltd.

Sun, Sun Microsystems, le logo Sun, docs.sun.com, AnswerBook, AnswerBook2, et Solaris sont des marques de fabrique ou des marques déposées, de Sun Microsystems, Inc. aux Etats-Unis et dans d'autres pays. Toutes les marques SPARC sont utilisées sous licence et sont des marques de fabrique ou des marques déposées de SPARC International, Inc. aux Etats-Unis et dans d'autres pays. Les produits portant les marques SPARC sont basés sur une architecture développée par Sun Microsystems, Inc.

L'interface d'utilisation graphique OPEN LOOK et Sun™ a été développée par Sun Microsystems, Inc. pour ses utilisateurs et licenciés. Sun reconnaît les efforts de pionniers de Xerox pour la recherche et le développement du concept des interfaces d'utilisation visuelle ou graphique pour l'industrie de l'informatique. Sun détient une licence non exclusive de Xerox sur l'interface d'utilisation graphique Xerox, cette licence couvrant également les licenciés de Sun qui mettent en place l'interface d'utilisation graphique OPEN LOOK et qui en outre se conforment aux licences écrites de Sun.

CETTE PUBLICATION EST FOURNIE "EN L'ETAT" ET AUCUNE GARANTIE, EXPRESSE OU IMPLICITE, N'EST ACCORDEE, Y COMPRIS DES GARANTIES CONCERNANT LA VALEUR MARCHANDE, L'APTITUDE DE LA PUBLICATION A REPONDRE A UNE UTILISATION PARTICULIERE, OU LE FAIT QU'ELLE NE SOIT PAS CONTREFAISANTE DE PRODUIT DE TIERS. CE DENI DE GARANTIE NE S'APPLIQUERAIT PAS, DANS LA MESURE OU IL SERAIT TENU JURIDIQUEMENT NUL ET NON AVENU.



041124@10536



Inhalt

Vorwort	5
1 Einführung in Sun Cluster	9
Herstellen der Hochverfügbarkeit von Anwendungen mithilfe von Sun Cluster	9
Verfügbarkeitsmanagement	10
Failover- und Scalable-Dienste sowie parallele Anwendungen	11
IP-Netzwerk-Multipathing	11
Speicherverwaltung	11
Campus-Cluster	13
Fehlerüberwachung	14
Verwaltungs- und Konfigurations-Tools	14
SunPlex-Manager	14
Befehlszeilenschnittstelle	15
Sun Management Center	16
Rollenbasierte Zugriffskontrolle (Role-Based Access Control, RBAC)	16
2 Schlüsselkonzepte für Sun Cluster	17
Cluster-Knoten	17
Cluster-Interconnect	18
Cluster-Mitgliedschaft	18
Cluster-Konfigurations-Repository (Cluster Configuration Repository, CCR)	19
Fehler-Monitore	20
Datendienstüberwachung	20
Plattenpfadüberwachung	20
IPMP-Überwachung	21
Quorum-Geräte	21

	Datenintegrität	21
	Fehlerschutz	22
	Failfast-Mechanismus für den Fehlerschutz	23
	Geräten	23
	Globale Geräte	24
	Lokale Geräte	25
	Plattengerätegruppen	25
	Datendienste	25
	Ressourcentypen	26
	Ressourcen	26
	Ressourcengruppen	27
	Datendiensttypen	27
3	Sun Cluster-Architektur	29
	Sun Cluster-Hardwareumgebung	29
	Sun Cluster-Softwareumgebung	30
	Cluster-Mitgliedschaftsüberwachung (Cluster Membership Monitor, CMM)	32
	Cluster-Konfigurations-Repository (Cluster Configuration Repository, CCR)	32
	Cluster-Dateisysteme	32
	Skalierbare Datendienste	33
	Lastausgleichsverfahren	34
	Multihost-Plattenspeicher	35
	Cluster-Interconnect	36
	IPMP-Gruppen	37
	Öffentliche Netzwerkschnittstellen	37
	Index	39

Vorwort

Sun™ Cluster Überblick für Solaris OS stellt das Sun Cluster-Produkt vor und erläutert dessen Zweck sowie die Mittel, mit denen Sun Cluster diesen Zweck verfolgt. Das Handbuch erläutert zudem Schlüsselkonzepte für Sun Cluster. Anhand der Informationen in diesem Dokument können Sie sich mit den Funktionen und der Funktionalität von Sun Cluster vertraut machen.

Verwandte Dokumentation

Informationen zu verwandten Sun Cluster-Themen können Sie der in der folgenden Tabelle aufgeführten Dokumentation entnehmen. Die gesamte Sun Cluster-Dokumentation steht unter <http://docs.sun.com> zur Verfügung.

Thema	Dokumentation
Überblick	Sun Cluster Überblick für Solaris OS
Konzepte	<i>Sun Cluster Konzepthandbuch für Solaris OS</i>
Hardware-Installation und -Verwaltung	<i>Sun Cluster 3.x Hardware Administration Manual for Solaris OS</i> Einzelne Hardwareverwaltungshandbücher
Softwareinstallation	<i>Sun Cluster Handbuch Softwareinstallation für Solaris OS</i>
Datendienstinstallation und -verwaltung	<i>Sun Cluster Data Services Planning and Administration Guide for Solaris OS</i> Einzelne Datendiensthandbücher
Datendienstentwicklung	<i>Sun Cluster Entwicklerhandbuch Datendienste für Solaris OS</i>

Thema	Dokumentation
Systemverwaltung	<i>Sun Cluster Handbuch Systemverwaltung für Solaris OS</i>
Fehlermeldungen	<i>Sun Cluster Error Messages Guide for Solaris OS</i>
Befehle und Funktionen	<i>Sun Cluster Reference Manual for Solaris OS</i>

Eine vollständige Liste der Sun Cluster-Dokumentation finden Sie in den Versionshinweisen zu Ihrer Sun Cluster-Version unter <http://docs.sun.com>.

Zugriff auf die Online-Dokumentation von Sun

Über die Website docs.sun.comSM erhalten Sie Zugriff auf die technische Online-Dokumentation von Sun. Sie können das Archiv unter docs.sun.com durchsuchen oder nach einem bestimmten Buchtitel oder Thema suchen. Die URL lautet <http://docs.sun.com>.

Bestellen von Sun-Dokumentation

Ausgewählte Produktdokumentationen bietet Sun Microsystems auch in gedruckter Form an. Eine Liste der Dokumente und Hinweise zum Bezug finden Sie unter "Buy printed documentation" auf der Website <http://docs.sun.com>.

Hilfe anfordern

Wenden Sie sich im Falle von Problemen bei der Installation und Verwendung des Sun Cluster-Systems an Ihren Dienstleister, und geben Sie folgende Informationen an:

- Ihren Namen und E-Mail-Adresse (sofern verfügbar)
- Firmennamen, Adresse, Telefonnummer
- Modell und Seriennummern des Systems
- Versionsnummer des Betriebssystems (z. B. Solaris 9)
- Versionsnummer der Sun Cluster-Software (z. B. 3.1 9/04)

Sammeln Sie für Ihren Dienstleister mithilfe folgender Befehle Informationen zu allen Knoten.

Befehl	Funktion
<code>prtconf -v</code>	Zeigt die Größe des Systemspeichers an und gibt Informationen zu Peripheriegeräten zurück.
<code>psrinfo -v</code>	Zeigt Informationen zu Prozessoren an.
<code>showrev -p</code>	Gibt die installierten Korrekturversionen zurück.
<code>prtdiag -v</code>	Zeigt Informationen zu Systemdiagnosen an.
<code>scinstall -pv</code>	Zeigt die Sun Cluster-Software- und Paketversion an.
<code>scstat</code>	Liefert einen Schnappschuss des Cluster-Status.
<code>scconf -p</code>	Listet die Cluster-Konfigurationsinformationen auf.
<code>scrgadm -p</code>	Zeigt Informationen zu installierten Ressourcen, Ressourcengruppen und Ressourcentypen an.

Halten Sie zudem den Inhalt der Datei `/var/adm/messages` bereit.

Typografische Konventionen

Die folgende Tabelle beschreibt die in diesem Buch verwendeten typografischen Änderungen.

TABELLE P-1 Typografische Konventionen

Schriftart oder Symbol	Bedeutung	Beispiel
AaBbCc123	Die Namen von Befehlen, Dateien, Verzeichnissen; Bildschirmausgabe.	Bearbeiten Sie Ihre <code>.login</code> -Datei. Verwenden Sie <code>ls -a</code> , um eine Liste aller Dateien zu erhalten. Rechnername% Sie haben eine neue Nachricht.
AaBbCc123	Die Eingaben des Benutzers, im Gegensatz zu den Bildschirmausgaben des Computers	Rechnername% su Passwort:

TABELLE P-1 Typografische Konventionen (Fortsetzung)

Schriftart oder Symbol	Bedeutung	Beispiel
<i>AaBbCc123</i>	Befehlszeilen-Variable: durch einen realen Namen oder Wert ersetzen	Der Befehl zum Entfernen einer Datei lautet <i>rm filename</i> .
<i>AaBbCc123</i>	Buchtitel, neue Begriffe und Ausdrücke, die hervorgehoben werden sollen.	Lesen Sie hierzu Kapitel 6 im <i>Benutzerhandbuch</i> . Führen Sie eine <i>Patch-Analyse</i> durch. Speichern Sie die Datei <i>nicht</i> . [Einige Hervorhebungen werden online in Fettbuchstaben dargestellt.]

Beispiele für Shell-Eingabeaufforderungen in Befehlen

Die folgende Tabelle zeigt die Standard-Systemeingabeaufforderung und die Superbenutzer-Eingabeaufforderung für die C-Shell, die Bourne-Shell und die Korn-Shell.

TABELLE P-2 Shell-Eingabeaufforderungen

Shell	Eingabeaufforderung
C Shell-Eingabeaufforderung	system%
C Shell-Superbenutzer-Eingabeaufforderung	system#
Bourne Shell- und Korn Shell-Eingabeaufforderung	\$
Bourne Shell- und Korn Shell-Superbenutzer-Eingabeaufforderung	#

Einführung in Sun Cluster

Das SunPlex-System ist eine integrierte Lösung aus Hardware und Sun Cluster-Software zur Erstellung von hoch verfügbaren Diensten und skalierbaren Diensten. Dieses Kapitel bietet einen Überblick auf hoher Ebene über die Sun Cluster-Funktionen.

Dieses Kapitel enthält die folgenden Abschnitte:

- „Herstellen der Hochverfügbarkeit von Anwendungen mithilfe von Sun Cluster “ auf Seite 9
- „Fehlerüberwachung “ auf Seite 14
- „Verwaltungs- und Konfigurations-Tools “ auf Seite 14

Herstellen der Hochverfügbarkeit von Anwendungen mithilfe von Sun Cluster

Ein Cluster besteht aus zwei oder mehr Systemen bzw. Knoten, die als ein einziges, ständig verfügbares System zusammenarbeiten, um den Benutzern Anwendungen, Systemressourcen und Daten bereitzustellen. Jeder Knoten im Cluster ist ein voll funktionsfähiges, eigenständiges System. In einer Cluster-Umgebung sind die Knoten jedoch über einen Interconnect verbunden und arbeiten zur Steigerung der Verfügbarkeit und Leistung als Einheit zusammen.

Hoch verfügbare Cluster stellen nahezu ständigen Zugriff auf Daten und Anwendungen bereit, indem sie den Cluster bei Ausfällen betriebsbereit halten, die ein System mit einem einzelnen Server normalerweise zum Absturz brächten. Ein einzelner Fehler (Hardware, Software oder Netzwerk) kann nicht zu einem Cluster-Ausfall führen. Fehlertolerante Hardwaresysteme sorgen demgegenüber für einen kontinuierlichen Zugriff auf Daten und Anwendungen, verursachen aber aufgrund der spezialisierten Hardware höhere Kosten. Fehlertolerante Systeme haben in der Regel keine Absicherung gegen Softwarefehler.

Jedes Sun Cluster-System ist eine Sammlung eng miteinander verbundener Knoten, die eine einzige Verwaltungsansicht für Netzwerkdienste und Anwendungen bietet. Das Sun Cluster-System erzielt seine Hochverfügbarkeit durch eine Kombination aus der folgenden Hardware und Software:

- Redundante Plattensysteme stellen Speicherplatz bereit. Diese Plattensysteme sind in der Regel gespiegelt, um eine Betriebsunterbrechung zu vermeiden, wenn eine Platte oder ein Teilsystem ausfällt. Redundante Verbindungen mit den Plattensystemen stellen sicher, dass keine Daten isoliert werden, wenn ein Server, Controller oder Kabel ausfällt. Ein Hochgeschwindigkeits-Interconnect zwischen Knoten sorgt für den Zugriff auf Ressourcen. Alle Knoten im Cluster sind auch mit einem öffentlichen Netzwerk verbunden, so dass Clients von mehreren Netzwerken aus auf den Cluster zugreifen können.
- Redundante Komponenten, die während des Betriebs gewechselt werden können, wie Stromkabel und Kühlsysteme, verbessern die Verfügbarkeit, da die Systeme nach einem Hardwarefehler den Betrieb aufrechterhalten können. Die Komponenten, die während des Betriebs gewechselt werden können, ermöglichen das Hinzufügen oder Entfernen von Hardwarekomponenten in einem laufenden System, ohne dass dieses heruntergefahren werden muss.
- Das Hochverfügbarkeits-Framework der Sun Cluster-Software stellt ein Knotenversagen rasch fest und migriert die Anwendung bzw. den Dienst auf einen anderen Knoten, der in einer identischen Umgebung läuft. Alle Anwendungen sind jederzeit verfügbar. Die von einem Knotenversagen nicht betroffenen Anwendungen stehen während der Wiederherstellung vollständig zur Verfügung. Darüber hinaus sind die Anwendungen des ausgefallenen Knotens wieder verfügbar, sobald sie wiederhergestellt sind. Eine wiederhergestellte Anwendung muss nicht die Wiederherstellung aller anderen Anwendungen abwarten.

Verfügbarkeitsmanagement

Eine Anwendung ist hoch verfügbar, wenn sie trotz eines einzelnen Software- oder Hardwarefehlers im System in Betrieb bleibt. Ausfälle durch Fehler oder Datenbeschädigung innerhalb der Anwendung selbst sind ausgeschlossen. Hoch verfügbare Anwendungen lassen sich wie folgt beschreiben:

- Die Wiederherstellung ist für die Anwendungen, die eine Ressource verwenden, transparent.
- Der Ressourcenzugriff bleibt während des Knotenversagens vollständig gewährleistet.
- Die Anwendungen können nicht erkennen, dass der Host-Knoten auf einen anderen Knoten verschoben wurde.
- Der Ausfall eines einzelnen Knotens ist für diejenigen Programme auf verbleibenden Knoten, die zu dem Knoten gehörende Dateien, Geräte und Plattendatenträger verwenden, vollständig transparent.

Failover- und Scalable-Dienste sowie parallele Anwendungen

Über Failover- und Scalable-Dienste sowie parallele Anwendungen können Ihre Anwendungen hoch verfügbar gemacht und die Leistung einer Anwendung im Cluster verbessert werden.

Ein Failover-Dienst sorgt mittels Redundanz für Hochverfügbarkeit. Wenn ein Fehler auftritt, kann eine laufende Anwendung so konfiguriert werden, dass sie entweder auf demselben Knoten neu gestartet wird, oder dass sie auf einen anderen Knoten im Cluster verschoben wird. Hierfür ist kein Bedienereingriff erforderlich.

Zur Leistungssteigerung setzt ein Scalable-Dienst die verschiedenen Knoten in einem Cluster für die gleichzeitige Ausführung einer Anwendung ein. In einer Scalable-Konfiguration kann jeder Cluster-Knoten Daten liefern und Client-Abfragen verarbeiten.

Bei parallelen Datenbanken können mehrere Instanzen des Datenbankservers folgende Aufgaben ausführen:

- Im Cluster teilnehmen
- Verschiedene Abfragen derselben Datenbank gleichzeitig abwickeln
- Bei langen Abfragen parallele Abfragemöglichkeit bereitstellen

Weitere Informationen zu Scalable- und Failover-Diensten erhalten Sie unter [„Datendiensttypen“](#) auf Seite 27.

IP-Netzwerk-Multipathing

Clients führen über das öffentliche Netzwerk Datenanforderungen an den Cluster durch. Jeder Cluster-Knoten ist über einen oder mehrere öffentliche Netzwerkadapter mit mindestens einem öffentlichen Netzwerk verbunden.

Bei IP-Netzwerk-Multipathing können mehrere Netzwerk-Ports eines Servers an dasselbe Teilnetz angeschlossen sein. Zunächst sorgt die IP-Netzwerk-Multipathing-Software für Spielraum bei Netzwerkadapterversagen, indem sie das Versagen bzw. die Reparatur eines Netzwerkadapters feststellt. Anschließend schaltet sie die Netzwerkdresse automatisch zum alternativen Adapter und wieder zurück. Wenn mehr als ein Netzwerkadapter in Betrieb ist, erhöht IP-Netzwerk-Multipathing den Datendurchsatz durch die Aufteilung von gesendeten Datenpaketen unter den Adaptern.

Speicherverwaltung

Durch den Multihost-Speicher werden die Platten durch Anschluss an mehrere Knoten hoch verfügbar gemacht. Mehrere Knoten aktivieren mehrere Pfade für den Datenzugriff. Wenn ein Pfad ausfällt, steht ein anderer Pfad als Ersatz zur Verfügung.

Multihost-Platten ermöglichen die folgenden Cluster-Prozesse:

- Toleranz von einzelnen Knotenausfällen.
- Zentralisierung von Anwendungsdaten, Anwendungsbinärdateien und Konfigurationsdateien.
- Schutz gegen Knotenversagen. Wenn Client-Anfragen über einen Knoten auf Daten zugreifen und dieser Knoten ausfällt, werden die Anfragen umgeschaltet und über einen anderen Knoten abgewickelt, der mit denselben Platten direkt verbunden ist.
- Der Zugriff erfolgt entweder global über einen Primärknoten, der die Platten "unterstützt" oder durch einen direkten gleichzeitigen Zugriff über lokale Pfade.

Unterstützung des Datenträgermanagements

Ein Datenträger-Manager ermöglicht dem Benutzer das Verwalten einer großen Plattenanzahl sowie der zugehörigen Plattendaten. Datenträger-Manager steigern durch folgende Funktionen die Speicherkapazität und Datenverfügbarkeit:

- Plattenlaufwerk-Striping und Verkettung
- Plattenspiegelung
- Plattenlaufwerk-Hot-Spares
- Abwicklung von Plattenausfällen und Plattenersatz

Sun Cluster-Systeme unterstützen die folgenden Datenträger-Manager:

- Solaris Volume Manager
- VERITAS Volume Manager

Sun StorEdge Traffic Manager

Die Sun StorEdge Traffic Manager-Software ist ab dem Solaris-Betriebssystem 8 E/A-Kern-Framework voll integriert. Mithilfe der Sun StorEdge Traffic Manager-Software können Sie Geräte, auf die über mehrere E/A-Controller-Schnittstellen innerhalb einer einzelnen Instanz der Solaris-Betriebsumgebung zugegriffen wird, besser darstellen und verwalten. Die Sun StorEdge Traffic Manager-Architektur kann Folgendes leisten:

- Schutz vor E/A-Ausfällen durch E/A-Controller-Versagen
- Automatisches Wechseln zu einem anderen Controller bei E/A-Controller-Versagen
- Verbesserte E/A-Leistung durch Lastausgleich über mehrere E/A-Kanäle

Unterstützung von Hardware-RAID (Redundant Array of Independent Disks)

Sun Cluster-Systeme unterstützen die Verwendung von Hardware-RAID (Redundant Array of Independent Disks) und Host-basiertem Software-RAID. Hardware-RAID verwendet die Redundanz des Speicher-Arrays bzw. die Hardwareredundanz des

Speichersystems, um sicherzustellen, dass unabhängige Hardwarefehler die Datenverfügbarkeit nicht beeinträchtigen. Beim Spiegeln über getrennte Speicher-Arrays stellt Host-basiertes Software-RAID sicher, dass unabhängige Hardwarefehler die Datenverfügbarkeit nicht beeinträchtigen, wenn ein ganzes Speicher-Array offline geschaltet ist. Hardware-RAID und Host-basiertes Software-RAID können gleichzeitig eingesetzt werden. Dennoch ist nur eine RAID-Lösung erforderlich, um einen hohen Grad an Datenverfügbarkeit sicherzustellen.

Dateisystemunterstützung

Eine der inhärenten Eigenschaften von Cluster-Systemen besteht in der gemeinsamen Nutzung von Ressourcen. Daher benötigt ein Cluster ein Dateisystem, das für Kohärenz bei der gemeinsamen Nutzung von Dateien sorgt. Über das Sun Cluster-Dateisystem können Benutzer oder Anwendungen auf jede beliebige Datei auf allen Cluster-Knoten über die Remote- bzw. lokalen Standard-UNIX-APIs zugreifen. Sun Cluster-Systeme unterstützen die folgenden Dateisysteme

- UNIX-Dateisysteme (UFS)
- Sun StorEdge QFS-Dateisysteme
- VERITAS-Dateisysteme (VxFS)

Wenn eine Anwendung von einem Knoten auf einen anderen verschoben wird, ist keine Änderung an der Anwendung erforderlich, um auf dieselben Dateien zugreifen zu können. Vorhandene Anwendungen können das Cluster-Dateisystem ohne Änderungen voll nutzen.

Campus-Cluster

Standardmäßige Sun Cluster-Systeme bieten eine hohe Verfügbarkeit und Zuverlässigkeit von einem einzelnen Standort aus. Wenn Ihre Anwendung nach einem unvorhersehbaren Ereignis, beispielsweise nach einem Erdbeben, einer Überschwemmung oder einem Stromausfall, zur Verfügung stehen muss, können Sie Ihren Cluster als Campus-Cluster konfigurieren.

Mithilfe von Campus-Clustern können Sie Cluster-Komponenten, wie beispielsweise Knoten und gemeinsam genutzten Speicher in getrennten Räumen, die Kilometer voneinander entfernt liegen, finden. Sie können Ihre Knoten und den gemeinsam genutzten Speicher voneinander trennen und von verschiedenen Einrichtungen innerhalb Ihres Unternehmensgeländes oder von einem anderen kilometerweit entfernten Standort aus suchen. Sollte an einem Standort ein Systemausfall eintreten, können die nicht betroffenen Knoten den Dienst des ausgefallenen Knotens übernehmen. Auf diese Weise bleiben die Anwendungen und Daten für Ihre Benutzer verfügbar.

Fehlerüberwachung

Das Sun Cluster-System macht den Pfad zwischen Benutzern und Daten über Multihostplatten, Multipathing und ein globales Dateisystem hoch verfügbar. Das Sun Cluster-System überwacht Fehler folgender Elemente:

- Anwendungen – Die meisten Sun Cluster-Datendienste bieten einen Fehler-Monitor, der den Zustand des Datendienstes in regelmäßigen Abständen überprüft. Ein Fehler-Monitor überprüft, ob der Anwendungsdaemon bzw. die -dämonen laufen und die Clients bedient werden. Basierend auf den von den Testsignalen zurückgegebenen Informationen wird eine vordefinierte Aktion ausgeführt, wie zum Beispiel der Neustart von Dämonen oder die Initialisierung eines Failover.
- Plattenpfade – Die Sun Cluster-Software unterstützt die Plattenpfadüberwachung (Disk-Path Monitoring, DPM). DPM verbessert insgesamt die Zuverlässigkeit von Failover bzw. Switchover, indem über Fehler bei sekundären Plattenpfaden berichtet wird.
- Internet Protocol (IP) Multipath – Solaris IP-Netzwerk-Multipathing-Software auf Sun Cluster-Systemen stellt den Grundmechanismus für die Überwachung von öffentlichen Netzwerkadaptern bereit. IPMP ermöglicht auch das Failover von IP-Adressen von einem Adapter zu einem anderen, wenn ein Fehler festgestellt wird.

Verwaltungs- und Konfigurations-Tools

Sie können das Sun Cluster-System entweder über die SunPlex-Manager-GUI oder über die Befehlszeilenschnittstelle (Command-Line Interface, CLI) installieren, konfigurieren und verwalten.

Das Sun Cluster-System verfügt auch über ein Modul, das als Teil der Sun Management Center-Software ausgeführt wird und eine GUI für bestimmte Cluster-Aufgaben bereitstellt.

SunPlex-Manager

SunPlex-Manager ist ein browserbasiertes Tool für die Verwaltung von Sun Cluster-Systemen. Mithilfe der SunPlex-Manager-Software können Verwalter Systeme verwalten und überwachen, Software installieren und Systeme konfigurieren.

Die SunPlex Manager-Software enthält folgende Funktionen:

- Integrierte Sicherheits- und Autorisierungsmechanismen
- SLL-Unterstützung (Secure Sockets Layer)
- RBAC (Role-based access control, rollenbasierte Zugriffskontrolle)
- PAM (Pluggable Authentication Module, sofort integrierbares Authentisierungsmodul)
- Gruppenverwaltungsfunktionen für NAFO und IP-Netzwerk-Multipathing
- Verwaltung von Quorum-Geräten, Transporten, gemeinsam genutzten Speichergeräten und Ressourcengruppen
- Fortgeschrittene Fehlerprüfung und automatische Erkennung von privaten Interconnects.

Befehlszeilenschnittstelle

Die Befehlszeilenschnittstelle von Sun Cluster besteht aus einem Satz von Dienstprogrammen, die zum Installieren und Verwalten von Sun Cluster-Systemen sowie zum Verwalten des Datenträger-Manager-Teils der Sun Cluster-Software eingesetzt werden können.

Folgende SunPlex-Verwaltungsaufgaben können über die Sun Cluster-CLI ausgeführt werden:

- Validieren einer Sun Cluster-Konfiguration
- Installieren und Konfigurieren der Sun Cluster-Software
- Aktualisieren einer Sun Cluster-Konfiguration
- Verwalten der Registrierung von Ressourcentypen, der Erstellung von Ressourcengruppen und der Aktivierung von Ressourcen innerhalb einer Ressourcengruppe
- Ändern der Knotenunterstützung und Zustände von Ressourcengruppen und Plattengerätegruppen
- Steuern des Zugriffs über rollenbasierte Zugriffskontrolle (Role-Based Access Control, RBAC)
- Herunterfahren des gesamten Clusters

Sun Management Center

Das Sun Cluster-System verfügt auch über ein Modul, das als Teil der Sun Management Center-Software ausgeführt wird. Sun Management Center dient im Cluster als Grundlage für Verwaltungs- und Überwachungsvorgänge und ermöglicht es Systemverwaltern, die folgenden Aufgaben über eine GUI bzw. CLI auszuführen:

- Konfigurieren eines Remote-Systems
- Überwachen der Leistung
- Erkennen und Isolieren von Hardware- und Softwarefehlern

Sun Management Center kann auch als Schnittstelle zum Verwalten einer dynamischen Rekonfiguration innerhalb von Sun Cluster-Servern eingesetzt werden. Die dynamische Rekonfiguration umfasst Domänenerstellung, dynamischen Platinenanschluss und dynamische Trennung.

Rollenbasierte Zugriffskontrolle (Role-Based Access Control, RBAC)

In herkömmlichen UNIX-Systemen ist der Root-Benutzer, auch Superbenutzer genannt, omnipotent. Er kann jede beliebige Datei lesen und darin schreiben, alle Programme ausführen und Signale zum erzwungenen Abbruch eines beliebigen Prozesses senden. Die rollenbasierte Zugriffskontrolle (RBAC) von Solaris ist eine Alternative zu dem Superbenutzer-Modell mit seinem "Alles-oder-Nichts"-Ansatz. RBAC verwendet das Sicherheitsprinzip der niedrigsten Berechtigung. Das heißt, dass keinem Benutzer mehr Berechtigungen erteilt werden, als zum Ausführen einer bestimmten Aufgabe erforderlich sind.

RBAC ermöglicht es einer Organisation, die Superbenutzer-Fähigkeiten aufzuteilen und sie speziellen Benutzerkonten bzw. Rollen zuzuordnen, die dann einzelnen Personen zugewiesen werden. Diese Abtrennung und Paketbildung ermöglicht eine Reihe von Sicherheitsrichtlinien. Es können Konten für Verwalter mit bestimmten Funktionen eingerichtet werden, und zwar in Bereichen wie Sicherheit, Netzwerk, Firewall, Sicherungen und Systembetrieb.

Schlüsselkonzepte für Sun Cluster

In diesem Kapitel werden die Schlüsselkonzepte für die Hardware- und Softwarekomponenten des Sun Cluster-Systems erläutert, mit denen Sie vor der Arbeit mit Sun Cluster-Systemen vertraut sein sollten.

Dieses Kapitel enthält die folgenden Abschnitte:

- „Cluster-Knoten ” auf Seite 17
- „Cluster-Interconnect ” auf Seite 18
- „Cluster-Mitgliedschaft ” auf Seite 18
- „Cluster-Konfigurations-Repository (Cluster Configuration Repository, CCR) ” auf Seite 19
- „Fehler-Monitore” auf Seite 20
- „Quorum-Geräte ” auf Seite 21
- „Geräten” auf Seite 23
- „Datendienste ” auf Seite 25

Cluster-Knoten

Ein Cluster-Knoten ist ein Rechner, auf dem sowohl die Solaris-Software als auch die Sun Cluster-Software ausgeführt wird. Mit der Sun Cluster-Software können Sie zwei bis acht Knoten in einen Cluster einbinden.

Cluster-Knoten sind im Allgemeinen an mindestens eine Platte angeschlossen. Nicht an Platten angeschlossene Knoten verwenden das Cluster-Dateisystem zum Zugriff auf die Multihostplatten. Knoten in Konfigurationen paralleler Datenbanken teilen den Zugriff auf einige oder alle Platten.

Jeder Knoten im Cluster nimmt wahr, wenn ein anderer Knoten dem Cluster beiträgt oder diesen verlässt. Daneben nimmt jeder Knoten sowohl die lokalen Ressourcen als auch die Ressourcen auf den anderen Cluster-Knoten wahr.

Knoten innerhalb desselben Clusters sollten eine vergleichbare Verarbeitungs-, Speicher- und E/A-Kapazität aufweisen, so dass ein Failover ohne nennenswerten Leistungsabfall möglich ist. Da Failover vorkommen können, muss jeder Knoten für den Fall eines Knotenversagens über ausreichende Kapazität zur Erfüllung der Dienstebenenvereinbarungen (Service Level Agreements, SLA) verfügen.

Cluster-Interconnect

Der Cluster-Interconnect ist die reale Konfiguration der Geräte, die für die Übertragung von Cluster-privaten Kommunikationen und Datendienstkommunikationen zwischen Cluster-Knoten eingesetzt werden.

Über redundante Interconnects wird der Betrieb auf funktionsfähigen Interconnects aufrechterhalten, während die Systemverwalter Fehler isolieren und die Kommunikation wieder herstellen. Die Sun Cluster-Software erkennt und repariert Fehler und startet die Kommunikation automatisch über einen reparierten Interconnect neu.

Weitere Informationen finden Sie unter „[Cluster-Interconnect](#)“ auf Seite 36.

Cluster-Mitgliedschaft

Der Cluster-Mitglieder-Monitor (CMM) ist ein verteilter Satz von Agenten, die Nachrichten über den Cluster-Interconnect austauschen, um folgende Aufgaben auszuführen:

- Erzwingen einer konsistenten Mitgliederansicht auf allen Knoten (Quorum)
- Steuern einer synchronisierten Rekonfiguration als Reaktion auf Mitgliedschaftsänderungen
- Abwickeln der Cluster-Partitionierung
- Sicherstellen von vollständiger Konnektivität zwischen allen Cluster-Mitgliedern, indem fehlerhafte Knoten bis zu deren Reparatur aus dem Cluster ausgeschlossen werden

Die Hauptfunktion des CMM besteht im Festlegen der Cluster-Mitgliedschaft. Hierfür ist eine Cluster-weite Einigung über den Knotensatz, der zum jeweiligen Zeitpunkt am Cluster teilnimmt, erforderlich. Der CMM stellt wichtige Cluster-Statusänderungen auf den Knoten fest, wie zum Beispiel den Kommunikationsverlust zwischen einem oder mehreren Knoten. Der CMM nutzt das

Transport-Kernel-Modul zum Generieren von Heartbeats über das Transportmedium an andere Knoten im Cluster. Wenn der CMM innerhalb einer definierten Zeitüberschreitungsperiode keinen Heartbeat eines Knotens feststellt, wird davon ausgegangen, dass der Knoten ausgefallen ist, und der CMM startet eine Cluster-Rekonfiguration, um die Cluster-Mitgliedschaft neu zu verhandeln.

Der CMM führt folgende Aufgaben aus, um die Cluster-Mitgliedschaft festzulegen und die Datenintegrität sicherzustellen:

- Festhalten von Änderungen bei der Cluster-Mitgliedschaft, zum Beispiel ein Knoten, der dem Cluster beitrifft oder diesen verlässt.
- Sicherstellen, dass ein fehlerhafter Knoten den Cluster verlässt.
- Sicherstellen, dass ein fehlerhafter Knoten bis zu seiner Reparatur inaktiv bleibt.
- Verhindern, dass der Cluster sich selbst in Knoten-Teilsätze partitioniert.

Weitere Informationen darüber, wie sich der Cluster vor der Partitionierung in mehrere getrennte Cluster schützt, finden Sie unter „[Datenintegrität](#)“ auf Seite 21.

Cluster-Konfigurations-Repository (Cluster Configuration Repository, CCR)

Das Cluster-Konfigurations-Repository (CCR) ist eine private, Cluster-weite, verteilte Datenbank zur Speicherung von Informationen über Konfiguration und Zustand des Clusters. Um eine Beschädigung der Konfigurationsdaten zu vermeiden, muss jedem Knoten der aktuelle Zustand der Cluster-Ressourcen bekannt sein. Das CCR stellt sicher, dass alle Knoten ein konsistentes Bild des Clusters haben. Das CCR wird aktualisiert, wenn Fehler- bzw. Wiederherstellungssituationen auftreten bzw. wenn sich der allgemeine Status des Clusters ändert.

Die CCR-Strukturen enthalten folgende Informationstypen:

- Cluster- und Knotennamen
- Cluster-Transportkonfiguration
- Namen der Solaris Volume Manager-Plattensätze bzw. VERITAS-Plattengruppen
- Eine Liste der Knoten, die die einzelnen Plattengruppen unterstützen können
- Betriebsparameterwerte für Datendienste
- Pfade zu Datendienst-Rückmeldemethoden
- DID-Gerätekonfiguration
- Aktueller Cluster-Status

Fehler-Monitore

Das Sun Cluster-System macht alle Komponenten auf dem "Pfad" zwischen Benutzern und Daten hoch verfügbar, indem die Anwendungen selbst, das Dateisystem und die Netzwerkschnittstellen überwacht werden.

Die Sun Cluster-Software stellt ein Knotenversagen rasch fest und erstellt einen äquivalenten Server für die Ressourcen auf dem ausgefallenen Knoten. Die Sun Cluster-Software stellt sicher, dass die vom Knotenversagen nicht betroffenen Ressourcen während der Wiederherstellungsphase konstant verfügbar bleiben und dass die Ressourcen des ausgefallenen Knotens unmittelbar nach ihrer Wiederherstellung wieder verfügbar werden.

Datendienstüberwachung

Jeder Sun Cluster-Datendienst stellt einen Fehler-Monitor zur Verfügung, der den Datendienst regelmäßig auf einwandfreies Funktionieren prüft. Ein Fehler-Monitor überprüft, ob der Anwendungsdämon bzw. die -dämonen laufen und die Clients bedient werden. Aufgrund der von den Testsignalen zurückgegebenen Informationen können vordefinierte Aktionen wie ein Neustart eines Dämons oder das Einleiten eines Failovers ausgelöst werden.

Plattenpfadüberwachung

Die Sun Cluster-Software unterstützt Plattenpfadüberwachung (Disk-Path Monitoring, DPM). DPM verbessert insgesamt die Zuverlässigkeit von Failover- bzw. Switchover-Vorgängen, indem das Versagen eines sekundären Plattenpfads gemeldet wird. Eine der beiden folgenden Methoden kann zur Plattenpfadüberwachung verwendet werden. Die erste Methode wird mit dem `scdpm`-Befehl zur Verfügung gestellt. Mit diesem Befehl können Sie die Plattenpfade im Cluster überwachen, die Überwachung aufheben oder deren Status anzeigen. Weitere Informationen zu Befehlszeilenoptionen finden Sie in der Online-Dokumentation unter `scdpm(1M)`.

Die zweite Methode zur Überwachung von Plattenpfaden im Cluster wird von der grafischen Benutzeroberfläche (Grafical User Interface, GUI) von SunPlex-Manager zur Verfügung gestellt. SunPlex-Manager bietet eine topologische Ansicht der überwachten Plattenpfade. Die Ansicht wird alle 10 Minuten aktualisiert, um Informationen zur Anzahl der fehlgeschlagenen Pings zu liefern.

IPMP-Überwachung

Jeder Cluster-Knoten hat seine eigene IP-Netzwerk-Multipathing-Konfiguration, die sich von der auf anderen Cluster-Knoten unterscheiden kann.

IP-Netzwerk-Multipathing überwacht die folgenden Netzwerk-Kommunikationsfehler:

- Der Übertragungs- und Empfangspfad des Netzwerkadapters überträgt keine Pakete mehr.
- Die Verbindung des Netzwerkadapters mit der Verknüpfung ist zusammengebrochen.
- Der Port des Schalters überträgt bzw. empfängt keine Pakete.
- Die reale Schnittstelle in einer Gruppe ist beim Booten des Systems nicht vorhanden.

Quorum-Geräte

Ein Quorum-Gerät ist eine Platte, die von zwei oder mehr Knoten gemeinsam genutzt wird und Stimmen abgibt. Die Stimmen dienen der Feststellung des Quorums für den Betrieb des Clusters. Der Cluster kann nur arbeiten, wenn ein Quorum von Stimmen verfügbar ist. Das Quorum-Gerät wird verwendet, wenn ein Cluster in separate Knotensätze partitioniert wird, um festzulegen, welcher Knotensatz den neuen Cluster bildet.

Sowohl Cluster-Knoten als auch Quorum-Geräte geben eine Stimme für das Quorum ab. Standardmäßig erhalten Cluster-Knoten eine Stimmenanzahl von Eins für das Quorum, sobald sie booten und Cluster-Mitglieder werden. Knoten können eine Stimmenanzahl von Null haben, wenn der Knoten gerade installiert wird oder wenn ein Verwalter den Knoten in Wartungszustand versetzt hat.

Quorum-Geräte erhalten eine Stimmenanzahl für das Quorum, die sich nach der Anzahl von Knotenverbindungen mit dem Gerät richtet. Wenn ein Quorum-Gerät konfiguriert wird, erhält es eine maximale Stimmenanzahl von $N-1$, wobei N der Anzahl der mit dem Quorum-Gerät verbundenen Stimmen entspricht. Ein Quorum-Gerät, das zum Beispiel mit zwei Knoten mit einer Stimmenanzahl von nicht Null verbunden ist, hat einen Quorum-Zählwert von Eins (Zwei minus Eins).

Datenintegrität

Das Sun Cluster-System versucht, Datenbeschädigung zu verhindern und Datenintegrität sicherzustellen. Da Cluster-Knoten Daten und Ressourcen gemeinsam nutzen, darf ein Cluster nie in gleichzeitig aktive, getrennte Partitionen unterteilt werden. Der CMM stellt sicher, dass jeweils nur ein Cluster in Betrieb ist.

Zwei Arten von Problemen können aufgrund der Partitionierung von Clustern auftreten: Split Brain und Amnesie. Zum Split Brain kommt es, wenn der Cluster-Interconnect zwischen den Knoten verloren geht und der Cluster in Teil-Cluster zerfällt, die sich jeweils als die einzige Partition wahrnehmen. Ein Teil-Cluster, der keine weiteren Teil-Cluster wahrnimmt, kann Konflikte bei gemeinsam genutzten Ressourcen verursachen, wie zum Beispiel duplizierte Netzwerkadressen und Datenbeschädigung.

Amnesie tritt ein, wenn alle Knoten den Cluster in gestaffelten Gruppen verlassen. Beispiel: Ein Cluster hat zwei Knoten, A und B. Wenn A ausfällt, werden die Konfigurationsdaten im CCR nur auf Knoten B, nicht aber auf Knoten A aktualisiert. Wenn später Knoten B ausfällt, wird Knoten A neu gestartet und mit alten CCR-Inhalten ausgeführt. Dieser Zustand wird als Amnesie bezeichnet und kann dazu führen, dass ein Cluster mit veralteten Konfigurationsinformationen läuft.

Split Brain und Amnesie können vermieden werden, indem jeder Knoten eine Stimme erhält und eine Stimmenmehrzahl für den Betrieb eines Clusters vorgeschrieben wird. Eine Partition mit der Mehrheit der Stimmen hat ein *Quorum* und wird für den Betrieb aktiviert. Dieser Mechanismus der Stimmenmehrzahl funktioniert gut, wenn ein Cluster über mehr als zwei Knoten verfügt. In einem Zwei-Knoten-Cluster ist zwei eine Mehrheit. Wenn ein solcher Cluster in Partitionen zerfällt, sorgt eine externe Stimme für ein Quorum bei einer der Partitionen. Diese externe Stimme wird von einem Quorum-Gerät beigesteuert. Ein Quorum-Gerät kann jede Platte sein, die von beiden Knoten gemeinsam genutzt wird.

Tabelle 2-1 beschreibt, wie die Sun Cluster-Software Quorum zur Vermeidung von Split-Brain und Amnesie verwendet.

TABELLE 2-1 Cluster-Quorum und Split Brain- und Amnesie-Probleme

Partitionstyp	Quorum-Lösung
Split Brain	Ermöglicht nur der Partition (Teil-Cluster) mit Stimmenmehrzahl die Ausführung als Cluster. Nur eine Partition mit einer solchen Mehrheit ist möglich. Nachdem ein Knoten den Kampf um das Quorum verloren hat, gerät er in Panik.
Amnesie	Stellt sicher, dass beim Booten eines Clusters mindestens ein Knoten zum Cluster gehört, der Mitglied der letzten Cluster-Mitgliedschaft war (und somit über die neuesten Konfigurationsdaten verfügt).

Fehlerschutz

Ein wichtiges Thema bei Clustern ist ein Fehler, der zur Partitionierung des Clusters führt (als *Split Brain* bezeichnet). In diesem Fall können nicht mehr alle Knoten miteinander kommunizieren, so dass einzelne Knoten oder Knoten-Teilsätze ggf. versuchen, Einzel- oder Untermengen-Cluster zu bilden. Jede Untermenge oder Partition kann davon “überzeugt” sein, alleinigen Zugriff auf die Multihostplatten und die Eigentümerschaft zu haben. Wenn mehrere Knoten versuchen, auf die Platten zu schreiben, kann dies zu Datenbeschädigung führen.

Der Fehlerschutz schränkt den Knotenzugriff auf die Multihostplatten ein, indem der Zugriff auf die Platten verhindert wird. Wenn ein Knoten den Cluster verlässt (aufgrund eines Ausfalls oder Partitionierung), wird mit dem Fehlerschutz sichergestellt, dass der Knoten keinen Zugriff mehr auf die Platte hat. Nur aktuelle Mitgliederknoten haben Zugriff auf die Platten. Das sichert die Datenintegrität.

Das Sun Cluster-System verwendet SCSI-Plattenreservierungen zur Implementierung des Fehlerschutzes. Mit den SCSI-Reservierungen werden die Multihostplatten vor den ausgefallenen Knoten "geschützt" und der Zugriff auf diese Platten wird verhindert.

Wenn ein Cluster-Mitglied erkennt, dass ein anderer Knoten nicht mehr über den Cluster-Interconnect kommuniziert, leitet es ein Fehlerschutzverfahren ein, um den ausgefallenen Knoten am Zugriff auf die gemeinsam genutzten Platten zu hindern. Wenn dieser Fehlerschutz eintritt, gerät der geschützte Knoten in Panik, und eine Meldung zum "Reservierungskonflikt" wird auf seiner Konsole angezeigt.

Failfast-Mechanismus für den Fehlerschutz

Der Failfast-Mechanismus versetzt einen fehlerhaften Knoten in Panik, hindert ihn aber nicht an einem Neustart. Anschließend kann der Knoten neu booten und versuchen, wieder dem Cluster beizutreten.

Wenn ein Knoten die Konnektivität mit anderen Knoten im Cluster verliert und nicht zu einer Partition gehört, die ein Quorum erzielen kann, wird er erzwungenermaßen von einem anderen Knoten aus dem Cluster entfernt. Ein anderer Knoten, der Teil der Partition ist, die ein Quorum erzielen kann, belegt die gemeinsam genutzten Platten mit Reservierungen. Der Knoten ohne Quorum gerät dann infolge des Failfast-Mechanismus in Panik.

Geräten

Das globale Dateisystem macht alle Dateien im Cluster auf allen Knoten gleichermaßen verfügbar und sichtbar. Auf die gleiche Art und Weise macht die Sun Cluster-Software alle Geräte eines Clusters Cluster-weit verfügbar und sichtbar. Das heißt, dass das E/A-Teilsystem Zugriff auf ein beliebiges Gerät im Cluster von jedem Knoten aus ermöglicht, unabhängig davon, wo das Gerät real angeschlossen ist. Dieser Zugriff wird als globaler Gerätezugriff bezeichnet.

Globale Geräte

Sun Cluster-Systeme verwenden globale Geräte, um Cluster-weiten, hoch verfügbaren Zugriff auf alle Geräte im Cluster von jedem Knoten aus zu ermöglichen. Wenn ein Knoten ausfällt, während er Zugriff auf ein globales Gerät gewährt, schaltet die Sun Cluster-Software normalerweise auf einen anderen Pfad zu dem Gerät um und leitet den Zugriff auf diesen Pfad um. Diese Umleitung lässt sich mit globalen Geräten problemlos ausführen, weil pfadunabhängig immer der gleiche Gerätenamen verwendet wird. Der Zugriff auf ein Remote-Gerät erfolgt auf die gleiche Weise wie auf ein gleichnamiges lokales Gerät. Auch die API für den Zugriff auf ein globales Gerät auf dem Cluster ist die gleiche wie diejenige, die für den lokalen Zugriff auf ein Gerät verwendet wird.

Zu den globalen Geräten bei Sun Cluster gehören Platten, CD-ROMs und Bänder. Platten sind jedoch die einzigen unterstützten globalen Multiport-Geräte. Diese eingeschränkte Unterstützung bedeutet, dass CD-ROM- und Bandgeräte derzeit keine hoch verfügbaren Geräte sind. Die lokalen Platten auf jedem Server sind ebenfalls keine Multiport-Geräte und deswegen nicht hoch verfügbar.

Der Cluster weist jeder Platte, jedem CD-ROM-Laufwerk und jedem Bandgerät im Cluster eine einmalige ID zu. Diese Zuweisung ermöglicht einen konsistenten Zugriff auf jedes Gerät von jedem Cluster-Knoten aus.

Geräte-ID

Die Sun Cluster-Software verwaltet globale Geräte über einen so genannten Geräte-ID-Treiber (DID-Treiber). Dieser Treiber wird verwendet, um automatisch jedem Gerät im Cluster einschließlich Multihostplatten, Bandlaufwerken und CD-ROM-Laufwerken eine einmalige ID zuzuweisen.

Der DID-Treiber ist fester Bestandteil der Zugriffsfunktion auf globale Geräte des Clusters. Der DID-Treiber sendet ein Testsignal an alle Knoten des Clusters und erstellt eine Liste der einmaligen Plattengeräte. Er weist auch jedem Gerät eine einmalige Geräteklassen- und Gerätenummer zu, die über alle Cluster-Knoten hinweg konsistent ist. Der Zugriff auf die globalen Geräte erfolgt über die einmalige, vom DID-Treiber zugewiesene DID, nicht über die üblichen Solaris-DIDs.

Damit wird sichergestellt, dass alle Anwendungen, die auf Platten zugreifen, wie zum Beispiel Solaris Volume Manager oder Sun Java System Directory Server, Cluster-weit einen konsistenten Pfad verwenden. Diese Konsistenz ist besonders für Multihost-Platten wichtig, da die lokalen Geräteklassen- und Gerätenummern für die einzelnen Geräte von Knoten zu Knoten unterschiedlich sein können. Die Nummern können auch die Solaris-Konventionen zur Gerätebenennung ändern.

Lokale Geräte

Die Sun Cluster-Software verwaltet auch lokale Geräte. Auf solche Geräte kann nur auf einem Knoten zugegriffen werden, der einen Dienst ausführt und real an den Cluster angeschlossen ist. Lokale Geräte erbringen möglicherweise bessere Leistung als globale Geräte, da sie keine Zustandsinformationen auf mehreren Knoten gleichzeitig replizieren müssen. Wenn die Gerätedomäne versagt, entfällt auch der Zugriff auf das Gerät, wenn es nicht von mehreren Knoten gemeinsam genutzt werden kann.

Plattengerätegruppen

Durch Plattengerätegruppen werden Datenträger-Manager-Plattengruppen "global", da hierdurch den zugrunde liegenden Platten Multipath- und Multihost-Unterstützung zur Verfügung gestellt wird. Jeder mit den Multihostplatten real verbundener Cluster-Knoten stellt einen Pfad zur Plattengerätegruppe zur Verfügung.

Im Sun Cluster-System können Multihostplatten von der Sun Cluster-Software gesteuert werden, wenn sie als Plattengerätegruppen registriert werden. Durch die Registrierung erhält das Sun Cluster-System Informationen darüber, welche Knoten über Pfade zu welchen Datenträger-Manager-Plattengruppen verfügen. Die Sun Cluster-Software erstellt eine im raw-Modus betriebene Plattengerätegruppe für jedes Plattengerät und Bandlaufwerk im Cluster. Diese Cluster-Gerätegruppen bleiben so lange im Zustand "Offline", bis auf sie als globale Geräte zugegriffen wird. Dies geschieht entweder durch Einhängen eines globalen Dateisystems oder durch Zugriff auf eine raw-Datenbankdatei.

Datendienste

Ein Datendienst besteht aus einer Kombination aus Software- und Konfigurationsdateien, mit deren Hilfe eine Anwendung ohne Änderung in einer Sun Cluster-Konfiguration ausgeführt werden kann. Wenn eine Sun Cluster-Konfiguration ausgeführt wird, läuft eine Anwendung als Ressource unter der Steuerung von Ressourcengruppen-Manager (RGM). Mit einem Datendienst kann eine Anwendung wie Sun Java System Web Server oder eine Oracle-Datenbank für die Ausführung auf einem Cluster anstelle eines einzelnen Servers konfiguriert werden.

Die Datendienstsoftware implementiert Sun Cluster-Verwaltungsmethoden, die folgende Vorgänge an der Anwendung ausführen:

- Starten der Anwendung
- Stoppen der Anwendung
- Überwachen der Fehler in der Anwendung und Wiederherstellen nach Fehlern

Die Konfigurationsdateien eines Datendienstes definieren die Eigenschaften der Ressource, die eine Anwendung für RGM darstellt.

RGM steuert die Anordnung der Failover- und Scalable-Datendienste im Cluster. RGM ist für das Starten und Stoppen der Datendienste auf ausgewählten Cluster-Knoten zuständig, als Reaktion auf Änderungen bei der Cluster-Mitgliedschaft. RGM aktiviert Datendienstanwendungen für die Nutzung des Cluster-Frameworks.

RGM steuert Datendienste als Ressourcen. Diese Implementierungen werden entweder von Sun bereitgestellt oder von einem Entwickler erstellt. Der Entwickler kann eine generische Datendienstvorlage, die Datendienst-Entwicklungsbibliotheks-API (DSDL-API) oder die Ressourcenverwaltungs-API (RMAPI) nutzen. Der Cluster-Verwalter erstellt und verwaltet Ressourcen in Containern, die als Ressourcengruppen bezeichnet werden. Durch RGM- oder Verwalteraktionen werden Ressourcen und Ressourcengruppen in den Zustand "Online" bzw. "Offline" versetzt.

Ressourcentypen

Ein Ressourcentyp ist eine Sammlung von Eigenschaften, die eine Anwendung für den Cluster beschreiben. Diese Sammlung enthält Informationen darüber, wie die Anwendung auf den Cluster-Knoten gestartet, gestoppt und überwacht wird. Ein Ressourcentyp enthält auch anwendungsspezifische Eigenschaften, die für die Verwendung der Anwendung im Cluster definiert werden müssen. Sun Cluster-Datendienste verfügen über mehrere vordefinierte Ressourcentypen. Sun Cluster HA für Oracle ist zum Beispiel der Ressourcentyp `SUNW.oracle-server` und Sun Cluster HA für Apache der Ressourcentyp `SUNW.apache`.

Ressourcen

Eine Ressource ist eine Instanz eines Ressourcentyps, der Cluster-weit definiert ist. Der Ressourcentyp aktiviert mehrere Instanzen einer auf dem Cluster zu installierenden Anwendung. Nach Initialisieren einer Ressource weist RGM den anwendungsspezifischen Eigenschaften Werte zu, und die Ressource übernimmt alle auf Ressourcentypebene vorhandenen Eigenschaften.

Datendienste verwenden mehrere Typen von Ressourcen. Anwendungen wie Apache Web Server oder Sun Java System Web Server verwenden Netzwerkadressen (logische Hostnamen und gemeinsam genutzte Adressen), von denen die Anwendungen abhängen. Anwendung und Netzwerkressourcen bilden eine grundlegende Einheit, die vom RGM verwaltet wird.

Ressourcengruppen

Vom RGM verwaltete Ressourcen werden in Ressourcengruppen zusammengefasst und können so als Einheit verwaltet werden. Eine Ressourcengruppe besteht aus verwandten oder voneinander abhängigen Ressourcen. Eine Ressource, die vom Ressourcentyp `SUNW.LogicalHostname` abgeleitet wurde, kann zum Beispiel in derselben Ressourcengruppe wie eine Ressource des Oracle-Datenbank-Ressourcentyps abgelegt werden. Eine Ressourcengruppe migriert als Einheit, wenn ein Failover oder ein Switchover für die Ressourcengruppe eingeleitet wird.

Datendiensttypen

Mithilfe von Datendiensten können Anwendungen hoch verfügbar gemacht werden. Scalable-Dienste tragen zur Vermeidung signifikanter Anwendungsausfälle nach einzelnen Fehlern in einem Cluster bei.

Datendienste müssen als einer der folgenden Datendiensttypen konfiguriert werden:

- Failover-Datendienst
- Skalierbarer Datendienst
- Paralleler Datendienst

Failover-Datendienste

Failover ist der Vorgang, bei dem der Cluster automatisch einen Dienst von einem ausgefallenen Primärknoten auf einen dafür vorgesehenen redundanten Sekundärknoten verschiebt. Failover-Anwendungen weisen folgende Merkmale auf:

- Sie können auf einem einzigen Cluster-Knoten ausgeführt werden.
- Keine Cluster-Unterstützung
- Vom Cluster-Framework für hohe Verfügbarkeit abhängig

Wenn der Fehler-Monitor einen Fehler erkennt, versucht er entweder, die Instanz auf demselben Knoten erneut zu starten oder die Instanz auf einem anderen Knoten zu starten (Failover), je nachdem, wie der Datendienst konfiguriert wurde.

Failover-Dienste verwenden eine Failover-Ressourcengruppe, die Ressourcen für Anwendungsinstanzen und Netzwerkressourcen enthält (logische Hostnamen). Logische Hostnamen sind IP-Adressen, die auf einem Knoten als aktiv konfiguriert werden können. Später werden sie auf dem Originalknoten automatisch als inaktiv und auf einem anderen Knoten als aktiv konfiguriert.

Der Client nimmt möglicherweise eine kurze Unterbrechung im Dienst wahr und muss ggf. die Verbindung neu herstellen, nachdem das Failover abgeschlossen ist. Er nimmt jedoch den Wechsel des realen Servers, der den Dienst zur Verfügung stellt, nicht wahr.

Skalierbare Datendienste

Der Scalable-Datendienst ermöglicht die gleichzeitige Ausführung der Anwendungsinstanzen auf mehreren Knoten. Scalable-Dienste verwenden zwei Ressourcengruppen. Die Scalable-Ressourcengruppe enthält die Anwendungsressourcen, während die Failover-Ressourcengruppe die Netzwerkressourcen (gemeinsam genutzte Adressen) enthält, von denen der Scalable-Dienst abhängt. Die Scalable-Ressourcengruppe kann auf mehreren Knoten online sein, so dass mehrere Instanzen dieses Dienstes gleichzeitig ausgeführt werden können. Die Failover-Ressourcengruppe, welche die gemeinsam genutzten Adressen hostet, ist jeweils nur auf einem Knoten online. Alle Knoten, die einen Scalable-Dienst hosten, verwenden dieselbe gemeinsam genutzte Adresse, um den Dienst zu hosten.

Der Cluster erhält die Dienstanforderungen über eine einzige Netzwerkschnittstelle (die globale Schnittstelle). Diese Anforderungen werden auf Grundlage mehrerer vordefinierter Algorithmen, die in den Lastausgleichsverfahren festgelegt sind, an die Knoten verteilt. Der Cluster kann die Dienstlast mithilfe des Lastausgleichsverfahrens zwischen mehreren Knoten ausgleichen.

Parallele Anwendungen

Sun Cluster-Systeme stellen eine Umgebung zur Verfügung, in der parallele Datenbanken für die gemeinsame, parallele Ausführung von Anwendungen auf allen Cluster-Knoten verwendet werden. Sun Cluster Support für Oracle Parallel Server/Real Application Clusters ist ein Satz Pakete, deren Installation die Ausführung von Oracle Parallel Server/Real Application Clusters auf Sun Cluster-Knoten ermöglicht. Dieser Datendienst ermöglicht auch Sun Cluster Support für Oracle Parallel Server/Real Application Clusters für die Verwaltung mit Sun Cluster-Befehlen.

Eine parallele Anwendung wurde für das Ausführen in einer Cluster-Umgebung eingerichtet, so dass die Anwendung von zwei oder mehr Knoten gleichzeitig unterstützt werden kann. In einer Oracle Parallel Server/Real Application Clusters-Umgebung arbeiten mehrere Oracle-Instanzen zusammen, um Zugriff auf dieselbe, gemeinsam genutzte Datenbank zu gewähren. Die Oracle-Clients können jede beliebige dieser Instanzen für den Datenbankzugriff verwenden. Wenn also eine oder mehrere Instanzen ausfallen, können die Clients Verbindungen mit einer der noch laufenden Instanzen herstellen und weiterhin auf die Datenbank zugreifen.

Sun Cluster-Architektur

Mithilfe der Sun Cluster-Architektur können Systemgruppen als ein einziges, großes System bereitgestellt, verwaltet und angezeigt werden.

Dieses Kapitel enthält die folgenden Abschnitte:

- „Sun Cluster-Hardwareumgebung ” auf Seite 29
- „Sun Cluster-Softwareumgebung ” auf Seite 30
- „Skalierbare Datendienste ” auf Seite 33
- „Multihost-Plattenspeicher ” auf Seite 35
- „Cluster-Interconnect ” auf Seite 36
- „IPMP-Gruppen ” auf Seite 37

Sun Cluster-Hardwareumgebung

Ein Cluster setzt sich aus den folgenden Hardwarekomponenten zusammen:

- Cluster-Knoten mit lokalen, nicht gemeinsam genutzten Platten bilden die Hauptcomputerplattform des Clusters.
- Über Multihost-Speicher werden Platten bereitgestellt, die von den Knoten gemeinsam genutzt werden.
- Wechselmedien wie Bänder und CD-ROM sind als globale Geräte konfiguriert.
- Cluster-Interconnect ist der Kommunikationskanal zwischen den Knoten.
- Öffentliche Netzwerkschnittstellen aktivieren die Netzwerkschnittstellen, die von den Client-Systemen für den Zugriff auf Datendienste im Cluster verwendet werden.

Abbildung 3-1 zeigt, wie die Hardwarekomponenten zusammenarbeiten.

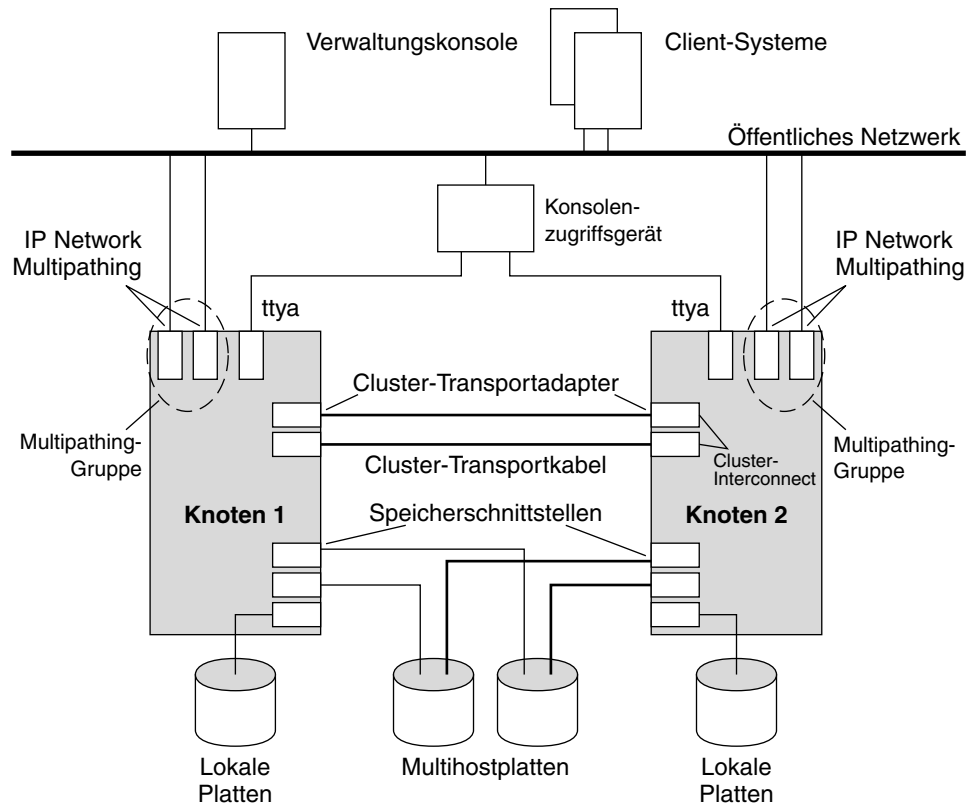


ABBILDUNG 3-1 Sun Cluster-Hardwarekomponenten

Sun Cluster-Softwareumgebung

Auf einem Knoten, der als Cluster-Mitglied eingesetzt werden soll, muss folgende Software installiert sein:

- Solaris-Software
- Sun Cluster-Software
- Datendiensteanwendung
- Datenträgerverwaltung (Solaris™ Volume Manager oder VERITAS Volume Manager)

Davon ausgenommen ist eine Konfiguration, die vorinstallierte Datenträgerverwaltung verwendet. Für eine solche Konfiguration ist möglicherweise kein Software-Datenträger-Manager erforderlich.

Abbildung 3-2 zeigt auf höchster Ebene die Softwarekomponenten, die gemeinsam die Sun Cluster-Softwareumgebung bilden.

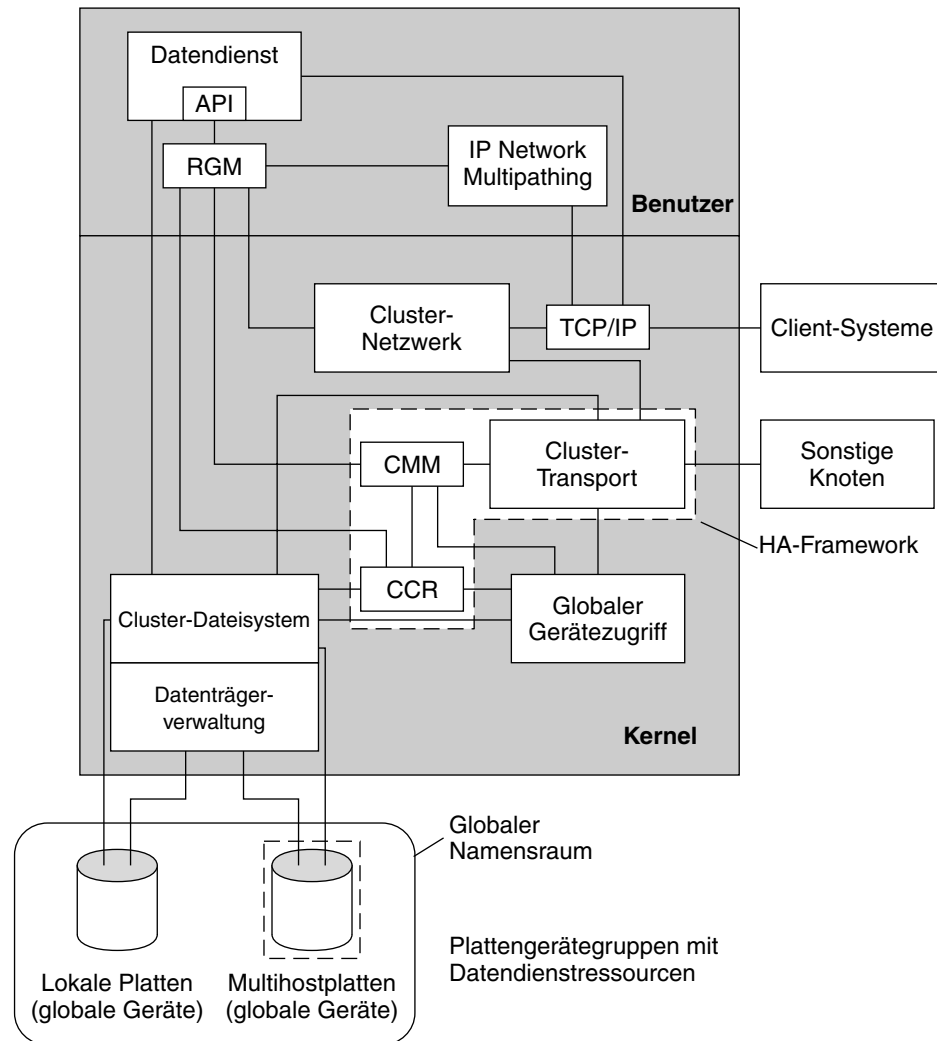


ABBILDUNG 3-2 Sun Cluster-Softwarearchitektur

Cluster-Mitgliedschaftsüberwachung (Cluster Membership Monitor, CMM)

Alle Knoten müssen eine konsistente Einigung hinsichtlich der Cluster-Mitgliedschaft erzielen, um die Daten vor Beschädigung zu schützen. Der CMM koordiniert gegebenenfalls eine Cluster-Rekonfiguration der Cluster-Dienste infolge eines Fehlers.

Der CMM erhält Informationen über die Konnektivität mit anderen Knoten aus der Cluster-Transportschicht. Der CMM verwendet während einer Rekonfiguration den Cluster-Interconnect zum Austauschen von Statusinformationen.

Nach dem Erkennen einer Änderung bei der Cluster-Mitgliedschaft führt der CMM eine synchronisierte Konfiguration des Clusters aus. Bei dieser Konfiguration können Cluster-Ressourcen auf Grundlage der neuen Mitgliedschaft im Cluster neu verteilt werden.

Der CMM wird vollständig im Kernel ausgeführt.

Cluster-Konfigurations-Repository (Cluster Configuration Repository, CCR)

Das CCR stellt mithilfe des CMM sicher, dass ein Cluster nur mit einem festgelegten Quorum läuft. Das CCR ist dafür zuständig, die Datenkonsistenz im ganzen Cluster zu überprüfen, eine ggf. erforderliche Wiederherstellung durchzuführen und Aktualisierungen für die Daten bereitzustellen.

Cluster-Dateisysteme

Ein Cluster-Dateisystem ist ein Proxy zwischen den folgenden Komponenten:

- Dem Kernel auf einem Knoten und dem zugrunde liegenden Dateisystem.
- Dem Datenträger-Manager, der auf einem Knoten mit realer Verbindung mit der Platte bzw. den Platten ausgeführt wird.

Cluster-Dateisysteme hängen von globalen Geräten (Platten, Bänder, CD-ROMs) ab. Der Zugriff auf globale Geräte ist von einem beliebigen Knoten im Cluster über den gleichen Dateinamen möglich, zum Beispiel `/dev/global/`. Für den Knoten ist keine reale Verbindung mit dem Speichergerät erforderlich. Sie können ein globales Gerät genau so verwenden wie ein normales Gerät, das heißt, Sie können mit den Befehlen `newfs` bzw. `mkfs` Dateisysteme darauf erstellen.

Das Cluster-Dateisystem hat folgende Merkmale:

- Die Speicherorte für den Dateizugriff sind transparent. Ein Prozess kann eine Datei an einem beliebigen Speicherort im System öffnen. Zudem können die Prozesse die Datei auf allen Knoten mithilfe desselben Pfadnamens suchen.

Hinweis – Wenn das Cluster-Dateisystem Dateien liest, wird die Zugriffszeit für diese Dateien nicht aktualisiert.

- Mit Kohärenzprotokollen wird die UNIX-Dateizugriffssemantik auch dann bewahrt, wenn mehrere Knoten gleichzeitig auf die Datei zugreifen.
- Umfassendes Caching wird zusammen mit Bulk-E/A-Bewegungen ohne Zwischenspeicherung (Zero-Copy) eingesetzt, um Dateidaten effizient zu verschieben.
- Das Cluster-Dateisystem stellt mit den `fcntl(2)`-Schnittstellen eine hoch verfügbare kooperative Dateisperrfunktion zur Verfügung. Anwendungen, die auf mehreren Cluster-Knoten ausgeführt werden, können den Datenzugriff synchronisieren, indem sie die kooperative Dateisperrung auf eine Cluster-Dateisystemdatei anwenden. Dateisperren werden sofort von allen den Cluster verlassenden Knoten und von allen gesperrten und fehlgeschlagenen Anwendungen aufgehoben.
- Kontinuierlicher Datenzugriff ist auch bei Ausfällen gesichert. Anwendungen sind von den Ausfällen nicht betroffen, solange noch ein Pfad zu den Platten funktionsfähig ist. Diese Garantie gilt für den Plattenzugriff im raw-Modus und für alle Dateisystemvorgänge.
- Cluster-Dateisysteme sind vom zugrunde liegenden Dateisystem und von der Datenträgerverwaltung unabhängig. Cluster-Dateisysteme machen jedes unterstützte, auf den Platten vorhandene Dateisystem global.

Skalierbare Datendienste

Das wichtigste Ziel eines Cluster-Netzwerks besteht darin, Skalierbarkeit der Datendienste zu ermöglichen. Skalierbarkeit bedeutet, dass ein Datendienst trotz steigender Belastung eine konstante Antwortzeit beibehalten kann, weil dem Cluster neue Knoten hinzugefügt und neue Serverinstanzen ausgeführt werden. Ein gutes Beispiel für einen Scalable-Datendienst ist ein Webdienst. In der Regel besteht ein Scalable-Datendienst aus mehreren Instanzen, von denen jede auf unterschiedlichen Cluster-Knoten ausgeführt wird. Insgesamt verhalten sich diese Instanzen für einen Remote-Client wie ein einziger Dienst und führen die Funktionen dieses Dienstes aus. Bei einem Scalable-Webdienst mit mehreren `httpd`-Dämonen, die auf

unterschiedlichen Knoten ausgeführt werden, kann jeder beliebige Dämon zum Bedienen einer Client-Anforderung eingesetzt werden. Welcher Dämon die Anforderung bedient, hängt vom *Lastausgleichsverfahren* ab. Die Antwort an den Client kommt scheinbar vom Dienst und nicht vom konkreten Dämon, der die Anforderung bedient hat, so dass der Eindruck eines einzigen Dienstes gewahrt wird.

Die nachstehende Abbildung zeigt eine Scalable-Dienst-Architektur.

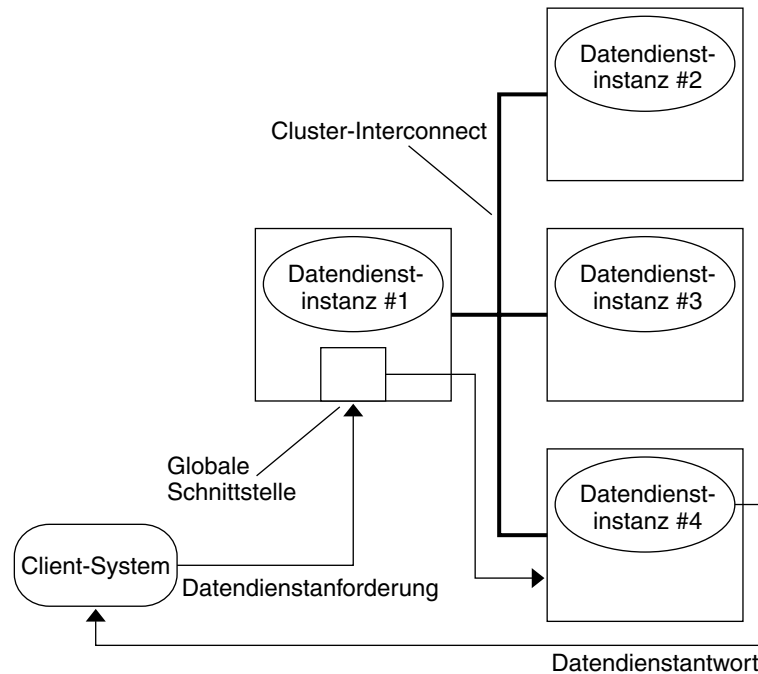


ABBILDUNG 3-3 Scalable-Datendienst-Architektur

Die Knoten, welche die globale Schnittstelle nicht hosten (Proxy-Knoten), hosten die gemeinsam genutzte Adresse auf ihren Schleifenschnittstellen. Die an der globalen Schnittstelle eintreffenden Pakete werden anhand konfigurierbarer Lastausgleichsverfahren auf andere Cluster-Knoten verteilt. Die möglichen Lastausgleichsverfahren werden im Folgenden beschrieben.

Lastausgleichsverfahren

Der Lastausgleich verbessert die Leistung der Scalable-Dienste sowohl bei der Antwortzeit als auch beim Durchsatz.

Es gibt zwei Klassen von Scalable-Datendiensten: *reine* Dienste und *Sticky*-Dienste. Ein *reiner* Dienst ist ein Dienst, bei dem jede Instanz Client-Anforderungen beantworten kann. Bei einem *Sticky*-Dienst sorgt der Cluster für den Lastausgleich bei Anforderungen an den Knoten. Diese Anforderungen werden nicht an andere Instanzen umgeleitet.

Ein *reiner* Dienst verwendet ein gewichtetes Lastausgleichsverfahren. Bei diesem Lastausgleichsverfahren werden Client-Anforderungen standardmäßig gleichmäßig auf die Serverinstanzen im Cluster verteilt. Wenn zum Beispiel bei einem Drei-Knoten-Cluster jedem Knoten die Gewichtung 1 zugewiesen ist, bedient jeder Knoten ein Drittel der Anforderungen eines beliebigen Clients für diesen Dienst. Gewichtungen können jederzeit über die `scrgadm (1M)`-Befehlschnittstelle bzw. über die GUI von SunPlex-Manager geändert werden.

Es gibt zwei Typen von *Sticky*-Diensten: *Normal-Sticky* und *Platzhalter-Sticky*. Mit *Sticky*-Diensten können Sitzungen auf Anwendungsebene gleichzeitig über mehrere TCP-Verbindungen ausgeführt werden, um den Zustandsspeicher (Anwendungssitzungszustand) gemeinsam zu nutzen.

Mit normalen *Sticky*-Diensten kann ein Client den Zustand zwischen mehreren gleichzeitigen TCP-Verbindungen teilen. Ein Client wird als "*Sticky*" in Bezug auf die Serverinstanz bezeichnet, die einen einzigen Port abhört. Der Client hat die Sicherheit, dass alle seine Anforderungen an dieselbe Serverinstanz gehen. Voraussetzung hierfür ist, dass die Instanz aktiv und zugänglich bleibt und das Lastausgleichsverfahren nicht geändert wird, solange der Dienst online ist.

Platzhalter-Sticky-Dienste verwenden dynamisch zugewiesene Port-Nummern, erwarten jedoch trotzdem, dass die Client-Anforderungen an denselben Knoten geleitet werden. Der Client ist in Bezug auf dieselbe IP-Adresse "*Sticky-Platzhalter*" bei den Ports.

Multihost-Plattenspeicher

Die Sun Cluster-Software macht Platten mithilfe des Multihost-Plattenspeichers hoch verfügbar. Dieser Plattenspeicher kann an mehrere Knoten gleichzeitig angeschlossen werden. Mithilfe von Datenträgerverwaltungs-Software können diese Platten in einem gemeinsam genutzten Speicher mit einem Cluster-Knoten als Master angeordnet werden. Daraufhin werden die Platten so konfiguriert, dass sie bei Ausfällen auf einen anderen Knoten verschoben werden. Die Verwendung von Multihost-Platten für Sun Cluster-Systeme bietet zahlreiche Vorteile, unter anderem die folgenden:

- Globaler Zugriff auf Dateisysteme
- Mehrere Zugriffspfade auf Dateisysteme und Daten
- Toleranz einzelner Knotenausfälle

Cluster-Interconnect

Alle Knoten müssen mit dem Cluster-Interconnect über mindestens zwei redundante, real unabhängige Netzwerke oder Pfade verbunden sein, um einen Single Point of Failure (Ausfallpunkt) zu vermeiden. Für Redundanz sind zwei Interconnects erforderlich. Es können jedoch bis zu sechs verwendet werden, um den Datenverkehr zu verteilen, Engpässe zu vermeiden und Redundanz und Skalierbarkeit zu verbessern. Der Sun Cluster-Interconnect verwendet Fast Ethernet, Gigabit-Ethernet, Sun Fire Link oder Scalable Coherent Interface (SCI, IEEE 1596-1992) und ermöglicht so hoch leistungsfähige Cluster-private Kommunikation.

In Cluster-Umgebungen sind Hochgeschwindigkeits-Interconnects und -Protokolle mit geringer Verzögerung für die Kommunikation zwischen den Knoten entscheidend. SCI-Interconnect in Sun Cluster-Systemen bietet verbesserte Leistung im Vergleich zu üblichen Netzwerk-Schnittstellenkarten (Network Interface Cards, NICs). Sun Cluster verwendet die RSM™-Schnittstelle (Remote Shared Memory) für die Kommunikation zwischen den Knoten über ein Sun Fire Link-Netzwerk. RSM ist eine Sun-Schnittstelle zur Meldungsübertragung, die bei Remote-Speichervorgängen sehr effizient ist.

Der RSMRDT-Treiber (RSM Reliable Datagram Transport) umfasst einen Treiber, der auf der RSM-API basiert und eine Bibliothek, die die RSMRDT-API-Schnittstelle exportiert. Dieser Treiber verbessert die Leistung der Oracle Parallel Server/Real Application-Cluster. Darüber hinaus verbessert er die Lastausgleichs- und Hochverfügbarkeits-Funktionen (HA) durch direkte Bereitstellung im Treiber und somit transparente Verfügbarkeit für die Clients.

Der Cluster-Interconnect besteht aus folgenden Hardwarekomponenten:

- *Adapter* – Die Netzwerk-Schnittstellenkarten, die sich in jedem Cluster-Knoten befinden. Ein Netzwerkadapter mit mehreren Schnittstellen kann zu einem Single Point of Failure werden, wenn der ganze Adapter ausfällt.
- *Verbindungspunkte* – Die Schalter, die sich außerhalb des Cluster-Knotens befinden. Verbindungspunkte übernehmen Pass-Through- und Umschaltfunktionen, damit Sie mehr als zwei Knoten zusammenschließen können. In einem Zwei-Knoten-Cluster benötigen Sie keine Verbindungspunkte, weil die Knoten mit redundanten realen Kabeln direkt miteinander verbunden werden können. Diese redundanten Kabel werden auf jedem Knoten mit redundanten Adaptern verbunden. Für Konfigurationen mit mehr als zwei Knoten sind Verbindungspunkte erforderlich.
- *Kabel* – Die realen Verbindungen, entweder zwischen zwei Netzwerkadaptern oder zwischen einem Adapter und einem Verbindungspunkt.

Abbildung 3–4 zeigt, wie die drei Komponenten miteinander verbunden sind.

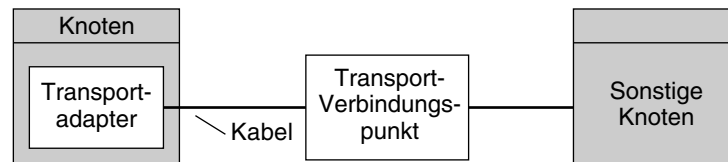


ABBILDUNG 3-4 Cluster-Interconnect

IPMP-Gruppen

Öffentliche Netzwerkadapter sind als IPMP-Gruppen (Multipathing-Gruppen) organisiert. Jede Multipathing-Gruppe hat mindestens einen öffentlichen Netzwerkadapter. Jeder Adapter in einer Multipathing-Gruppe kann aktiv sein oder Sie können Standby-Schnittstellen konfigurieren, die bis zum Auftreten eines Failovers inaktiv bleiben.

Multipathing-Gruppen bilden die Grundlage für logische Hostnamen und gemeinsam genutzte Adressen. Dieselbe Multipathing-Gruppe kann auf einem Knoten eine beliebige Anzahl logischer Hostnamen oder gemeinsam genutzter Adressen hosten. Multipathing kann erstellt werden, um die Konnektivität von Cluster-Knoten mit dem öffentlichen Netzwerk zu überwachen.

Weitere Informationen zu logischen Hostnamen und gemeinsam genutzten Adressen finden Sie im *Sun Cluster Data Services Planning and Administration Guide for Solaris OS*.

Öffentliche Netzwerkschnittstellen

Die Verbindung der Clients mit dem Cluster erfolgt über die öffentlichen Netzwerkschnittstellen. Jede Netzwerkadapterkarte kann an ein oder mehrere öffentliche Netzwerke angeschlossen sein, je nachdem, ob die Karte mehrere Hardwareschnittstellen hat. Sie können Knoten mit mehreren öffentlichen Netzwerk-Schnittstellenkarten einrichten, die so konfiguriert sind, dass mehrere Karten aktiv sind und dadurch gegenseitig als Failover-Sicherung dienen. Wenn einer der Adapter ausfällt, wird die Solaris-IP-Multipathing-Software für Sun Cluster aufgerufen, um für die fehlerhafte Schnittstelle ein Failover zu einem anderen Adapter in der Gruppe auszuführen.

Index

A

Adapter, *Siehe* Netzwerk, Adapter
Agenten, *Siehe* Datendienste
Amnesie, 21-22
Anwendungen
 Siehe auch Datendienste
 Fehlertoleranz, 9-13
 Hohe Verfügbarkeit, 9-13
 parallel, 11, 28
 Überwachung, 14

B

Befehlszeilenschnittstelle (CLI), 15

C

Cluster
 Campus, 13
 Dateisysteme, 13, 32-33
 Interconnects, 18, 36-37
 Knoten, 17-18
 Kommunikation, 18
 Konfiguration, 19, 32
 Mitglieder, 17, 32
 Mitgliedschaft, 18-19
 öffentliches Netzwerk, 37
 Partitionierung, 21-22
Cluster Configuration Repository (CCR), 19, 32
Cluster Membership Monitor (CMM), 18-19, 32

D

Data Service Development Library API
 (DSDL-API), 25-28
Dateisperrung, 33
Dateisysteme
 Cluster, 13, 32-33
 einhängen, 32-33
Datenbanken, 11
Datendienste
 Definition, 25-28
 Failover, 27-28
 Fehlerüberwachung, 14
 parallel, 28
 Ressourcen, 26-27
 Ressourcengruppen, 27
 Ressourcentypen, 26
 skalierbar
 Architektur, 33-35
 rein, 34-35
 Ressourcen, 28
 sticky, 34-35
 Typen, 27-28
Datenintegrität, 21-22
Datenträgermanagement, 12
Datenträgerverwaltung, 35
Dienste, *Siehe* Datendienste

E

Einhängen, 32-33

F

- Failfast, 23
- Failover
 - Bereitstellung durch Oracle Parallel Server/Real Application Clusters-Software, 28
 - Datendienste, 27-28
 - Transparent, 10
- Failover-, Dienste, 11
- Fehler
 - Erkennung, 14
 - Hardware und Software, 14
 - Schutz, 22-23
- Fehlertoleranz, 9-13

G

- Gemeinsam genutzte Adresse, skalierbare Datendienste, 28
- Gemeinsam genutzte Plattengruppen, 28
- Geräte
 - globale, 24
 - Gruppen, 25
 - IDs (DIDs), 24
 - lokal, 25
 - Quorum, 21-23
- Globale Geräte
 - Beschreibung, 24
 - einhängen, 32-33
 - Plattengerätegruppen, 25
- Globaler Namensraum, 24

H

- Hardware
 - Cluster-Interconnects, 36
 - Cluster-Knoten, 17-18
 - Fehler, 14
 - Hohe Verfügbarkeit, 9-13
 - Redundant Array of Independent Disks (RAID), 12-13
 - Sun StorEdge Traffic Manager, 12
 - Umgebung, 29-30
- Hohe Verfügbarkeit, 9-13

I

- IDs, Gerät, 24
- Interconnects, *Siehe* Cluster, Interconnects
- Internet Protocol (IP), 28
- IP-Netzwerk-Multipathing, 11, 21, 37
- IPMP
 - Siehe* IP-Netzwerk-Multipathing

K

- Knoten, 17-18
- Komponenten
 - Hardware, 29-30
 - Software, 30-33
- Konfiguration
 - Parallele Datenbank, 17
 - Repository, 19, 32
- Konfigurations, tools, 14-16

L

- Lastausgleich
 - Beschreibung, 33
 - Verfahren, 34-35
- Logischer Hostname,
 - Failover-Datendienste, 27-28
- Lokale Geräte, 25

M

- Mitgliedschaft, 17, 18-19, 32
- Multihost-Speicher, 11-13
- Multipathing, 11, 14, 21, 37

N

- Netzwerk
 - Adapter, 11, 21, 37
 - Lastausgleich, 33, 34-35
 - öffentlich
 - Beschreibung, 37
 - Öffentlich
 - IP-Netzwerk-Multipathing, 11

Netzwerk (Fortsetzung)
 öffentlich
 IP-Netzwerk-Multipathing, 21, 37
 Öffentlich
 Überwachung, 14
 Schnittstellen, 11, 37

O

Öffentliches Netzwerk, *Siehe* Netzwerk,
 öffentlich
Oracle Parallel Server/Real Application
 Clusters, 11-13

P

Panik, 23
Parallel
 Anwendungen, 28
 Datenbanken, 11, 17
Parallele, Anwendungen, 11
Partitionierung, Cluster, 21-22
Platten
 Fehlerschutz, 22-23
 Gerätegruppen, 25
 globale Geräte, 24
 lokale, 24
 Management, 12
 Multihost, 11-13, 24, 25, 35
 Quorum, 21-23
 Spiegelung, 12
Plattenpfadüberwachung (DPM), 20

Q

Quorum, 21-23

R

Redundant Array of Independent Disks
 (RAID), 12-13
Redundanz
 Hardware, 12-13
 Plattensysteme, 9-13

Repository, 19, 32
Reservierungskonflikt, 23
Resource Group Manager (RGM)
 Funktionalität, 25-28
 Ressourcengruppen, und, 27
Resource Management API (RMAPI), 25-28
Ressourcen
 Definition, 26-27
 Gemeinsame Nutzung, 13
 Gruppen
 Beschreibung, 27
 Failover, 27-28
 Typen, 26
 Wiederherstellung, 10
Rollenbasierte Zugriffskontrolle (RBAC), 16

S

Scalable-, Dienste, 11
scdpm Befehl, 20
Schnittstellen, 21, 37
Schutz, 22-23
SCSI, 22-23, 23
Skalierbar
 Datendienste, 28
skalierbar
 Datendienste
 Architektur, 33-35
Skalierbar
 Ressourcengruppen, 28
Skalierbarkeit, *Siehe* Scalable
Software
 Fehler, 14
 Hohe Verfügbarkeit, 9-13
 Host-basiert, 12-13
 Komponenten, 30-33
 Redundant Array of Independent Disks
 (RAID), 12-13
Solaris Volume Manager, 12
Speicher
 Array, 12-13
 Multihost, 11-13, 35
 Verwaltung, 11-13
Split-Brain, 21-22, 22-23
Stimmenanzahl, Quorum, 21-23
Sun Cluster Support für Oracle Parallel
 Server/Real Application Clusters, 28

Sun Management Center, 16
Sun StorEdge Traffic Manager, 12
SunPlex-Manager, 14-15, 20

T

Tools, 14-16
Traffic Manager, 12
Treiber, *Siehe* Geräte, IDs (DIDs)

U

Überwachung
 Fehler, 14
 Netzwerkschnittstellen, 21
 Plattenpfade, 20
Umgebung
 Hardware, 29-30
 Software, 30-33

V

Verfügbarkeitsmanagement, 10
VERITAS Volume Manager (VxVM), 12
Verwaltungs, tools, 14-16

W

Wiederherstellung, 9-13

Z

Zugriffskontrolle, 16