



Sun Cluster 3.0 U1 개념

Sun Microsystems, Inc.
901 San Antonio Road
Palo Alto, CA 94303-4900
U.S.A. 650-960-1300

부품번호: 816-1955-10
2001년 8월, Revision A

Copyright 2001 Sun Microsystems, Inc. 901 San Antonio Road, Palo Alto, California 94303-4900 U.S.A. 모든 권리는 저작권자의 소유입니다. 이 제품이나 문서는 저작권에 의해 보호 받고, 사용, 복사, 배포 및 역컴파일을 제한하는 사용권에 따라 배포됩니다. Sun과 권리 양도 사용자(있는 경우)의 사전 서면 승인이 없으면 이 제품이나 문서의 일부를 어떠한 수단이나 어떠한 형태로도 전제할 수 없습니다. 글꼴 기술을 포함한 협력업체 소프트웨어에 대한 저작권 및 사용권은 Sun 공급업체에 있습니다.

제품 중에는 캘리포니아 대학에 사용권이 있는 Berkeley BSD 시스템에서 파생된 부분이 포함되어 있을 수 있습니다. UNIX는 미국 및 다른 나라에서 X/Open Company, Ltd.를 통해 독립적으로 사용권이 부여되는 등록 상표입니다. Netscape Communicator™에는 다음 저작권이 적용됩니다. (c) Copyright 1995 Netscape Communications Corporation. All rights reserved.

Sun, Sun Microsystems, Sun 로고, AnswerBook2, docs.sun.com, Sun Management Center, Solstice DiskSuite, Sun StorEdge 및 Solaris는 미국 및 다른 나라에서 사용하는 Sun Microsystems, Inc.의 상표, 등록 상표 또는 서비스 마크입니다. 모든 SPARC 등록 상표는 미국 및 다른 나라에서 사용하는 SPARC International, Inc.의 등록 상표이므로 사용권을 부여 받아 사용해야 합니다. SPARC 등록 상표가 있는 제품은 Sun Microsystems, Inc.에서 개발한 아키텍처를 기반으로 합니다.

OPEN LOOK 및 Sun™ GUI(그래픽 사용자 인터페이스)는 사용자와 권리 양도 사용자를 위해 Sun Microsystems, Inc.에서 개발하였습니다. Sun에서는 컴퓨터 업계 최초로 비주얼 또는 그래픽 사용자 인터페이스의 개념을 연구하여 개발한 Xerox의 노력을 높이 평가하고 사의를 표합니다. Sun은 Xerox와 Xerox 그래픽 사용자 인터페이스에 대한 비독점적 사용권을 보유하고 있습니다. 이 사용권은 OPEN LOOK GUI를 구현하거나 Sun의 서면 사용권 계약을 따르는 Sun의 권리 양도 사용자에게도 적용됩니다.

권한 제한: 미국 정부에서 사용하거나 복사하거나 공개할 때는 FAR 52.227-14(g)(2)(6/87) 및 FAR 52.227-19(6/87) 또는 DFAR 252.227-7015(b)(6/95) 및 DFAR 227.7202-3(a)의 제한을 받습니다.

이 문서에서는 본문의 내용을 "의미 그대로" 제공하며, 상품성, 특정 목적에 대한 적합성 또는 비위반과 같이 함축된 보증을 포함하여 모든 명시적 또는 함축적 조건, 표현 및 보증은 이 문서에 포함되어 있지 않습니다. 이 책임 부인 항목은 법률을 위반하지 않는 범위 내에서만 적용됩니다.

Copyright 2001 Sun Microsystems, Inc., 901 San Antonio Road, Palo Alto, Californie 94303 Etats-Unis. 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 des systèmes 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. La notice suivante est applicable à Netscape Communicator™:(c) Copyright 1995 Netscape Communications Corporation. Tous droits réservés.

Sun, Sun Microsystems, le logo Sun, AnswerBook2, docs.sun.com, Sun Management Center, Solstice DiskSuite, Sun StorEdge, et Solaris sont des marques de fabrique ou des marques déposées, ou marques de service, 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.



목차

	머리말	7
1.	소개 및 개요	11
	SunPlex 시스템 소개	11
	높은 가용성과 고장 허용	12
	SunPlex 시스템의 페일오버 및 확장성	12
	SunPlex 시스템에 대한 세 가지 관점	13
	하드웨어 설치 및 서비스 관점	13
	시스템 관리자 관점	14
	응용프로그램 프로그래머 관점	16
	SunPlex 시스템 작업	17
2.	주요 개념 - 하드웨어 서비스 제공업체	19
	SunPlex 시스템 하드웨어 구성 요소	19
	클러스터 노드	20
	멀티호스트 디스크	22
	로컬 디스크	23
	이동식 미디어	24
	클러스터 상호 연결	24
	공용 네트워크 인터페이스	25

	클라이언트 시스템	25
	콘솔 액세스 장치	25
	관리 콘솔	26
	Sun Cluster 토폴로지	27
	클러스터 쌍 토폴로지	27
	Pair+M 토폴로지	28
	N+1(방사형) 토폴로지	29
3.	주요 개념 - 관리 및 응용프로그램 개발	31
	클러스터 관리 및 응용프로그램 개발	31
	관리 인터페이스	32
	클러스터 시간	33
	고가용성 프레임워크	33
	글로벌 장치	37
	디스크 장치 그룹	38
	글로벌 이름 공간	39
	클러스터 파일 시스템	41
	정족수 및 정족수 장치	43
	블룸 관리자	47
	데이터 서비스	48
	새로운 데이터 서비스 개발	55
	응용프로그램 트래픽을 위한 클러스터 상호 연결 사용법	57
	자원, 자원 그룹 및 자원 유형	58
	PNM(Public Network Management) 및 NAFO(Network Adapter Failover)	61
4.	자주 문의되는 사항(FAQ)	63
	고가용성 FAQ	63
	파일 시스템 FAQ	64
	블룸 관리 FAQ	65

데이터 서비스 FAQ	65
공용 네트워크 FAQ	66
클라이언트 시스템 FAQ	67
클러스터 기억 장치 FAQ	68
클러스터 상호 연결 FAQ	68
클라이언트 시스템 FAQ	69
관리 콘솔 FAQ	69
터미널 콘센트레이터 및 시스템 서비스 프로세서 FAQ	70
용어해설	73

머리말

Sun™ Cluster 3.0 U1 개념에는 *SunPlex™* 시스템에 대한 개념 및 참조 정보가 있습니다. *SunPlex* 시스템에는 *Sun*의 클러스터 솔루션을 구성하는 모든 하드웨어 및 소프트웨어 구성 요소가 포함되어 있습니다.

이 문서는 *Sun Cluster* 소프트웨어에 대한 교육을 받은 전문 시스템 관리자를 위한 내용입니다. 이 책을 계획 또는 사전 판매 안내서로 사용해서는 안 됩니다. 이 책을 읽기 전에 이미 시스템 요구사항을 관별하고 해당되는 장비와 소프트웨어를 구입해야 합니다.

이 책에 설명된 개념을 이해하려면, *Solaris™* 운영 환경에 대한 지식이 있어야 하고, *SunPlex* 시스템에서 사용되는 볼륨 관리자 소프트웨어에 익숙해야 합니다.

활자체 규약

글자체 또는 기호	의미	예
AaBbCc123	명령, 파일 및 디렉토리의 이름, 화면상의 컴퓨터 출력	.login 파일을 편집하십시오. 모든 파일을 나열하려면 <code>ls -a</code> 를 사용하십시오. % You have mail.
AaBbCc123	화면상 컴퓨터 출력과 대비될 경우의 입력할 내용	% su Password:

글자체 또는 기호	의미	예
<i>AaBbCc123</i>	책 제목, 새 단어 및 용어, 강조표시할 단어	<i>User's Guide</i> 의 제 6장을 읽으십시오. 이를 <i>class</i> 옵션이라고 합니다. 이를 수행하려면 반드시 슈퍼유저여야 합니다.
	명령행 변수. 실제 이름 값으로 대체됩니다.	파일을 삭제하려면, rm <i>filename</i> 을 입력하십시오.

셸 프롬프트

셸	프롬프트
C 셸	<i>machine_name%</i>
C 셸 슈퍼유저	<i>machine_name#</i>
Bourne 셸 및 Korn 셸	\$
Bourne 셸 및 Korn 셸 슈퍼유저	#

관련 문서

주제	제목	제품 번호
설치	<i>Sun Cluster 3.0 U1</i> 설치 안내서	816-1962-10
하드웨어	<i>Sun Cluster 3.0 U1 Hardware Guide</i>	806-7070

주제	제목	제품 번호
데이터 서비스	<i>Sun Cluster 3.0 U1 Data Services Installation and Configuration Guide</i>	806-7071
API 개발	<i>Sun Cluster 3.0 U1 Data Services Developer's Guide</i>	806-7072
관리	<i>Sun Cluster 3.0 U1</i> 시스템 관리 안내서	816-1967-10
오류 메시지 및 문제점 해결	<i>Sun Cluster 3.0 U1 Error Messages Manual</i>	806-7076
릴리스 정보	<i>Sun Cluster 3.0 U1</i> 릴리스 정보	816-1972-10

Sun 문서 주문

인터넷 전문 서점인 **Fatbrain.com**에서 **Sun Microsystems** 사의 정선된 제품 참고문서를 공급합니다. 문서 목록과 주문 방법에 대해서는 다음 주소에 있는 **Fatbrain.com**의 **Sun Documentation Center**를 방문하십시오.

<http://www1.fatbrain.com/documentation/sun>

Sun 문서 온라인 액세스

docs.sun.comSM 웹 사이트에서 **Sun** 기술 문서에 액세스할 수 있습니다. 다음 주소에서 docs.sun.com 아카이브를 살펴보고 특정 서적 제목이나 주제에 대해 검색할 수 있습니다.

<http://docs.sun.com>

도움말 받기

SunPlex 시스템 설치 및 사용에 문제가 있으면, 서비스 담당자에게 문의하십시오. 문의시 다음 정보를 제공하십시오.

- 이름 및 전자 우편 주소(사용가능할 경우)
- 회사 이름, 주소 및 전화 번호
- 시스템 모델 및 일련 번호
- 운영 환경의 릴리스 번호(예를 들어, Solaris 8)
- Sun Cluster 소프트웨어의 릴리스 번호(예를 들어, Sun Cluster 3.0)

서비스 제공업체에게 시스템에 있는 각 노드에 대한 정보를 제공하려면 다음 명령을 사용하십시오.

명령	기능
<code>prtconf -v</code>	시스템 메모리의 크기를 표시하고 주변 장치에 대한 정보를 보고합니다.
<code>psrinfo -v</code>	프로세서에 대한 정보를 표시합니다.
<code>showrev --p</code>	설치된 패치를 알려줍니다.
<code>prtdiag -v</code>	시스템 진단 정보를 표시합니다.
<code>scinstall -pv</code>	Sun Cluster 소프트웨어 릴리스 및 패키지 버전 정보를 표시합니다.

`/var/adm/messages` 파일의 내용을 사용할 수도 있습니다.

소개 및 개요

SunPlex 시스템은 가용성과 확장성이 높은 서비스를 제공할 수 있도록 통합된 하드웨어 및 소프트웨어 솔루션입니다.

Sun Cluster 3.0 U1 개념에서는 **SunPlex** 설명서의 주요 사용자에게 필요한 개념을 설명합니다. 이 설명서의 대상은 다음과 같습니다.

- 클러스터 하드웨어를 설치하고 서비스를 제공하는 서비스 제공업체.
- **Sun Cluster** 소프트웨어를 설치, 구성 및 관리하는 시스템 관리자.
- 응용프로그램에 사용할 수 있도록 현재 **Sun Cluster** 제품에 포함되지 않은 확장 가능한 파일오버 서비스를 개발하는 응용프로그램 개발자.

이 설명서를 읽은 후에 나머지 **SunPlex** 설명서를 보면 **SunPlex** 시스템을 완전하게 이해할 수 있습니다.

이 장의 내용은 다음과 같습니다.

- **SunPlex** 시스템을 소개하고 높은 수준의 개요를 제공합니다.
- **SunPlex** 사용자의 몇 가지 관점에 대해 설명합니다.
- **SunPlex** 시스템을 사용하기 전에 알아야 할 주요 개념을 설명합니다.
- 절차와 관련 정보가 포함된 **SunPlex** 문서와 주요 개념을 연결합니다.
- 클러스터 관련 작업을 각 작업 수행 절차가 포함된 문서에 연결합니다.

SunPlex 시스템 소개

SunPlex 시스템은 **Solaris** 운영 환경을 클러스터 운영 체제로 확장합니다. 클러스터 또는 플렉스는 데이터베이스, 웹 서비스 및 파일 서비스를 포함한 네트워크 서비스나 응용 프로그램을 단일 클라이언트 환경으로 만들어주는 느슨한 결합의 컴퓨팅 노드 모음입니다.

각 클러스터 노드는 독립적으로 자체 프로세스를 실행하는 서버입니다. 이 프로세스는 다른 프로세스와 통신을 통해 사용자에게 응용프로그램, 시스템 자원 및 데이터를 제공하고 네트워크 클라이언트에 하나로 표시되는 단일 시스템을 구성할 수 있습니다.

클러스터에는 기존의 단일 서버 시스템보다 좋은 여러 가지 장점이 있습니다. 즉, 페일 오버(failover) 및 확장 가능한 서비스 지원, 모듈 단위로 확장할 수 있는 용량, 기존 하드웨어의 고장 허용 시스템에 비해 저렴한 항목 가격 등의 장점이 있습니다.

SunPlex 시스템의 목표는 다음과 같습니다.

- 소프트웨어 또는 하드웨어 결함으로 인한 시스템 작동 중지 시간을 줄이거나 제거합니다.
- 단일 서버 시스템의 작동이 중지시키는 장애가 발생 하더라도 일반 사용자가 데이터와 응용 프로그램을 사용할 수 있도록 합니다.
- 클러스터에 노드를 추가하는 방법으로 프로세서를 추가하여 서비스를 확장하고 응용프로그램 처리량을 증가시킵니다.
- 전체 클러스터를 종료하지 않고도 유지 보수를 수행할 수 있도록 하여 높은 시스템 가용성을 제공합니다.

높은 가용성과 고장 허용

SunPlex 시스템은 가용성이 높은(HA) 시스템으로 설계되어 데이터 및 응용프로그램에 거의 연속적으로 액세스할 수 있습니다.

그러나 고장 허용 하드웨어 시스템은 데이터 및 응용프로그램에 대한 지속적인 액세스를 제공하지만 특수한 하드웨어를 사용해야 하기 때문에 비용이 많이 듭니다. 또한 소프트웨어 문제로 인한 고장은 일반적으로 고장 허용 시스템에서 해결할 수 없습니다.

SunPlex 시스템은 하드웨어와 소프트웨어 결함을 통해 높은 가용성을 제공합니다. 중복된 클러스터 상호 연결, 기억 장치 및 공용 네트워크를 사용하여 한 위치에서 고장이 발생해도 전체 시스템을 안전하게 보호합니다. 클러스터 소프트웨어는 클러스터 구성 노드의 상태를 계속 모니터링하고, 노드에 장애가 발생하면 클러스터에서 제외하여 데이터가 손상되지 않도록 보호합니다. 또한 클러스터는 서비스와 서비스에 사용되는 시스템 자원을 모니터링하여 장애가 발생하면 페일오버하거나 다시 시작합니다.

높은 가용성에 대한 질문과 대답은 “고가용성 FAQ”(63페이지)를 참조하십시오.

SunPlex 시스템의 페일오버 및 확장성

SunPlex 시스템을 사용하면 페일오버 서비스나 확장 가능한 서비스를 구현할 수 있습니다. 일반적으로 페일오버 서비스는 높은 가용성(중복)만 제공하지만 확장 가능한 서비스는 높은 가용성과 함께 높은 성능을 제공합니다. 단일 클러스터에서 페일오버 서비스와 확장 가능한 서비스를 모두 지원할 수 있습니다.

페일오버 서비스

페일오버는 1차 노드에 장애가 발생할 경우에 자동으로 1차 노드의 서비스를 지정된 2차 노드로 이동시키는 프로세스입니다. 페일오버를 사용하면 Sun Cluster의 가용성이 높아집니다.

페일오버가 발생하면 클라이언트의 서비스 연결이 일시 중단될 수 있으며, 페일오버가 종료된 후에 다시 연결해야 할 수도 있습니다. 그러나 클라이언트에서는 서비스를 제공하는 물리적 서버를 알 수 없습니다.

확장 가능한 서비스

페일오버가 중복성과 관련이 있는 반면, 확장성은 로드와 관계없이 일정한 응답 시간 또는 처리량을 제공합니다. 확장 가능한 서비스를 사용하면 클러스터의 여러 노드가 동시에 응용프로그램을 실행할 수 있기 때문에 성능이 향상됩니다. 확장 가능한 구성에서는 클러스터의 각 노드가 데이터를 제공하고 클라이언트 요청을 처리할 수 있습니다.

페일오버 및 확장 가능 서비스에 대한 자세한 내용은 “데이터 서비스”(48페이지)를 참조하십시오.

SunPlex 시스템에 대한 세 가지 관점

이 단원에서는 SunPlex에 대한 서로 다른 세 가지 관점과 주요 개념 및 각 관점과 관련된 문서에 대하여 설명합니다. 각 관점은 다음과 같습니다.

- 하드웨어 설치 및 서비스 담당자
- 시스템 관리자
- 응용프로그램 프로그래머

하드웨어 설치 및 서비스 관점

하드웨어 서비스 담당자의 경우에는 SunPlex를 서버, 네트워크 및 기억 장치가 포함된 완성된 하드웨어 모음으로 생각할 수 있습니다. 이 구성 요소는 모든 구성 요소를 백업하여 단일 지점의 장애가 발생하지 않도록 케이블로 연결되어 있습니다.

주요 개념 – 하드웨어

하드웨어 서비스 담당자는 다음과 같은 클러스터 개념을 이해해야 합니다.

- 클러스터 하드웨어 구성 및 케이블 설치
- 설치 및 서비스 제공(추가, 제거, 교체)

- 네트워크 인터페이스 구성 요소(어댑터, 연결 장치, 케이블)
 - 디스크 인터페이스 카드
 - 디스크 배열
 - 디스크 드라이브
 - 관리 콘솔 및 콘솔 액세스 장치
- 관리 콘솔 및 콘솔 액세스 장치 설정

권장 하드웨어 개념 참조서

다음 단락에는 위의 주요 개념과 관련된 자료가 있습니다.

- “클러스터 노드”(20페이지)
- “멀티호스트 디스크”(22페이지)
- “로컬 디스크”(23페이지)
- “클러스터 상호 연결”(24페이지)
- “공용 네트워크 인터페이스”(25페이지)
- “클라이언트 시스템”(25페이지)
- “관리 콘솔”(26페이지)
- “콘솔 액세스 장치”(25페이지)
- “클러스터 쌍 토폴로지”(27페이지)
- “N+1(방사형) 토폴로지”(29페이지)

관련 SunPlex 문서

다음 SunPlex 문서에는 하드웨어 서비스 개념과 관련된 절차 및 정보가 있습니다.

- *Sun Cluster 3.0 U1 Hardware Guide*

시스템 관리자 관점

시스템 관리자의 경우에는 SunPlex가 기억 장치 장치를 공유하고 케이블로 연결된 서버(노드) 세트라고 생각할 수 있습니다. 시스템 관리자는 다음 내용을 이해해야 합니다.

- 클러스터 노드 사이의 연결을 모니터링할 수 있도록 Solaris 소프트웨어에 통합된 특수 클러스터 소프트웨어
- 클러스터 노드에서 실행되는 사용자 응용프로그램의 상태를 모니터링할 수 있는 특수 소프트웨어
- 디스크를 설정하고 관리하는 볼륨 관리 소프트웨어

- 직접 디스크에 연결되지는 않지만 모든 노드가 모든 기억 장치에 액세스할 수 있도록 하는 특수 클러스터 소프트웨어
- 노드에 로컬로 연결된 경우에도 모든 노드에 파일을 표시할 수 있도록 하는 특수 클러스터 소프트웨어

주요 개념 – 시스템 관리

시스템 관리자는 다음 개념과 프로세스를 이해해야 합니다.

- 하드웨어 및 소프트웨어 구성 요소 사이의 상호 작용
- 클러스터 설치 및 구성 방법에 대한 일반적인 흐름
 - Solaris 운영 환경 설치
 - Sun Cluster 소프트웨어 설치 및 구성
 - 볼륨 관리자 설치 및 구성
 - 클러스터에서 사용할 수 있도록 응용프로그램 소프트웨어 설치 및 구성
 - Sun Cluster 데이터 서비스 소프트웨어 설치 및 구성
- 클러스터 하드웨어 및 소프트웨어 구성 요소를 추가, 제거 및 교체하고 서비스를 제공하기 위한 클러스터 관리 절차
- 성능을 개선하기 위한 구성 변경

권장하는 시스템 관리자 개념 참조서

다음 단원에는 위의 주요 개념과 관련된 자료가 있습니다.

- “관리 인터페이스”(32페이지)
- “고가용성 프레임워크”(33페이지)
- “글로벌 장치”(37페이지)
- “디스크 장치 그룹”(38페이지)
- “글로벌 이름 공간”(39페이지)
- “클러스터 파일 시스템”(41페이지)
- “정족수 및 정족수 장치”(43페이지)
- “볼륨 관리자”(47페이지)
- “데이터 서비스”(48페이지)
- “자원, 자원 그룹 및 자원 유형”(58페이지)
- “PNM(Public Network Management) 및 NAFO(Network Adapter Failover)”(61페이지)
- 4장

관련 SunPlex 문서 – 시스템 관리자

다음의 SunPlex 문서에는 시스템 관리 개념과 관련된 절차와 정보가 있습니다.

- *Sun Cluster 3.0 U1* 설치 안내서
- *Sun Cluster 3.0 U1* 시스템 관리 안내서
- *Sun Cluster 3.0 U1 Error Messages Manual*

응용프로그램 프로그래머 관점

SunPlex 시스템은 Oracle, NFS, DNS, iPlanet Web Server, Apache Web Server, Netscape Directory Server 등과 같은 응용프로그램을 위한 데이터 서비스를 제공합니다. 별도 구입한 응용프로그램이 Sun Cluster 소프트웨어의 제어를 통해 실행되도록 구성하면 데이터 서비스가 완성됩니다. Sun Cluster 소프트웨어는 응용프로그램을 시작하고, 중지하고, 모니터하는 구성 파일과 관리 방법을 제공합니다. 페일오버 서비스나 확장 가능한 서비스를 새로 만들어야 하는 경우에 SunPlex API(Application Programming Interface) 및 DSET API(Data Service Enabling Technologies API)를 사용하면 클러스터에서 응용프로그램을 데이터 서비스로 실행할 수 있도록 필요한 구성 파일과 관리 방법을 개발할 수 있습니다.

주요 개념 – 응용프로그램 프로그래머

응용프로그램 프로그래머는 다음을 이해해야 합니다.

- 사용하는 응용프로그램이 페일오버 서비스나 확장 가능한 데이터 서비스로 실행될 수 있는지를 판단하는 응용프로그램 특성
- Sun Cluster API, DSET API 및 “범용” 데이터 서비스. 프로그래머가 클러스터 환경에 맞게 응용프로그램을 구성하려면 프로그램이나 스크립트를 작성할 때 가장 적합한 도구를 결정해야 합니다.

권장하는 응용프로그램 프로그래머 개념 참조서

다음 단원에는 위의 주요 개념과 관련된 자료가 있습니다.

- “데이터 서비스”(48페이지)
- “자원, 자원 그룹 및 자원 유형”(58페이지)
- 4장

관련 SunPlex 문서 – 응용프로그램 프로그래머

다음의 SunPlex 문서에는 응용프로그램 프로그래머 개념과 관련된 절차 및 정보가 있습니다.

- *Sun Cluster 3.0 U1 Data Services Developer's Guide*
- *Sun Cluster 3.0 U1 Data Services Installation and Configuration Guide*

SunPlex 시스템 작업

모든 SunPlex 시스템 작업에 약간의 배경 개념이 필요합니다. 다음 표에 고급 작업 보기 및 작업 단계를 설명하는 문서가 있습니다. 이 책의 개념 단원에서는 개념과 작업 사이의 매핑 관계를 설명합니다.

표1-1 작업 맵: 문서에 사용자 작업 매핑

수행할 작업	참조할 문서
클러스터 하드웨어 설치	<i>Sun Cluster 3.0 U1 Hardware Guide</i>
클러스터에 Solaris 소프트웨어 설치	<i>Sun Cluster 3.0 U1</i> 설치 안내서
Sun™ Management Center 소프트웨어 설치	<i>Sun Cluster 3.0 U1</i> 설치 안내서
Sun Cluster 소프트웨어 설치 및 구성	<i>Sun Cluster 3.0 U1</i> 설치 안내서
볼륨 관리 소프트웨어 설치 및 구성	<i>Sun Cluster 3.0 U1</i> 설치 안내서 볼륨 관리 문서
Sun Cluster 데이터 서비스 설치 및 구성	<i>Sun Cluster 3.0 U1 Data Services Installation and Configuration Guide</i>
서비스 클러스터 하드웨어	<i>Sun Cluster 3.0 U1 Hardware Guide</i>
Sun Cluster 소프트웨어 관리	<i>Sun Cluster 3.0 U1</i> 시스템 관리 안내서
볼륨 관리 소프트웨어 관리	<i>Sun Cluster 3.0 U1</i> 시스템 관리 안내서 및 볼륨 관리 문서
응용프로그램 소프트웨어 관리	응용프로그램 문서

표1-1 작업 맵: 문서에 사용자 작업 매핑 계속

수행할 작업	참조할 문서
문제 식별 및 권장하는 사용자 조치	<i>Sun Cluster 3.0 U1 Error Messages Manual</i>
새 데이터 서비스 만들기	<i>Sun Cluster 3.0 U1 Data Services Developer's Guide</i>

주요 개념 - 하드웨어 서비스 제공업체

이 장에서는 SunPlex 시스템을 구성하는 하드웨어 구성 요소와 관련된 주요 개념에 대하여 설명합니다.

SunPlex 시스템 하드웨어 구성 요소

이 정보는 기본적으로 하드웨어 서비스 제공업체를 위한 내용입니다. 이 개념은 서비스 제공업체에서 클러스터 하드웨어를 설치, 구성하거나 서비스를 제공하기 전에 하드웨어 구성 요소 사이의 관계를 이해하는 데 도움이 됩니다. 클러스터 시스템 관리자는 클러스터 소프트웨어 설치, 구성 및 관리에 대한 배경 정보로 이 정보를 사용할 수 있습니다.

클러스터는 다음과 같은 몇 가지 하드웨어 구성 요소로 구성됩니다.

- 로컬 디스크(비공유)가 있는 클러스터 노드
- 멀티호스트 기억 장치(노드 사이에 디스크)
- 이동식 미디어(테이프 및 CD-ROM)
- 클러스터 상호 연결
- 공용 네트워크 인터페이스
- 클라이언트 시스템
- 관리 콘솔
- 콘솔 액세스 장치

SunPlex 시스템을 사용하면 “Sun Cluster 토폴로지”(27페이지)에서 설명하는 다양한 구성 방법으로 이 구성 요소를 결합할 수 있습니다.

다음 그림은 클러스터 구성 샘플입니다.

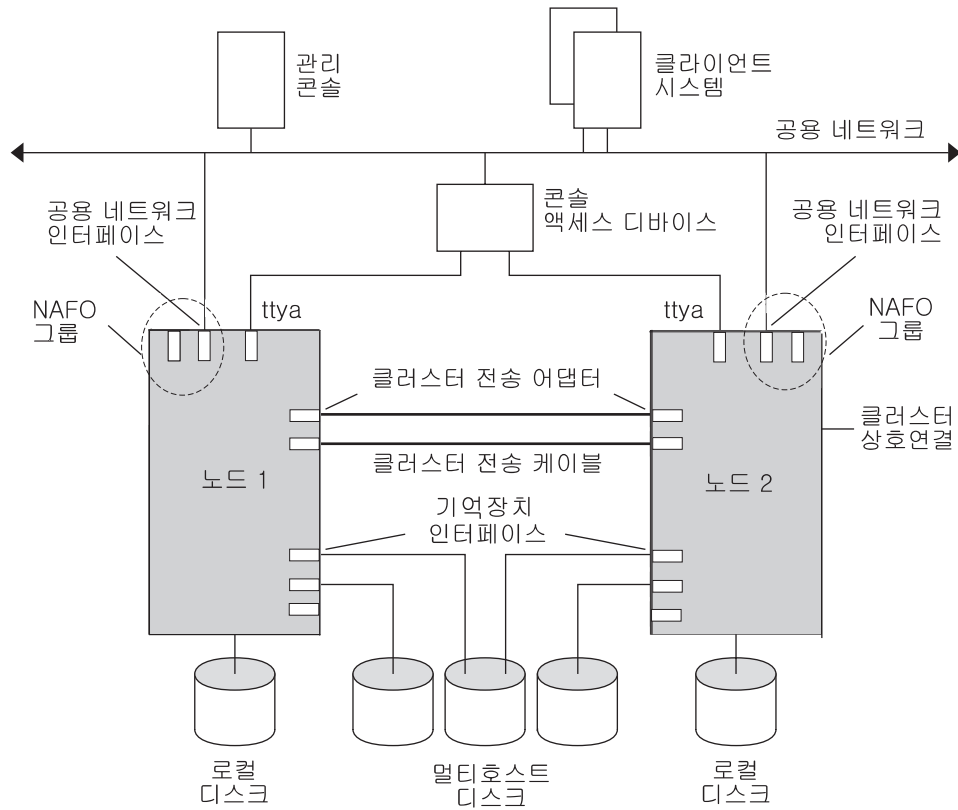


그림 2-1 2-노드 클러스터 구성 샘플

클러스터 노드

클러스터 노드는 **Solaris** 운영 환경 및 **Sun Cluster** 소프트웨어 모두 실행하는 시스템으로 클러스터의 현재 구성원(클라이언트 시스템)이거나 구성원이 될 수 있는 시스템입니다. **Sun Cluster** 소프트웨어가 있으면 클러스터에 2 - 8개의 노드가 있을 수 있습니다. 지원되는 노드 구성은 “**Sun Cluster** 토폴로지”(27페이지)을 참조하십시오.

클러스터 노드는 일반적으로 하나 이상의 멀티호스트 디스크에 연결됩니다. 멀티호스트 디스크에 연결되지 않은 노드는 멀티호스트 디스크에 액세스하기 위해 클러스터 파일 시스템을 사용합니다. 예를 들어, 하나의 확장 가능한 서비스를 구성하면 멀티호스트 디스크에 직접 연결되지 않아도 노드가 요청에 응할 수 있습니다.

병렬 데이터베이스 구성의 노드는 모든 디스크에 대한 동시 액세스를 공유합니다. 병렬 데이터베이스 구성에 대한 자세한 내용은 “멀티호스트 디스크”(22페이지) 및 3장을 참조하십시오.

클러스터의 모든 노드는 클러스터에 액세스하여 관리하는 데 사용하는 공용 이름(클러스터 이름)으로 그룹화됩니다.

공용 네트워크 어댑터는 노드를 공유 네트워크에 연결하여 클러스터에 대한 클라이언트 액세스를 제공합니다.

클라이언트 시스템은 물리적으로 독립된 하나 이상의 네트워크를 통해 클러스터의 다른 노드와 통신을 합니다. 이렇게 물리적으로 독립된 네트워크 세트를 클러스터 상호 연결이라고 합니다.

다른 노드가 클러스터에 결합되거나 클러스터에서 제거될 때 클러스터의 모든 노드가 이것을 인식합니다. 또한 클러스터의 모든 노드가 로컬로 실행되는 자원뿐 아니라 다른 클러스터 노드에서 실행되는 자원을 인식합니다.

성능이 크게 떨어지지 않고 페일오버가 발생하도록 하려면 동일한 클러스터의 노드가 모두 유사한 프로세싱, 메모리 및 I/O 기능을 사용해야 합니다. 페일오버를 위해 모든 노드에 다른 노드의 워크로드를 백업하거나 2차 노드가 될 수 있을 만큼 충분한 용량이 있어야 합니다.

각 노드는 개별 루트(/) 파일 시스템을 부팅합니다.

클라이언트 시스템용 높은 수준의 소프트웨어 구성 요소

클라이언트 시스템의 기능을 하려면 다음 소프트웨어가 설치되어야 합니다.

- Solaris 운영 환경
- Sun Cluster 소프트웨어
- 데이터 서비스 응용프로그램
- 볼륨 관리(Solstice DiskSuite™ 또는 VERITAS Volume Manager)

독립된 디스크로 구성된 하드웨어 중복 배열(RAID)을 사용하는 Oracle Parallel Server(OPS) 구성에서는 예외입니다. 이러한 구성에서는 Oracle 데이터를 관리하기 위해 Solstice DiskSuite 또는 VERITAS Volume Manager와 같은 소프트웨어 볼륨 관리자가 필요 없습니다.

Solaris 운영 환경, Sun Cluster 및 볼륨 관리 소프트웨어를 설치하는 방법은 *Sun Cluster 3.0 U1* 설치 안내서를 참조하십시오.

데이터 서비스 설치 및 구성 방법은 *Sun Cluster 3.0 U1 Data Services Installation and Configuration Guide*를 참조하십시오.

이전의 소프트웨어 구성 요소에 대한 개념은 3장을 참조하십시오.

다음 그림은 Sun Cluster 소프트웨어 환경을 만들기 위해 사용하는 소프트웨어 구성 요소입니다.

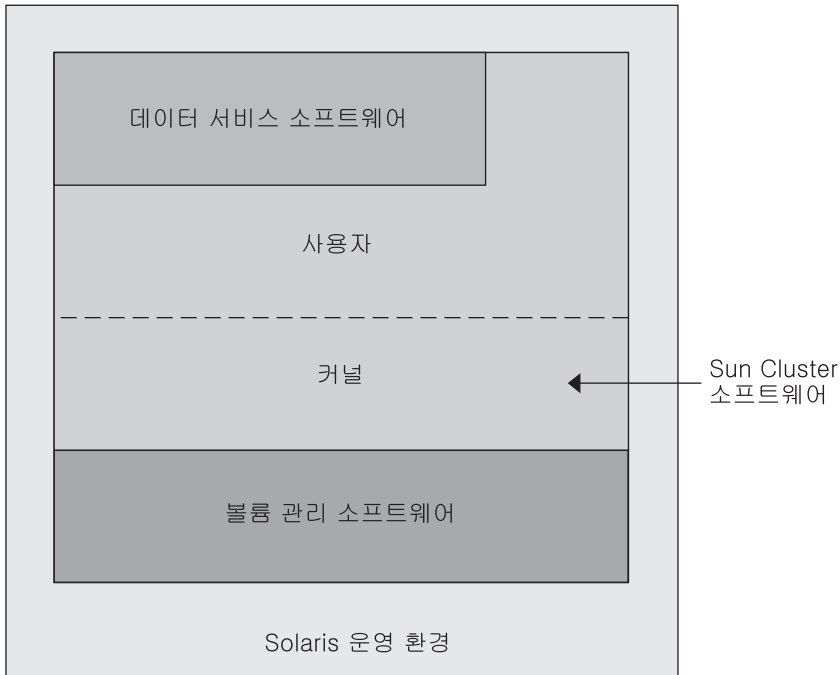


그림 2-2 높은 수준의 Sun Cluster 소프트웨어 구성 요소 관계
클라이언트 시스템에 대한 질문 및 응답은 4장을 참조하십시오.

멀티호스트 디스크

Sun Cluster에는 멀티호스트 디스크 기억 장치가 필요합니다. 멀티호스트 디스크 기억 장치는 동시에 둘 이상의 노드에 연결할 수 있는 디스크입니다. Sun Cluster 환경에 멀티호스트 기억 장치를 사용하면 디스크 가용성을 높일 수 있습니다.

멀티호스트 디스크에는 다음과 같은 기능이 있습니다.

- 노드 하나에 장애가 발생해도 계속 작동합니다.
- 응용프로그램 데이터를 저장하고 응용프로그램 바이너리와 구성 파일도 저장할 수 있습니다.
- 노드 장애로부터 보호합니다. 클라이언트 요청이 하나의 노드를 통해 데이터에 액세스할 때 노드에 장애가 발생하면 동일한 디스크에 직접 연결된 다른 노드를 사용하도록 요청이 스위치오버됩니다.
- 멀티호스트 디스크는 디스크를 “마스터”하는 1차 노드를 통하거나 로컬 경로를 통한 직접 동시 액세스에 의해 글로벌하게 액세스됩니다. 현재는 OPS만이 직접 동시 액세스를 사용합니다.

볼륨 관리자는 멀티호스트 디스크의 데이터 중복을 위해 이중화 또는 RAID-5 구성을 제공합니다. 현재 Sun Cluster에서는 Solstice DiskSuite 및 VERITAS Volume Manager를 볼륨 관리자로 사용하고 Sun StorEdge™ A3x00 기억 장치 유닛에 RDAC RAID-5 하드웨어 컨트롤러를 사용할 수 있습니다.

멀티호스트 디스크를 디스크 이중화 및 스트라이핑과 결합하면 노드 장애와 각 디스크 장애로부터 보호할 수 있습니다.

클라이언트 시스템에 대한 질문 및 응답은 4장을 참조하십시오.

Multi-Initiator SCSI

이 단원의 내용은 SCSI 기억 장치 장치에만 적용되고 멀티호스트 디스크에 사용되는 광섬유 채널 기억 장치에는 적용되지 않습니다.

독립형 서버에서는 서버 노드가 서버를 특정 SCSI 버스에 연결하는 SCSI 호스트 어댑터 회로를 사용하여 SCSI 버스 작동을 제어합니다. 이러한 SCSI 호스트 어댑터 회로를 *SCSI initiator*라고 합니다. 이 회로가 SCSI 버스에 대한 모든 버스 작업을 시작합니다. Sun 시스템에서 SCSI 호스트 어댑터의 기본 SCSI 주소는 7입니다.

클러스터 구성은 멀티호스트 디스크를 사용하여 여러 서버 노드 사이의 기억 장치를 공유합니다. 클러스터 기억 장치가 종단 장치가 하나인 SCSI 장치나 차동 SCSI 장치로 구성된 경우에 이러한 구성을 *multi-initiator SCSI*라고 합니다. 이 용어가 의미하는 것처럼 SCSI 버스에는 하나 이상의 SCSI initiator가 있습니다.

SCSI 사양에서는 SCSI 버스에 있는 각 장치에 고유한 SCSI 주소가 있어야 합니다. 호스트 어댑터도 SCSI 버스에 있는 장치입니다. *Multi-initiator* 환경의 기본 하드웨어 구성에서는 모든 SCSI 호스트 어댑터가 7로 기본 설정되기 때문에 충돌합니다.

이러한 충돌을 해결하려면 각 SCSI 버스에서 SCSI 주소가 7인 SCSI 호스트 어댑터 중 하나만 남겨 두고 나머지 호스트 어댑터는 사용하지 않은 SCSI 주소로 설정해야 합니다. 적절한 계획 방법은 이러한 “사용되지 않는” SCSI 주소에 현재 사용되지 않는 주소와 최종적으로 사용되지 않을 주소 둘 다를 포함시키는 것입니다. 이후 사용되지 않을 주소의 예는 새로운 드라이브를 빈 드라이브 슬롯에 설치하여 기억 장치를 추가할 경우의 주소입니다. 대부분의 구성에서 2차 호스트 어댑터에 대해 사용가능한 SCSI 주소는 6입니다.

`scsi-initiator-id` OBP(Open Boot PROM) 등록 정보를 설정하여 이러한 호스트 어댑터에 대해 선택된 SCSI 주소를 변경할 수 있습니다. 하나의 노드 또는 호스트 어댑터 기준으로 이 등록 정보를 글로벌로 설정할 수 있습니다. 각 SCSI 호스트 어댑터에 대한 고유한 `scsi-initiator-id`를 설정하는 방법에 대한 지시사항은 *Sun Cluster 3.0 U1 Hardware Guide*에서 각각의 디스크 인클로저에 포함되어 있습니다.

로컬 디스크

로컬 디스크는 하나의 노드에만 연결된 디스크입니다. 그러므로, 노드 장애로부터 보호되지 않습니다(고가용성이 아님). 그러나 로컬 디스크를 포함한 모든 디스크가 클로

별 이름공간에 포함되어 있고 글로벌 장치로 구성되어 있습니다. 그러므로 디스크 자체는 모든 클러스터 노드에서 볼 수 있습니다.

이러한 디스크의 파일 시스템을 글로벌 마운트 지점 아래에 놓아서 다른 노드가 사용할 수 있도록 만들 수 있습니다. 현재 이러한 글로벌 파일 시스템 중 하나를 가지고 있는 노드가 마운트하는 데 실패하면, 모든 노드가 그 파일 시스템에 대한 액세스를 유실하게 됩니다. 볼륨 관리자를 사용하면 디스크 오류가 있어도 이러한 파일 시스템에 액세스할 수 있도록 디스크를 이중화할 수 있지만, 노드 장애를 보호하지는 않습니다.

글로벌 장치에 대한 자세한 내용은 “글로벌 장치”(37페이지) 단원을 참조하십시오.

이동식 미디어

테이프 드라이브와 CD-ROM 드라이브와 같은 이동식 미디어가 클러스터에서 지원됩니다. 일반적으로, 클러스터링되지 않은 환경에서와 동일한 방법으로 이러한 장치를 설치, 구성하고 서비스를 제공할 수 있습니다. 이 장치는 Sun Cluster에 글로벌 장치로 구성되므로 클러스터의 모든 노드에서 액세스할 수 있습니다. 이동식 미디어 설치와 구성에 대한 자세한 내용은 *Sun Cluster 3.0 U1 Hardware Guide*를 참조하십시오.

글로벌 장치에 대한 자세한 내용은 “글로벌 장치”(37페이지) 단원을 참조하십시오.

클러스터 상호 연결

클러스터 상호 연결은 클러스터 개인 통신과 데이터 서비스 통신을 클러스터 노드들 사이에 전송하기 위해 사용되는 장치들의 물리적 구성입니다. 상호 연결은 클러스터 개인 통신에 포괄적으로 사용되므로 성능을 제한할 수 있습니다.

클러스터 노드만 개인 상호 연결에 연결될 수 있습니다. Sun Cluster 보안 모델은 클러스터 노드만 개인용 상호 연결에 대한 물리적 액세스를 갖고 있다고 가정합니다.

모든 노드는 단일 실패 지점을 피하기 위해, 최소한 두 개의 중복된 관리 콘솔 또는 경로를 통해 클러스터 상호 연결에 의해 연결되어야 합니다. 두 노드 사이에 여러개의 관리 콘솔(2 - 6개)을 사용할 수 있습니다. 클러스터 상호 연결은 어댑터, 연결 장치, 케이블 등의 세 가지 하드웨어로 구성됩니다.

다음 목록은 이러한 하드웨어 구성 요소 각각에 대해 설명합니다.

- 어댑터 - 각 클러스터 노드에 있는 물리적 네트워크 카드. 이름은 **qfe2**와 같이 장치 이름 뒤에 물리적인 장치 번호를 붙여서 만듭니다. 일부 어댑터에는 단 하나의 물리적 네트워크 연결이 있지만 다른 어댑터들에는 **qfe** 카드처럼 하나 이상의 물리적 연결이 있을 수 있습니다. 또한 일부 어댑터에는 네트워크 인터페이스와 기억 장치 인터페이스 둘 다 들어 있습니다.

다중 인터페이스가 있는 네트워크 카드는 전체 카드가 실패할 경우, 단일 실패 지점이 될 수 있습니다. 최대 가용성을 위해서는, 두 노드 사이의 경로에만 단일 네트워크 카드에 의존하지 않도록 클러스터를 계획하십시오.

- 연결장치 - 클러스터 노드 외부에 상주하는 스위치. 이 연결장치는 바로 전달 기능과 전환 기능을 수행하여 두 개를 초과하는 노드를 함께 연결할 수 있게 합니다. 2-노드 클러스터에서는 노드들이 각 노드에서 중복 어댑터에 연결된 물리적 중복 케이블을 통해 서로 직접 연결될 수 있으므로, 연결장치가 필요하지 않습니다. 일반적으로 두 개를 초과하는 노드로 된 구성에서 연결장치가 요구됩니다.
 - 케이블 - 두개의 네트워크 어댑터간 및 어댑터와 연결장치 간의 물리적 연결.
- 클러스터 상호 연결에 대한 질문 및 응답에 대해서는 4장을 참조하십시오.

공용 네트워크 인터페이스

클라이언트는 공용 네트워크 인터페이스를 통해 클러스터에 연결합니다. 각 네트워크 어댑터 카드는 카드에 여러 하드웨어 인터페이스가 있는지에 따라 하나 이상의 공용 네트워크에 연결할 수 있습니다. 하나의 카드가 활동 중이고 다른 카드들은 백업으로 작동되도록 구성된, 하나 이상의 공용 네트워크 인터페이스 카드를 포함하도록 노드를 설정할 수 있습니다. “공용 네트워크 관리”(PNM)라는 Sun Cluster 소프트웨어 서브시스템은 활동 중인 인터페이스를 모니터링합니다. 활동 중인 어댑터가 실패할 경우, 백업 어댑터 중 하나로 인터페이스를 페일오버하기 위해 NAFO(Network Adapter Failover) 소프트웨어를 호출합니다.

공용 네트워크 인터페이스에 대한 클러스터링에 관련되는 특수 하드웨어 고려사항은 없습니다.

클라이언트 시스템에 대한 질문 및 응답에 대해서는 4장을 참조하십시오.

클라이언트 시스템

클라이언트 시스템은 공용 네트워크를 통해 클러스터에 액세스하는 다른 서버나 워크스테이션을 포함합니다. 클라이언트측 프로그램은 클러스터에서 실행되는 서버측 응용프로그램에 의해 제공되는 데이터나 다른 서비스를 사용합니다.

클라이언트 시스템은 고가용성 시스템이 아닙니다. 클러스터의 데이터 및 응용프로그램은 고가용성입니다.

클라이언트 시스템에 대한 질문 및 응답에 대해서는 4장을 참조하십시오.

콘솔 액세스 장치

모든 클러스터 노드에 대해 콘솔 액세스가 있어야 합니다. 콘솔 액세스를 확보하려면, 클러스터 하드웨어와 함께 구입한 터미널 콘센트레이터, Sun Enterprise E10000 서버상의 SSP 또는 각 노드에서 ttya에 액세스할 수 있는 또 다른 장치를 사용하십시오.

Sun에서는 사용할 수 있는 터미널 콘센트레이터가 하나만 지원됩니다. 지원되는 Sun 터미널 콘센트레이터를 사용할 것인지는 사용자가 선택할 수 있습니다. 터미널 콘센트레이터는 TCP/IP 네트워크를 사용하여 각 노드에서 /dev/console에 액세스할

수 있게 합니다. 그 결과, 네트워크 상에 있는 어떤 원격 워크스테이션에서도 각 노드에 콘솔 레벨로 액세스 할 수 있습니다.

SSP는 Sun Enterprise E10000 서버에 대한 콘솔 액세스를 제공합니다. SSP는 Sun Enterprise E10000 서버를 지원하기 위해 구성된 이더넷 네트워크의 상의 시스템입니다. SSP는 Sun Enterprise E10000 서버용 관리 콘솔입니다. Sun Enterprise E10000 네트워크 콘솔 기능을 사용하면, 네트워크상의 어떤 워크스테이션에서도 호스트 콘솔 세션을 열 수 있습니다.

다른 콘솔 액세스 방법에는 다른 터미널 콘센트레이터가 포함됩니다(tip(1) 다른 노드나 단순 단말기(터미널)로부터의 직렬 포트 액세스). Sun™ 키보드 및 모니터를 사용하거나 하드웨어 서비스 제공업체에서 지원하는 다른 직렬 포트 장치를 사용할 수 있습니다.

관리 콘솔

관리 콘솔이라는 전용 SPARCstation™ 시스템을 사용하여 활동 중인 클러스터를 관리할 수 있습니다. 일반적으로, CCP(Cluster Control Panel) 및 Sun Management Center 제품용 Sun Cluster 모듈과 같은 관리 도구 소프트웨어를 관리 콘솔에 설치하고 실행합니다. CCP에서 cconsole을 사용하면 하나 이상의 노드 콘솔을 한번에 연결할 수 있습니다. CCP 사용에 대해서는 Sun Cluster 3.0 U1 시스템 관리 안내서의 내용을 참조하십시오.

관리 콘솔은 클러스터 노드가 아닙니다. 공용 네트워크나 혹은 네트워크용 터미널 콘센트레이터를 거쳐 클러스터 노드에 원격 액세스하기 위해 관리 콘솔을 사용합니다. 클러스터가 Sun™ Enterprise E10000 플랫폼으로 구성될 경우, 관리 콘솔에서 SSP(System Service Processor)로 로그인하고 netcon(1M) 명령으로 연결할 수 있는 기능이 있어야 합니다.

일반적으로, 노드를 모니터 없이 구성합니다. 그러면, 터미널 콘센트레이터에 연결되어 있는 관리 콘솔과 노드의 직렬 포트에 연결된 터미널 콘센트레이터에서 telnet 세션을 통해 노드 콘솔에 액세스합니다. Sun Enterprise E10000 서버의 경우에는 SSP에서 연결합니다. 자세한 내용은 “콘솔 액세스 장치”(25페이지)를 참조하십시오.

Sun Cluster 는 전용 관리 콘솔을 요구하지는 않지만 전용 관리 콘솔을 사용하면 다음과 같은 이점이 있습니다.

- 동일한 시스템에서 콘솔과 관리 도구를 그룹화하여 중앙집중화된 클러스터 관리가 가능하도록 합니다.
- 하드웨어 서비스 제공업체에 의해 더 빠르게 문제점을 분석할 수 있게 합니다.

관리콘솔에 대한 질문 및 응답에 대해서는 4장을 참조하십시오.

Sun Cluster 토폴로지

토폴로지는 클러스터 노드를 클러스터에서 사용되는 기억 장치 플랫폼에 연결하는 연결 구조입니다.

Sun Cluster 는 다음과 같은 토폴로지를 지원합니다.

- 클러스터 쌍
- N+1(방사형)

다음 절에서는 각각의 토폴로지에 대해 설명합니다.

클러스터 쌍 토폴로지

클러스터 쌍 토폴로지는 단일 클러스터 관리 프레임워크 하에 작동되는 두 개 이상의 노드 쌍입니다. 이 구성에서는 쌍 사이에서만 페일오버가 발생합니다. 그러나 모든 노드는 관리 콘솔에 의해 연결되고 Sun Cluster 소프트웨어 제어 하에 작동됩니다. 이 토폴로지를 사용하여, 하나의 쌍에서 병렬 데이터베이스 응용프로그램을 실행하고 다른 쌍에서 페일오버나 고가용성 응용프로그램을 실행할 수 있습니다.

클러스터 파일 시스템을 사용하면, 모든 노드가 응용프로그램 데이터를 저장하는 디스크에 직접 연결되어 있지 않아도 둘 이상의 노드가 확장 가능 서비스나 병렬 데이터베이스를 실행하는 두 쌍 구성도 사용할 수 있습니다.

다음 그림은 클러스터 쌍 구성을 보여줍니다.

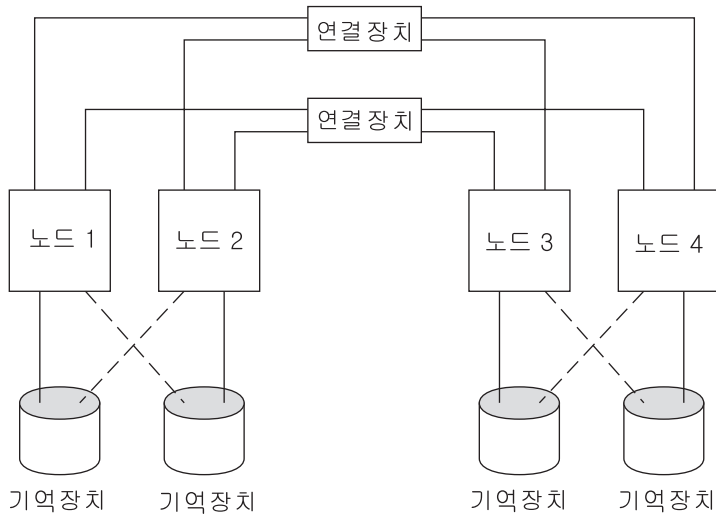


그림 2-3 클러스터 쌍 토폴로지

Pair+M 토폴로지

pair+M 토폴로지에는 공유 기억 장치 및 추가 노드 세트에 직접 연결된 노드의 쌍이 포함되어 있습니다. 이 추가 노드 세트는 자체적인 직접 연결이 없이 공유 기억 장치에 액세스하기 위해 클러스터 상호 연결을 사용합니다. 이 구성에 있는 모든 노드는 여전히 불륨 관리자로 구성됩니다.

다음 그림은 네 개의 노드 중 두 개의 노드(노드 3 및 노드 4)가 클러스터 상호 연결을 사용하여 기억 장치에 액세스하는 pair+M 토폴로지를 보여줍니다. 이 구성은 공유 기억 장치에 직접 액세스하지 않은 추가 노드를 포함하도록 확장될 수 있습니다.

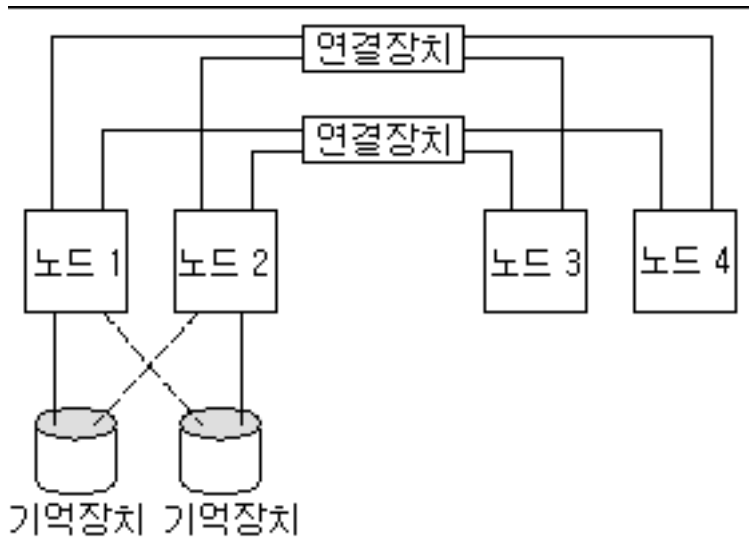


그림 2-4 Pair+M 토폴로지

N+1(방사형) 토폴로지

N+1 토폴로지에는 몇 개의 1차 노드와 하나의 2차 노드가 들어 있습니다. 1차 노드와 2차 노드를 동일하게 구성할 필요는 없습니다. 1차 노드는 활동적으로 응용프로그램 서비스를 제공합니다. 2차 노드는 1차 노드의 실패가 있을 때까지 비활동 상태일 필요는 없습니다.

2차 노드는 이러한 구성에서 모든 멀티호스트 기억 장치에 물리적으로 연결된 유일한 노드입니다.

1차 노드에서 장애가 발생하면 Sun Cluster가 자원을 2차 노드로 페일오버합니다. 그곳에서 자원은 (자동 또는 수동으로) 다시 1차 노드로 전환될 때까지 기능을 수행합니다.

2차 노드는 1차 노드 중 하나가 실패할 경우, 부하를 처리할 수 있을 만큼 충분히 많은 CPU 용량이 있어야 합니다.

다음 그림은 N+1 구성을 보여줍니다.

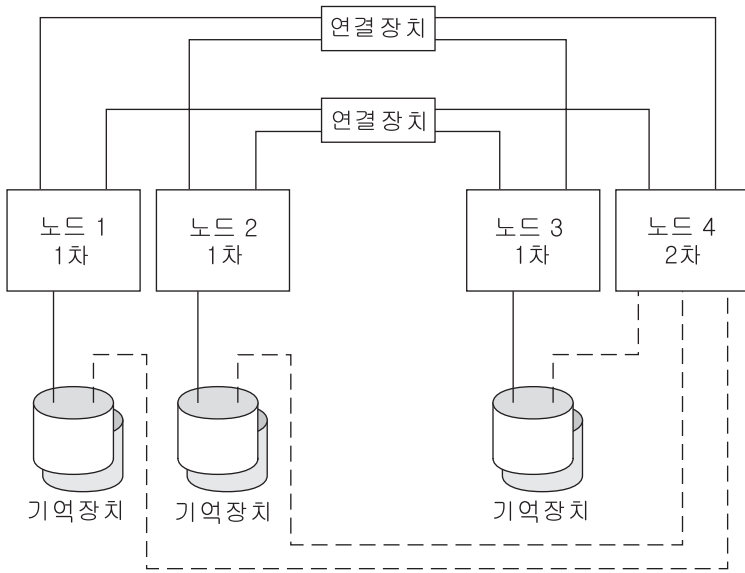


그림 2-5 N+1 토폴로지

주요 개념 - 관리 및 응용프로그램 개발

이 장에서는 SunPlex 시스템을 구성하는 하드웨어 구성 요소에 관련되는 주요 개념에 대해 설명합니다. 주요 내용은 다음과 같습니다.

- “관리 인터페이스”(32페이지)
- “클러스터 시간”(33페이지)
- “고가용성 프레임워크”(33페이지)
- “글로벌 장치”(37페이지)
- “디스크 장치 그룹”(38페이지)
- “글로벌 이름 공간”(39페이지)
- “클러스터 파일 시스템”(41페이지)
- “정족수 및 정족수 장치”(43페이지)
- “블룸 관리자”(47페이지)
- “데이터 서비스”(48페이지)
- “새로운 데이터 서비스 개발”(55페이지)
- “자원, 자원 그룹 및 자원 유형”(58페이지)
- “PNM(Public Network Management) 및 NAFO(Network Adapter Failover)”(61페이지)

클러스터 관리 및 응용프로그램 개발

이 정보는 기본적으로 SunPlex API 및 SDK를 사용하는 시스템 관리자 및 응용프로그램 개발자를 위한 것입니다. 클러스터 시스템 관리자는 클러스터 소프트웨어 설치, 구성 및 관리에 대한 배경 정보로 이 정보를 사용할 수 있습니다. 응용프로그램 개발자는 이 정보를 사용하여 작업할 클러스터 환경을 알 수 있습니다.

다음 그림은 클러스터 관리 개념이 클러스터 구조에 매핑되는 방법에 대한 고급 보기를 보여줍니다.

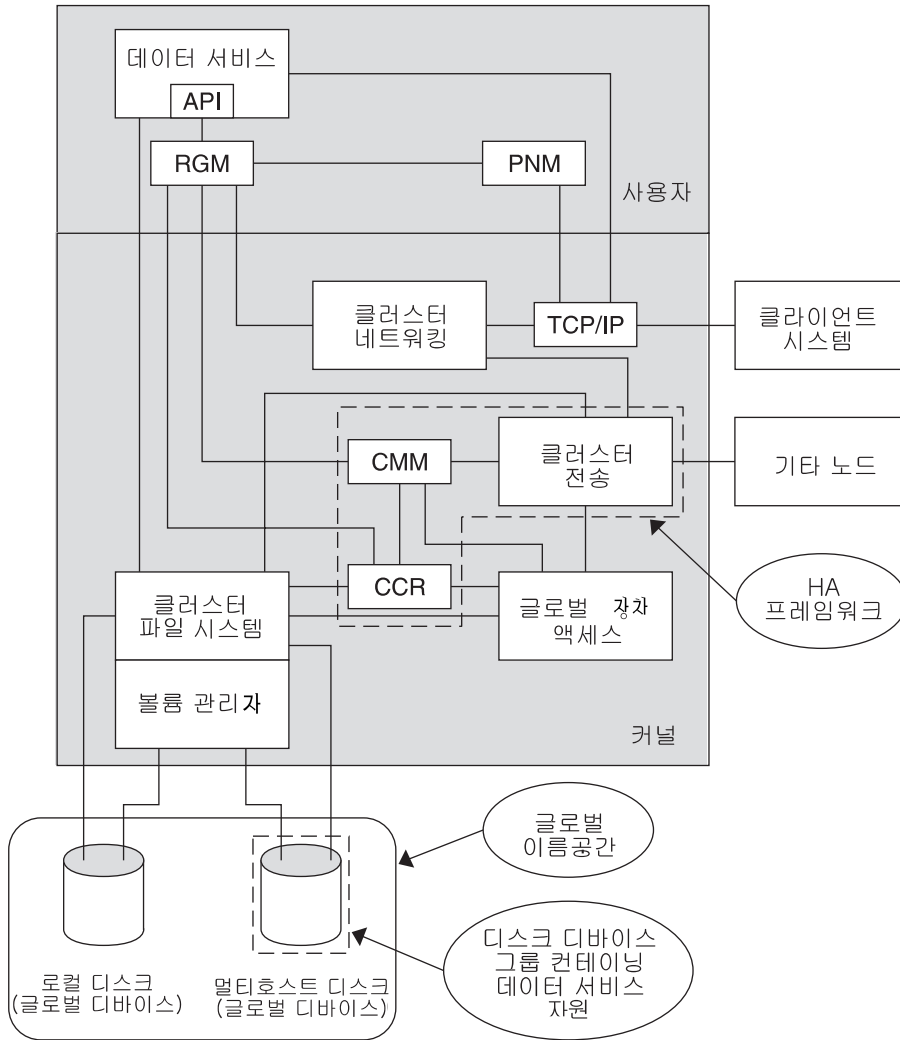


그림 3-1 Sun Cluster 소프트웨어 구조

관리 인터페이스

여러 가지 사용자 인터페이스에서 SunPlex 시스템을 설치하고 구성하고 관리하는 방법을 선택할 수 있습니다. 문서화된 명령줄 인터페이스를 통해 시스템 관리 작업을 수행할 수 있습니다. 명령줄 인터페이스의 맨 위에는 선택한 구성 작업을 쉽게 할 수 있

는 몇 가지 유틸리티가 있습니다. SunPlex 시스템에는 Sun Management Center의 일부로 실행되면서 일정한 클러스터 작업을 할 수 있는 GUI를 제공하는 모듈도 있습니다. 관리 인터페이스의 자세한 설명은 *Sun Cluster 3.0 U1* 시스템 관리 안내서의 소개 장을 참조하십시오.

클러스터 시간

클러스터에서 모든 노드들 사이의 시간은 동기화되어야 합니다. 클러스터 작동에서 외부 시간 소스를 사용하여 클러스터 노드를 동기화할 것인지는 중요하지 않습니다. SunPlex 시스템은 NTP(Network Time Protocol)를 사용하여 노드 사이의 시간을 동기화합니다.

일반적으로 초의 끝자리 수 부분에 대해 시스템 시계를 변경해도 문제는 없습니다. 그러나 활동 중인 클러스터에서 `date(1)`, `rdate(1M)` 또는 `xntpdate(1M)`를 실행할 경우(대화식으로 또는 `cron` 스크립트 내에서), 시스템 시계를 시간 소스에 동기화하기 위해 시간에서 초의 끝자리 수 부분 뿐만 아니라 그 이상의 부분도 변경하게 할 수 있습니다. 이러한 강제 변경으로 파일을 수정한 타임스탬프에 문제가 되거나 NTP 서비스에 혼란이 올 수 있습니다.

각 클러스터 노드에 Solaris 운영 환경을 설치할 경우, 노드에 대한 기본 시간 및 날짜 설정을 변경할 수 있는 기회가 주어집니다. 일반적으로 출하 시의 기본값을 승인할 수 있습니다.

`scinstall(1M)`을 사용하여 Sun Cluster를 설치할 경우, 클러스터에 대해 NTP를 구성하는 프로세스의 단계가 포함됩니다. Sun Cluster 소프트웨어는 템플릿 파일인 `ntp.cluster`를 제공합니다(설치된 클러스터 노드에서 `/etc/inet/ntp.cluster` 참조). 이 파일은 모든 클러스터 노드 사이의 피어 관계를 확립하는데, 이 때 한 노드는 “선호하는” 노드가 됩니다. 노드들은 개인 호스트 이름으로 식별되고 시간 동기화가 클러스터 상호 연결을 거쳐 발생합니다. NTP에 대해 클러스터를 구성하는 방법에 대한 지시사항은 *Sun Cluster 3.0 U1* 설치 안내서에 들어 있습니다.

또한 클러스터 밖에서 하나 이상의 NTP 서버를 설정하고 `ntp.conf` 파일을 변경하여 그 구성을 반영할 수도 있습니다.

정상적인 작동 하에서는 클러스터에서 시간을 조정할 필요가 없습니다. 그러나, Solaris 운영 환경을 설치할 때 잘못 설정된 시간을 변경하려면, *Sun Cluster 3.0 U1* 시스템 관리 안내서에 포함된 절차를 참조하십시오.

고가용성 프레임워크

SunPlex 시스템은 네트워크 인터페이스를 포함하여 사용자들 및 데이터 고가용성, 자체 응용프로그램, 파일 시스템 그리고 멀티호스트 디스크들 사이의 “경로”에 모든 구성 요소를 만듭니다. 일반적으로, 클러스터 구성 요소는 시스템에서 단일(소프트웨어 또는 하드웨어) 실패를 극복할 경우, 가용성이 높습니다.

다음 테이블은 SunPlex 구성 요소 실패의 종류(하드웨어 및 소프트웨어 모두)와 고가용성 프레임워크에 형성된 복구 종류를 보여줍니다.

표3-1 SunPlex 실패 감지 및 복구 레벨

장애가 발생한 클러스터 구성 요소	소프트웨어 복구	하드웨어 복구
데이터 서비스	HA API, HA 프레임워크	없음
공용 네트워크 어댑터	네트워크 어댑터 페일오버(NAFO)	다중 공용 네트워크 어댑터 카드
클러스터 파일 시스템	1차 및 2차 복제본	멀티호스트 디스크
이중화된 멀티호스트 디스크	볼륨 관리(Solstice DiskSuite 및 VERITAS Volume Manager)	하드웨어 RAID-5(예: Sun StorEdge A3x00)
글로벌 장치	1차 및 2차 복제본	장치, 클러스터 전송 연결장치에 대한 다중 경로
독립 또는 개인 네트워크	HA 전송 소프트웨어	여러 개인용 하드웨어 독립 네트워크
노드	CMM, 페일페스트 드라이버	다중 노드

Sun Cluster 소프트웨어의 프레임워크는 가용성이 높기 때문에 노드 장애를 즉시 감지하고 프레임워크 자원을 제공할 새로운 서버를 클러스터의 남은 노드에 만듭니다. 모든 프레임워크 자원을 동시에 사용하지 못하는 경우는 없습니다. 훼손된 노드의 영향을 받지 않는 프레임워크 자원은 복구시에도 완전한 사용이 가능합니다. 또한 실패한 노드의 프레임워크 자원은 복구되는 대로 사용할 수 있게 됩니다. 복구된 프레임워크 자원은 다른 모든 프레임워크 자원이 복구를 완료할 때까지 기다리지 않아도 됩니다.

대부분의고가용성 프레임워크 자원은 복구될 때 이 자원을 사용하는 응용프로그램(데이터 서비스)에 투명하게 복구됩니다. 프레임워크 자원 액세스의 시멘틱은 노드장애로부터 완전하게 보존됩니다. 응용프로그램은 프레임워크 자원 서버가 다른 노드로 이동되었다는 것조차 알아채지 못합니다. 대체 하드웨어 경로가 다른 노드의 디스크에 대해 존재하는 한, 단일노드의 장애는 이 노드에 접속된 파일, 장치 및 디스크 볼륨을 사용하는 나머지 노드에 있는 프로그램에 완전하게 드러납니다(나타납니다). 예를 들어, 여러 노드에 대한 포트를 갖고 있는 멀티호스트 디스크를 사용할 경우가 있습니다.

클러스터 멤버십 모니터

CMM(Cluster Membership Monitor)은 클라이언트 시스템당 하나씩 분산된 에이전트 세트입니다. 에이전트들은 클러스터 상호 연결을 거쳐 메시지를 교환하여 다음을 수행합니다.

- 모든 노드(정족수)에 대한 일관된 멤버십 보기 시행
- 등록된 콜백을 사용하여 멤버십 변경사항에 대한 응답으로 동기화된 재구성 구동
- 클러스터 파티션 처리(기억분리, 기억상실)
- 모든 클라이언트 시스템 사이의 완전한 연결성 보장

이전 **Sun Cluster** 소프트웨어 릴리스와는 달리, **CMM**은 커널에서 전체적으로 실행됩니다.

클러스터 멤버십

CMM의 주요 기능은 클러스터에 포함된 노드세트 상에서 항상 클러스터 전역의 일치를 구현하는 것입니다. 이러한 제약 조건을 클러스터 멤버십이라고 합니다.

클러스터 멤버십을 판별하고, 궁극적으로 데이터 무결성을 보장하기 위해 **CMM**은 다음을 수행합니다.

- 노드 조인 또는 클러스터를 벗어나는 것과 같은 클러스터 멤버십에서의 변경을 설명합니다.
- “불량” 노드는 클러스터에서 나가도록 합니다.
- “불량” 노드는 수리될 때까지 클러스터 외부에 있게 합니다.
- 클러스터가 자체를 노드들의 서브세트로 파티션하지 않도록 합니다.

클러스터가 별도의 여러 클러스터로 스스로를 파티션하지 않도록 보호하는 방법에 대해서는 “정족수 및 정족수 장치”(43페이지)을 참조하십시오.

클러스터 멤버십 모니터 재구성

데이터를 훼손하지 않고 안전하게 보존하려면, 모든 노드가 클러스터 멤버십에서 일관되게 일치해야 합니다. **CMM**은 노드장애 대처시 필요한 경우 클러스터 서비스(응용 프로그램)의 클러스터 재구성을 조정합니다.

CMM은 클러스터 전송 계층으로부터 다른 노드에 대한 연결 정보를 수신합니다. **CMM**은 클러스터 상호 연결을 사용하여 재구성 동안의 상태 정보를 교환합니다.

클러스터 멤버십에서의 변경을 발견하면, **CMM**은 클러스터의 동기화된 구성을 수행하며, 이 때 클러스터 자원은 클러스터의 새로운 멤버십을 기초로 재분배될 수 있습니다.

페일패스트 메커니즘

CMM이 노드에서 치명적인 문제를 확인하면 클러스터 프레임워크를 호출하여 강제로 해당 노드를 종료시키고 클러스터 멤버십에서 제거합니다. 이러한 메커니즘을 페일패스트라고 합니다. 페일패스트 메커니즘은 다음 두 가지 방법으로 노드를 종료시킵니다.

- 노드가 클러스터에서 제거되고 노드 수가 부족한 상태로 새로운 클러스터를 시작하면 공유 디스크에 대한 액세스가 “금지”됩니다. 페일패스트 사용에 대한 자세한 내용은 (46페이지)“실패 방지”를 참조하십시오.
- 하나 이상의 클러스터 관련 데몬이 중단되면(clexecd, rpc.pmfcd, rgmd 또는 rpc.ed) CMM에서 장애가 확인되어 노드가 종료됩니다.

클러스터 데몬이 중단되어 노드가 종료되면 해당 노드의 콘솔에 다음과 같은 메시지가 나타납니다.

```
panic[cpu0]/thread=40e60: Failfast: Aborting because "pmfcd" died 35 seconds ago.
409b8 cl_runtime: __0FZsc_syslog_msg_log_no_argsPviTCPcTB+48 (70f900, 30, 70df54, 407acc, 0)
%10-7: 1006c80 000000a 000000a 10093bc 406d3c80 7110340 0000000 4001 fbf0
```

노드가 종료된 후에 다시 부팅되어 클러스터에 다시 연결될 수도 있고 OpenBoot PROM (OBP) 프롬프트 상태로 있을 수도 있습니다. 어떤 작업을 할 것인지는 OBP의 auto-boot? 매개 변수 설정에 따라 결정됩니다.

CCR(Cluster Configuration Repository)

CCR(Cluster Configuration Repository)은 클러스터의 구성 및 상태에 대한 정보를 저장하기 위한 전체 클러스터-범위의 개인용 데이터베이스입니다. CCR은 분배 데이터베이스입니다. 각 노드는 데이터베이스의 전체 사본을 관리합니다. CCR을 사용하면 모든 노드에서 일관된 화면으로 클러스터 “전체”를 볼 수 있습니다. 데이터가 손상되지 않도록 하려면 각 노드에서 클러스터 자원의 현재 상태를 알아야 합니다.

CCR은 2단계 실행 알고리즘을 사용하여 갱신합니다. 갱신은 모든 클러스터 구성원에 성공적으로 이루어져야 하며, 그렇지 않을 경우는 갱신되지 않습니다(롤백됩니다). CCR은 클러스터 상호 연결을 사용하여 분배된 갱신사항을 적용합니다.



주의: CCR이 텍스트 파일로 구성되어도, 수동으로 CCR 파일들을 편집하지 마십시오. 각 파일에는 일관성이 유지되도록 체크섬 레코드가 포함됩니다. 수동으로 CCR 파일을 갱신하면 노드나 전체 클러스터가 기능 수행을 정지시킬 수 있습니다.

CCR은 정족수가 확립될 때만 클러스터가 실행되도록 하기 위해 CMM에 의존합니다. CCR은 클러스터에서 데이터 일관성을 확인해야 하는 책임을 갖고 있으므로 필요에 따라 복구를 수행하고 데이터를 갱신합니다.

글로벌 장치

SunPlex 시스템은 글로벌 장치를 사용하여, 장치가 실제로 접속되어 있는 곳에 관계없이 어떤 노드에서나 클러스터의 장치를 사용할 수 있도록 클러스터 전반의고가용성 액세스를 제공합니다. 일반적으로, 글로벌 장치에 대한 액세스를 제공할 때 노드가 실패하면, **Sun Cluster** 소프트웨어는 자동으로 그 장치에 대한 경로를 찾아서 해당되는 경로로 액세스를 보냅니다. **SunPlex** 글로벌 장치로는 디스크, **CD-ROM** 및 테이프가 있습니다. 그러나, 디스크는 유일하게 지원되는 멀티포트 글로벌 장치입니다. 즉, **CD-ROM** 및 테이프 장치는 현재고가용성 장치가 아닙니다. 각 서버의 로컬 디스크 역시 멀티포트 상태가 아니므로고가용성 장치가 아닙니다.

클러스터는 자동으로 클러스터 내의 각 디스크, **CD-ROM** 및 테이프 장치에 고유한 **ID**를 할당합니다. 이러한 할당으로 인해 클러스터의 임의의 노드에서 각 장치에 일관되게 액세스할 수 있습니다. 글로벌 장치 이름 공간은 `/dev/global` 디렉토리에 보유되어 있습니다. 자세한 내용은 “글로벌 이름 공간”(39페이지)을 참조하십시오.

멀티포트 글로벌 장치는 장치에 대한 한 개 이상의 경로를 제공합니다. 멀티호스트 디스크의 경우, 디스크는 여러 노드에 의해 호스팅된 디스크 장치 그룹의 일부이므로 멀티호스트 디스크는 가용성이 높아집니다.

장치 ID(DID)

Sun Cluster 소프트웨어는 장치 **ID(DID)** 의사 드라이버로 알려진 구성을 통해 글로벌 장치를 관리합니다. 이 드라이버는 멀티호스트 디스크, 테이프 드라이브 및 **CD-ROM**과 같은 클러스터 내의 모든 장치에 자동으로 고유 **ID**를 할당하는 데 사용됩니다.

장치 **ID (DID)** 의사 드라이버는 글로벌 액세스 기능의 필수적인 부분입니다. **DID** 드라이버는 클러스터의 모든 노드를 규명하고 고유한 디스크 장치 목록을 만들어, 클러스터의 모든 노드에서 일관되는 고유한 메이저 번호와 마이너 번호를 각각에 할당합니다. 글로벌 장치에 대한 액세스가 수행되어 디스크에 대한 `c0t0d0`과 같은 기존 **Solaris** 장치 **ID** 대신 **DID** 드라이버에 의해 할당된 고유 장치 **ID**를 이용합니다.

이러한 방법을 사용하면 디스크 장치를 이용하는 응용프로그램 액세스 장치(원시 장치를 사용하는 볼륨 관리자나 응용 프로그램)가 일관된 경로를 사용할 수 있도록 합니다. 이러한 일관성은 멀티호스트 디스크에 대해 특히 유용합니다. 각 장치의 로컬 주 번호 및 부 번호는 노드마다 다를 수 있으므로 **Solaris** 장치 이름지정 규칙도 변경될 수 있기 때문입니다. 예를 들어, **node1**은 멀티호스트 디스크를 `c1t2d0`으로, **node2**는 `c3t2d0`으로 보는 등 동일한 디스크를 완전히 다르게 볼 수 있습니다. **DID** 드라이버는 글로벌 이름을 할당하여, **d10**처럼 각 노드에 멀티호스트 디스크에 대해 일관된 매핑을 제공합니다.

`scdidadm(1M)` 및 `scgdevs(1M)`를 통해 장치 **ID**를 갱신하고 관리합니다. 자세한 정보는 관련된 설명서 페이지를 참조하십시오.

디스크 장치 그룹

SunPlex 시스템에서 모든 멀티호스트 디스크는 Sun Cluster 소프트웨어의 제어 하에 있어야 합니다. 먼저 멀티호스트 디스크에서 볼륨 관리자 디스크 그룹(Solstice DiskSuite 디스크 세트 또는 VERITAS Volume Manager 디스크 그룹)을 작성합니다. 그리고 나서, 볼륨 관리자 디스크 그룹을 디스크 장치 그룹으로 등록합니다. 디스크 장치 그룹은 글로벌 장치의 유형입니다. 추가로, Sun Cluster는 모든 개별 디스크를 디스크 장치 그룹으로 등록합니다.

등록은 어느 노드가 어떤 볼륨 관리자 디스크 그룹에 대한 경로를 가질 것인지에 대한 정보를 SunPlex 시스템에 제공합니다. 그러면 클러스터 전체에서 볼륨 관리자 디스크 그룹에 액세스할 수 있습니다. 둘 이상의 노드를 (마스터) 디스크 장치 그룹에 쓸 수 있으면 해당 디스크 장치 그룹에 저장된 데이터의 가용성이 높아집니다. 가용성이 높은 디스크 장치 그룹을 사용하면 클러스터 파일 시스템을 하우징할 수 있습니다.

주: 디스크 장치 그룹은 자원 그룹의 영향을 받지 않습니다. 하나의 노드가 데이터 서비스에 의해 액세스되는 디스크 그룹을 마스터할 때, 다른 노드가 (데이터 서비스 프로세스 그룹을 나타내는) 자원 그룹을 마스터할 수 있습니다. 그러나 가장 실용적인 것은 동일한 노드에서 응용프로그램의 자원(응용프로그램 디먼)을 포함하는 자원 그룹과 특수 응용프로그램의 데이터를 저장하는 디스크 장치 그룹을 보존하는 것입니다. 디스크 장치 그룹과 자원 그룹 사이의 연관에 대한 자세한 정보는 *Sun Cluster 3.0 U1 Data Services Installation and Configuration Guide*의 개요 장을 참조하십시오.

디스크 장치 그룹을 사용하면 볼륨 관리자 디스크 그룹이 “글로벌”이 됩니다. 즉, 구성 디스크에 복수 경로 지원을 제공합니다. 실제로 멀티호스트 디스크에 접속된 각 클러스터 노드는 디스크 장치 그룹에 대한 경로를 제공합니다.

디스크 장치 그룹 페일오버

디스크 인클로저는 여러 개의 노드에 연결되어 있으므로 그 인클로저에 있는 모든 디스크 장치 그룹은 현재 장치 그룹을 마스터하는 노드가 실패할 경우에 대체 경로를 통해 액세스할 수 있습니다. 장치 그룹을 마스터하는 노드의 실패는 복구 및 일관성 검사를 수행하는데 시간이 소요되는 것을 제외하고는 장치 그룹에 대한 액세스에 영향을 주지 않습니다. 이 시간 동안, 모든 요청은 시스템이 장치 그룹을 사용가능하게 할 때까지 정체됩니다(응용프로그램에서 알 수 있음).

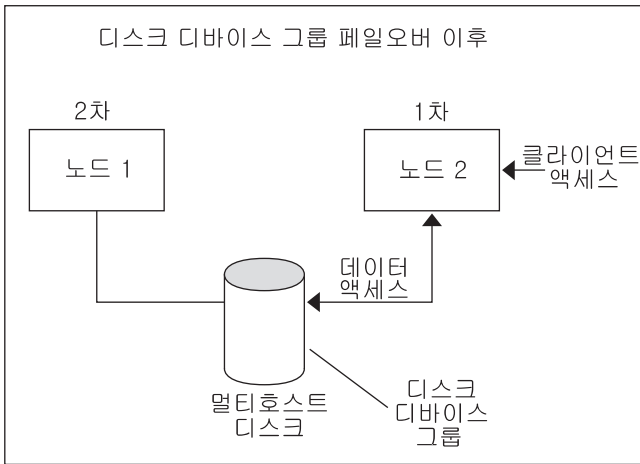
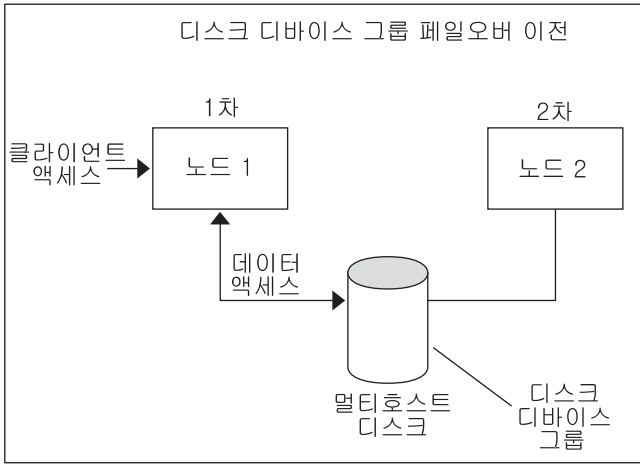


그림 3-2 디스크 장치 그룹 페일오버

글로벌 이름 공간

글로벌 장치를 작동 가능하게 하는 Sun Cluster 소프트웨어 메커니즘은 글로벌 이름 공간입니다. 글로벌 이름 공간에는 볼륨 관리자 이름 공간 뿐 아니라 `/dev/global/` 계층이 포함됩니다. 글로벌 이름 공간은 멀티호스트 디스크와 로컬 디스크(그리고 CD-ROM 및 테이프와 같은 다른 클러스터 장치) 둘 다를 반영하며, 멀티호스트 디스크에 대한 여러 페일오버 경로를 제공합니다. 실제로 멀티호스트 디스크에 연결된 각 노드는 모든 클러스터의 노드에 대한 기억 장치 경로를 제공합니다.

일반적으로 볼륨 관리자 이름 공간은 Solstice DiskSuite의 경우, `/dev/md/diskset/dsk` (그리고 `rdsk`) 디렉토리에 상주하고, VxVM의 경우에는 `/dev/vx/dsk/disk-group` 및 `/dev/vx/rdsk/disk-group` 디렉토리에 상주합니다. 이름 공간들은 각각

클러스터를 통해 가져온 각 **Solstice DiskSuite** 디스크 세트와 각 **VxVM** 디스크 그룹에 대한 디렉토리로 구성됩니다. 이 디렉토리 각각은 해당되는 디스크 세트나 디스크 그룹의 볼륨이나 메타 장치 각각에 대해 장치 노드를 하우징합니다.

SunPlex 시스템에서 로컬 볼륨 관리자 이름 공간의 장치 노드 각각은 `/global/.devices/node@nodeID` 파일 시스템의 장치 노드에 대한 상징적 링크에 의해 대체됩니다. 여기서, **nodeID**는 클러스터에서 노드를 나타내는 정수입니다. **Sun Cluster** 소프트웨어는 계속해서 표준 위치에서 상징적 링크로 볼륨 관리자 장치를 제시합니다. 글로벌 이름 공간과 표준 볼륨 관리자 이름 공간 둘 다 임의의 클러스터 노드에서 사용할 수 있습니다.

글로벌 이름 공간의 장점은 다음과 같습니다.

- 각 노드는 장치 관리 모델에서 거의 변경되지 않고 독립적으로 남아 있습니다.
- 장치는 선택적으로 글로벌이 될 수 있습니다.
- 협력업체 링크 생성기는 계속 작동될 수 있습니다.
- 로컬 장치 이름이 제공되면, 해당되는 글로벌 이름을 확보하도록 쉬운 매핑이 제공 됩니다.

로컬 및 글로벌 이름 공간 예제

다음 테이블은 `c0t0d0s0` 멀티호스트 디스크에 대한 로컬 및 글로벌 이름 공간 사이의 매핑을 보여줍니다.

표3-2 로컬 및 글로벌 이름 공간 매핑

Component/ Path	로컬 노드 이름 공간	글로벌 이름 공간
Solaris 논리 이 름	<code>/dev/dsk/ c0t0d0s0</code>	<code>/global/.devices/node@ID /dev/dsk/ c0t0d0s0</code>
DID 이름	<code>/dev/did/ dsk/d0s0</code>	<code>/global/.devices/node@ID /dev/dsk/ c0t0d0s0</code>
Solstice DiskSuite	<code>/dev/md/ diskset/dsk/d0</code>	<code>/global/.devices/node@ID /dev/md/diskset/ dsk/d0</code>
VERITAS Volume Manager	<code>/dev/vx/dsk/ disk-group /v0</code>	<code>/global/.devices/node@ID /dev/vx/dsk/ disk-group/v0</code>

글로벌 이름 공간은 설치시 자동으로 생성되어 재구성을 다시 시동할 때마다 갱신됩니다. `scgdevs (1M)` 명령을 실행하여 글로벌 이름 공간을 생성할 수도 있습니다.

클러스터 파일 시스템

클러스터 파일 시스템은 한 노드의 커널과 디스크에 물리적으로 연결되어 있는 노드에서 실행되는 볼륨 관리자 및 기초 파일 시스템 사이의 프록시입니다.

클러스터 파일 시스템은 하나 이상의 노드에 실제로 연결되어 있는 글로벌 장치(디스크, 테이프, CD-ROM)에 종속됩니다. 글로벌 장치는 노드가 기억 장치에 실제 연결되었는지 여부에 관계없이 동일한 파일 이름(예: /dev/global/)을 통해 클러스터의 노드로부터 액세스할 수 있습니다. 글로벌 장치를 일반 장치와 동일하게 사용할 수 있습니다. 즉, `newfs`와(또는) `mkfs`를 이용하여 파일 시스템을 작성할 수 있습니다.

`mount -g` 명령을 사용하여 글로벌 장치의 파일 시스템을 전역으로 마운트할 수도 있고 `mount` 명령을 사용하여 로컬로 마운트할 수도 있습니다.

클러스터의 모든 노드에서 프로그램이 동일한 파일 이름(예: /global/foo)을 사용하여 클러스터 파일 시스템의 파일에 액세스할 수 있습니다.

클러스터 파일 시스템은 모든 클라이언트 시스템에 마운트됩니다. 클라이언트 시스템의 서버세트에서 클러스터 파일 시스템을 마운트할 수 없습니다.

클러스터 파일 시스템은 별도로 구분되는 파일 시스템 형식이 아닙니다. 즉, 클라이언트는 하부 파일 시스템(예: UFS)을 봅니다.

클러스터 파일 시스템 사용

SunPlex 시스템에서 모든 멀티호스트 디스크는 디스크 장치 그룹으로서 구성됩니다. 이 그룹은 **Solstice DiskSuite** 디스크 세트, **VxVM** 디스크 그룹 또는 소프트웨어 기반 볼륨 관리자의 제어 하에 있지 않은 개인 디스크가 될 수 있습니다.

클러스터 파일 시스템의 가용성을 높이려면 하부 디스크 기억 장치를 둘 이상의 노드에 연결해야 합니다. 따라서 클러스터 파일 시스템에 만든 로컬 파일 시스템(노드의 로컬 디스크에 저장된 파일 시스템)은 가용성이 높지 않습니다.

일반적인 파일 시스템과 같이 두 가지 방법으로 클러스터 파일 시스템을 마운트할 수 있습니다.

- 수동—다음과 같이 `mount` 명령과 `-g` 또는 `-o global` 마운트 옵션을 사용하여 명령줄에서 클러스터 파일 시스템을 마운트합니다.

```
# mount -g /dev/global/dsk/d0s0 /global/oracle/data
```

- 자동—부팅할 때 클러스터 파일 시스템을 마운트하도록 `global` 마운트 옵션을 사용하여 `/etc/vfstab` 파일에 항목을 만듭니다. 그런 다음 모든 노드에서 `/global` 디렉토리에 마운트 지점을 작성할 수 있습니다. `/global` 디렉토리는 권장 위치이며 반드시 지정해야 하는 것은 아닙니다. 다음은 `/etc/vfstab` 파일에서 클러스터 파일 시스템에 대한 예문입니다.

```
/dev/md/oracle/dsk/d1 /dev/md/oracle/rdsk/d1 /global/oracle/data  
ufs 2 yes global,logging
```

주: Sun Cluster 소프트웨어가 클러스터 파일 시스템에 대한 이름지정 정책을 제시하지는 않지만, `/global/ disk-device-group`과 같은 동일한 디렉토리 아래의 모든 클러스터 파일 시스템에 대한 마운트 지점을 작성하여 관리를 용이하게 할 수는 있습니다. 자세한 정보는 **Sun Cluster 3.0 U1** 설치 안내서 및 **Sun Cluster 3.0 U1** 시스템 관리 안내서의 내용을 참조하십시오.

클러스터 파일 시스템의 기능

클러스터 파일 시스템에는 다음과 같은 기능이 있습니다.

- 파일 액세스 위치가 투명합니다. 프로세스는 위치에 관계없이 시스템에 있는 파일을 열 수 있으므로 모든 노드의 프로세스들은 동일한 경로 이름을 사용하여 파일을 찾을 수 있습니다.
- 동기 프로토콜을 사용하여 파일이 동시에 여러 노드로부터 액세스될 경우에도 UNIX 파일 액세스 시맨틱을 보존합니다.
- 효율적으로 파일 데이터를 이동하기 위하여 무복사 벌크 I/O 이동과 함께 확장 캐싱이 사용됩니다.
- 클러스터 파일 시스템은 `fcntl(2)` 인터페이스를 사용하여 가용성이 높은 권고 파일 잠금 기능을 제공합니다. 여러 클러스터 노드에서 실행되는 응용프로그램은 클러스터 파일 시스템의 파일에 대하여 권고 파일 잠금 기능을 사용하여 데이터 액세스를 동기화할 수 있습니다. 클러스터에서 제거되는 노드와 잠금을 유지하는 동안 장애가 발생하는 응용프로그램에서는 즉시 파일 잠금이 복구됩니다.
- 장애가 발생할 경우에도 데이터에 대한 액세스는 계속할 수 있습니다. 응용프로그램은 디스크에 대한 경로가 계속 작동하면 실패하지 않습니다. 이러한 보장은 원시 디스크 액세스와 모든 파일 시스템 조작에 대해 유지됩니다.
- 클러스터 파일 시스템은 하부 파일 시스템 및 볼륨 관리 소프트웨어와 독립적으로 작동합니다. 클러스터 파일 시스템은 지원되는 디스크의 파일 시스템을 모두 글로벌로 만듭니다.

Syncdir 마운트 옵션

`syncdir` 마운트 옵션은 UFS를 하부 파일 시스템으로 사용하는 클러스터 파일 시스템에 사용할 수 있습니다. 그러나 `syncdir`을 지정하지 않을 경우 중요한 성능 향상이 있습니다. `syncdir`을 지정한 경우, 쓰기의 POSIX 호환은 보증됩니다. 그렇게 하지 않으면, NFS 파일 시스템에서와 같은 동작을 수반하게 됩니다. 예를 들어, `syncdir`이 없으면, 파일을 닫을 때까지 공간 부족 상태를 발견하지 못하는 경우가 있습니다. `syncdir` (및 POSIX 동작)이 있으면, 쓰기 조작 동안 공간 부족 상태가 발견되었을 것입니다. `syncdir`을 지정하지 않아서 생기는 문제점이 간혹 있지만, 이를 지정하지 않으면 성능 향상 이점을 얻을 수 있으므로 지정하지 않는 것을 권장합니다.

글로벌 장치와 클러스터 파일 시스템에 대한 FAQ는 “파일 시스템 FAQ”(64페이지)를 참조하십시오.

정족수 및 정족수 장치

클러스터 노드는 데이터와 자원을 공유하기 때문에 클러스터가 동시에 활성중인 별도의 파티션으로 분리되지 않도록 해야 합니다. CMM을 사용하면 클러스터 상호 연결이 분할된 경우에도 하나 이상의 클러스터가 항상 작동합니다.

클러스터 파티션에서는 기억 분리 및 기억 상실 두 가지 문제가 발생합니다. 기억 분리는 노드들 사이의 클러스터 상호 연결이 유실되고 클러스터가 서브 클러스터로 분할될 경우에 발생합니다. 이 때, 각 서브 클러스터는 스스로를 유일한 파티션이라고 간주합니다. 이 문제는 클러스터 노드 사이의 통신 문제 때문에 발생합니다. 기억 상실은 클러스터가 시스템 종료 시간보다 이전에 클러스터 데이터에 대해 시스템을 종료한 후 재시작할 경우에 발생합니다. 프레임워크 데이터 버전이 여러 개 있고 최신 버전을 사용할 수 없을 때 새로 구성된 클러스터가 시작되면 이 문제가 발생할 수 있습니다.

각 노드에 한 표만 부여하고 대부분의 표가 작동 클러스터에 참여하도록 지시하면 기억 분리 및 기억 상실 문제를 방지할 수 있습니다. 대부분의 표가 있는 파티션은 정족수가 있기 때문에 작동할 수 있습니다. 클러스터에 노드가 두 개 이상이면 이러한 다수 표 메커니즘을 사용할 수 있습니다. 노드가 두 개인 클러스터에서는 다수가 둘입니다. 이러한 클러스터가 분할되면 각 파티션이 정족수를 얻기 위해 외부 표가 필요합니다. 필요한 외부 표는 정족수 장치에서 제공합니다. 두 노드 사이에 공유하는 디스크를 정족수 장치로 사용할 수 있습니다. 정족수 장치로 사용되는 디스크에는 사용자 데이터가 저장될 수 있습니다.

표 3-3에서는 Sun Cluster 소프트웨어에서 정족수를 사용하여 기억 분리 및 기억 상실 문제를 해결하는 방법을 설명합니다.

표 3-3 클러스터 정족수와 브레인 분할 및 엠네시아 문제점

파티션 유형	정족수 해결 방법
기억 분리	대다수의 투표가 있는 파티션(서브 클러스터)만 실행되도록 합니다(그러한 대다수의 파티션은 하나만 존재합니다).
기억 상실	클러스터가 부트될 경우, 최근 클러스터 멤버십의 구성원이었던 최소한 하나의 노드가 존재하도록 보증합니다.(그러므로 최근 구성 데이터를 수반함).

정족수 알고리즘은 동적으로 작동됩니다. 클러스터 이벤트가 계산을 트리거하면 클러스터 수명 동안 계산 결과가 변경될 수 있습니다.

정족수 투표 계산

클러스터 노드와 정족수 장치가 모두 투표하여 정족수를 채웁니다. 기본적으로, 클러스터 노드는 부트하여 클라이언트 시스템이 될 때 하나의 정족수 투표수를 확보합니다. 또한 노드가 설치되거나 관리자가 노드를 유지보수 상태에 둘 경우, 노드는 투표 계수를 확보하지 못합니다.

정족수 장치는 장치에 대한 노드 연결 수를 기초로 정족수 투표수를 확보할 수 있습니다. 정족수 장치가 설정되면, 최대 투표수 $N-1$ 을 확보합니다. 여기서 N 은 정족수 장치에 대한 포트를 갖고 있는, 투표수가 0이 아닌 노드들의 수입니다. 예를 들어, 투표수가 0이 아닌 두 노드에 연결된 정족수 장치는 정족수가 1입니다(2 - 1).

클러스터 설치 동안 정족수 장치를 구성하거나 나중에 *Sun Cluster 3.0 U1* 시스템 관리 안내서에서 설명된 절차를 사용하여 정족수 장치를 구성합니다.

주: 정족수 장치는 현재 접속되어 있는 노드들 중에서 최소한 하나의 노드가 클라이언트 시스템일 경우에만 투표수에 포함됩니다. 또한 클러스터 부트 시, 정족수 장치는 현재 접속되어 있는 노드들 중에서 최소한 하나의 노드가 부팅되고 시스템이 종료되었을 때 최근에 부트된 클러스터의 구성원일 경우에만 투표수에 포함됩니다.

정족수 구성

정족수 구성은 클러스터의 노드 수에 따라 달라집니다.

- **2-노드 클러스터 - 2-노드 클러스터를 형성하려면 두 개의 정족수 투표가 필요합니다.** 이러한 두 개의 투표는 두 개의 클러스터 노드 또는 하나의 노드와 정족수 장치로부터 제공됩니다. 그러나, 다른 노드가 실패할 경우 하나의 노드가 계속될 수 있도록 하려면 2-노드 클러스터에 정족수 장치가 구성되어 있어야 합니다.
- **노드가 세 개 이상인 클러스터 - 디스크 기억 장치 인클로저에 대한 액세스를 공유하는 모든 노드 쌍 사이에 정족수 장치를 지정해야 합니다.** 예를 들어, 그림 3-3에서 보여지는 클러스터와 비슷한 3-노드 클러스터가 있는 것으로 가정됩니다. 이 그림에서, **nodeA** 및 **nodeB**는 동일한 디스크 인클로저에 대한 액세스를 공유하고 **nodeB** 및 **nodeC**는 다른 디스크 인클로저에 대한 액세스를 공유합니다. 총 5개의 정족수 투표가 있습니다. 노드에서 세 개와 노드들 사이에 고유하는 정족수 장치에서 두 개의 투표가 제공됩니다. 클러스터를 형성하려면 정족수 투표에서 다수가 필요합니다.

디스크 기억 장치 인클로저에 대한 액세스를 공유하는 모든 노드 쌍 사이에 정족수 장치를 지정할 필요는 없습니다. 그렇지 않으면 *Sun Cluster*에 의해 수행됩니다. 그러나, $N+1$ 구성이 2-노드 클러스터로 변화한 곳의 경우 필요한 정족수 투표를 제공할 수 있으며, 두 개의 모든 디스크 인클로저에 액세스하는 노드 또한 실패합니다. 모든 쌍에 정족수 장치를 구성하는 경우, 남은 노드는 클러스터로서 작동할 수 있습니다.

이러한 구성의 예는 그림 3-3를 참조하십시오.

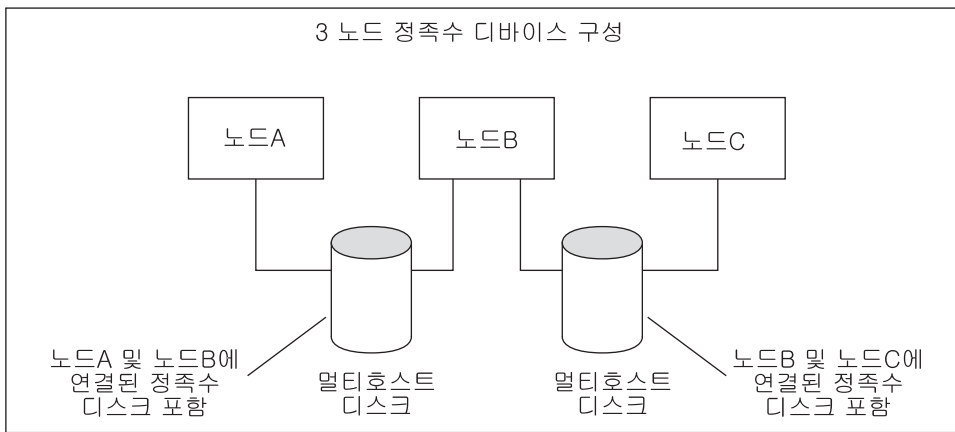
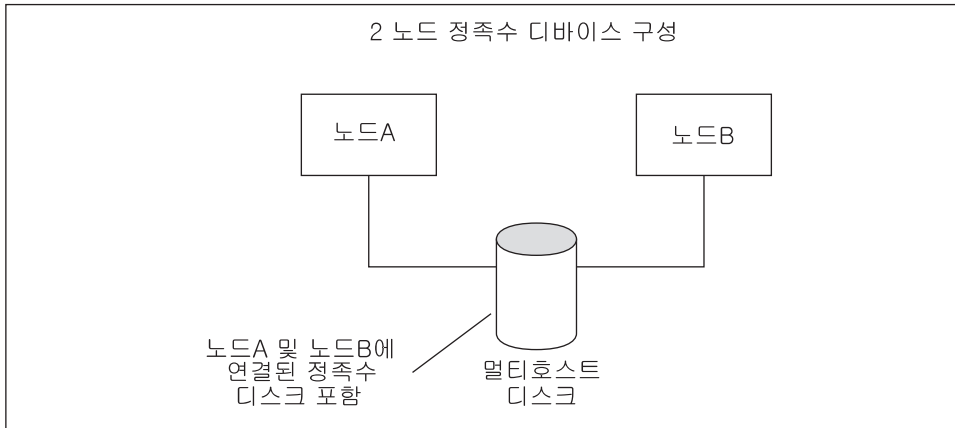


그림 3-3 정족수 장치 구성 예

정족수 지침

정족수 장치를 설정할 때 다음 지침을 사용하십시오.

- 동일한 공유 디스크 기억 장치 인클로저에 접속된 모든 노드 사이에 정족수 장치를 설정하십시오. 공유 인클로저 내에서 정족수 장치로 하나의 디스크를 추가하여, 노드가 실패할 경우 다른 노드가 정족수를 유지하여 공유 인클로저의 디스크 장치 그룹을 마스터하도록 하십시오.
- 최소한 두 개의 노드에서 정족수 장치를 연결해야 합니다.
- 정족수 장치는 이중 포트 정족수 장치로 사용되는 **SCSI-2** 또는 **SCSI-3** 디스크가 될 수 있습니다. 세 개 이상의 노드에 연결된 디스크는 정족수 장치로 사용되는지에 관계없이 **SCSI-3 PGR(Persistent Group Reservation)**을 지원해야 합니다. 자세한 정보는 **Sun Cluster 3.0 U1** 설치 안내서에서 계획 장을 참조하십시오.
- 정족수 장치로 사용자 데이터를 포함하는 디스크를 사용할 수 있습니다.

조언: 각 정족수 장치에 장애가 발생하지 않도록 보호하려면 노드 세트 사이에 두 개 이상의 정족수 장치를 구성하십시오. 다른 인클로저의 디스크를 사용하고, 각 노드 사이에 홀수 개의 정족수 장치를 구성하십시오.

실패 방지

클러스터에 대한 주요 문제점은 클러스터가 파티션되는(브레인 분할이라고 함) 실패입니다. 이러한 상황이 발생하면, 모든 노드가 통신할 수 있는 것은 아니므로 개인 노드나 노드 서브세트가 개인 또는 서브세트 클러스터를 형성할 수도 있습니다. 각 서브세트 또는 파티션은 멀티호스트 디스크에 대해 단 하나의 액세스 및 소유권을 갖고 있는 것으로 인식할 수도 있습니다. 디스크에 기록하려고 하는 여러 노드들로 데이터가 훼손될 수 있습니다.

실패 방지는 실제로 디스크에 대한 액세스를 금지하여 멀티호스트 디스크에 대한 노드 액세스를 제한합니다. 노드가 클러스터에서 나갈 경우(실패하거나 파티션되어), 실패 방지는 그 노드가 더이상 디스크에 액세스할 수 없게 만듭니다. 현재 구성원 노드들만 디스크에 대한 액세스를 갖게 되므로, 데이터 무결성이 유지됩니다.

디스크 장치 서비스는 멀티호스트 디스크를 사용하는 서비스에 대한 페일오버 기능을 제공합니다. 현재 디스크 장치 그룹의 1차(소유자) 노드로서 서비스를 제공하는 클라이언트 시스템이 실패하거나 도달할 수 없게 되면, 새로운 1차 노드가 선택되고, 계속해서 부수적인 인터럽트만으로 디스크 장치 그룹에 액세스할 수 있게 합니다. 이 프로세스 동안, 이전의 1차 노드는 새로운 1차 노드가 시작되기 전에 장치에 대한 액세스를 멈춰야만 합니다. 그러나 구성원이 클러스터 밖으로 제거되어 도달할 수 없게 되면, 클러스터는 1차 노드였던 장치들을 해제하도록 노드에 알릴 수 없습니다. 그러므로 살아남은 구성원이 실패한 구성원으로부터 글로벌 전역을 제어하고 액세스할 수 있도록 하는 수단이 필요합니다.

SunPlex 시스템은 SCSI 디스크 예약을 사용하여 실패 방지를 구현합니다. SCSI 예약을 사용하면, 실패한 노드는 멀티호스트 디스크로부터 “금지되어” 디스크에 액세스할 수 없게 됩니다.

SCSI-2 디스크 예약은 디스크에 접속된 모든 노드에 대한 액세스를 부여하거나(어떤 예약도 없을 경우) 단일 노드(예약이 있는 노드)에 대한 액세스로 제한하는 예약 양식을 지원합니다.

다른 노드가 더이상 클러스터 상호 연결을 통해 통신할 수 없음을 클라이언트 시스템이 발견하면, 그 구성원은 실패 방지 절차를 초기화하여 다른 노드가 공유 디스크에 액세스하지 못하도록 합니다. 이러한 실패 방지가 발생할 경우, 콘솔의 “예약 충돌” 메시지와 함께 금지된 노드가 공황 상태가 되는 것이 정상입니다.

예약 충돌은 노드가 더이상 클라이언트 시스템이 아님을 발견한 후 SCSI 예약이 이 노드와 다른 노드 사이에 공유하는 모든 디스크에 놓이므로 발생합니다. 금지된 노드는 금지되고 있음을 인식하지 못할 수 있으므로 공유 디스크 중 하나에 액세스하려면 예약을 발견하여 공황 상태에 빠집니다.

장애 방지를 위한 파일패스트 메커니즘

장애가 발생한 노드가 다시 부팅되어 공유 기억 장치에 쓰지 못하도록 하기 위하여 클러스터 프레임워크에서 사용하는 메커니즘을 파일패스트라고 합니다.

클러스터를 구성하는 노드는 정족수 디스크를 포함하여 액세스할 수 있는 디스크에 대하여 특정 **ioctl**, **MHIOCENFAILFAST**를 계속 사용할 수 있도록 합니다. 이 **ioctl**은 디스크 드라이버에 대한 지시어이고, 디스크가 다른 노드에 예약되어 디스크에 액세스할 수 없을 경우에 노드가 종료될 수 있도록 합니다.

MHIOCENFAILFAST ioctl은 노드에서 **Reservation_Conflict** 오류 코드에 대하여 디스크에 실행하는 모든 읽기 및 쓰기 작업에서 반환되는 오류를 검사하도록 합니다. **ioctl**은 주기적으로 백그라운드에서 디스크에 대한 테스트 작업을 실행하여 **Reservation_Conflict**를 검사합니다. **Reservation_Conflict**가 반환되면 포그라운드 및 백그라운드 제어 흐름 경로가 모두 중단됩니다.

SCSI-2 디스크의 경우 예약이 지속되지 않습니다. 즉, 노드를 다시 부팅하면 없어집니다. **PGR(Persistent Group Reservation)**이 있는 **SCSI-3** 디스크의 경우에는 예약 정보가 디스크에 저장되어 노드를 다시 부팅한 후에도 유지됩니다. 파일패스트 메커니즘은 **SCSI-2** 디스크를 사용하는 경우나 **SCSI-3** 디스크를 사용하는 경우에 모두 동일하게 작동합니다.

노드가 클러스터의 다른 노드와 연결이 끊어지고 정족수를 채울 수 있는 파티션에 포함되지 않은 경우에는 다른 노드에 의해 강제로 클러스터에서 제거됩니다. 정족수를 채울 수 있는 파티션에 포함된 다른 노드가 공유 디스크에 예약을 설정하고, 정족수가 채워지지 않은 노드가 예약된 공유 디스크에 액세스하려고 시도하면 파일패스트 메커니즘에 의해 예약 충돌이 발생하여 종료됩니다.

노드가 종료된 후에 다시 부팅되어 클러스터에 다시 연결될 수도 있고 **OpenBoot PROM (OBP)** 프롬프트 상태로 있을 수도 있습니다. 어떤 작업을 할 것인지는 **OBP**의 **auto-boot?** 매개 변수 설정에 따라 결정됩니다.

볼륨 관리자

SunPlex 시스템은 미러 및 긴급 예비 디스크를 사용하여 데이터 가용성을 증가시키고 디스크 실패 및 교체를 처리하기 위해 볼륨 관리 소프트웨어를 사용합니다.

SunPlex 시스템에는 고유한 내부 볼륨 관리자 구성 요소가 없지만, 다음과 같은 볼륨 관리자의 영향을 받습니다.

- Solstice DiskSuite
- VERITAS Volume Manager

클러스터의 볼륨 관리 소프트웨어는 다음에 대한 지원을 제공합니다.

- 노드 실패에 대한 페일오버 처리
- 다른 노드로부터의 다중 경로 지원

■ 디스크 장치 그룹에 대한 투명한 원격 액세스

볼륨 관리 객체가 클러스터의 제어를 받으면 디스크 장치 그룹이 됩니다. 볼륨 관리자
에 대한 자세한 내용은 볼륨 관리자 소프트웨어 설명서를 참조하십시오.

주: 디스크 세트나 디스크 그룹을 계획할 때 고려해야 할 중요한 사항은 연관된 디스크
장치 그룹이 클러스터 내의 응용프로그램 자원(데이터)과 어떻게 연관되는 지를 이해하
는 것입니다. 이러한 문제에 대해서는 **Sun Cluster 3.0 U1 설치 안내서** 및 **Sun Cluster
3.0 U1 Data Services Installation and Configuration Guide**의 내용을 참조하십시오.

데이터 서비스

데이터 서비스라는 용어는 단일 서버에서 보다는 클러스터에서 실행되도록 구성된
Oracle 또는 **iPlanet Web Server**와 같은 다사 응용프로그램을 설명하기 위해 사용됩니
다. 데이터 서비스는 응용프로그램, 특수한 **Sun Cluster** 구성 파일, 다음과 같은 응용
프로그램 작업을 제어하는 **Sun Cluster** 관리 방법 등으로 구성됩니다.

- 시작
- 중지
- 모니터 및 수정 조치

그림 3-4에서는 단일 응용프로그램 서버에서 실행되는 응용프로그램(단일 서버 모
델)을 클러스터에서 실행되는 동일한 응용프로그램(클러스터 서버 모델)가 비교합니
다. 사용자의 관점에서 보면 클러스터 응용프로그램이 더 빠르게 실행될 수 있고 가용
성이 높다는 것 외에는 두 구성 사이에 큰 차이가 없습니다.

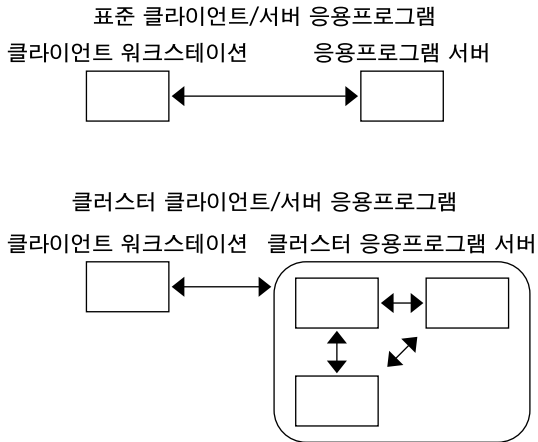


그림 3-4 표준 및 클러스터 클라이언트/서버 구성 비교

단일 서버 모델에서는 응용프로그램이 특정 공용 네트워크 인터페이스(호스트 이름)
를 통해 서버에 액세스하도록 구성합니다. 따라서 호스트 이름이 물리적 서버와 연결
됩니다.

클러스터 서버 모델에서는 공용 네트워크 인터페이스가 논리적 호스트 이름이나 공유 주소입니다. 네트워크 자원은 논리적 호스트 이름과 공유 자원을 모두 나타내는 용어입니다.

일부 데이터 서비스에서는 논리적 호스트 이름 또는 공유 주소를 네트워크 인터페이스로 지정해야 합니다. 두 가지를 동일하게 사용할 수 없습니다. 그 외의 데이터 서비스에서는 논리적 호스트 이름이나 공유 주소를 지정할 수 있습니다. 지정하는 인터페이스 종류에 대한 자세한 내용은 각 데이터 서비스에 대한 설치 및 구성을 참조하십시오.

네트워크 자원은 특정 물리적 서버와 연결되지 않고 물리적 서버 사이에 이주할 수 있습니다.

네트워크 자원은 처음에 하나의 노드(1차)와 연결됩니다. 1차 노드에 장애가 발생하면 네트워크 자원과 응용프로그램 자원이 다른 클러스터 노드(2차)로 페일오버합니다. 네트워크 자원이 페일오버하면 잠시 지연된 후에 응용프로그램 자원이 2차 노드에서 계속 실행됩니다.

그림 3-5에서는 단일 서버 모델과 클러스터 서버 모델을 비교합니다. 클러스터 서버 모델에서는 네트워크 자원(이 예에서는 논리적 호스트 이름)이 둘 이상의 클러스터 노드 사이에 이동할 수 있습니다. 응용프로그램은 특정 서버와 연결된 호스트 이름 대신 논리적 호스트 이름을 사용하도록 구성됩니다.

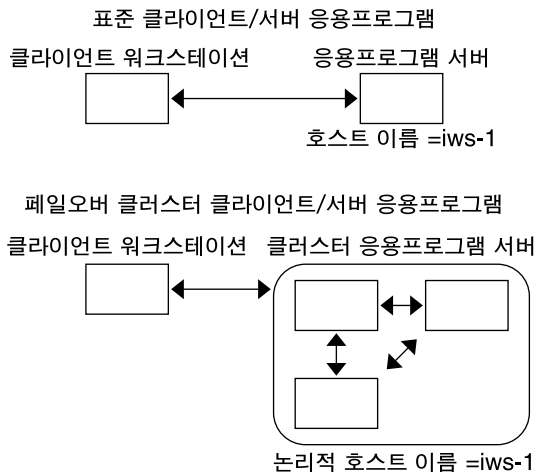


그림 3-5 고정 호스트 이름과 논리적 호스트 이름 비교

공유 주소도 처음에는 하나의 노드와 연결됩니다. 이 노드를 **GIN(Global Interface Node)**이라고 합니다. 공유 주소는 클러스터에 대한 단일 네트워크 인터페이스로 사용됩니다. 이러한 인터페이스를 글로벌 인터페이스라고 합니다.

논리적 호스트 이름 모델과 확장 가능한 서비스 모델 사이의 차이는 확장 가능한 서비스 모델에서는 각 노드에도 루프백 인터페이스에 활성으로 구성된 공유 주소가 있다는 것입니다. 이러한 구성 때문에 여러 노드에 대하여 여러 데이터 서비스 인스턴스를 동시에 실행할 수 있습니다. “확장 가능한 서비스”라는 용어는 클러스터 노드를 추가하여 응용프로그램에 CPU 기능을 추가하는 방법으로 성능을 확장할 수 있다는 의미입니다.

GIN에 장애가 발생하면 응용프로그램 인스턴스를 실행하는 다른 노드로 공유 주소를 변경하고 변경한 노드를 새 GIN으로 만들 수 있습니다. 아니면 이전에 응용프로그램을 실행하지 않던 다른 클러스터 노드로 페일오버할 수 있습니다.

그림 3-6에서는 단일 서버 구성과 확장 가능한 클러스터 서비스 구성을 비교합니다. 확장 가능한 서비스 구성에서는 모든 노드에 대한 공유 주소가 있습니다. 페일오버 데이터 서비스에 논리적 호스트 이름을 사용하는 방법과 유사한 방법으로 특정 서버와 연결된 호스트 이름 대신 이 공유 주소를 사용하도록 응용프로그램이 구성됩니다.

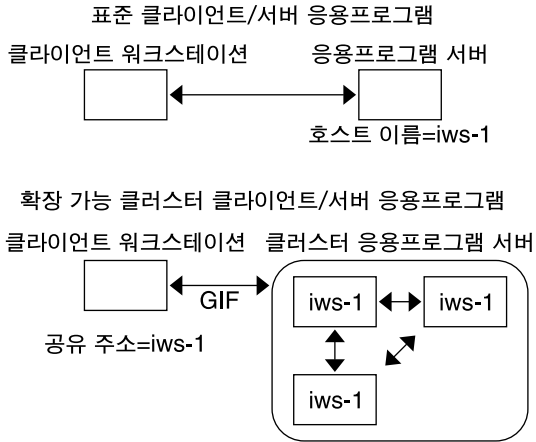


그림 3-6 고정 호스트 이름과 공유 주소 비교

데이터 서비스 방법

Sun Cluster 소프트웨어는 여러 가지 서비스 관리 방법을 지원합니다. 이 방법은 클러스터 노드에서 응용프로그램을 시작, 정지 및 모니터링하기 위해 이 방법을 사용하는 Resource Group Manager(RGM)의 제어 하에 실행됩니다. 이 메소드들은 클러스터 프레임워크 소프트웨어와 멀티호스트 디스크와 함께 응용프로그램이 고가용성 데이터 서비스가 되도록 합니다.

또한 RGM은 응용프로그램의 인스턴스와 네트워크 자원(논리 호스트이름 및 공유 주소)과 같은 클러스터 내의 자원도 관리합니다.

Sun Cluster 소프트웨어에서 제공하는 메소드에 추가로 SunPlex 시스템에서도 API 및 여러 가지 데이터 서비스 개발 도구를 제공합니다. 응용프로그램 프로그래머가 이 도구를 사용하면 다른 응용프로그램이 Sun Cluster 소프트웨어에서 가용성이 높은 데이터 서비스로 실행될 수 있도록 필요한 데이터 서비스 메소드를 개발할 수 있습니다.

Resource Group Manager(RGM)

RGM은 데이터 서비스(응용프로그램)를 자원 유형 구현에 의해 관리되는 자원으로 제어합니다. 이 구현은 Sun에서 제공되거나 개발자가 일반 데이터 서비스 템플릿,

DSDL API(데이터 서비스 개발 라이브러리 API) 또는 RMAPI(Resource Management API)를 사용하여 작성합니다. 클러스터 관리자는 *resource groups*라는 컨테이너에 자원을 만들고 관리합니다. RGM은 클러스터 멤버십 변경에 대한 응답으로 선택된 노드에서 자원을 정지하였다가 시작합니다.

RGM은 자원 및 자원 그룹에 대하여 작동합니다. RGM 작업으로 자원 및 자원 그룹이 온라인 및 오프라인 상태로 전환될 수 있습니다. 자원 및 자원 그룹에 적용할 수 있는 상태와 설정에 대한 자세한 설명은 (59페이지)“자원 및 자원 그룹의 상태와 설정” 단원을 참조하십시오.

페일오버 데이터 서비스

데이터 서비스가 실행되는 노드가 실패할 경우, 서비스는 사용자 간섭 없이 다른 작업 노드로 이주됩니다. 페일오버 서비스는 응용프로그램 인스턴스 자원과 네트워크 자원(논리 호스트 이름)에 대한 컨테이너인 페일오버 자원 그룹을 이용합니다. 논리 호스트 이름은 하나의 노드에서 구성될 수 있는 IP 주소로, 나중에 원래 노드에서 자동으로 구성이 중지되고 다른 노드에서 구성이 시작됩니다.

페일오버 데이터 서비스의 경우, 응용프로그램 인스턴스는 단일 노드에서만 실행됩니다. 결함 모니터가 오류를 발견하면, 데이터 서비스가 구성된 방법에 따라 동일한 노드에서 인스턴스를 재시작하려고 하거나 다른 노드에서 인스턴스를 시작하려고 합니다(페일오버).

확장 가능 데이터 서비스

확장 가능 데이터 서비스는 여러 노드에서 인스턴스가 사용되고 있을 가능성을 수반합니다. 확장 가능한 서비스는 두 가지 자원 그룹을 사용합니다. 확장 가능 자원 그룹은 확장 가능한 서비스가 영향을 받는 네트워크 자원(공유 주소)을 포함하도록 페일오버 자원 그룹과 응용프로그램 자원을 포함시킵니다. 확장 가능 자원 그룹은 여러 노드에서 온라인 상태로 있을 수 있으므로 서비스의 여러 인스턴스가 한번에 실행될 수 있습니다. 공유 주소를 호스팅하는 페일오버 자원 그룹은 한번에 한 노드에서만 온라인 상태입니다. 확장 가능 서비스에 호스팅하는 모든 노드가 동일한 공유 주소를 사용하여 서비스를 호스팅합니다.

서비스 요청은 단일 네트워크 인터페이스(글로벌 인터페이스)를 통해 클러스터에 제공되고 로드 밸런싱 정책에 의해 설정된 사전 정의된 몇 가지의 알고리즘 중 하나를 기초로 노드에 분산됩니다. 클러스터는 로드 밸런싱 정책을 사용하여 몇몇 노드 사이의 서비스 부하 균형을 맞추는 로드 밸런싱 정책을 사용할 수 있습니다. 다른 공유 주소에 호스트하는 다른 노드에 여러 개의 GIF가 있을 수 있으므로 유의하십시오.

확장 가능 서비스의 경우, 응용프로그램 인스턴스는 몇 개의 노드에서 동시에 실행될 수 있습니다. 글로벌 인터페이스를 호스트하는 노드가 실패할 경우, 글로벌 인터페이스는 다른 노드로 페일오버합니다. 응용프로그램 인스턴스 실행이 실패하는 경우, 인스턴스는 동일한 노드에서 재시작을 시도합니다.

응용프로그램 인스턴스는 동일한 노드에서 재시작될 수 없으며, 사용되지 않는 다른 노드는 서비스를 실행하도록 구성된 경우, 서비스는 사용되지 않는 노드로 페일오버

됩니다. 그렇지 않으면, 나머지 노드에서 실행을 계속하여, 서비스 처리량이 줄어들 수 있습니다.

주: 각 응용프로그램 인스턴스에 대한 TCP 상태는 GIF 노드가 아니라 인스턴스가 있는 노드에 보존됩니다. 그러므로 GIF 노드의 실패는 연결에 영향을 주지 않습니다.

그림 3-7 에는 페일오버, 확장 가능한 리소스 그룹, 확장 가능한 서비스를 위한 둘 사이의 관계 등에 대한 예가 있습니다. 이 예에는 세 가지 자원 그룹이 있습니다. 페일오버 자원 그룹에는고가용성 DNS에 대한 응용프로그램 자원과고가용성 DNS 및고가용성 Apache 웹 서버에서 사용되는 네트워크 자원이 포함됩니다. 확장 가능 자원 그룹에는 Apache 웹 서버의 응용프로그램 인스턴스만 포함됩니다. 확장 가능 및 페일오버 자원 그룹 사이에 자원 그룹 종등록 정보(실선)이 있고 모든 Apache 응용프로그램 자원이 공유 주소(실선)인 네트워크 자원 schost-2에 종속된다는 점에 유의하십시오.

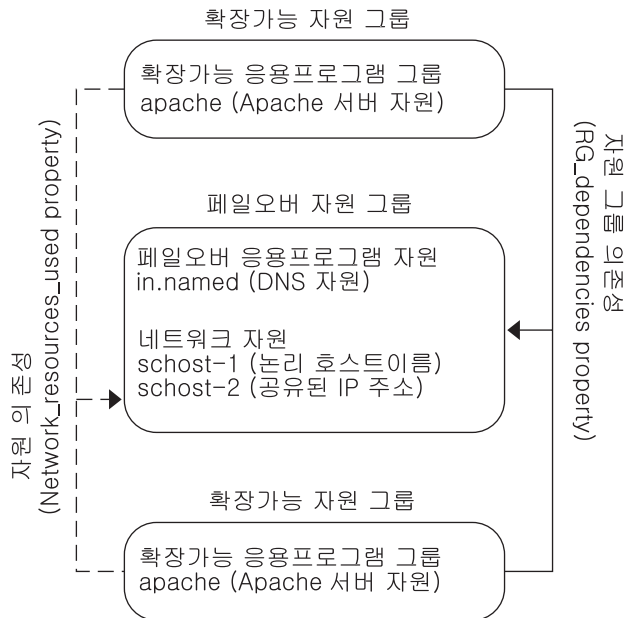


그림 3-7 페일오버 및 확장 가능 자원 그룹 예

확장 가능 서비스 구조

클러스터 네트워킹의 기본 목적은 데이터 서비스에 대한 확장성을 제공하는 것입니다. 확장성은 서비스에 대해 제공되는 로드가 증가하는 대로 데이터 서비스는 새로운 노드가 클러스터에 추가되고 새로운 서버 인스턴스가 실행되는 것처럼 이러한 작업로드 증가 시 일정한 응답 시간을 유지할 수 있음을 의미합니다. 그러한 서비스를 확장 가능 데이터 서비스라고 합니다. 확장 가능 데이터 서비스의 좋은 예는 웹 서비스입니다.

다. 일반적으로, 확장 가능 데이터 서비스는 몇 가지의 인스턴스로 구성되며, 각 인스턴스는 클러스터의 서로 다른 노드에서 실행됩니다. 이러한 인스턴스들은 해당 서비스의 원격 클라이언트 관점에서는 하나의 서비스로 작동하여 서비스 기능을 구현합니다. 예를 들어, 서로 다른 노드에서 실행되는 httpd 데몬으로 구성된 확장 가능한 웹 서비스가 있을 수 있습니다. httpd 데몬은 클라이언트 요청에 서비스를 제공할 수도 있습니다. 요청에 서비스를 제공하는 디먼은 로드 밸런싱 정책에 따라 다릅니다. 요청에 서비스를 제공한 특정 디먼이 아니라 서비스에서 제공될 클라이언트에 대한 응답이 표시되므로 단일 서비스 형태가 유지됩니다.

확장 가능 서비스는 다음으로 구성됩니다.

- 확장 가능 서비스에 대한 네트워킹 기반 구조 지원
- 로드 밸런싱
- 네트워킹 및 데이터 서비스에 대한 HA 지원(Resource Group Manager 사용)

다음 그림은 확장 가능 서비스 구조를 보여줍니다.

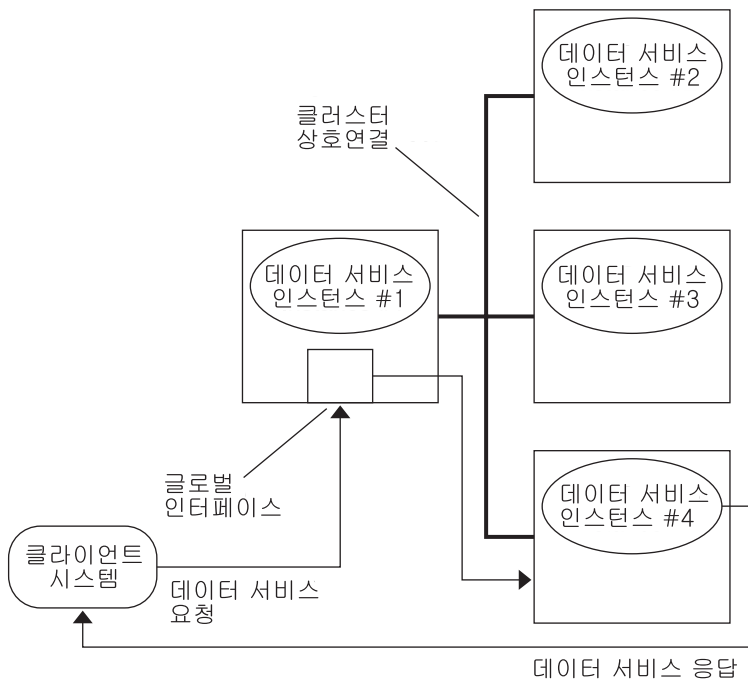


그림 3-8 확장 가능 서비스 구조

글로벌 인터페이스를 호스트하지 않는 노드(프록시 노드)에는 해당되는 루프백 인터페이스에 호스트된 공유 주소가 있습니다. 글로벌 인터페이스의 패킷은 구성 가능한 로드 밸런싱 정책을 기초로 다른 클러스터 노드에 분산됩니다. 가능한 로드 밸런싱 정책은 다음 부분에 설명되어 있습니다.

로드 밸런싱 정책

로드 밸런싱은 응답 시간과 처리량에서 확장 가능 서비스의 성능을 개선합니다.

확장 가능한 데이터 서비스에는 *pure* 및 *sticky* 두 가지가 있습니다. *pure* 서비스는 인스턴스가 클라이언트 요청에 응답할 수 있는 서비스입니다. *sticky* 서비스는 클라이언트가 같은 인스턴스에 요청을 보내는 서비스입니다. 그러한 요청은 다른 인스턴스에 보내지 않아도 됩니다.

pure 서비스는 가중된 로드 밸런싱 정책을 사용합니다. 이 로드 밸런싱 정책에서 클라이언트 요청은 기본적으로 클러스터의 서버 인스턴스에서 일정하게 분산됩니다. 예를 들어, 3-노드 클러스터에서 각 노드의 가중치가 1이라고 가정합니다. 각 노드는 해당 서비스 대신 클라이언트의 요청 중 1/3에 서비스를 제공합니다. 가중치는 언제든지 `scrgadm(1M)` 명령 인터페이스나 **SunPlex Manager GUI**를 통해 관리자에 의해 변경될 수 있습니다.

sticky 서비스에는 두 가지의 주요 기능인 보통 *sticky*와 와일드 카드 *sticky*가 있습니다. *Sticky* 서비스는 여러 **TCP** 연결을 거쳐 동시 응용프로그램 레벨 세션이 **in-state** 메모리(응용프로그램 세션 상태)를 공유할 수 있게 합니다.

보통 *sticky* 서비스는 클라이언트가 여러 개의 동시 **TCP** 연결 사이에 상태를 공유할 수 있게 합니다. 클라이언트는 단일 포트에서 청취하는 서버 인스턴스에 대해 “*sticky*”된다고 합니다. 클라이언트는 인스턴스가 남아 있고 액세스가 가능한 경우에 자체의 모든 요청이 동일한 서버 인스턴스로 가도록 보장받으며, 로드 밸런싱 정책은 서비스가 온라인인 동안에는 변경되지 않습니다.

예를 들어, 클라이언트의 웹 브라우저가 세 개의 서로 다른 **TCP** 연결을 사용하여 포트 80에서 고유 **IP** 주소에 연결되지만, 그 연결들은 서비스에서 캐시된 세션 정보를 교환합니다.

sticky 정책의 일반화는 동일한 인스턴스의 장면 뒤에서 세션 정보를 교환하는 여러 개의 확장 가능 서비스로 확장됩니다. 이러한 서비스가 같은 인스턴스의 장면뒤에서 세션 정보를 교환할 경우, 클라이언트는 서로 다른 포트에서 청취하는 동일 노드의 여러 서버 인스턴스에 대해 “*sticky*”된다고 합니다.

예를 들어, **e-commerce** 사이트의 고객이 포트 80에서 일반 **HTTP**를 사용하여 상품들로 시장 바구니를 채우지만, 바구니의 상품 대금을 신용 카드 지불할 경우 보안 데이터를 보내기 위해 포트 443의 **SSL**로 전환합니다.

와일드 카드 *sticky* 서비스는 동적으로 할당된 포트 번호를 사용하지만, 여전히 클라이언트 요청이 같은 노드로 갈 것으로 예상합니다. 클라이언트는 동일한 **IP** 주소에 대해 포트를 거치는 “*sticky* 와일드 카드”입니다.

이 정책의 좋은 예는 수동 모드 **FTP**입니다. 클라이언트는 포트 21의 **FTP** 서버에 연결되고, 동적 포트 범위의 청취자 포트 서버에 다시 연결하도록 서버에서 알립니다. **IP** 주소에 대한 모든 요청은 제어 정보를 통해 서버가 클라이언트를 알린 동일 노드로 전송됩니다.

이 *sticky* 정책 각각에 대해 가중된 로드 밸런싱 정책이 기본적으로 적용되므로 클라이언트의 초기 요청은 로드 밸런서에 의해 지시된 인스턴스에 보내집니다. 인스턴스가 실

행되는 노드에 대한 유사성을 클라이언트가 확립하고 나면, 차후 요청은 액세스할 수 있는 경우 그 인스턴스로 보내집니다. 그리고 로드 밸런싱 정책은 변경되지 않습니다.

특정 로드 밸런싱 정책에 대한 추가 세부사항이 아래에 설명되어 있습니다.

- **Weighted.** 로드는 지정된 가중치에 따라 다양한 노드들 사이에 분배됩니다. 이 정책은 `Load_balancing_weights` 등록 정보에 대해 `LB_WEIGHTED`를 사용하여 설정됩니다. 노드에 대한 가중치가 명백히 설정되지 않은 경우, 노드에 대한 가중치는 기본값인 하나가 됩니다.

이 정책은 라운드 로빈(**round robin**)이 아닙니다. 라운드 로빈 정책은 클라이언트의 각 요청이 항상 서로 다른 노드에 가도록 합니다. 첫번째 요청은 노드 1에, 두 번째 요청은 노드 2에 갑니다. 가중된 정책은 클라이언트로부터의 트래픽 중 특정 비율 만큼만 특정 노드에 보내지도록 합니다. 이 정책은 주소 개인 요청이 아닙니다.

- **Sticky.** 이 정책에서 포트 세트는 응용프로그램 자원이 구성된 시간에 알려집니다. 정책은 `Load_balancing_policy` 자원 등록 정보에 대해 `LB_STICKY` 값을 사용하여 설정됩니다.
- **Sticky-wildcard.** 이 정책은 일반 “sticky” 정책의 수퍼세트입니다. IP 주소에 의해 식별된 확장 가능 서비스의 경우, 포트는 서버에 의해 할당됩니다(미리 알려지지 않습니다). 포트는 변경될 수도 있습니다. 정책은 `Load_balancing_policy` 자원 등록 정보에 대해 `LB_STICKY_WILD` 설정을 사용하여 설정됩니다.

페일백 설정

자원 그룹은 한 노드에서 다른 노드로 페일오버됩니다. 그러면 초기의 2차 노드가 새로운 1차 노드가 됩니다. 페일백 설정은 초기 1차 노드가 다시 온라인 상태가 될 때 수행되는 작업을 지정합니다. 초기 1차 노드가 다시 1차 노드가 되는 페일백 옵션 또는 현재 1차 노드를 그대로 유지하는 옵션이 있습니다. `Failback` 리소스 그룹 등록 정보 설정을 사용하여 원하는 옵션을 지정합니다.

특정 인스턴스에서, 예를 들어 자원 그룹을 호스트하는 원래 노드가 반복적으로 실패하고 재부트될 경우, 페일백을 설정하면 자원 그룹에 대한 가용성이 감소될 수 있습니다.

데이터 서비스 결합 모니터

각각 **SunPlex** 데이터 서비스는 정기적으로 데이터 서비스를 규명하여 상태를 판별하는 결합 모니터를 제공합니다. 결합 모니터는 응용프로그램 디먼이 실행되는지, 클라이언트에 서비스가 제공되는지를 확인합니다. 프로브에 의해 반환된 정보를 기초로, 디먼을 재시작하고 페일오버를 야기하는 것과 같은 사전에 정의된 조치가 초기화될 수 있습니다.

새로운 데이터 서비스 개발

Sun에서는 클러스터에서 여러 응용프로그램이 페일오버 또는 확장 가능한 서비스로 작동하도록 할 수 있는 구성 파일과 관리 메소드 템플릿을 제공합니다.고가용성 데이

터 서비스로 실행할 응용프로그램이 현재 Sun에서 제공되는 것이 아니면, API 또는 DSDL API를 사용하여 응용프로그램을 취하고 이를 고가용성 데이터 서비스로 실행 되도록 구성할 수 있습니다.

응용프로그램이 페일오버 서비스가 될 수 있도록 할 것인지를 결정하는 기준이 있습니다. 특정 기준은 응용프로그램에 대해 사용할 수 있는 API를 설명하는 SunPlex 문서에 설명됩니다.

여기서 사용하는 서비스가 확장 가능 데이터 서비스 구조의 장점을 취할 수 있는지 알 수 있도록 도와주는 일부 지침을 제시하기로 하겠습니다. 확장 가능한 서비스에 대한 일반적인 내용은 (51페이지)“확장 가능 데이터 서비스” 절을 참조하십시오.

다음 지침을 만족시키는 새로운 서비스는 확장 가능 서비스로 사용할 수 있습니다. 기존의 서비스가 이러한 지침을 정확하게 따르지 않으면, 서비스가 지침을 따르도록 부분을 재작성해야 할 수도 있습니다.

확장 가능 데이터 서비스는 다음 특징을 가집니다. 먼저, 그러한 서비스는 하나 이상의 서버 인스턴스로 구성됩니다. 각 인스턴스는 클러스터의 다른 노드에서 실행됩니다. 동일한 서비스에서 두 개 이상의 인스턴스가 동일한 노드에서 실행할 수는 없습니다.

두 번째, 서비스가 외부 논리 데이터 저장소를 제공할 경우, 여러 서버 인스턴스로부터 이 저장소로의 동시 액세스를 동기화하여, 데이터가 변경되는 대로 데이터를 읽거나 갱신사항을 유실하는 일이 없도록 해야 합니다. 메모리내 상태와 저장소를 구별하기 위해 “외부”라고 말하며, 저장소 자체가 복제될 수 있어도 그 저장소가 단일 엔티티로 표시되므로 “논리”라고 합니다. 게다가, 이 논리 데이터 저장소에는 서버 인스턴스가 저장소를 갱신할 때마다 갱신사항이 즉시 다른 인스턴스에 의해 보여지는 등록정보가 있습니다.

SunPlex 시스템은 클러스터 파일 시스템과 해당되는 글로벌 원시 파티션을 통해 그러한 외부 저장장치를 제공합니다. 예를 들면, 서비스가 새로운 데이터를 외부 로그 파일에 기록하거나 기존 데이터를 제 위치에서 수정한다고 가정합니다. 이 서비스의 여러 인스턴스가 실행될 경우, 각 인스턴스는 이 외부 로그에 대한 액세스를 가지므로 각각은 동시에 이 로그에 액세스할 수도 있습니다. 각 인스턴스는 해당되는 액세스를 이 로그에 동기화해야 합니다. 그렇지 않으면, 인스턴스는 서로 간섭합니다. 서비스는 `fcntl(2)` 및 `lockf(3C)`를 통해 일반 Solaris 파일 잠금을 사용하여 원하는 동기화를 수행할 수 있습니다.

그러한 저장소의 다른 예로는 고가용성 Oracle이나 Oracle Parallel Server와 같은 백엔드 데이터베이스를 들 수 있습니다. 그러한 백엔드 데이터베이스 서버는 데이터베이스 조회와 갱신 트랜잭션을 사용하여 내장된 동기화를 제공하므로, 여러 서버 인스턴스가 자체의 고유 동기화를 구현하지 않아도 됩니다.

현재 모습에서 확장 가능 서비스가 아닌 서비스의 예는 Sun의 IMAP 서버입니다. 서비스는 저장소를 갱신하지만, 그 저장소는 개인용이므로 여러 IMAP 인스턴스가 저장소에 기록할 경우, 그 인스턴스들은 갱신사항이 동기화되지 않으므로 서로 겹쳐씹니다. IMAP 서버는 동시 액세스를 동기화하기 위해 다시 작성해야 합니다.

마지막으로, 인스턴스에는 다른 인스턴스의 데이터에서 분리된 개인용 데이터가 있을 수 있다는 점에 유의하십시오. 그러한 경우, 데이터는 개인용이고, 해당되는 인스턴스만 이를 조작할 수 있으므로 서비스는 자체를 동시 액세스에 동기화하는데 관여하지

않아도 됩니다. 이 경우, 글로벌로 액세스할 수 있게 될 수도 있으므로 클러스터 파일 시스템 아래에 개인 데이터를 저장하지 않도록 주의해야 합니다.

Data Service API 및 Data Service Development Library API

SunPlex 시스템은 다음을 제공하여 응용프로그램을 고가용성으로 만듭니다

- SunPlex 시스템의 일부로 제공되는 데이터 서비스
- Data Service API
- Data Service Development Library API
- “일반” 데이터 서비스

*Sun Cluster 3.0 U1 Data Services Installation and Configuration Guide*에서는 SunPlex 시스템에 제공되는 데이터 서비스를 설치하고 구성하는 방법에 대해 설명합니다. *Sun Cluster 3.0 U1 Data Services Developer's Guide*에서는 Sun Cluster 프레임워크 하에서 고가용성이 될 다른 응용프로그램을 측정하는 방법에 대해 설명합니다.

응용프로그램 프로그래머가 Sun Cluster API를 사용하면 데이터 서비스 인스턴스를 시작하고 중지하는 스크립트와 결합 모니터를 개발할 수 있습니다. 이러한 도구를 사용하면, 응용프로그램에 페일오버 또는 확장 가능 데이터 서비스 중 어느 것을 제공할 것인지 측정할 수 있습니다. 또한 SunPlex 시스템은 응용프로그램의 필요한 시작 및 정지 메소드를 빨리 생성하여 고가용성 데이터 서비스로 실행되도록 하는데 사용될 수 있는 “일반” 데이터 서비스를 제공합니다

응용프로그램 트래픽을 위한 클러스터 상호 연결 사용법

클러스터에는 클러스터 상호 연결을 형성하는 노드간 여러 네트워크 연결이 있어야 합니다. 클러스터 소프트웨어는 고가용성 및 성능 향상 둘 모두를 위해 다중 상호 연결을 사용합니다. 내부 트래픽(예를 들어, 파일 시스템 데이터 또는 확장 가능 서비스 데이터)의 경우, 메시지는 사용 가능한 모든 상호 연결을 통해 라운드 로빈 방식으로 스트리핑됩니다.

클러스터 상호 연결은 노드사이의 고가용 통신을 위해 응용프로그램에도 사용 가능합니다. 예를 들어, 분산 응용프로그램에는 통신을 필요로 하는 다른 노드에서 실행하는 구성 요소가 있을 수 있습니다. 공용 상호 연결이 아닌 클러스터 상호 연결을 사용하여, 이 연결은 각 링크에 대한 실패로부터 안전합니다.

노드간 통신을 위해 클러스터 상호 연결을 사용하려면, 응용프로그램은 클러스터가 설치되었을 때 구성된 개인용 호스트 이름을 사용해야 합니다. 예를 들어, 노드 1의 개인용 호스트 이름이 `clusternode1-priv`인 경우, 해당 이름을 사용하여 클러스터 상호 연결을 통해 노드 1로 통신하십시오. 이 이름을 사용하여 열린 TCP 소켓은 클러스터 상호 연결을 통해 라우트되며 네트워크 실패의 경우 투명하게 다시 라우트될 수 있습니다.

개인용 호스트 이름이 설치 중에 구성될 수 있기 때문에, 클러스터 상호 연결은 해당 시간에 선택된 이름을 사용할 수 있다는 것에 유의하십시오. 실제 이름은

scha_privatelink_hostname_node 인수를 사용하여 scha_cluster_get(3HA)에서 얻을 수 있습니다.

클러스터 상호 연결에 대한 응용프로그램 레벨 사용의 경우, 단일 상호 연결은 각 노드 쌍 간에 사용됩니다. 예를 들어, 세 개의 노드에서 실행하고 클러스터 상호 연결을 통해 통신하는 응용프로그램을 고려하십시오. 노드 1과 2 사이의 통신이 인터페이스 hme0에서 일어날 수 있는 반면, 노드 1과 3 사이의 통신은 인터페이스 qfe1에서 일어날 수 있습니다. 즉, 두 노드간 응용프로그램 통신이 단일 상호 연결로 제한되는 반면, 내부 클러스터 통신은 모든 상호 연결을 통해 스트리밍됩니다.

응용프로그램이 내부 클러스터 트래픽과 상호 연결을 공유하므로, 응용프로그램에 사용 가능한 대역폭은 다른 클러스터 트래픽에 사용되는 대역폭에 따라 다릅니다. 실패할 경우, 내부 트래픽이 나머지 상호 연결을 통해 라운드 로빈될 수 있는 반면, 실패한 상호 연결의 응용프로그램 연결은 작업하는 상호 연결로 전환될 수 있습니다.

두 가지 유형의 주소가 클러스터 상호 연결을 지원하고, 개인용 호스트 이름의 gethostbyname(3N)은 정상적으로 두 개의 IP 주소를 반환합니다. 첫번째 주소는 논리 *pairwise* 주소라 하고, 두 번째 주소는 논리 *pernode* 주소라 합니다.

별도의 논리 *pairwise* 주소는 각 노드 쌍에 할당됩니다. 이 작은 논리 네트워크는 연결에 대한 페일오버를 지원합니다. 각 노드는 수정된 *pernode* 주소로도 할당됩니다. 즉, clusternode1-priv에 대한 논리 *pairwise* 주소가 각 노드마다 다른 반면, clusternode1-priv에 대한 논리 *pernode* 주소는 각 노드마다 동일합니다. 그러나 노드에는 노드 스스로에 대한 *pairwise* 주소가 없으므로, 노드 1에 있는 gethostbyname(clusternode1-priv)은 논리 *pernode* 주소만 반환합니다.

클러스터 상호 연결을 통해 연결을 승인하고 보안을 이유로 IP 주소를 확인하는 응용프로그램은 첫번째 IP 주소만이 아니라 gethostbyname에서 반환되는 모든 IP 주소에 대해 확인해야 합니다.

어느 점에 있어서도 응용프로그램에서 일관된 IP 주소를 필요로 하는 경우, 모든 연결이 *pernode* 주소에서 왕래하는 것으로 나타날 수 있도록 응용프로그램을 구성하여 클라이언트와 서버 측 모두에 대한 *pernode* 주소로 바인드하십시오.

자원, 자원 그룹 및 자원 유형

데이터 서비스는 여러 종류의 자원을 사용합니다. Apache Web Server 또는 iPlanet Web Server와 같은 응용프로그램은 응용프로그램에서 사용하는 네트워크 주소(논리적 호스트 이름 및 공유 주소)를 사용합니다. 응용프로그램과 네트워크 자원은 RGM에서 관리되는 기본 단위를 형성합니다.

데이터 서비스는 자원 유형입니다. 예를 들어, Sun Cluster HA for Oracle는 자원 유형 SUNW.oracle이고 Sun Cluster HA for Apache는 자원 유형 SUNW.apache입니다.

자원은 클러스터에 전반적으로 정의된 자원 유형의 인스턴스화입니다. 몇가지의 정의된 자원 유형이 있습니다.

네트워크 자원은 SUNW.LogiclaHostname 또는 SUNW.SharedAddress 자원 유형 중 하나입니다. 이들 두 자원 유형은 Sun Cluster 제품에 의해 사전에 등록됩니다.

SUNW.HAStorage 자원 유형은 자원에 따라 자원 시동 및 디스크 장치 그룹이 동기화되어 사용됩니다. 데이터 서비스를 시작하기 전에 클러스터 파일 시스템 마운트 포인트 경로, 글로벌 장치 및 장치 그룹 이름이 사용가능한지 확인하십시오.

RGM에서 관리되는 자원은 자원 그룹이라고 하는 그룹에 위치되므로 그 자원들은 하나의 단위로 관리될 수 있습니다. 자원 그룹은 페일오버나 스위치오버가 자원 그룹에서 초기화된 경우에 하나의 단위로 이주됩니다.

주: 응용프로그램 자원을 포함하는 자원 그룹을 온라인으로 가져오면, 응용프로그램이 시작됩니다. 데이터 서비스 시작 메소드는 성공적으로 종료되기 전에 응용프로그램이 가동되어 실행 중 상태가 될 때까지 기다립니다. 응용프로그램이 가동되어 실행되는 시기를 판별하는 것은 데이터 서비스가 클라이언트에 서비스를 제공하는 데이터 서비스 결합 모니터가 판별하는 것과 동일한 방법으로 수행됩니다. 이 프로세스에 대한 자세한 정보는 *Sun Cluster 3.0 U1 Data Services Installation and Configuration Guide*의 내용을 참조하십시오

자원 및 자원 그룹의 상태와 설정

관리자는 자원과 자원 그룹에 정적 설정을 적용합니다. 이 설정은 관리 작업을 통해서만 변경될 수 있습니다. RGM은 동적인 “상태” 사이에 자원 그룹을 이동합니다. 이 설정과 상태는 다음 목록에서 설명합니다.

- **Managed** 또는 **unmanaged** - 이것은 자원 그룹에만 적용되는 클러스터 범위의 설정입니다. 자원 그룹은 RGM에 의해 관리됩니다. `scrgadm(1M)` 명령을 사용하면 RGM이 자원 그룹을 관리하거나 관리하지 않도록 할 수 있습니다. 이 설정은 클러스터를 재구성해도 변경되지 않습니다.

자원 그룹을 처음 만들 때는 관리되지 않습니다. 그룹에 있는 자원이 활성화되기 전에 관리되도록 해야 합니다.

확장 가능한 웹 서버와 같은 일부 데이터 서비스에서는 네트워크 자원을 시작하기 전과 중지한 후에 작업을 해야 합니다. 이 작업은 시작(INIT) 및 종료(FINI) 데이터 서비스 메소드에 의해 수행됩니다. INIT 메소드는 자원이 있는 자원 그룹이 관리되는 상태인 경우에만 실행됩니다.

자원 그룹이 관리되지 않는 상태에서 관리되는 상태로 변경되면 그룹에 대하여 등록된 INIT 메소드가 그룹의 자원에 대하여 실행됩니다.

자원 그룹이 관리되는 상태에서 관리되지 않는 상태로 변경되면 등록된 FINI 메소드가 호출되어 삭제를 수행합니다.

INIT 및 FINI 메소드의 가장 일반적인 용도는 확장 가능한 서비스를 위해 네트워크 자원에 사용되지만 응용프로그램에 의해 수행되지 않는 초기화 또는 삭제 작업에도 사용할 수 있습니다.

- **Enabled** 또는 **disabled** - 이것은 자원에 적용되는 클러스터 범위의 설정입니다. `scrgadm(1M)` 명령을 사용하면 자원을 활성화하거나 비활성화할 수 있습니다. 이 설정은 클러스터를 재구성해도 변경되지 않습니다.

자원에 대한 정상 설정은 시스템에서 활성화되어 실행되는 것입니다.

모든 클러스터 노드에서 자원을 사용하지 못하도록 하려면 자원을 비활성화하면 됩니다. 비활성화된 자원은 일반적인 용도로 사용할 수 없습니다.

- **Online** 또는 **offline** - 이것은 자원 및 자원 그룹에 모두 적용되는 동적 상태입니다.

이 상태는 스위치오버 또는 페일오버 중에 클러스터 재구성 단계를 통해 클러스터 트랜지션으로 변경됩니다. 이 설정은 관리 작업을 통해서만 변경될 수 있습니다. `scswitch(1M)` 명령을 사용하면 자원이나 자원 그룹의 온라인 또는 오프라인 상태를 변경할 수 있습니다.

페일오버 자원이나 자원 그룹은 항상 한 노드에서는 온라인 상태가 될 수 있습니다. 확장 가능한 자원이나 자원 그룹은 일부 노드에서는 온라인 상태이고 다른 노드에서는 오프라인 상태일 수 있습니다. 스위치오버나 페일오버 중에 자원 그룹과 포함된 자원이 한 노드에서 오프라인 상태가 되고 다른 노드에서는 온라인 상태가 됩니다.

자원 그룹이 오프라인이면 모든 자원이 오프라인 상태가 됩니다. 자원 그룹이 온라인이면 모든 자원이 온라인 상태가 됩니다.

자원 그룹에는 여러 자원이 포함될 수 있고, 자원 사이에는 의존 관계가 있습니다. 이러한 의존성을 위해서는 자원이 특정 순서로 온라인 및 오프라인 상태가 되어야 합니다. 자원을 온라인 및 오프라인 상태로 변경하는 메소드의 실행 시간은 자원마다 다를 수 있습니다. 자원의 의존성과 시작 및 중지 시간의 차이 때문에 클러스터 재구성 중에 단일 자원 그룹 내에 있는 자원의 온라인 및 오프라인 상태가 서로 다를 수 있습니다.

자원 및 자원 그룹 등록 정보

SunPlex 데이터 서비스에 대해 자원 및 자원 그룹의 등록 정보 값을 구성할 수 있습니다. 표준 등록 정보는 모든 데이터 서비스에 공통입니다. **Extension** 등록 정보는 각 데이터 서비스에만 적용됩니다. 일부 표준 및 확장 등록 정보는 수정하지 않아도 되도록 기본 설정으로 구성됩니다. 다른 등록 정보들은 자원 작성 및 구성 프로세스의 일부로 설정해야 합니다. 각 데이터 서비스에 대한 문서에서는 설정할 수 있는 자원 등록 정보와 설정 방법을 지정합니다.

표준 등록 정보는 보통 특정 데이터 서비스와 독립적인 자원 및 자원 그룹 등록 정보를 구성하는데 사용됩니다. 일련의 표준 등록 정보들에 대해서는 *Sun Cluster 3.0 U1 Data Services Installation and Configuration Guide*의 부록에 설명되어 있습니다.

확장 등록 정보는 응용프로그램 바이너리 위치, 구성 파일 및 자원 종등록 정보와 같은 정보를 제공합니다. 사용하는 데이터 서비스를 구성하는 것처럼 확장 등록 정보를 수정할 수 있습니다. 일련의 확장 등록 정보들에 대해서는 *Sun Cluster 3.0 U1 Data Services Installation and Configuration Guide*에서 데이터 서비스에 대한 개별 장에 설명되어 있습니다.

PNM(Public Network Management) 및 NAFO(Network Adapter Failover)

클라이언트는 공용 네트워크 인터페이스를 통해 클러스터에 데이터 요청을 합니다. 각 클러스터 노드는 공용 네트워크 어댑터를 통해 최소한 하나의 공용 네트워크에 연결됩니다.

Sun Cluster PNM(Public Network Management) 소프트웨어는 공용 네트워크 어댑터를 모니터링하고 결함이 발견될 경우에 어댑터 사이에 IP 주소를 페일오버하기 위한 기본 메커니즘을 제공합니다. 각 클러스터 노드에는 다른 클러스터 노드에서는 다를 수도 있는 자체의 고유 구성이 있습니다.

공용 네트워크 어댑터는 네트워크 어댑터 페일오버 그룹(**NAFO 그룹**)으로 구성됩니다. 각 **NAFO** 그룹에는 하나 이상의 공용 네트워크 어댑터가 있습니다. 주어진 **NAFO** 그룹에 대해 언제든지 단 하나의 그룹만 활성화될 수 있는 반면, 활동 중인 어댑터에서 **PNM** 디먼에 의해 결함이 발견된 경우에 어댑터 페일오버 동안 사용되는 백업 어댑터로는 동일한 그룹의 여러 어댑터가 제공됩니다. 페일오버는 활동 중인 어댑터와 연관되는 IP 주소가 백업 어댑터로 이동되어, 노드에 대한 공용 네트워크 연결을 유지하도록 합니다. 페일오버는 어댑터 인터페이스 레벨에서 발생하므로 **TCP**와 같은 고급 연결은 페일오버 동안의 간단한 임시 지연을 제외하고는 영향을 받지 않습니다.

주: **TCP**의 경합 복구 특성으로, 일부 세그먼트가 페일오버 동안 유실되어 **TCP**에서 경합 제어 메커니즘을 활성화하므로 성공적인 페일오버 후에는 **TCP** 엔드포인트에서 추가 지연이 발생할 수도 있습니다.

NAFO 그룹은 논리 호스트 이름 및 공유 주소 자원에 대한 빌딩 블록을 제공합니다. `scrgadm(1M)` 명령은 필요하면 사용자를 위해 자동으로 **NAFO** 그룹을 작성합니다. 또한 사용자도 논리 호스트 이름과 공유 주소 자원의 **NAFO** 그룹을 독립적으로 만들어서 클러스터 노드에 대한 공용 네트워크 연결을 모니터링할 수 있습니다. 한 노드의 동일한 **NAFO** 그룹이 여러 논리 호스트 이름이나 공유 주소 자원을 호스팅할 수 있습니다. 논리 호스트 이름과 공유 주소 자원에 대한 자세한 정보는 *Sun Cluster 3.0 U1 Data Services Installation and Configuration Guide*의 내용을 참조하십시오.

주: **NAFO** 메커니즘의 설계는 어댑터 장애를 발견하고 마스킹하기 위한 것입니다. 이것은 관리자가 논리(또는 공유) IP 주소 중 하나를 제거하기 위해 `ifconfig(1M)`를 사용하여 복구하도록 설계된 것이 아닙니다. **Sun Cluster** 소프트웨어는 논리 및 공유 IP 주소를 **RGM**에 의해 관리되는 자원으로 처리합니다. 관리자가 정확하게 IP 주소를 추가하거나 제거하려면 `scrgadm(1M)` 명령을 사용하여 자원이 포함된 자원 그룹을 수정하면 됩니다.

PNM 결함 발견 및 페일오버 프로세스

PNM은 활동 중인 어댑터의 패킷 카운터를 정기적으로 검사하는데, 이 때 건강한 어댑터의 패킷 카운터가 어댑터를 통한 정상적인 네트워크 트래픽으로 변경할 것으로 가정합니다. 패킷 카운터가 잠깐 동안 변경되지 않으면 **PNM**은 ping 순서로 가서 활

동 중인 어댑터를 통해 트래픽을 강요합니다. PNM은 각 순서의 끝에서 패킷 카운터 변경을 검사하고, 핑 순서가 여러 번 반복된 후에도 패킷 카운터가 변경되지 않은 채로 있으면 어댑터 결함을 선언합니다. 이러한 이벤트가 발생하면 페일오버는 백업 어댑터(사용가능한 것이 있으면)로 트리거됩니다.

입력 및 출력 패킷 카운터는 PNM에 의해 모니터됩니다. 따라서 입출력 모두 또는 한 가지가 일정 시간 동안 변경되지 않으면 ping 시퀀스가 시작됩니다.

핑 순서는 ALL_ROUTER 멀티캐스트 주소(224.0.0.2), ALL_HOST 멀티캐스트 주소(224.0.0.1) 및 로컬 서브넷 브로드캐스트 주소의 핑으로 구성됩니다.

핑은 최소 비용 우선(least-costly-first) 방식으로 구조화되므로 비용이 덜 드는 핑이 성공하면 비용이 더 드는 핑은 실행되지 않습니다. 또한 핑은 어댑터에서 트래픽을 생성하기 위한 수단으로서만 사용됩니다. 해당되는 종료 상태는 어댑터가 기능하는지, 아니면 결함이 있는지 결정하는데 도움이 되지 않습니다.

이 알고리즘에는 네 가지의 조정가능한 매개변수가 있습니다. 이 매개변수는 inactive_time, ping_timeout, repeat_test, slow_network 등입니다. 이 매개변수는 결함 감지 속도와 정확도 사이에 조정가능한 타협을 제공합니다. 매개변수에 대한 세부사항과 변경 방법에 대해서는 Sun Cluster 3.0 U1 시스템 관리 안내서에서 공용 네트워크 매개변수를 변경하는 절차를 참조하십시오.

NAFO 그룹의 활동 중인 어댑터에서 결함이 감지된 후, 백업 어댑터를 사용할 수 없으면 그 그룹은 DOWN으로 선언되지만, 모든 백업 어댑터 테스트는 계속됩니다. 그렇지 않고, 백업 어댑터를 사용할 수 있으면 백업 어댑터에 대해 페일오버가 발생합니다. 논리 주소와 해당되는 연관 플래그는 결함이 있는 활동 어댑터가 다운되어 측정할 수 없을 때 백업 어댑터로 "전송"됩니다.

IP 주소의 페일오버가 성공적으로 완료되면, ARP 브로드캐스트가 전송됩니다. 따라서 원격 클라이언트 연결은 유지됩니다.

자주 문의되는 사항(FAQ)

이 장에는 SunPlex 시스템에 대하여 자주 문의되는 사항에 대한 응답이 포함되어 있습니다. 응답은 주제별로 조직되어 있습니다.

고가용성 FAQ

- 고가용성 시스템이란 정확히 무엇입니까?

SunPlex 시스템은 서버 시스템을 정상적으로 사용할 수 없게 만드는 실패가 발생한 경우에도 응용프로그램이 실행되도록 클러스터에서 유지하는 능력을 고가용성(HA)이라고 정의합니다.

- 클러스터가 고가용성을 제공하는 프로세스는 무엇입니까?

페일오버라고 하는 프로세스를 통해, 클러스터 프레임워크는 고가용성 환경을 제공합니다. 페일오버는 클러스터에서 실패한 노드로부터 또다른 작동 중인 노드로 데이터 서비스 자원을 이주하기 위해 클러스터에서 수행되는 일련의 단계입니다.

- HA와 확장 가능 서비스간의 차이점은 무엇입니까?

데이터 서비스에는 페일오버와 확장 가능 두 가지의 주요 기능이 있습니다.

HA 서비스는 응용프로그램이 클러스터에서 한번에 하나의 1차 노드에서만 실행됨을 의미합니다. 다른 노드는 다른 응용프로그램을 실행할 수도 있지만, 각 응용프로그램은 하나의 노드에서만 실행됩니다. 1차 노드가 실패할 경우, 실패한 노드에서 실행되는 응용프로그램은 다른 노드로 페일오버하여 실행을 계속합니다.

확장 가능 서비스는 하나의 응용프로그램을 여러 노드에 분산시켜서 하나의 논리 서비스를 작성합니다. 확장 가능 서비스는 실행되는 전체 클러스터에서 여러 노드와 프로세스를 조정합니다.

각 응용프로그램에 대하여 한 노드가 클러스터에 대한 물리적 인터페이스를 호스트합니다. 이 노드를 GIN(Global Interface Node)이라고 합니다. 클러스터에는 여러

개의 GIN이 있을 수 있습니다. 각 GIN은 확장 가능한 서비스에서 사용할 수 있는 하나 이상의 논리적 인터페이스를 호스트합니다. 이러한 논리적 인터페이스를 글로벌 인터페이스라고 합니다. 하나의 GIN의 특정 응용프로그램에 보내는 모든 요청에 대한 글로벌 인터페이스를 호스트하고 응용프로그램 서버가 실행되는 여러 노드에 이것을 디스패치합니다. GIN에 장애가 발생하면 글로벌 인터페이스가 생존 노드로 페일오버합니다.

실행 중인 응용프로그램에 있는 노드가 실패할 경우, 응용프로그램은 실패한 노드가 클러스터로 반환할 때까지 다른 노드에서 계속 실행됩니다. 이 때 약간의 성능 저하가 발생합니다.

파일 시스템 FAQ

- 클라이언트로서 다른 클러스터 노드를 가지고 있는 상태에서 하나 이상의 클러스터 노드를 고가용성 NFS 서버로 실행할 수 있습니까?

아닙니다. 루프백 마운트를 하지 않습니다.

- Resource Group Manager 제어 하에 없는 응용프로그램에 대해 클러스터 파일 시스템을 사용할 수 있습니까?

예. 그러나 RGM 제어가 없으면 응용프로그램은 실행되고 있는 노드가 실패할 경우 실행될 수 없습니다.

- 모든 클러스터 파일 시스템의 마운트 포인트가 /global 디렉토리에 있어야 합니까?

아니오. 그러나 클러스터 파일 시스템을 /global과 같은 동일 마운트 지점에 두면, 파일 시스템의 구성 및 관리가 쉬워집니다.

- 클러스터 파일 시스템을 사용하는 것과 NFS 파일 시스템을 내보내는 것 사이의 차이점은 무엇입니까?

몇 가지 차이점이 있습니다.

1. 클러스터 파일 시스템은 글로벌 장치를 지원합니다. NFS는 장치에 대한 원격 액세스를 지원하지 않습니다.
2. 클러스터 파일 시스템에는 글로벌 이름 공간이 있습니다. 하나의 마운트 명령만 필요합니다. NFS를 사용할 경우, 각 노드에서 파일 시스템을 마운트해야 합니다.
3. 클러스터 파일 시스템은 NFS를 수행하는 경우보다 많이 파일을 캐시합니다. 예를 들어, 읽기, 쓰기, 파일 잠금 및 비동기 I/O를 위해 여러 노드에서 파일에 액세스합니다.
4. 클러스터 파일 시스템은 하나의 서버가 실패할 경우, 끊어짐이 없는 페일오버를 지원합니다. NFS는 여러 서버를 지원하지만, 페일오버는 읽기 전용 파일 시스템에 대해서만 가능합니다.
5. 클러스터 파일 시스템은 원격 DMA 및 zero-copy 기능을 제공하는 미래의 빠른 클러스터 상호 연결을 개발하기 위해 만들어졌습니다.

6. 클러스터 파일 시스템에서 파일에 대한 등록 정보를 변경할 경우(예를 들어, `chmod(1M)`을 사용), 변경은 모든 노드에서 즉시 반영됩니다. 내보낸 NFS 파일 시스템에서는 이를 수행하는데 더 많은 시간이 소요될 수 있습니다.

- 클러스터 노드에 파일 시스템 `/global/devices/<node>@<node ID>`가 나타납니다. 이 파일 시스템을 사용하여 가용성이 높은 글로벌 데이터를 저장할 수 있습니까?

이 파일 시스템은 글로벌 장치 이름 공간을 저장합니다. 이것은 일반적인 용도에 사용하는 파일 시스템이 아닙니다. 데이터는 글로벌이지만 글로벌 방식으로 액세스할 수 없습니다. 각 노드는 자체 글로벌 장치 이름 공간만 액세스합니다. 노드가 중단되면 다른 노드가 중단된 노드에 대한 이름 공간에 액세스할 수 있습니다. 이 파일 시스템은 가용성이 높지 않습니다. 글로벌 방식으로 액세스하거나 가용성이 높아야 하는 데이터를 저장하는 데는 이 파일 시스템을 사용하면 안됩니다.

볼륨 관리 FAQ

- 모든 디스크 장치를 이중화해야 합니까?

고가용성으로 간주되는 디스크 장치의 경우에는 이중화해야 합니다. 그렇지 않으면, RAID-5 하드웨어를 사용하십시오. 모든 데이터 서비스는 고가용성 디스크 장치나 고가용성 디스크 장치에 마운트된 클러스터 파일 시스템을 사용해야 합니다. 그러한 구성에서는 단일 디스크 실패가 허용될 수 있습니다.

- 로컬 디스크(부트 디스크)에 하나의 볼륨 관리자를 사용하고 멀티호스트 디스크에 다른 볼륨 관리자를 사용할 수 있습니까?

로컬 디스크를 관리하는 Solstice DiskSuite 소프트웨어와 멀티호스트 디스크를 관리하는 VERITAS Volume Manager에서 이 구성이 지원됩니다. 다른 조합은 지원되지 않습니다.

데이터 서비스 FAQ

- 어떤 SunPlex 데이터 서비스를 사용할 수 있습니까?

지원되는 데이터 서비스 목록은 *Sun Cluster 3.0 U1* 릴리스 정보에 포함되어 있습니다.

- SunPlex 데이터 서비스에서는 어떤 응용프로그램 버전이 지원됩니까?

지원되는 응용프로그램 버전 목록은 *Sun Cluster 3.0 U1* 릴리스 정보에 포함되어 있습니다.

- 자체 데이터 서비스를 작성할 수 있습니까?

예. 자세한 정보는 Data Service Development Library API와 함께 제공되는 *Sun Cluster 3.0 U1 Data Services Developer's Guide* 및 *Data Service Enabling Technologies* 문서를 참조하십시오.

- 네트워크 자원을 작성할 때, 숫자 IP 주소 또는 호스트 이름을 지정해야 합니까?
네트워크 자원을 지정하는 데 선호하는 방법은 숫자 IP 주소 보다는 UNIX 호스트 이름을 사용하는 것입니다.
- 네트워크 자원을 작성할 때, 논리 호스트 이름(LogicalHostname 자원)이나 공유 주소(SharedAddress 자원)를 사용하는 것의 차이점은 무엇입니까?

Sun Cluster HA for NFS의 경우가 아니면 문서가 페일오버 모드 자원 그룹의 LogicalHostname 자원을 사용하기 위해 호출할 때마다 SharedAddress 자원 또는 LogicalHostname 자원 중에 한 가지를 사용할 수 있습니다. SharedAddress 자원을 사용하면 클러스터 네트워킹 소프트웨어가 SharedAddress에 대해 구성되며, LogicalHostname에 대해 구성되지 않으므로 일부 추가 오버헤드가 발생합니다.

SharedAddress를 사용하면 좋은 경우는 확장 가능 및 페일오버 데이터 서비스 둘 다에 대해 구성하여 클라이언트가 동일한 호스트 이름을 사용하여 두 서비스 모두에 액세스할 수 있도록 할 경우입니다. 이 경우, SharedAddress 자원은 페일오버 응용프로그램 자원과 함께 자원 그룹에 포함되지만, 확장 가능 서비스 자원은 별도의 자원 그룹에 포함되어 SharedAddress를 사용하도록 구성됩니다. 그러면 확장 가능 서비스와 페일오버 서비스가 모두 SharedAddress 자원에 구성된 동일한 호스트 이름/주소 세트를 사용할 수 있습니다.

공용 네트워크 FAQ

- 어떤 공용 네트워크 어댑터가 SunPlex 시스템을 지원합니까?
현재, SunPlex 시스템은 이더넷(10/100BASE-T 및 1000BASE-SX Gb) 공용 네트워크 어댑터를 지원합니다. 앞으로 새로운 인터페이스가 지원될 수 있으므로 Sun 판매 영업대표와 함께 대부분의 현재 정보에 대해 확인하도록 하십시오.
- 페일오버에서 MAC 주소의 역할은 무엇입니까?
페일오버가 발생할 경우, 새로운 ARP(Address Resolution Protocol) 패킷이 생성되어 전체에 브로드캐스팅됩니다. 이러한 ARP 패킷에는 새로운 MAC 주소(노드가 페일오버한 새로운 실제 어댑터와 이전 IP 주소가 있습니다. 네트워크의 다른 기체가 이 패킷 중 하나를 수신할 경우, 그 기체는 이전 MAC-IP 매핑을 ARP 캐쉬에서 지우고 새 것을 사용합니다.
- SunPlex 시스템이 호스트 어댑터에 대한 OpenBoot PROM에서 local-mac-address?=true 설정을 지원합니까?
아니오. 이 변수는 지원되지 않습니다.
- NAFO가 활성화 및 백업 어댑터 사이에 스위치오버를 수행할 때 얼마나 지연될 수 있습니까?

몇 분 동안 지연될 수 있습니다. 이것은 NAFO 스위치오버가 수행될 때 ARP를 외부로 전송하기 때문입니다. 그러나 클라이언트와 클러스터 사이의 라우터가 ARP를 반드시 사용하는 것은 아닙니다. 따라서 라우터에서 이 IP 주소에 대한 ARP 캐시 항목의 시간이 만료될 때까지 MAC 주소를 사용할 수 있습니다. 양쪽 NAFO 어댑터가 이더넷 스위치에 연결되어 지연이 발생할 수도 있습니다. NAFO 스위치오버가 수행되면 NAFO 어댑터 중 하나는 unplumb되고 두 번째 어댑터는 plumb됩니다. 그러면 이더넷 스위치가 포트를 비활성화하고 다른 포트를 활성화해야 합니다. 약간의 시간이 걸릴 수도 있습니다. 또한, 이더넷에서는 스위치와 새로 활성화된 어댑터 사이에 속도 협상이 진행되어 시간이 걸립니다. 최종적으로 스위치오버가 완료되면 NAFO가 새로 활성화된 어댑터에 대하여 몇 가지 sanity 검사를 수행하여 모든 것이 제대로 작동하는지 확인합니다.

클라이언트 시스템 FAQ

- 모든 클라이언트 시스템이 같은 루트 암호를 갖고 있어야 합니까?

각 클라이언트 시스템에서 동일한 루트 암호를 가질 필요는 없습니다. 그러나 모든 노드에서 같은 루트 암호를 사용하면 클러스터 관리가 간단해질 수 있습니다.

- 노드가 부트되는 순서가 중요합니까?

대부분의 경우, 그렇지 않습니다. 그러나 암네시아(“정족수 및 정족수 장치”(43페이지) 참조)를 막으려면 부트 순서가 중요합니다. 예를 들어, 노드 2가 정족수 장치를 소유하고 노드 1이 다운된 후, 사용자가 노드 2를 다운할 경우, 노드 1을 다시 가져오기 전에 노드 2를 가져와야 합니다. 그러면, 클러스터 구성 정보 날짜가 지난 노드를 가져오는 일이 없어집니다.

- 클러스터 노드에서 로컬 디스크를 이중화해야 합니까?

예. 이 이중화가 요구사항은 아니지만, 클러스터 노드 디스크를 이중화하면 노드를 다운하는 이중화하지 않은 디스크 실패를 예방할 수 있습니다. 클러스터 노드의 로컬 디스크를 이중화하면 시스템 관리에 오버헤드가 부가됩니다.

- 클라이언트 시스템 백업 문제점은 무엇입니까?

하나의 클러스터에 대해 몇 개의 백업 방법을 사용할 수 있습니다. 한 가지 방법은 테이프 드라이브/라이브러리가 첨부된 백업 노드로 하나의 노드를 사용하는 것입니다. 그리고 나서, 데이터를 백업하기 위해 클러스터 파일 시스템을 사용합니다. 이 노드를 공유 디스크에 연결하지 마십시오.

백업 및 복원 절차에 대해서는 *Sun Cluster 3.0 U1* 시스템 관리 안내서에서 추가 정보를 참조하십시오.

클러스터 기억 장치 FAQ

- 멀티호스트 기억 장치를 고가용성으로 만드는 것은 무엇입니까?

멀티호스트 기억 장치는 이중화(또는 하드웨어 기반 RAID-5 제어기)으로 인한 단일 디스크 유실을 살릴 수 있으므로 고가용성입니다. 멀티호스트 기억 장치는 여러 개의 연결을 갖고 있으므로, 연결된 단일 노드의 유실에 잘 견딜 수 있습니다.

클러스터 상호 연결 FAQ

- 어떤 클러스터 상호 연결이 SunPlex 시스템을 지원합니까?

현재, SunPlex 시스템은 이더넷(100BASE-T Fast Ethernet 및 1000BASE-SX Gb) 클러스터 상호 연결을 지원합니다.

- “케이블”과 전송 “경로” 사이에는 어떤 차이가 있습니까?

클러스터 전송 케이블은 전송 어댑터와 스위치를 사용하여 구성됩니다. 케이블은 컴포넌트끼리 연결하는 방식으로 어댑터와 스위치를 결합시킵니다. 클러스터 토폴로지 관리자는 사용 가능한 케이블을 사용하여 노드 사이에 종단 간 전송 경로를 만듭니다. 케이블은 직접 전송 경로에 매핑하지 않습니다.

케이블은 관리자에 의해 정적으로 “활성화”되고 “비활성화”됩니다. 케이블에는 활성화 또는 비활성화 “상태”가 있지만 항상 존재하는 “상태”는 아닙니다. 케이블이 비활성화되면 구성되지 않은 것과 동일합니다. 비활성화된 케이블은 전송 경로로 사용할 수 없습니다. 비활성화된 경로는 프로브되지 않기 때문에 상태를 확인할 수 없습니다. 케이블의 상태는 `scconf -p` 명령을 사용하여 볼 수 있습니다..

전송 경로는 클러스터 토폴로지 관리자에 의해 동적으로 구성됩니다. 전송 경로의 “상태”는 토폴로지 관리자에 의해 결정됩니다. 경로의 상태는 “온라인” 또는 “오프라인”이 될 수 있습니다. 전송 경로의 상태는 `scstat (1M)` 명령을 사용하여 볼 수 있습니다.

다음은 케이블이 네 개인 2-노드 클러스터의 예입니다.

```
node1:adapter0      to switch1, port0
node1:adapter1      to switch2, port0
node2:adapter0      to switch1, port1
node2:adapter1      to switch2, port1
```

네 개의 케이블로 두 개의 전송 경로를 만들 수 있습니다.

```
node1:adapter0    to node2:adapter0
node2:adapter1    to node2:adapter1
```

클라이언트 시스템 FAQ

- 클러스터에서 사용할 경우 특수 클라이언트 요구사항이나 제한사항을 고려해야 합니까?

클라이언트 시스템은 다른 서버에서처럼 클러스터에 연결합니다. 어떤 경우에는 데이터 서비스 응용프로그램에 따라, 클라이언트가 데이터 서비스 응용프로그램에 연결할 수 있도록 클라이언트측 소프트웨어를 설치하거나 다른 구성 변경사항을 수행해야 할 수도 있습니다. 클라이언트측 구성 요구사항에 대한 자세한 정보는 *Sun Cluster 3.0 U1 Data Services Installation and Configuration Guide*에서 해당되는 별도의 장을 참조하십시오.

관리 콘솔 FAQ

- SunPlex 시스템에 관리 콘솔이 필요합니까?

예.

- 관리 콘솔은 클러스터 전용이어야 합니까? 아니면, 다른 작업에도 사용할 수 있습니까?

SunPlex 시스템은 전용 관리 콘솔을 요구하지는 않지만 전용 관리 콘솔을 사용하면 다음과 같은 이점이 있습니다.

- 동일한 시스템에서 콘솔과 관리 도구를 그룹화하여 중앙집중화된 클러스터 관리가 가능하도록 합니다.
- 하드웨어 서비스 제공업체에 의해 더 빠르게 문제점을 분석할 수 있게 합니다.
- 관리 콘솔은 클러스터에 “가까이”(예를 들어, 같은 방)에 위치해야 합니까?
하드웨어 서비스 제공업체에 확인해 보십시오. 제공업체에서 클러스터 자체에 근접하게 콘솔이 위치되도록 요구할 수도 있습니다. 콘솔이 같은 방에 위치되어야 하는 기술적인 이유는 없습니다.
- 거리 요구사항이 먼저 만족될 경우, 관리 콘솔이 여러 개의 클러스터에 서비스를 제공할 수 있습니까?

예. 단일 관리 콘솔에서 여러 클러스터를 제어할 수 있습니다. 또한 클러스터 간의 하나의 터미널 콘센트레이터를 공유할 수도 있습니다.

터미널 콘센트레이터 및 시스템 서비스 프로세서 FAQ

■ SunPlex 시스템에 터미널 콘센트레이터가 필요합니까?

Sun Cluster 3.0으로 시작하는 모든 소프트웨어 버전은 터미널 콘센트레이터 없이 실행됩니다. 실패에 대비하기 위해 터미널 콘센트레이터를 요구했던 Sun Cluster 2.2 제품과는 달리, 이후 버전은 터미널 콘센트레이터의 영향을 받지 않습니다.

■ 대부분의 SunPlex 서버에서 터미널 집신 장치를 사용하는데 E10000는 이것을 사용하지 않습니다. 그 이유는 무엇입니까?

터미널 콘센트레이터는 대부분의 서비스를 위한 효율적인 이더넷 직렬 변환기입니다. 해당되는 콘솔 포트는 직렬 포트입니다. Sun Enterprise E10000 server에는 직렬 콘솔이 없습니다. System Service Processor(SSP)는 이더넷 또는 jtag 포트를 통한 콘솔입니다. Sun Enterprise E10000 server의 경우, 항상 콘솔용으로 SSP를 사용합니다.

■ 터미널 콘센트레이터를 사용하는 이점은 무엇입니까?

터미널 콘센트레이터를 사용하면 노드가 OBP(OpenBoot PROM)에 있는 경우처럼, 네트워크 임의의 곳에서 원격 워크스테이션에서 각 노드에 대한 콘솔 레벨 액세스를 제공합니다.

■ Sun에서 지원되지 않는 터미널 콘센트레이터를 사용하지 않을 경우, 사용하려는 것을 규정하기 위해 알아야 하는 것은 무엇입니까?

Sun과 다른 콘솔 장치에서 지원되는 터미널 콘센트레이터 사이의 주요 차이점은 Sun 터미널 콘센트레이터가 부트할 때 콘솔로 분기하는 것을 막는 특수 펌웨어를 가지고 있다는 것입니다. 분기 또는 콘솔로 분기되는 것으로 해석될 수 있는 신호를 보낼 수 있는 콘솔 장치를 가지고 있으면, 노드가 종료된다는 점에 유의하십시오.

■ 재부트하지 않고 Sun에서 지원되는 터미널 콘센트레이터에서 잠긴 포트를 해제할 수 있습니까?

예. 재설정이 필요한 포트 번호를 확인하여 다음 명령을 사용하십시오.

```
telnet tc
Enter Annex port name or number: cli
annex: su -
annex# admin
admin : reset port_number
admin : quit
annex# hangup
#
```

Sun에서 지원되는 터미널 집신기를 구성하고 관리하는 작업에 대해서는 Sun Cluster 3.0 U1 시스템 관리 안내서에서 자세한 내용을 참조하십시오.

- 터미널 콘센트레이터 자체가 실패할 경우에는 어떻습니까? 다른 터미널 콘센트레이터를 준비해야 합니까?

아니오. 터미널 콘센트레이터가 실패할 경우, 클러스터 가용성은 유실되지 않습니다. 콘센트레이터가 다시 서비스를 제공할 때까지 노드 콘솔에 연결할 수 없게 됩니다.

- 터미널 콘센트레이터를 사용할 경우, 보안은 어떻습니까?

일반적으로, 터미널 콘센트레이터는 시스템 관리자가 사용되는 소규모 네트워크에 접속되며, 다른 클라이언트 액세스에 사용되는 네트워크에는 접속되지 않습니다. 특수 네트워크에 대한 액세스를 제한하여 보안을 제어할 수 있습니다.

용어해설

이 용어집은 SunPlex 3.0 문서에서 사용됩니다.

A

관리 콘솔 앰네시아	클러스터 관리 소프트웨어를 실행하기 위해 사용되는 워크스테이션. 잘못된 CCR(Cluster Configuration Data)로 시스템 종료된 후에 클러스터가 재시작되는 상태. 예를 들어, 노드 1만 작동되는 2-노드 클러스터에서 클러스터 구성이 노드 1에서 변경되면 노드 2의 CCR은 잘못됩니다. 클러스터가 시스템 종료된 후 노드 2에서 재시작될 경우, 노드 2의 잘못된 CCR로 인해 앰네시아 상태가 발생합니다.
자동 페일백	1차 노드가 실패한 후 나중에 클라이언트 시스템으로 재시작될 때 자원 그룹이나 장치 그룹을 해당되는 1차 노드로 반환시키는 프로세스.

B

백업 그룹	“네트워크 어댑터 페일오버 그룹”을 참조하십시오.
-------	-----------------------------

C

체크포인트	1차 노드와 2차 노드 사이에 동기화된 소프트웨어 상태를 보존하기 위해 1차 노드에서 2차 노드로 보내지는 통지. 또한 “1차” 및 “2차”를 참조하십시오.
클러스터	클러스터 파일 시스템을 공유하고 페일오버, 병렬 또는 확장 가능 자원을 실행하도록 함께 구성된 상호 연결된 둘 이상의 노드 또는 도메인.
CCR(Cluster Configuration Repository)	Sun Cluster 소프트웨어에서 클러스터 구성 정보를 지속적으로 저장하기 위해 사용되는 고가용성의 복제된 데이터 저장소.

클러스터 파일 시스템	클러스터 범위에서 기존의 로컬 파일 시스템에 대하여 가용성이 높은 액세스를 제공하는 클러스터 서비스.
클러스터 상호 연결	케이블, 클러스터 전송 연결장치 및 클러스터 전송 어댑터를 포함하는 하드웨어 네트워킹 기반 구조. Sun Cluster 및 데이터 서비스 소프트웨어는 클러스터간의 통신에 이 기반 구조를 사용합니다.
클라이언트 시스템	현재 클러스터 실현에서 활동 중인 구성원. 이 구성원은 다른 클라이언트 시스템과 자원을 공유하고 다른 클라이언트 시스템과 클러스터의 클라이언트 둘 다에 대해 서비스를 제공할 수 있습니다. “클러스터 노드”를 참조하십시오.
CMM(Cluster Membership Monitor)	일관성 있는 클러스터 멤버쉽 등록부를 유지보수하는 소프트웨어. 이 멤버쉽 정보는 클러스터 소프트웨어의 나머지 부분에서 고가용성 서비스를 찾을 곳을 결정하는 데 사용됩니다. CCM 은 클러스터 이외 구성원이 데이터를 손상시킬 수 없으며 일관성이 없거나 손상된 데이터를 클라이언트에 전송하지 못하도록 합니다.
클러스터 노드	클라이언트 시스템이 되도록 구성된 노드. 클러스터 노드는 현재 구성원이거나 그렇지 않을 수도 있습니다. “클러스터 구성원”을 참조하십시오.
클러스터 전송 어댑터	노드에 상주하며 노드를 클러스터 상호 연결에 연결하는 네트워크 어댑터. “클러스터 상호 연결”을 참조하십시오.
클러스터 전송 케이블	엔드포인트에 연결하는 네트워크 연결. 클러스터 전송 어댑터와 클러스터 전송 연결장치 사이 또는 두 개의 클러스터 전송 어댑터 사이의 연결. “클러스터 상호 연결”을 참조하십시오.
클러스터 전송 연결장치	클러스터 상호 연결의 일부로 사용되는 하드웨어 스위치. “클러스터 상호 연결”을 참조하십시오.
배열	동일한 노드에 있는 등록 정보. 이 개념은 성능 개선을 위해 클러스터를 구성하는 동안 사용됩니다.

D

데이터 서비스	RGM(Resource Group Manager) 제어 하에 고가용성 자원으로 실행 되도록 설계된 응용프로그램.
기본 마스터	페일오버 자원 유형을 온라인 상태로 가져오는 기본 클라이언트 시스템.

장치 그룹	디스크와 같이 클러스터 HA 구성에 있는 다른 노드에서 마스터될 수 있는 사용자 정의된 장치 리소스 그룹. 이 그룹에는 디스크의 장치 자원, Solstice DiskSuite 디스켓 및 VERITAS Volume Manager 디스크 그룹의 장치 자원이 포함될 수 있습니다.
장치 ID	<p>Solaris를 통해 사용가능한 장치를 식별하는 메커니즘. 장치 ID는 <code>devid_get(3DEVID)</code> man 페이지에 설명되어 있습니다.</p> <p>Sun Cluster DID 드라이버는 장치 ID를 사용하여 서로 다른 클러스터 노드에 있는 Solaris 논리적 이름 사이의 관계를 결정합니다. DID 드라이버는 해당되는 장치 ID에 대해 각 장치를 규명합니다. 장치 ID가 클러스터의 다른 곳에서 다른 장치와 일치하면, 두 장치 모두에 동일한 DID 이름이 부여됩니다. 전에 클러스터에 표시되었던 장치가 아니면 새로운 DID 이름이 할당됩니다. “Solaris 논리 이름”과 “DID 드라이버”를 참조하십시오.</p>
DID 드라이버	클러스터에서 일관성 있는 장치 이름 공간을 제공하기 위해 사용되는 Sun Cluster 에서 구현된 드라이버. “ DID 이름”을 참조하십시오.
DID 이름	SunPlex 시스템에서 글로벌 장치를 식별하기 위해 사용됩니다. 이것은 Solaris 논리 이름과의 일대일 또는 일대다 관계를 가지고 있는 클러스터 식별자입니다. 양식은 d XsY 로, 여기서 X 는 정수이고 Y 는 슬라이스 이름입니다. “ Solaris 논리 이름”을 참조하십시오.
디스크 장치 그룹	“장치 그룹”을 참조하십시오.
DLM(Distributed Lock Manager)	공유 디스크 Oracle Parallel Server(OPS) 환경에서 사용되는 잠금 소프트웨어. DLM 은 다른 노드에서 실행되는 Oracle 프로세스별 데이터베이스 액세스를 동기화하도록 합니다. DLM 은 고가용성을 위해 설계되었습니다. 프로세스나 노드가 손상될 경우, 나머지 노드를 종료하고 재시작하지 않아도 됩니다. 그러한 실패에서 복구하기 위해 DLM 의 빠른 재구성이 수행됩니다.
디스크 세트	“장치 그룹”을 참조하십시오.
디스크 그룹	“장치 그룹”을 참조하십시오.

E

엔드포인트 이벤트	클러스터 전송 어댑터나 클러스터 전송 연결장치의 실제 포트. 관리 오브젝트의 상태, 소유권, 심각도 또는 설명에서의 변경.
-----------	--

F

페일백 페일페스트	“자동 페일백”을 참조하십시오. 잘못된 작동으로 손상이 야기되기 전에 결함 노드의 클러스터에서 순차적으로 종료되고 제거되는 것.
페일오버	실패 발생 후 현재 1차 노드에서 새로운 1차 노드로 자원 그룹이나 장치 그룹을 자동으로 다시 위치시키는 것.
페일오버 자원	한번에 한 노드에서만 올바르게 마스터될 수 있는 각각의 자원. “단일 인스턴스 자원”과 “확장 가능 자원”도 참조하십시오.
결함 모니터	데이터 서비스의 다양한 부분과 취한 조치를 규명하기 위해 사용되는 프로그램과 결함 디먼. “자원 모니터”를 참조하십시오.

G

일반 자원 유형	데이터 서비스에 대한 서식 파일. 일반 자원 유형을 사용하여 페일오버 데이터 서비스에 간단한 응용프로그램을 만들 수 있습니다(한 노드에서 정지하고, 다른 노드에서 시작). 이러한 유형에서는 SunPlex API 에 의한 프로그래밍이 필요하지 않습니다.
일반 자원	일반 자원 유형의 일부로 자원 그룹 관리자의 제어 하에 있는 응용프로그램 디먼과 해당되는 하위 프로세스.
글로벌 장치	디스크, CD-ROM 및 테이프와 같은 모든 클라이언트 시스템에서 액세스할 수 있는 장치.
글로벌 장치 이름 공간	글로벌 장치에 대한 전체 클러스터의 논리적 이름을 포함하는 이름 공간. Solaris 환경에서 로컬 장치는 /dev/dsk, /dev/rdisk 및 /dev/rmt 디렉토리에 정의됩니다. 글로벌 장치 이름 공간은 /dev/global/dsk, /dev/global/rdisk 및 /dev/global/rmt 디렉토리에 글로벌 장치를 정의합니다.
글로벌 인터페이스	실제로 공유된 주소를 호스트하는 글로벌 네트워크 인터페이스. “공유 주소”를 참조하십시오.
글로벌 인터페이스 노드	글로벌 인터페이스를 호스트하는 노드.
글로벌 자원	Sun Cluster 소프트웨어의 커널 레벨에서 제공되는고가용성 자원. 글로벌 자원으로는 디스크(HA 장치 그룹), 클러스터 파일 시스템 및 글로벌 네트워킹이 있습니다.

H

HA 데이터 서비스 하트비트(heartbeat)

“데이터 서비스”를 참조하십시오.
모든 사용가능한 클러스터 상호 연결 전송 경로를 거쳐 보내지는 정기적인 메시지. 지정된 간격 및 재시도 횟수 후에 하트비트가 없으면 전송 통신의 내부 페일오버가 다른 경로로 트리거될 수 있습니다. 클라이언트 시스템에 대한 모든 경로가 실패하면 CMM이 클러스터 정족수를 재평가합니다.

I

인스턴스

“자원 호출”을 참조하십시오.

L

로드 밸런싱

확장 가능한 서비스에만 적용됩니다. 클라이언트 요청에 시간제 방식으로 서비스가 제공되도록 클러스터의 노드에서 응용프로그램 로드를 분배하는 프로세스. 자세한 내용은 (51페이지)“확장 가능 데이터 서비스”을 참조하십시오.

로드 밸런싱 정책

확장 가능한 서비스에만 적용됩니다. 응용프로그램 요청 로드가 노드들 사이에 분배되는 선호하는 방식. 자세한 내용은 (51페이지)“확장 가능 데이터 서비스”을 참조하십시오.

로컬 디스크

주어진 클러스터 노드에 실제로 속해 있는 디스크.

논리 호스트

응용프로그램, 디스크 세트 또는 응용프로그램 데이터가 상주하는 디스크 그룹과 클러스터에 액세스하기 위해 사용되는 네트워크 주소를 포함하는 Sun Cluster 2.0(최소) 개념. 이제는 SunPlex 시스템에서 이 개념을 사용하지 않습니다. SunPlex 시스템에서 이 개념을 구현하는 방법은 “디스크 장치 그룹”(38페이지) 및 “자원, 자원 그룹 및 자원 유형”(58페이지)를 참조하십시오.

논리 호스트 이름 자원

네트워크 주소를 나타내는 논리 호스트 이름들의 컬렉션을 포함하는 자원. 논리 호스트 이름 자원은 한번에 하나의 노드에서만 마스터될 수 있습니다. “논리 호스트”를 참조하십시오.

논리 네트워크 인터페이스

인터넷 구조에서, 호스트에는 하나 이상의 IP 주소가 있을 수 있습니다. Sun Cluster 소프트웨어는 추가 논리 네트워크 인터페이스를 구성하여 여러 논리 네트워크 인터페이스와 단일 실제 네트워크 인터페이스 사이의 매핑을 설정합니다. 각 논리 네트워크 인터페이스에는 단일 IP 주소가 있습니다. 이러한 매핑으로 하나의 실제 네트워크 인터페이스가 여러 IP 주소에 응답할 수 있습니다. 또한 추가 하드웨어

어 인터페이스 없이 테이크오버 또는 스위치오버 이벤트에서 클라이언트 시스템 사이에 IP 주소가 이동될 수 있게 합니다.

M

마스터 메타 장치 상태 데이터베이스 복제본(복제본)	“1차”를 참조하십시오. 모든 메타 장치와 오류 조건의 상태와 구성을 기록하는, 디스크에 저장된 데이터베이스. 이 정보는 Solstice DiskSuite 디스크 세트의 올바른 작동에 중요하므로 복제됩니다.
멀티홈 호스트	여러 공용 네트워크에 있는 호스트.
멀티호스트 디스크	실제로 여러 노드에 연결되어 있는 디스크.

N

NAFO(Network Adapter Failover) 그룹 네트워크 주소 자원	어댑터 장애 이벤트에서 서로 백업하도록 구성된, 동일 노드와 동일 서브네트에 있는 하나 이상의 네트워크 어댑터 세트. “네트워크 자원”을 참조하십시오.
네트워크 자원	하나 이상의 논리 호스트 이름이나 공유 주소를 포함하는 자원. “논리 호스트 이름 자원”과 “공유 주소 자원”을 참조하십시오.
노드	SunPlex 시스템의 일부일 수 있는 실제 시스템 또는 도메인(Sun Enterprise E10000 server 에서). “호스트”라고도 합니다.
비클러스터 모드	클라이언트 시스템을 부팅할 때 -x 부트 옵션을 사용한 결과 상태. 이 상태에서 노드는 더이상 클러스터 구성원이 아니지만 여전히 클러스터 노드입니다. “클라이언트 시스템”과 “클러스터 노드”를 참조하십시오.

P

병렬 자원 유형	여러 노드(두 개 이상)에서 동시에 마스터될 수 있도록 클러스터 환경에서 실행되도록 설계된, 병렬 데이터베이스와 같은 자원 유형.
병렬 서비스 인스턴스	개별 노드에서 실행되는 병렬 자원 유형의 인스턴스.
가상 마스터	“가상 1차”를 참조하십시오.
가상 1차	1차 노드가 실패할 경우, 페일오버 자원 유형을 마스터할 수 있는 클라이언트 시스템. “기본 마스터”를 참조하십시오.

1차	자원 그룹이나 장치 그룹이 현재 온라인 상태인 노드. 즉, 1차 노드는 자원과 연관되는 서비스를 구현하거나 호스팅하는 노드입니다. “2차”를 참조하십시오.
1차 호스트 이름	1차 공용 네트워크의 노드 이름. 이 이름은 항상 /etc/nodename에 지정된 노드 이름입니다. “2차 호스트 이름”을 참조하십시오.
개인 호스트 이름	클러스터 상호 연결을 거쳐 노드와 통신하기 위해 사용되는 호스트 이름 별명.
PNM(Public Network Management)	단일 네트워크 어댑터나 케이블 장애로 노드 가용성이 없어지는 것을 피하기 위해 결합 모니터링 및 페일오버를 사용하는 소프트웨어. PNM 페일오버는 네트워크 어댑터 페일오버 그룹이라고 하는 네트워크 어댑터 세트를 사용하여 클러스터 노드와 공용 네트워크 사이의 중복 연결을 제공합니다. 결합 모니터링 및 페일오버 기능은 함께 작동되어 자원 가용성을 보장합니다. “네트워크 어댑터 페일오버 그룹”을 참조하십시오.

Q

정족수 장치	클러스터 실행을 위한 정족수를 설정하기 위해 사용되는 투표에 기여하는 두 개 이상의 노드에서 공유하고 있는 디스크. 클러스터는 투표의 정족수가 사용가능할 경우에만 작동될 수 있습니다. 정족수 장치는 클러스터가 별도의 노드 세트로 분할되어 새 클러스터를 구성하는 노드 세트를 설정할 때 사용됩니다.
--------	---

R

자원	자원 유형의 인스턴스. 동일한 유형의 많은 자원이 존재할 수 있습니다. 각 자원은 기초가 되는 응용프로그램의 여러 인스턴스가 클러스터에서 실행될 수 있도록 고유한 이름과 등록 정보 값 세트를 가지고 있습니다.
자원 그룹	하나의 단위로서 RGM에 의해 관리되는 자원들의 컬렉션. RGM에 의해 관리될 각 자원은 자원 그룹으로 구성되어야 합니다. 일반적으로, 관련되어 서로 영향을 받는 자원들이 그룹화됩니다.
RGM(Resource Group Manager)	선택된 클러스터 노드에서 클러스터 자원을 자동으로 시작하고 정지하여 그러한 자원이 고가용성 및 확장성을 갖도록 하는데 사용되는 소프트웨어 기능. RGM은 하드웨어 또는 소프트웨어 장애 또는 재부트 이벤트에서 사전에 구성된 정책에 따라 작동합니다.
자원 그룹 상태	주어진 노드에서의 자원 그룹 상태.

자원 호출 노드에서 실행되는 자원 유형의 인스턴스. 노드에서 시작된 자원을 나타내는 추상적 개념.

RMAPI(Resource Management API)

클러스터 환경에서 응용프로그램이 고가용성을 갖도록 만드는 SunPlex 시스템 내의 응용프로그램 프로그래밍 인터페이스.

자원 모니터 올바르게 실행되는지의 여부와 수행 방법을 판별하기 위해 자원에 대해 정기적 결함 프로브를 실행하는 자원 유형 구현의 선택적 부품.

자원 상태(state) 주어진 노드에서의 Resource Group Manager 자원 상태.

자원 상태(status) 결함 모니터에서 보고된 자원 상태.

자원 유형 데이터 서비스, LogicalHostname 또는 SharedAddress 클러스터 오브젝트에 부여된 고유한 이름. 데이터 서비스 자원 유형은 페일오버 유형이거나 확장 가능한 유형이 될 수 있습니다. “데이터 서비스,” “페일오버 자원” 및 “확장 가능 자원”을 참조하십시오.

자원 유형 등록정보 주어진 유형의 자원을 설명하고 관리하기 위해 사용되는, 해당 자원 유형의 일부로 RGM에 의해 저장되는 키-값 쌍.

S

SCI(Scalable Coherent Interface)

클러스터 상호 연결로 사용되는 고속 상호 연결 하드웨어.

확장 가능한 자원 단일 서비스의 형태를 서비스의 원격 클라이언트에 제공하기 위해 클러스터 상호 연결을 사용하는, 여러 노드에서 실행되는(각 노드에 하나의 인스턴스) 자원.

확장 가능한 서비스 여러 노드에서 동시에 실행되는, 구현된 데이터 서비스.

2차 **1차** 노드가 실패할 때 디스크 장치 그룹 및 자원 그룹을 마스터할 수 있는 클라이언트 시스템. “**1차**”를 참조하십시오.

2차 호스트 이름 **2차** 공용 네트워크에서 노드에 액세스하기 위해 사용되는 이름. “**1차 호스트 이름**”을 참조하십시오.

공유 주소 자원 서비스가 해당 노드에서 확장 가능하도록 만들기 위해 클러스터 내의 노드에서 실행되는 확장 가능한 모든 서비스에 의해 바인드될 수 있는 네트워크 주소. 하나의 클러스터에 여러 공유 주소가 있을 수 있으며, 하나의 서비스가 여러 공유 주소에 바인드될 수 있습니다.

단일 인스턴스 자원 클러스터에서 최대 하나의 자원이 활동 중일 수 있는 자원.

Solaris 논리 이름	Solaris 장치를 관리하기 위해 일반적으로 사용되는 이름. 디스크의 경우 일반적으로 <code>/dev/rdisk/c0t2d0s2</code> 와 같이 표시됩니다. Solaris 논리 장치 이름마다 기초가 되는 Solaris 실제 장치 이름이 있습니다. “ DID 이름”과 “ Solaris 실제 이름”을 참조하십시오.
Solaris 실제 이름	Solaris 에서 해당되는 장치 드라이버에 의해 장치에 부여되는 이름. 이는 Solaris 시스템에서 <code>/devices</code> 트리 아래에 경로로 표시됩니다. 예를 들어, 일반적인 SCSI 디스크에는 다음과 같은 Solaris 실제 이름이 있습니다. <code>/devices/sbus@1f,0/SUNW,fas@e,8800000/sd@6,0:c,raw</code> “ Solaris 논리 이름”을 참조하십시오.
Solstice DiskSuite	SunPlex 에서 사용되는 볼륨 관리자. “볼륨 관리자”를 참조하십시오.
브레인 분할	하나의 클러스터가 여러 파티션으로 나뉘지는 상태로, 각 파티션은 다른 파티션들의 존재를 모르는 상태에서 형성됩니다.
스위치백	“페일백”을 참조하십시오.
스위치오버	자원 그룹이나 장치 그룹을 클러스터의 소유자(노드)에서 다른 소유자(또는 자원 그룹이 여러 1차 노드에 대해 구성된 경우에는 여러 소유자)에게 순차적으로 전송하는 것. 스위치오버는 관리자가 <code>sctswitch(1M)</code> 명령을 사용하여 초기화됩니다.
SSP(System Service Processor)	Enterprise 10000 구성에서 클라이언트 시스템과 통신하기 위해 특별히 사용되는, 클러스터에 대해 외부에 있는 장치.

T

테이크오버 터미널 콘센트레이터	“페일오버”를 참조하십시오. Enterprise 10000 이외 구성에서 클라이언트 시스템과 통신하기 위해 특별히 사용되는, 클러스터에 대해 외부에 있는 장치.
---------------------	---

V

VERITAS Volume Manager 볼륨 관리자	SunPlex 시스템에서 사용되는 볼륨 관리자. “볼륨 관리자”를 참조하십시오. 디스크 스트라이핑, 연결, 이중화, 메타 장치 또는 볼륨의 동적 증가를 통해 데이터 신뢰성을 제공하는 소프트웨어 제품.
---	---