



(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2010년04월12일
(11) 등록번호 10-0952553
(24) 등록일자 2010년04월05일

(51) Int. Cl.
G06F 11/16 (2006.01) G06F 15/16 (2006.01)
G06F 11/07 (2006.01)
(21) 출원번호 10-2008-0050330
(22) 출원일자 2008년05월29일
심사청구일자 2008년05월29일
(65) 공개번호 10-2008-0106065
(43) 공개일자 2008년12월04일
(30) 우선권주장
JP-P-2007-00143633 2007년05월30일 일본(JP)
(56) 선행기술조사문헌
JP2006048676 A
JP2006127398 A

(73) 특허권자
가부시키가이샤 히타치세이사쿠쇼
일본국 도쿄도 치요다쿠 마루노우치 1초메 6반 6고
(72) 발명자
हतत यूकूरु
일본 도쿄도 지요다쿠 마루노우찌 1쵸메 6반 1고
가부시키가이샤히타치세이사쿠쇼 지적재산권본부 내
우에노 히토시
일본 도쿄도 지요다쿠 마루노우찌 1쵸메 6반 1고
가부시키가이샤히타치세이사쿠쇼 지적재산권본부 내
(74) 대리인
박충범, 이중희, 장수길

전체 청구항 수 : 총 14 항

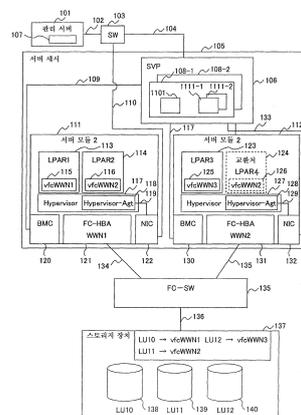
심사관 : 이경홍

(54) 가상 계산기 시스템 및 그 제어 방법

(57) 요약

SAN 환경 하의 임의의 물리 계산기 상의 LPAR에 장애가 발생한 경우, 다른 물리 계산기에 교대처의 LPAR을 설정하여, LPAR 단위의 교대를 가능하게 하고, RAID 장치측에서의 시큐리티 기능의 설정 변경을 불필요로 한다. SAN 환경 하의 물리 계산기 상에 생성된 LPAR에 장애가 발생하였을 때에, 관리 서버의 제어 하, 장애가 생긴 LPAR의 고유 ID(WWN)를 포함하는 구성 정보를 판독하고, 다른 물리 계산기 상에 교대용의 LPAR을 생성하고, 판독한 LPAR의 구성 정보를 교대처 LPAR에 설정함으로써, 장애 시의 LPAR 계승을 가능하게 한다.

대표도 - 도1



특허청구의 범위

청구항 1

제1 및 제2 물리 계산기를 포함하는 복수의 물리 계산기와, 상기 복수의 물리 계산기에 네트워크를 통하여 접속되는, 해당 물리 계산기 및 해당 논리 구간을 관리하는 관리 장치를 포함하고, 각 물리 계산기에 논리 구획을 생성하여 OS를 동작시킬 수 있는 가상 계산기 시스템으로서,

상기 제1 물리 계산기는,

상기 제1 물리 계산기 또는 거기에 형성된 제1 논리 구획에 장애가 발생한 것을 검출하는 장애 검출 수단과, 상기 제1 물리 계산기의 하드웨어 구성 정보 및 상기 제1 논리 구획에 할당된 고유의 구성 정보를 관리하는 제1 관리 수단을 갖고,

상기 관리 장치는,

상기 장애 검출 수단으로부터의 장애 발생의 보고를 받고, 상기 제1 관리 수단으로부터 상기 하드웨어 구성 정보 및 상기 고유의 구성 정보를 수신하는 수단과, 교대처의 제2 물리 계산기를 결정하고, 상기 제2 물리 계산기에 상기 하드웨어 구성 정보 및 상기 고유의 구성 정보를 송신하는 수단을 갖고,

상기 제2 물리 계산기는,

상기 관리 장치로부터 송신된 상기 하드웨어 구성 정보 및 상기 고유의 구성 정보를 수신하는 수단과, 상기 하드웨어 구성 정보 및 상기 고유의 구성 정보에 기초하여, 상기 제2 물리 계산기 상에 제2 논리 구획을 생성하는 것이 가능한지를 판정하는 수단과, 상기 판정 수단에 의해 상기 제2 논리 구획의 생성이 가능하다고 판정된 경우, 상기 고유의 구성 정보에 기초하여 제2 논리 구획을 생성하는 수단을 갖는 것을 특징으로 하는 가상 계산기 시스템.

청구항 2

제1항에 있어서,

상기 제1 물리 계산기는, 장애를 검출하였을 때, 상기 제1 논리 구획의 동작을 정지시키고, 상기 동작의 정지를 상기 관리 장치에 보고하는 수단을 갖고,

상기 제2 물리 계산기는, 상기 제2 논리 구획의 생성이 완료되면, 상기 완료의 보고를 상기 관리 장치에 보내는 수단을 갖고,

상기 관리 장치는, 상기 동작의 정지를 수취하였을 때, 상기 제1 논리 구획의 정지 상태를 표시하는 수단과, 상기 완료 보고를 수취하였을 때, 상기 제2 물리 계산기에 상기 제2 논리 구획을 기동시키는 커맨드를 송출하는 수단을 갖고,

상기 제2 물리 계산기는, 상기 기동 커맨드를 수신하였을 때, 상기 제2 논리 구획의 기동을 행하는 것을 특징으로 하는 가상 계산기 시스템.

청구항 3

제1항에 있어서,

상기 판정 수단은, 상기 제1 물리 계산기의 실효 CPU 성능과, 상기 제2 물리 계산기의 실효 CPU 성능을 계산하는 것을 특징으로 하는 가상 계산기 시스템.

청구항 4

제1항에 있어서,

상기 판정 수단은, 상기 제2 물리 계산기의 실효 CPU 성능이, 상기 제1 물리 계산기의 실효 CPU 성능 이상인 것을 조사하여, 상기 제2 논리 구획의 생성이 가능한지를 판정하는 것을 특징으로 하는 가상 계산기 시스템.

청구항 5

제1항에 있어서,

상기 판정 수단은, 상기 제2 물리 계산기의 메모리의 용량이 충분히 비어 있는지를 조사하여, 상기 제2 논리 구획의 생성이 가능한지를 판정하는 것을 특징으로 하는 가상 계산기 시스템.

청구항 6

제1항에 있어서,

상기 판정 수단은, 상기 제1 물리 계산기의 상기 제1 논리 구획이 사용하고 있었던 NIC 또는 HBA의 수와 동일한 수 이상의 NIC 또는 HBA가, 상기 제2 물리 계산기에서 미사용인지를 조사하여, 상기 제2 논리 구획의 생성이 가능한지를 판정하는 것을 특징으로 하는 가상 계산기 시스템.

청구항 7

제1항에 있어서,

상기 제1 및 제2 물리 계산기가 갖는 각 수단은, 제1 및 제2 논리 구획을 관리하는 하이퍼바이저 내에 포함되는 것을 특징으로 하는 가상 계산기 시스템.

청구항 8

제1 및 제2 물리 계산기를 포함하는 복수의 물리 계산기와, 상기 복수의 물리 계산기에 네트워크를 통하여 접속되는, 해당 물리 계산기 및 해당 논리 구간을 관리하는 관리 장치를 포함하고, 각 물리 계산기에 논리 구획을 생성하여 OS를 동작시킬 수 있는 가상 계산기 시스템에서의 논리 구획의 이행 제어 방법으로서,

상기 제1 물리 계산기에서, 상기 제1 물리 계산기 또는 거기에 형성된 제1 논리 구획에 장애가 발생한 것을 검출하는 장애 검출 스텝과, 상기 제1 물리 계산기의 하드웨어 구성 정보 및 상기 제1 논리 구획에 할당된 고유의 구성 정보를 관리하는 스텝과,

상기 관리 장치에서, 상기 장애 검출 수단으로부터의 장애 발생의 보고를 받고, 상기 제1 물리 계산기로부터 송신된, 상기 하드웨어 구성 정보 및 상기 고유의 구성 정보를 수신하는 스텝과, 교대처의 제2 물리 계산기를 결정하고, 상기 제2 물리 계산기에 상기 하드웨어 구성 정보 및 상기 고유의 구성 정보를 송신하는 스텝과,

상기 제2 물리 계산기에서, 상기 관리 장치로부터 송신된 상기 하드웨어 구성 정보 및 상기 고유의 구성 정보를 수신하는 스텝과, 상기 하드웨어 구성 정보 및 상기 고유의 구성 정보에 기초하여, 상기 제2 물리 계산기 상에 제2 논리 구획을 생성하는 것이 가능한지를 판정하는 스텝과,

상기 판정 스텝에 의해 상기 제2 논리 구획의 생성이 가능하다고 판정된 경우, 상기 고유의 구성 정보에 기초하여 제2 논리 구획을 생성하는 스텝을 갖는 것을 특징으로 하는 가상 계산기 시스템에서의 논리 구획의 이행 제어 방법.

청구항 9

복수의 물리 계산기와, 각 물리 계산기에서 복수의 LPAR(Logical Partition)을 생성하여 제어하는 하이퍼바이저와, 상기 복수의 물리 계산기 및 각 물리 계산기 상의 각 LPAR을 관리하는 관리 장치를 포함하고, 각 LPAR은 논리적인 Fibre Channel 포트를 갖고, 각 LPAR은 상기 논리적인 Fibre Channel 포트에 할당된 고유 ID에 의해 각 LPAR의 오퍼레이팅 시스템을 포함하는 소프트웨어 이미지의 부트를 스토리지 에리어 네트워크로부터 행하는 가상 계산기 시스템의 제어 방법으로서,

각 물리 계산기의 하이퍼바이저는, 상기 하이퍼바이저의 제어 하의 복수의 LPAR 중 어느 하나의 LPAR에서의 장애 발생을 검출한 경우에, 그 장애 발생을 상기 관리 장치에 통지하고,

상기 관리 장치는, 상기 장애 발생 LPAR의 상기 고유 ID를 포함하는 구성 정보를 상기 장애 발생을 검출한 하이퍼바이저로부터 수신하였을 때, 상기 장애 발생 LPAR의 상기 고유 ID를 포함하는 구성 정보를, 어느 하나의 물리 계산기의 하이퍼바이저에 송신하여 새로운 LPAR 생성을 지시하고, 상기 장애 발생 LPAR의 상기 고유 ID를 포함하는 구성 정보를 수신한 하이퍼바이저는, 상기 수신한 구성 정보와 동일한 구성 정보를 갖는 새로운 LPAR을 생성하는 가상 계산기 시스템의 제어 방법.

청구항 10

제9항에 있어서,

상기 관리 장치는, 상기 장애 발생 LPAR의 상기 고유 ID를 포함하는 구성 정보를, 상기 장애 발생 LPAR이 동작하고 있었던 물리 계산기와는 상이한 물리 계산기에 송신하는 가상 계산기 시스템의 제어 방법.

청구항 11

제9항에 있어서,

상기 관리 장치는, 각 물리 계산기의 CPU 성능, 메모리 용량, I/O 구성에 관한 정보를 저장하는 테이블을 포함하고, 상기 테이블을 참조하여 상기 장애 발생 LPAR과 동일한 구성 정보를 갖는 새로운 LPAR을 생성 가능한 물리 계산기를 선택하는 가상 계산기 시스템의 제어 방법.

청구항 12

제11항에 있어서,

상기 테이블에는 각 물리 계산기의 CPU 성능에 관한 정보로서, 미할당의 CPU 성능이 저장되고, 상기 관리 장치는, 상기 테이블을 참조하여 상기 미할당의 CPU 성능이 상기 장애 발생 LPAR의 실행 CPU 성능 이상인지를 조사하는 가상 계산기 시스템의 제어 방법.

청구항 13

제11항에 있어서,

상기 테이블에는 각 물리 계산기의 메모리 용량에 관한 정보로서, 미사용의 메모리 용량이 저장되고, 상기 관리 장치는, 상기 테이블을 참조하여 상기 미사용의 메모리 용량이 상기 장애 발생 LPAR이 사용하고 있었던 메모리 용량 이상인지를 조사하는 가상 계산기 시스템의 제어 방법.

청구항 14

제11항에 있어서,

상기 테이블에는 각 물리 계산기의 I/O 구성에 관한 정보로서, 미사용의 NIC수 및 HBA수가 저장되고, 상기 관리 장치는, 상기 테이블을 참조하여 상기 미사용의 NIC수 및 HBA수가 상기 장애 발생 LPAR이 사용하고 있었던 NIC수 및 HBA수 이상인지를 조사하는 가상 계산기 시스템의 제어 방법.

명세서

발명의 상세한 설명

기술분야

[0001] 본 발명은, 가상 계산기 시스템에 관한 것으로, 특히 임의의 물리 계산기 상의 논리 구간에 장애가 발생한 경우에, 다른 물리 계산기에 상기 논리 구간의 교대를 생성하여, 그 논리 구간의 처리를 이행하는 가상 계산기 시스템 및 논리 구획의 이행 제어 방법에 관한 것이다.

배경기술

[0002] 1대의 물리 계산기 상에 복수의 논리 계산기 또는 논리 구획(이하, LPAR(Logical Partition)이라고 함)을 구축하고, 각 논리 계산기에서 각각 OS(오퍼레이팅 시스템)를 동작시키고, 이에 의해 복수의 논리 계산기에서 복수의 고유의 OS를 동작시키는 것이 가능한 가상 계산기 시스템이 실용화되어 있다. 또한, 최근에는, 각각의 논리 계산기에 논리적인 FC(Fibre Channel) 확장 보드 또는 FC 포트를 갖게 한 가상 계산기 시스템을, RAID(Redundant Array of Inexpensive Disks) 장치를 포함하는 SAN(스토리지 에리어 네트워크) 환경에서 사용하는 예도 있다.

[0003] SAN 환경에서 부트를 실현하는 계산기 시스템으로서, OS가 인스톨되어 있는 RAID 장치 내의 로지컬 유닛의 데이터를 보호하기 위해, 각각의 계산기로부터만 액세스를 가능하게 하는 시큐리티 기능이 RAID 장치에 의해 유효하

게 되어 있다. 이 시큐리티 기능으로서는 일반적으로, 각각의 계산기에 탑재되는 FC 포트에 할당된 고유의 ID(World Wide Name)를 이용하고, OS가 인스톨된 로지컬 유닛과 계산기가 갖는 FC 포트에 할당된 고유의 ID(World Wide Name)를 관련지어, 그 ID(World Wide Name)를 갖는 FC 포트로부터의 액세스만을 허용하는 방법이 이용되고 있다. 또한, OS를 포함하는 소프트웨어에는, 장치 고유의 ID(World Wide Name)가 기록되어 있는 경우도 있다.

[0004] SAN으로부터의 부트를 행하는 계산기 시스템의 용장화 구성에서는, 현용계 계산기와 대기계 계산기에서 갖는 FC 포트에 할당된 고유의 ID(World Wide Name)가 서로 다르기 때문에, 현용계 계산기로부터 대기계 계산기로 교대할 때, OS를 포함하는 소프트웨어 이미지를 그대로 이용할 수 없어, SAN 관리 소프트웨어나 사람의 손에 의한 RAID 장치측에서의 시큐리티 기능의 설정 변경이 필요로 된다. 이것은, 현용계 계산기와 대기계 계산기라고 하는 물리 계산기에서 뿐만 아니라, 가상 계산기 시스템에서의 LPAR간에서도 마찬가지이다. 즉, 가상 계산기 시스템에서 물리 계산기 상에서 복수의 LPAR를 동작시켜, 현용 LPAR로부터 대기계 LPAR로 교대하는 경우에도, 각 LPAR의 논리적인 FC 포트에 할당된 고유의 ID(World Wide Name)가 서로 다르기 때문에, RAID 장치측에서의 시큐리티 기능의 설정 변경이 필요로 된다.

[0005] 복수의 물리 계산기 상에 각각 LPAR를 구축할 수 있는 가상 계산기 시스템으로서, 임의의 물리 계산기 상의 LPAR로부터 다른 물리 계산기에 LPAR의 구성 정보를 이동시켜 동작을 이어받게 하는 기술에 관해서는, 예를 들면 특허 문헌 1 및 특허 문헌 2에 개시되어 있다.

[0006] [특허 문헌 1] 일본 특허 공개 제2005-327279호 공보

[0007] [특허 문헌 2] 일본 특허 공개 평10-283210호 공보

발명의 내용

해결 하고자하는 과제

[0008] 상기 특허 문헌 1 및 2에는, 임의의 물리 계산기의 LPAR에 장애가 발생한 경우에서의, 다른 물리 계산기에 생성되는 LPAR를 예비기로서 이용하기 위한 LPAR의 이동에 대해서는 언급되어 있지 않다.

[0009] 또한, 상기 특허 문헌 1 및 2에는, SAN 환경 하의 가상 계산기 시스템으로서, 임의의 LPAR로부터 다른 LPAR로 교대하는 경우에, RAID 장치측에서의 시큐리티 기능의 설정 변경을 불필요로 하기 위해, LPAR의 논리적인 FC 포트에 할당된 고유의 ID(World Wide Name)를 이어받는 점에 대해서도 언급되어 있지 않다.

[0010] 본 발명의 목적은, SAN 환경 하에서 물리 계산기 상의 LPAR에 장애가 발생한 경우에, 다른 물리 계산기에 교대용 LPAR를 설정하여, RAID 장치측에서의 시큐리티 기능의 설정 변경을 불필요로 하는 LPAR의 이행을 가능하게 하는 가상 계산기 시스템을 제공하는 데에 있다.

과제 해결수단

[0011] 본 발명은, 바람직하게는, 제1 및 제2 물리 계산기를 포함하는 복수의 물리 계산기와, 그 복수의 물리 계산기에 네트워크를 통하여 접속되는, 해당 물리 계산기 및 해당 논리 구간을 관리하는 관리 장치를 포함하고, 각 물리 계산기에 논리 구획을 생성하여 OS를 동작시킬 수 있는 가상 계산기 시스템으로서,

[0012] 상기 제1 물리 계산기는, 그 제1 물리 계산기 또는 거기에 형성된 제1 논리 구획에 장애가 발생한 것을 검출하는 장애 검출 수단과, 상기 제1 물리 계산기의 하드웨어 구성 정보 및 상기 제1 논리 구획에 할당된 고유의 구성 정보를 관리하는 제1 관리 수단을 갖고,

[0013] 상기 관리 장치는, 상기 장애 검출 수단으로부터의 장애 발생의 보고를 받고, 상기 제1 관리 수단으로부터 상기 하드웨어 정보 및 상기 고유의 구성 정보를 수신하는 수단과, 교대처의 제2 물리 계산기를 결정하고, 그 제2 물리 계산기에 상기 하드웨어 정보 및 상기 고유의 구성 정보를 송신하는 수단을 갖고,

[0014] 상기 제2 물리 계산기는, 상기 관리 장치로부터 송신된 상기 하드웨어 정보 및 상기 고유의 구성 정보를 수신하는 수단과, 그 하드웨어 정보 및 그 고유의 구성 정보에 기초하여, 상기 제2 물리 계산기 상에 제2 논리 구획을 생성하는 것이 가능한지를 판정하는 수단과, 상기 판정 수단에 의해 상기 제2 논리 구획의 생성이 가능하다고 판정된 경우, 그 고유의 구성 정보에 기초하여 제2 논리 구획을 생성하는 수단을 갖는 가상 계산기 시스템으로서 구성된다.

효 과

[0015] 본 발명에 따르면, SAN 환경 하에서 물리 계산기 상의 LPAR에 장애가 발생한 경우에, 다른 물리 계산기에 고대용 LPAR를 설정하여, RAID 장치측에서의 시큐리티 기능의 설정 변경을 불필요로 하는 LPAR의 이행이 가능하게 된다. 또한, 관리 서버의 제어 하에, 이행처의 LPAR로 이행된 LPAR의 구성 정보 등을 이행하므로, 이행원의 물리 계산기에 장애가 발생한 경우에도, LPAR의 이행이 가능하다.

발명의 실시를 위한 구체적인 내용

- [0016] 이하, 본 발명의 실시 형태에 대해 도면을 참조하여 설명한다.
- [0017] 도 1을 참조하면, 본 실시예에 따른 계산기 시스템은, 1대의 서버 새시(105)에, 복수대의 서버 모듈(이하 간단히 서버라고 함)(111, 112)을 탑재할 수 있는 블레이드 서버의 형태를 이루고 있다. 서버 새시(105)에는, 서비스 프로세서(SVP)(106)가 탑재된다.
- [0018] 서버(111, 112)는, NIC(Network Interface Card)(122, 132)를 통하여 네트워크 SW(103) 경유로 관리 서버(101)에 접속되고, 또한 FC-HBA(Fibre Channel Host Bus Adapter)(121, 131)를 통하여 파이버 채널 스위치(FC-SW)(135) 경유로 스토리지 장치(137)에 접속된다.
- [0019] 서버(111, 112)는, 기본적으로 마찬가지로의 구성을 갖고, 각각 BMC(Base Management Controller)(120, 130), FC-HBA(121, 131), NIC(122, 132)를 갖고 있다. 하이퍼바이저(117, 127)는 물리적으로 1대의 서버를 논리적으로 복수의 서버로 보이게 하는 가상화 기구이다.
- [0020] 서버(111)에서는 1개의 하이퍼바이저(117) 상에 시뮬레이션된 2대의 LPAR(113, 114)이 구축되어 동작하고 있다. 하이퍼바이저(117, 127) 내의 Hypervisor-Agt(119, 129)는, LPAR의 장애를 검지하여 관리 서버(101)에 그 보고를 행하기 위한 에이전트이다.
- [0021] 본 실시예에서, 서버(112)에는 1대의 LPAR(123)이 동작하고 있지만, 후에 서버(111)의 LPAR2(114)의 고대 LPAR4(124)가 설정된다.
- [0022] FC-HBA(121, 131)는, 통신을 행하기 위해 그 HBA의 어드레스로서 FC 접속 포트 1개에 대해 1개의 WWN을 갖는다. LPAR(113, 114)은 논리적인 HBA를 1 포트(115, 116)씩 갖고, 각각 vfcWWN1(115), vfcWWN2(116)와 같은, 고유의 WWN(World Wide Name)이 부여된다. 논리적인 HBA도 물리적인 HBA와 마찬가지로의 WWN을 갖는다. 또한, 서버(112)에서의 LPAR3(123)도 마찬가지로 고유의 WWN이 부여된다.
- [0023] 스토리지 장치(137)는, 논리적으로 규정된 LU(논리 유닛)라고 불리는 다수의 Disk 유닛(138~140)을 갖고 있다. 어느 LU가 어느 서버에 접속되어 있는지를 나타내는 접속 정보는 스토리지 장치(137) 내의 컨트롤러에 의해 관리되고 있다. 예를 들면, LU10(138)은 vfcWWN1(115)의 WWN을 갖는 LPAR(113)에 접속되고, LU11(139)은 vfcWWN2(116)의 WWN을 갖는 LPAR(114)에 접속되어 있다. 이 접속 관계를 설정하는 기능을 LUN 시큐리티 설정 기능이라고 부른다.
- [0024] SVP(106)는 서버 새시 내의 모든 서버를 관리하고, 또한 서버의 전원 제어 및 장애 처리를 담당한다. 서버를 관리하기 위해, 서버의 하드웨어 구성 정보(1101)(도 11 참조) 및 하이퍼바이저 구성 정보(1111)(도 12 참조)를 SVP 내의 불휘발 메모리(도시 생략)에 기억하여 관리한다. 이들 구성 정보(1101, 1111)는 서버 단위로 관리되고, 도시한 예에서는 서버(111, 112)에 대응하여, 2면의 구성 정보(108-1, 108-2)를 갖는다. 또한, 하이퍼바이저 구성 정보(1111)에는 서버(111, 112)의 각각의 하이퍼바이저(117, 127)에 대응한 정보가 포함된다. 관리 서버(101)는, 서버(111, 112) 및 그것에 형성된 LPAR를 관리한다. 그를 위해서, 서버의 관리 정보(107)(도 13 참조)를 메모리(도시 생략)에 기억하여 관리한다. 본 실시예에서는, 또한 LPAR의 이행을 관리하는 기능을 갖는다.
- [0025] 다음으로, 도 11~도 13을 참조하여, 각 관리 정보의 내용에 대해 설명한다.
- [0026] 도 11에 도시한 바와 같이, 서버의 하드웨어 구성 정보(서버 모듈·하드웨어 구성 정보라고 하는 경우도 있음)(1101)는, 부트 설정 정보(1102), HBA-BIOS 정보(1103), addWWN 정보(1104), 물리 서버의 OS 종류 정보(1105), Hyper Treading의 무효 지정(1106), SVP가 보존하는 하이퍼바이저의 IP 어드레스(1107), 아키텍처(1108) 등의 물리 서버 정보를 유지한다. 이 하드웨어 구성 정보(1101)는 서버 모듈(파티션)마다 존재한다.
- [0027] 도 12에 도시한 바와 같이, 하이퍼바이저 구성 정보(1111)는 파티션 내의 LPAR 단위로 관리되는 정보로서,

LPAR(113, 114) 대응으로 존재한다(1111-1, 1111-2). 각 하이퍼바이저 구성 정보(1111)는, vfcWWN 정보(1112-1), LPAR이 가동 중인지의 여부를 나타내는 Active/NonActive(1113-1), CPU의 수 등을 포함하는 CPU 정보(1114-1), 메모리 용량(1115-1), HBA나 NIC 등을 포함하는 I/O 구성(1115-1) 등의 정보를 유지한다.

- [0028] 상기 서버의 하드웨어 구성 정보(1101) 및 하이퍼바이저 구성 정보(1111)는, SVP(106)에서 설정되어 관리되지만, 이들 정보는, 각 서버 상에서 동작하고 있는 하이퍼바이저에서도 유지하고 있다.
- [0029] 도 13에 도시한 바와 같이, 관리 서버(101)에서 관리되는 서버의 관리 정보(서버 모듈 관리 정보라고 하는 경우도 있음)(107)는, 서버 모듈 번호(1201), 하드웨어의 아키텍처 중별(1202), 실장 메모리 용량(1203), 가동 중인 LPAR의 합계 메모리 사용량(1204), 메모리의 빈 용량(1205), 실장 CPU 성능(1206), 할당 완료 CPU 성능의 합계(1207), 빈 CPU 성능(1208), 빈 NIC수(1209), 빈 HBA수(1210) 등의 정보를 유지한다.
- [0030] 본 실시예에 따르면, 서버(111)의 LPAR에 장애가 발생하였을 때에, 장애 보고를 접수한 관리 서버(101)는, 서버(112) 내에 교대용의 LPAR4(124)를 설정하고, 그 LPAR4(124)에, 장애가 발생한 LPAR의 LPAR 고유의 구성 정보를 이어받게 하기 위한 제어를 행한다.
- [0031] 이하, 도 2 및 도 3을 참조하여, 서버(111)의 LPAR에 장애가 발생하였을 때의 교대 LPAR의 설정 및 LPAR 고유의 구성 정보의 계승 처리에 대해, 상세하게 설명한다. 도시한 예는, 서버(111)의 LPAR2(114)에 장애가 발생한 경우에서의, 관리 서버(101), 서버(111)의 하이퍼바이저(117), 서버 모듈(112)의 하이퍼바이저(127)가 행하는 처리 동작을 나타낸다.
- [0032] LPAR2(114)에 장애가 발생하고, 서버(111)에서 동작하는 하이퍼바이저(117)가 그 장애를 검출하면(S201), 하이퍼바이저(117)는 관리 서버(101)에 장애 통지(Hypervisor-Agt 알러트)를 행한다(S202). 관리 서버(101)는 장애가 발생한 LPAR2를 정지하도록 정지 커맨드를 송출한다(S203). 하이퍼바이저(117)는, LPAR 정지 커맨드를 수신한 후, LPAR2의 가동 정지(deactivate 처리)를 행한다(S205). 그리고 deactivate 처리가 완료되면, 관리 서버(101)에 대해 Hypervisor-Agt 알러트를 송출하여, deactivate 완료를 전달한다(S206).
- [0033] Hypervisor-Agt 알러트를 받은 관리 서버(101)는, 관리 정보로서 장애가 발생한 LPAR의 정지 상태를 표시기에 표시하고(S207), LPAR2의 구성 정보 읽어들이기 커맨드를 송출한다(S208).
- [0034] 그 커맨드를 수신한 하이퍼바이저(117)는, 자신이 유지하고 있는, 서버 모듈·하드웨어 구성 정보 및 LPAR2의 하이퍼바이저 구성 정보를 관리 서버(101)에 송신한다(S209).
- [0035] 관리 서버(101)는, 데이터의 수신을 완료하면, 수신 완료를 표시한다(S210). 그 후, 교대처의 서버 모듈을 결정한다(S301). 예를 들면 교대처의 서버 모듈(112) 상에서 LPAR을 생성하고자 하는 하이퍼바이저(127)에 대해, 장애가 발생한 서버 모듈(111)의 서버 모듈·하드웨어 구성 정보 및 LPAR2의 하이퍼바이저 구성 정보를 수신하도록 지시한다(S302).
- [0036] 하이퍼바이저(127)는, 장애가 발생한 LPAR2에 관한 구성 정보를 수신하면(S303), 그 구성 정보에 기초하여, 교대처에서 LPAR이 생성 가능한지의 여부를 판정한다(S305). 이 판정에 대해서는 후에 상세하게 설명한다. 판정의 결과, 소정의 조건을 만족하고 있으면, 이행처의 서버(112)에 이행원인 LPAR2에 관한 구성 정보를 이어받은 LPAR이 생성된다(S306). 이 예에서는, LPAR4(124)가 이행처의 LPAR로 된다. LPAR4(124)의 생성이 완료되면, 하이퍼바이저(127)는 Hypervisor-Agt 알러트를 송출하여, LPAR의 생성 완료를 통지한다(S307).
- [0037] 관리 서버(101)는, Hypervisor-Agt 알러트를 수신하면, 하이퍼바이저(127)에 생성된 LPAR4를 기동하도록, 기동 커맨드를 송출한다(S308). 이 기동 커맨드를 수신한 하이퍼바이저(127)는, 생성한 LPAR(124)을 기동(activate)한다(S309). 그리고, Hypervisor-Agt 알러트를 송출하여, LPAR(124)의 기동 완료를 전달한다(S310). Hypervisor-Agt 알러트를 수취한 관리 서버(101)는, LPAR(124)의 기동 상태를 표시기에 표시한다(S311).
- [0038] 다음으로 도 4 및 도 5를 참조하여, LPAR2(114)에 장애가 발생하였을 때의 관리 서버(101)의 처리에 대해 설명한다.
- [0039] 하이퍼바이저(117)로부터 LPAR2에 장애가 발생하였다는 취지를 전달하는 Hypervisor-Agt 알러트를 받으면, 관리 서버(101)는 LPAR 장애 검출 시의 처리를 시작한다(S401).
- [0040] 우선, 장애가 발생한 LPAR2가 동작하는 서버 모듈(111)의 하이퍼바이저(117)에 대해, LPAR2의 가동을 정지하기 위한 정지 커맨드를 송출한다(S402). 그 후, LPAR2의 정지 처리가 완료될 때까지 대기하고(S403), 정지 처리가 정상적으로 완료되면, LPAR2의 표시 테이블을 「정지 상태」로 한다(S404). 한편, 정지 처리가 정상적으로 완

료되지 않으면, 콜드 스탠바이 실패를 표시하고(S411), 종료한다(S412).

- [0041] LPAR2의 표시 테이블이 「정지 상태」로 되었다면(S404), LPAR2의 구성 정보의 읽어들이기 커맨드를 송출한다(S405). LPAR2의 구성 정보를 수신하고(S406), 수신이 정상적으로 종료되었다면(S407), 수신 완료를 표시한다(S408). 한편, 수신이 정상적으로 종료되지 않으면 콜드 스탠바이 실패를 표시하고(S413), 종료한다(S414).
- [0042] 수신이 정상적으로 종료되고(S407), 수신 완료를 표시한(S408) 후에, LPAR2의 실효 CPU 성능과, LPAR2를 생성하는 서버 모듈 이외의 서버 모듈의 실효 CPU 성능을 계산한다.
- [0043] 여기서, LPAR2의 실효 CPU 성능은, (물리 CPU의 수)×(이행 전의 LPAR에서의 서비스율)로서 계산한다. 또한, LPAR2를 생성하는 서버 모듈 이외의 서버 모듈의 실효 CPU 성능은, (물리 CPU의 수)×(100%-(현재 가동하고 있는 모든 LPAR의 서비스율))로서 계산한다.
- [0044] 다음으로, 관리 서버(101)의 서버 모듈 관리 정보(107)를 이용하여, LPAR 생성을 위한 서버 모듈의 조건을 판정한다(S410). 이 조건이란, 예를 들면, 이하 (a)~(d)의 판정을 포함한다.
- [0045] (a) LPAR2와 동일한 아키텍처의 서버 모듈이 있는가? (b) LPAR2 이상의 메모리가 비어 있는 서버 모듈이 있는가? (c) LPAR2의 실효 CPU 성능 이상의 실효 CPU 성능을 갖는 서버 모듈이 있는가? (d) LPAR2가 사용하고 있었던 이상의 NIC, HBA가 비어 있는 서버 모듈이 있는가?
- [0046] 이들 4개의 조건을 모두 만족하고 있으면, 조건을 만족하고 있는 서버 모듈 중에서, 실효 CPU 성능이 최고인 것을 교대처의 서버 모듈로서 선택한다(S501). 4개의 조건 중 1개라도 만족하고 있지 않으면, 콜드 스탠바이 실패를 표시하고(S415), 종료한다(S416).
- [0047] 4개의 조건이 만족되는 교대처의 서버 모듈(이 예에서는 서버 모듈(112))이 선택되면, 교대처의 서버 모듈(112)의 하이퍼바이저(127)에 대해, 장애가 발생한 LPAR2에 관한 구성 정보를 전송하여, LPAR을 생성하도록 지시한다(S502). 그리고, 장애 발생원 서버 모듈(111)의 하이퍼바이저(117)로부터 수신한 데이터(장애 발생 LPAR2에 관한 구성 정보)를 하이퍼바이저(127)에 송신한다(S503). 이 데이터의 송신이 정상적으로 종료되면(S504), 송신 완료를 표시한다(S505). 한편, 데이터 송신이 정상적으로 완료되지 않으면(S504), 콜드 스탠바이 실패를 표시하고(S511), 종료한다(S512).
- [0048] 그 후, 교대처 서버 모듈(112)에서 LPAR이 생성되는 것을 대기한다(S506). 생성되는 LPAR4는, 장애가 발생한 LPAR2와 마찬가지로의 구성을 갖는 것이다. LPAR4의 생성이 정상적으로 종료되면, 교대처 서버 모듈(112)의 교대처 LPAR4(124)를 기동하는 커맨드를 송출한다(S507). 한편, LPAR4 생성이 정상적으로 종료되지 않으면, 콜드 스탠바이 실패를 표시하고(S513), 종료한다(S514).
- [0049] 교대 LPAR4(124)의 생성이 정상적으로 종료되고, 기동 커맨드를 송출하였다면(S507), 교대처 LPAR4(124)의 기동 완료를 대기한다(S508). 그리고 정상적으로 기동하였다면, 교대처 LPAR4(124)의 상태 표시를 「기동 상태」로 하고(S509), 종료한다(S510). 한편, LPAR4(124)의 기동이 정상적으로 기동하지 않는 경우에는, 콜드 스탠바이 실패를 표시하고(S515), 종료한다(S516).
- [0050] 이상과 같은 제어에 의해, 교대처 LPAR4(124)가 장애 발생 LPAR2(114)의 교대기로서 기동 가능하게 되는 것은 이하의 이유에 의한다. 스토리지 장치에의 액세스는 WWN에 의해 관리된다. WWN은 물리 디바이스의 포트마다 할당되지만, 본 실시예에서는 LPAR마다 논리 HBA를 설정하고, 논리 HBA의 포트마다 WWN을 할당하고 있다. 이하 이 논리 HBA의 WWN을 vfcWWN이라고 부른다. 도 1의 설명에서 기술한 바와 같이, LUN과 WWN의 접속 관계는 LUN 시큐리티 기능에 의해 설정되어 있다. 스토리지 장치측으로부터는, 논리/물리 WWN의 구별은 짓지 않으므로, LPAR 단위로의 LU에의 액세스권의 관리가 가능하게 된다(vfcWWN을 이용할 때에는, 스토리지 장치로부터 물리 디바이스의 WWN이 인식되지 않도록 설정한다). 이행처의 LPAR은, 장애 발생 시의 LPAR이 사용하고 있었던 vfcWWN과 동일한 vfcWWN을 사용하여 부팅함으로써, 이행 전과 동일한 시스템을 구동시킬 수 있다.
- [0051] 다음으로 도 6을 참조하여, LPAR2에 장애가 발생하였을 때의 하이퍼바이저의 처리에 대해 설명한다.
- [0052] LPAR2에 장애가 발생하면, 하이퍼바이저(117)는 LPAR 장애 검출 처리를 개시한다(S601). 장애 검출 처리에서, 장애 발생 요인을 해석하여, 회복 가능한 요인지의 여부를 판단한다(S602). 그 판단의 결과, LPAR 장애가 회복 불가능한 요인인 경우, Hypervisor-Agt(118)에 대해 LPAR 장애를 전달하기 위해, Hypervisor-Agt 얼러트 송출을 요구하고(S603), LPAR 장애 시의 로그 취득 등의 장애 처리를 실행하고(S604), 처리를 종료한다(S605).
- [0053] 한편, LPAR 장애가 회복 가능한 요인인 경우, 회복 처리를 행하고(S606), 종료한다(S607).

- [0054] 다음으로 도 7~도 8을 참조하여, 관리 서버(101)로부터의 커맨드 실행 요구에 수반하는 Hypervisor-Agt(118)에서의 커맨드의 처리에 대해 설명한다.
- [0055] 관리 서버(101)로부터 송신된 커맨드 실행 요구를 받으면, Hypervisor-Agt(118)는 수신 처리를 행한다(S701). 요구되는 커맨드에는 복수의 종류가 있으므로, 우선 커맨드의 종별을 해석한다(S702). 이 예에서는, LPAR의 정지를 행하는 LPAR 정지(deactivate) 커맨드와, LPAR 구성 정보 읽어들이기 커맨드와, LPAR 구성 정보 써넣기 커맨드와, LPAR의 기동을 행하는 LPAR 기동(activate) 커맨드와, LPAR 생성 커맨드의 5개의 커맨드의 처리를 행한다.
- [0056] LPAR deactivate 커맨드인 경우, 정지 대상 LPAR은 타당한지를 판정한다(S703). 타당하지 않다고 판단한 경우에는 에러 처리를 행하고(S707), 종료한다(S708). 정지 대상 LPAR2가 타당하다고 판단한 경우에는, 정지 대상 LPAR2의 정지 처리를 행한다(S704). 그리고, 정지 처리가 성공하였는지의 여부를 판단한다(S705). 정지 처리가 실패한 경우, 에러 처리하고(S707), 종료한다(S708). 한편, 정지 처리가 성공한 경우에는, LPAR2의 정지 완료료를 전달하기 위해 Hypervisor-Agt 얼러트 송신 요구를 행하고, 종료한다(S708).
- [0057] LPAR 구성 정보 읽어들이기 커맨드인 경우, 대상 LPAR2의 구성 정보를 관리 서버(101)에 전송한다. 그 후, 데이터 전송이 성공하였는지의 여부를 판단하고(S710), 데이터 전송이 성공하였다면 처리를 종료한다(S712). 한편, 실패하였다면, 에러 처리하고(S711), 종료한다(S712).
- [0058] LPAR 구성 정보 써넣기 커맨드인 경우, 대상 LPAR2의 구성 정보를 관리 서버(101)로부터 하이퍼바이저(127)에 전송한다. 그 후, 데이터 전송이 성공하였는지의 여부를 판단하고(S714), 데이터 전송이 성공하였다면 처리를 종료한다(S716). 한편, 실패하였다면, 에러 처리하고(S714), 종료한다(S716).
- [0059] 다음으로, LPAR 기동 커맨드인 경우(도 8 참조), 기동 대상의 LPAR2는 타당한지를 판정한다(S801). 그 결과, 타당하지 않다고 판단한 경우에는 에러 처리를 행하고(S805), 종료한다(S806). 한편, 기동 대상의 LPAR2가 타당하다고 판단한 경우에는, 기동 대상 LPAR2의 기동 처리를 행한다(S802). 그 후, 기동이 성공하였는지를 판단하고(S803), 기동에 실패한 경우에는 에러 처리를 행하고(S805), 종료한다(S806).
- [0060] 한편, 기동에 성공한 경우에는, LPAR의 activate 완료료를 전달하기 위해 Hypervisor-Agt 얼러트 송신 요구를 행하고(S804), 종료한다(S806).
- [0061] 다음으로, LPAR 생성 커맨드인 경우, 우선 이행 전 및 이행처의 실효 CPU 성능의 계산을 행한다(S807). 이행 전의 실효 CPU 성능은, (물리 CPU의 수)×(이행 전의 LPAR에서의 서비스율)로서 계산한다. 이행처의 실효 CPU 성능은, (물리 CPU의 수×(100%-(현재 기동하고 있는 모든 LPAR의 서비스율)))로서 계산한다.
- [0062] 그 후, 다음 3개의 조건의 판정을 행한다(S808). (1) 이행 전의 실효 성능과 이행처의 실효 CPU 성능을 비교하여 이행처의 실효 CPU 성능이 이행 전의 실효 CPU 성능 이상인 것. (2) 이행처의 메모리가 비어 있는 것. (3) 이행처에 이행원의 LPAR이 사용하고 있었던 수와 동일수 이상의 NIC, HBA가 비어 있는 것.
- [0063] 상기 3개의 조건 중 하나라도 만족하고 있지 않으면, LPAR 생성은 불가능하다고 간주하여, 에러 처리하고(S812), 종료한다(S813).
- [0064] 한편, 3개의 조건을 모두 만족하고 있으면, 대상 LPAR을 생성한다(S809). 이 예에서는, LPAR2의 교대처로서 LPAR4(124)를 생성한다.
- [0065] 그 후, LPAR의 생성이 성공하였는지를 판정하고(S810), 성공한 경우, LPAR 생성 완료료를 전달하기 위해 Hypervisor-Agt 얼러트 송신 요구를 행하고(S811), 종료한다(S813). 한편, LPAR의 생성이 실패한 경우에는, 에러 처리를 행하고(S812), 종료한다(S813).
- [0066] 다음으로, 도 9 및 도 10을 참조하여, Hypervisor-Agt 얼러트 송신 요구가 있었던 경우의 Hypervisor-Agt의 송신 처리에 대해 설명한다.
- [0067] Hypervisor-Agt 얼러트 송신 요구가 있었던 경우, Hypervisor-Agt(118)는 얼러트의 종별을 해석한다(S902).
- [0068] 그 결과, 얼러트의 종별이 LPAR 기동 완료인 경우에는, LPAR 기동 완료 얼러트를 송신하고(S903), 종료한다(S906).
- [0069] 얼러트의 종별이 LPAR 기동 실패인 경우에는, LPAR 기동 실패 얼러트를 송신하고(S904), 종료한다(S906).
- [0070] 얼러트의 종별이 LPAR 장애 발생인 경우에는, LPAR 장애 발생 얼러트를 송신하고(S905), 종료한다(S906).

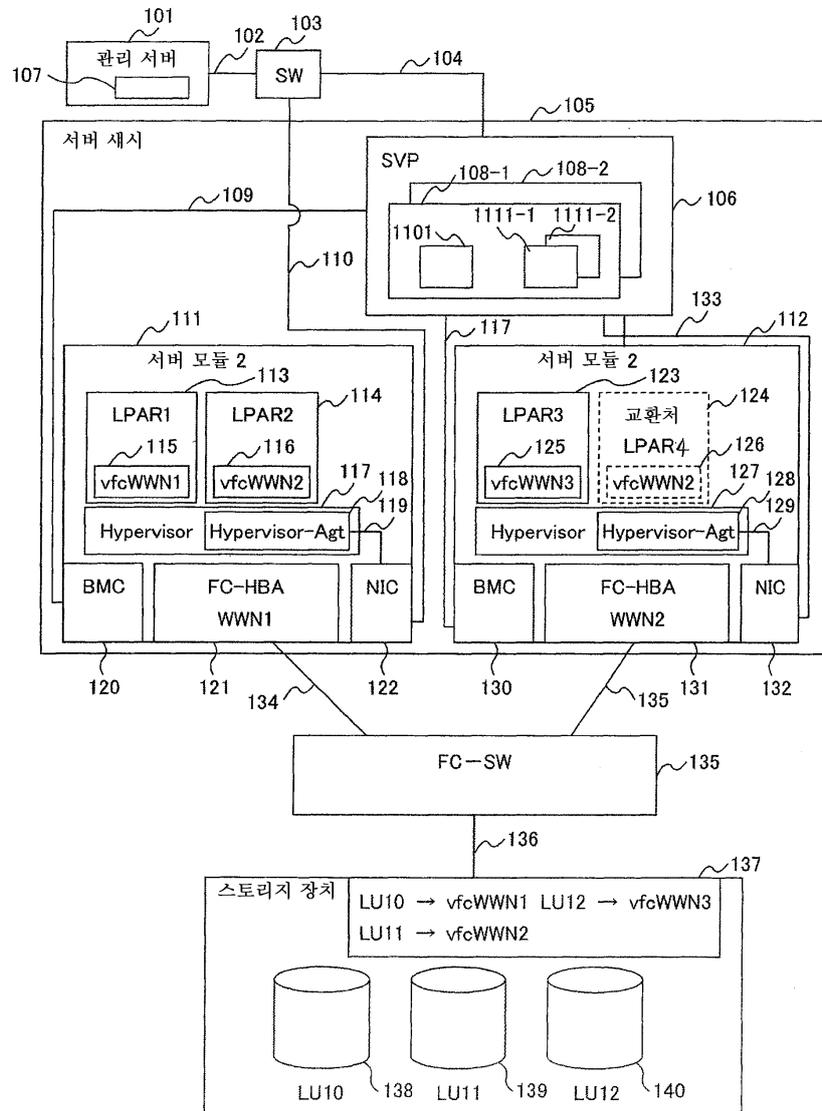
- [0071] 알러트의 중별이 LPAR 정지 완료인 경우에는, LPAR deactivate 완료 알러트를 송신하고(S1001), 종료한다(S906).
- [0072] 알러트의 중별이 LPAR 정지 실패인 경우에는, LPAR 정지 실패 알러트를 송신하고(S1002), 종료한다(S906).
- [0073] 알러트의 중별이 LPAR 생성 완료인 경우에는, LPAR 생성 완료 알러트를 송신하고(S1003), 종료한다(S906).
- [0074] 알러트의 중별이 LPAR 생성 실패인 경우에는, LPAR 생성 실패 알러트를 송신하고(S1004), 종료한다(S906).
- [0075] 상기한 예는, 서버(111)의 LPAR에 장애가 발생하였을 때에, 관리 서버(101)의 제어 하, 이행원 및 이행처의 하이퍼바이저간에서 다양한 정보를 주고 받아, LPAR의 이행 제어를 행하는 것이다.
- [0076] 또한, 서버 장애 시의 검출은 SVP로부터도 행할 수 있다. 이에 의해 하드웨어 장애 시에도 그 상에서 동작하고 있었던 LPAR을 각각의 물리 머신에 이행시킬 수 있다.
- [0077] 이상과 같이, 본 실시예에 따르면, 가상 계산기 시스템의 LPAR 장애 시에, LPAR 단위의 치밀한 교대를 실현할 수 있으므로, 효율이 요구되는 가상 계산기 시스템의 이용 업무에 적용할 수 있다. 또한, 복수의 물리 계산기 간에 성능상의 변동이 있는 경우, 특정한 LPAR의 물리 계산기간의 이동이 용이하게 가능하게 된다.

도면의 간단한 설명

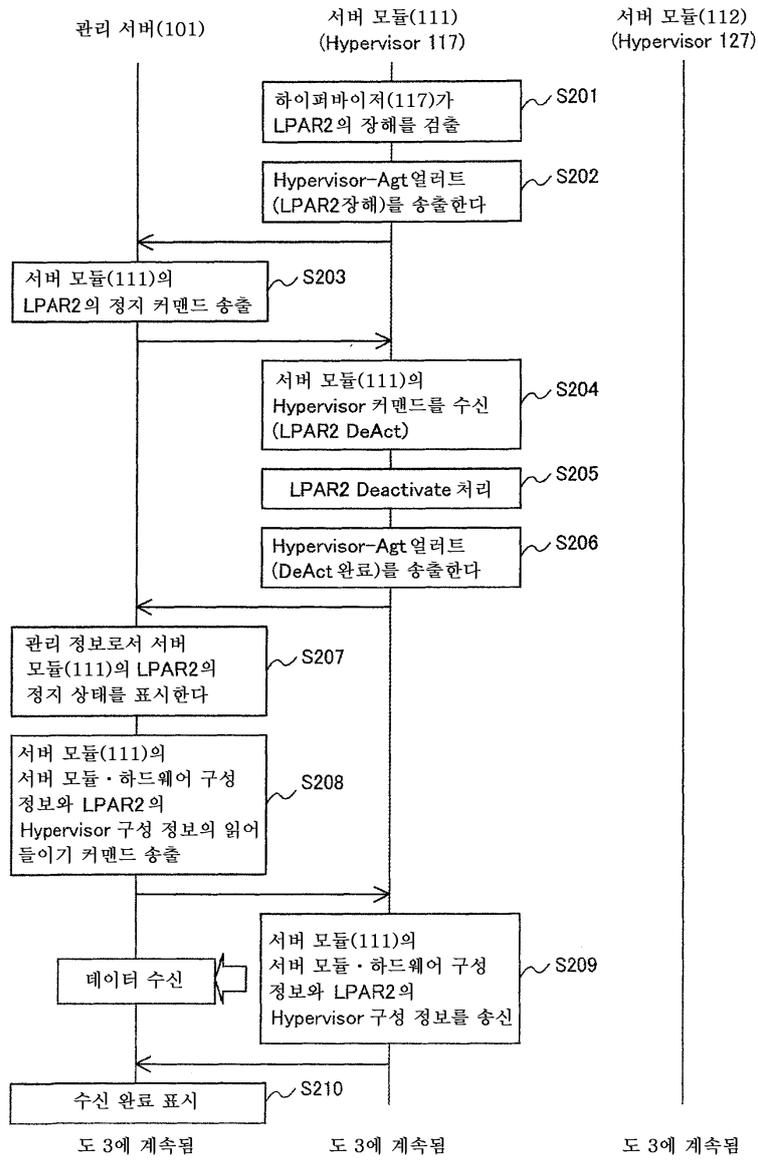
- [0078] 도 1은 일 실시예에서의 계산기 시스템의 구성을 도시하는 도면.
- [0079] 도 2는 장애 발생 시의 처리를 설명하는 플로우차트.
- [0080] 도 3은 장애 발생 시의 처리를 설명하는 플로우차트.
- [0081] 도 4는 장애 발생 시의 관리 서버의 처리를 설명하는 플로우차트.
- [0082] 도 5는 장애 발생 시의 관리 서버의 처리를 설명하는 플로우차트.
- [0083] 도 6은 장애 발생 시의 하이퍼바이저의 처리를 설명하는 플로우차트.
- [0084] 도 7은 Hypervisor-Agt에서의 커맨드의 처리를 설명하는 플로우차트.
- [0085] 도 8은 Hypervisor-Agt에서의 커맨드의 처리를 설명하는 플로우차트.
- [0086] 도 9는 Hypervisor-Agt의 송신 처리를 설명하는 플로우차트.
- [0087] 도 10은 Hypervisor-Agt의 송신 처리를 설명하는 플로우차트.
- [0088] 도 11은 서버의 하드웨어 구성 정보(1101)의 내용을 나타내는 도면.
- [0089] 도 12는 하이퍼바이저 구성 정보(1111)의 내용을 나타내는 도면.
- [0090] 도 13은 서버의 관리 정보(107)의 내용을 나타내는 도면.
- [0091] <도면의 주요 부분에 대한 부호의 설명>
- [0092] 105 : 서버 새시
- [0093] 106 : 서비스 프로세서(SVP)
- [0094] 111, 112 : 서버
- [0095] 113, 114 : LPAR
- [0096] 117, 127 : 하이퍼바이저
- [0097] 120, 130 : BMC(Base Management Controller)
- [0098] 121, 131 : FC-HBA
- [0099] 122, 132 : NIC
- [0100] 135 : 파이버 채널 스위치(FC-SW)
- [0101] 137 : 스토리지 장치

도면

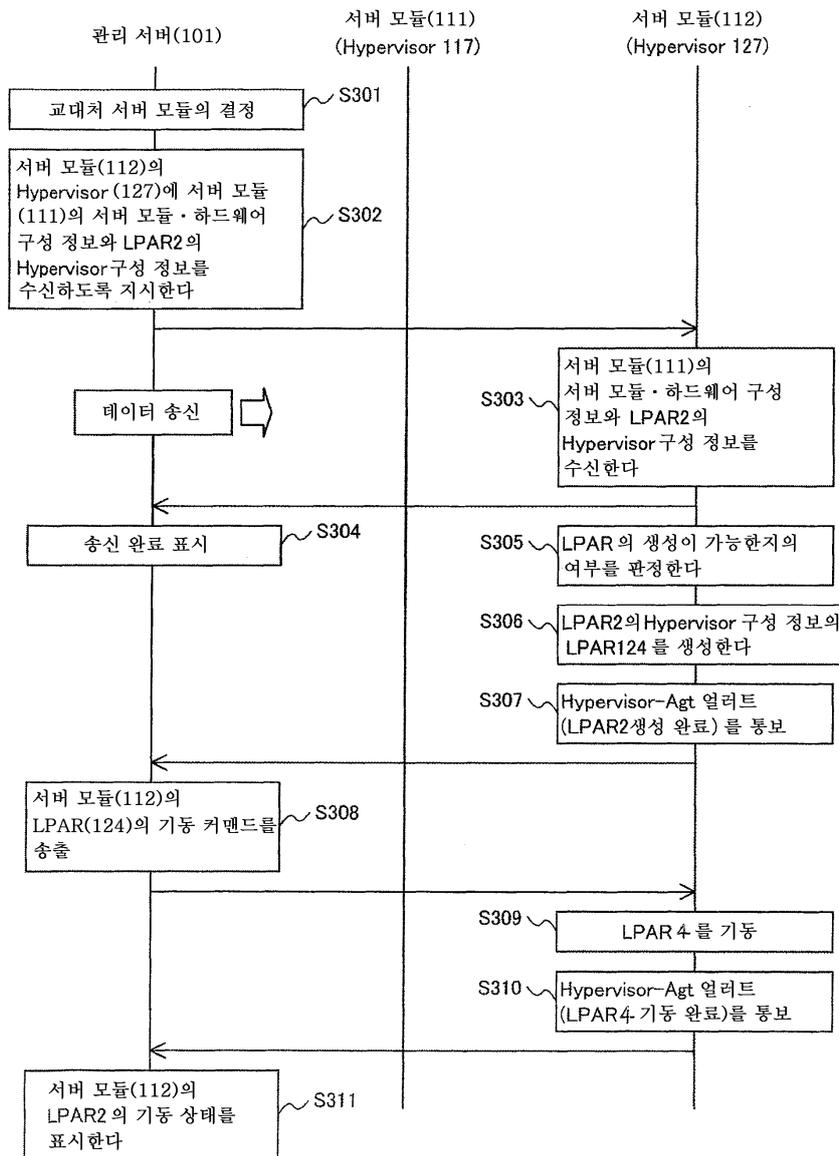
도면1



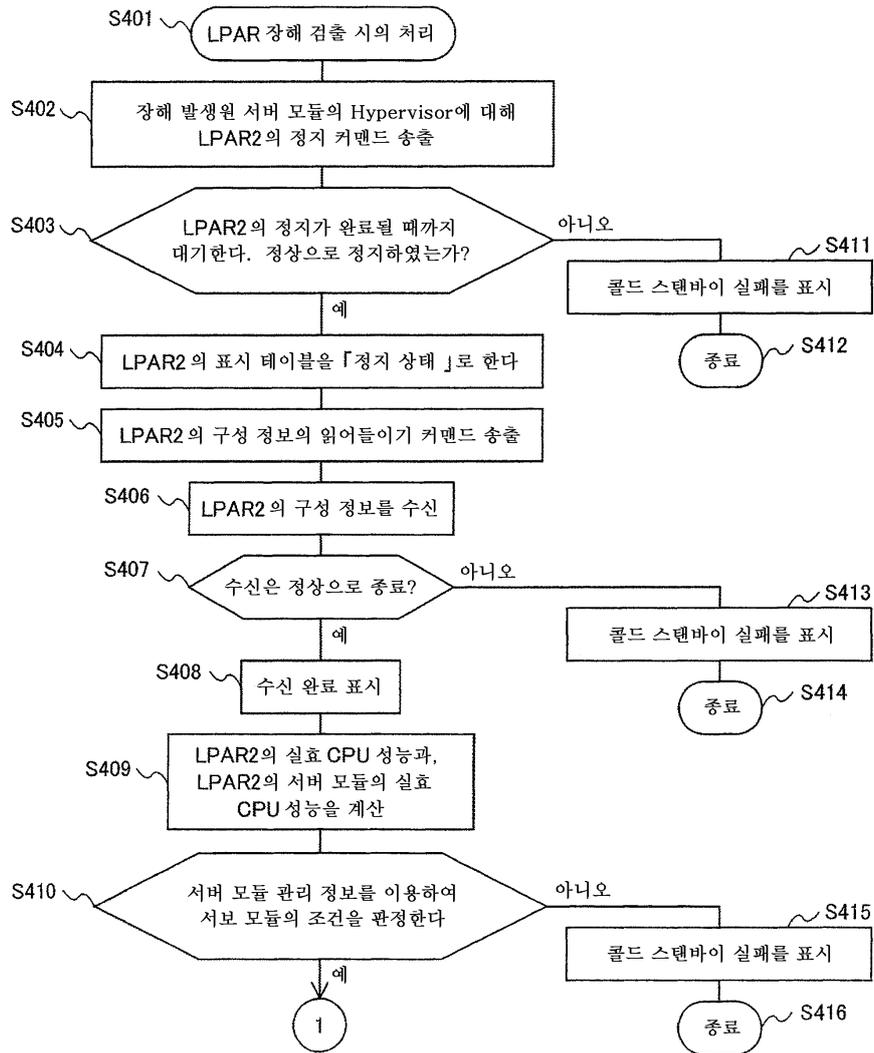
도면2



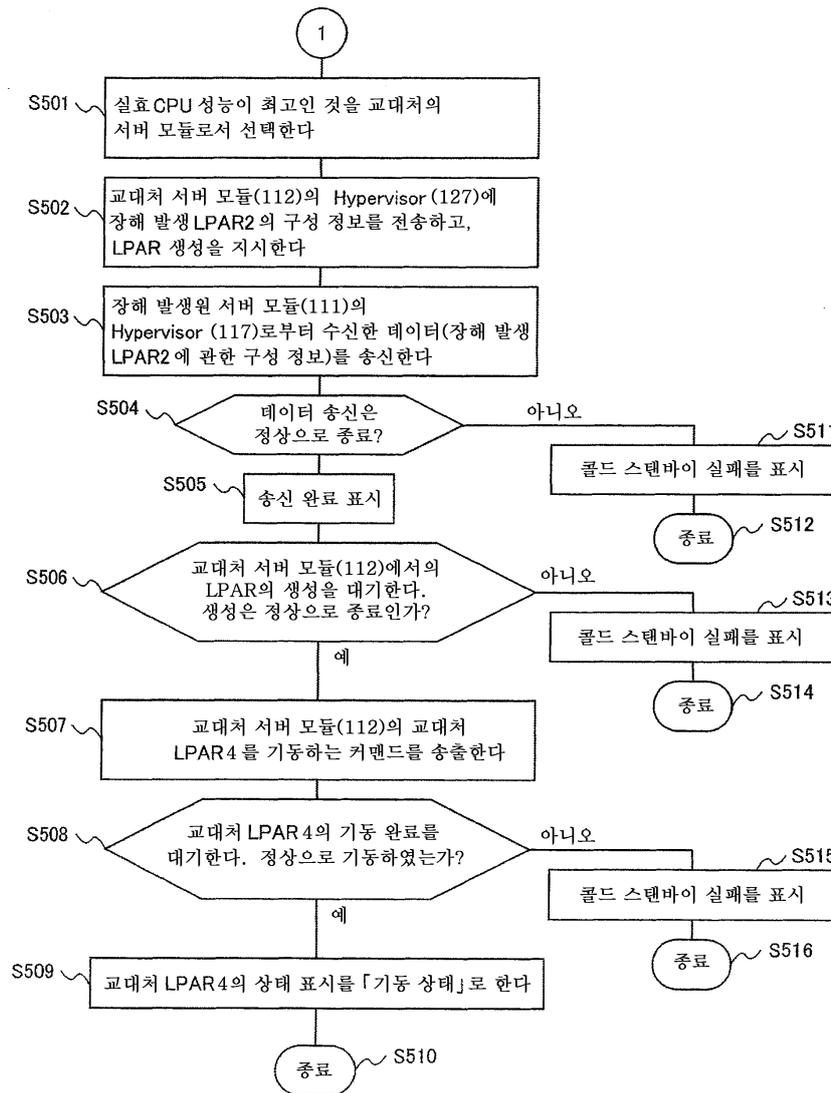
도면3



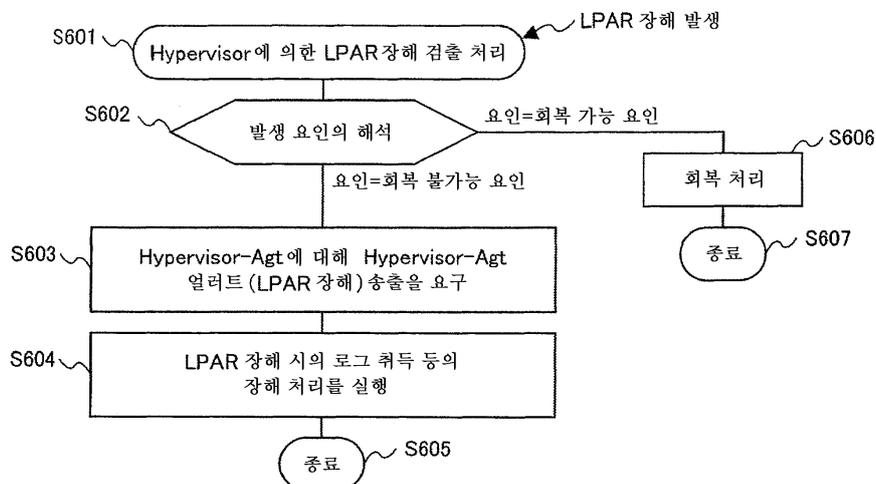
도면4



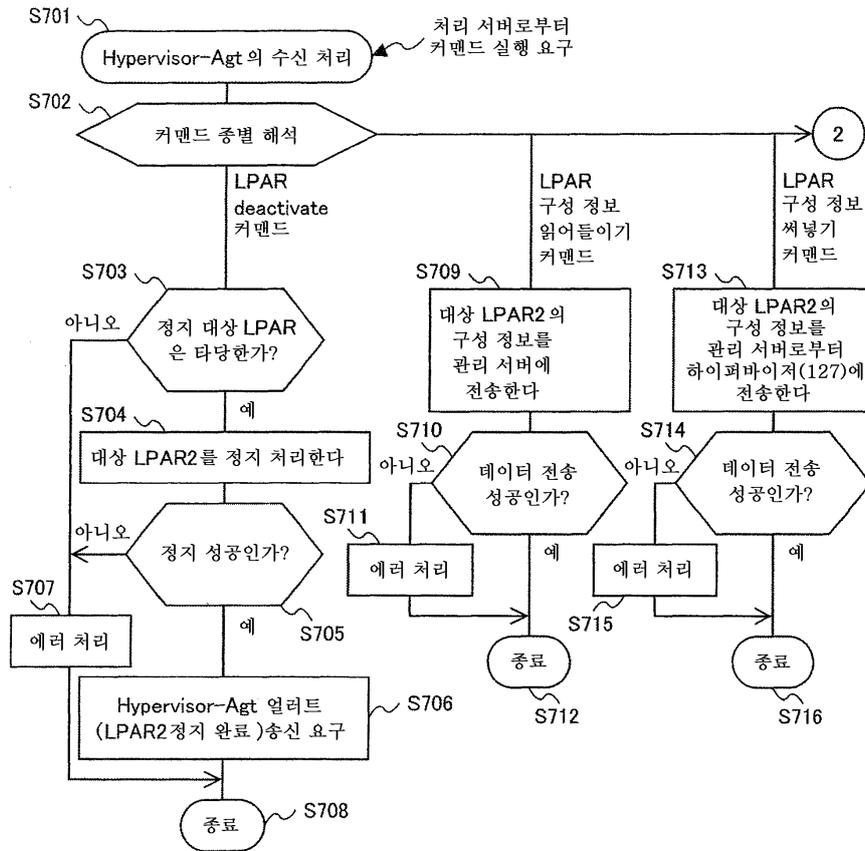
도면5



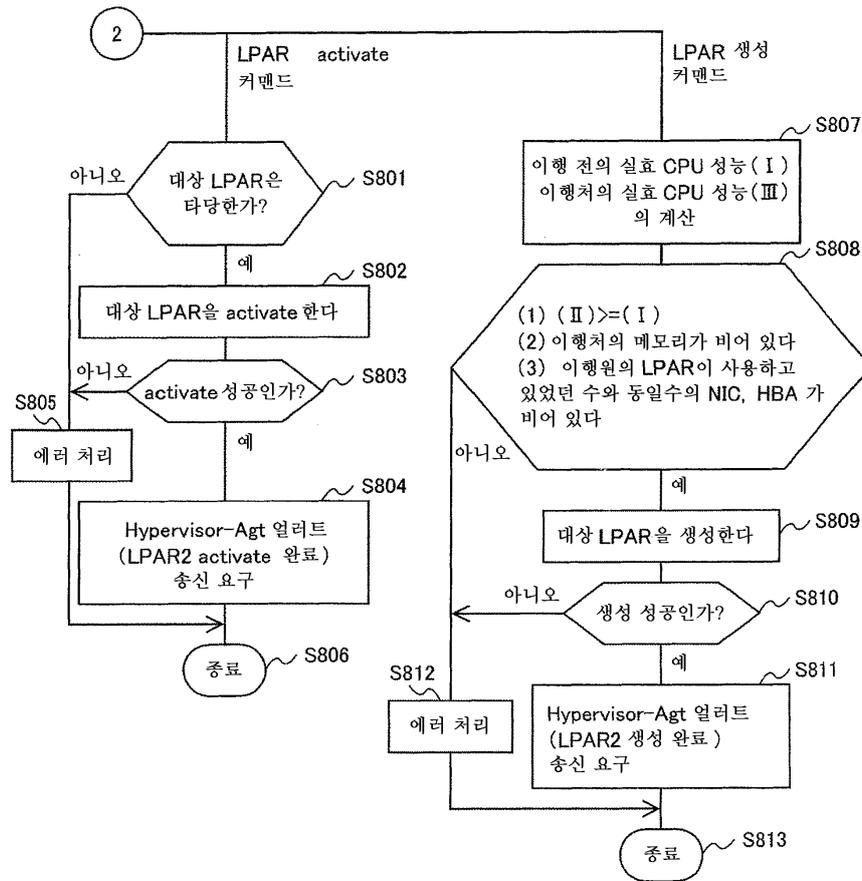
도면6



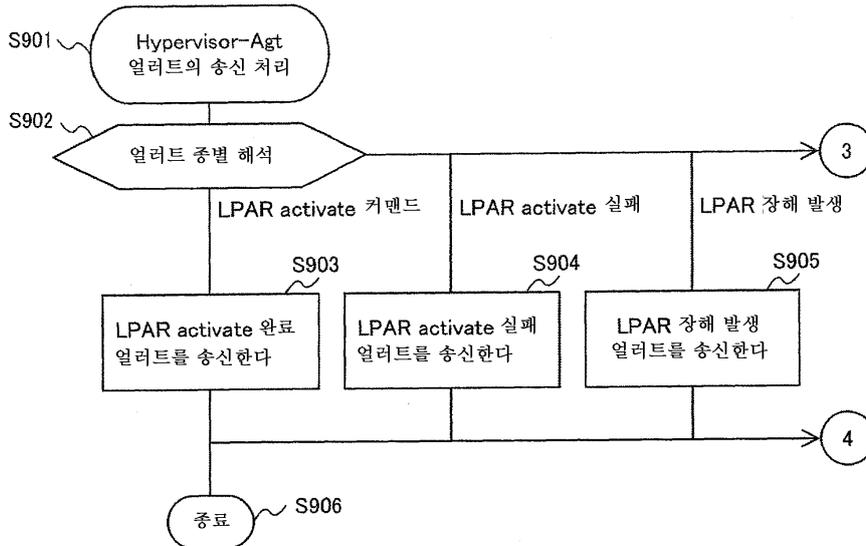
도면7



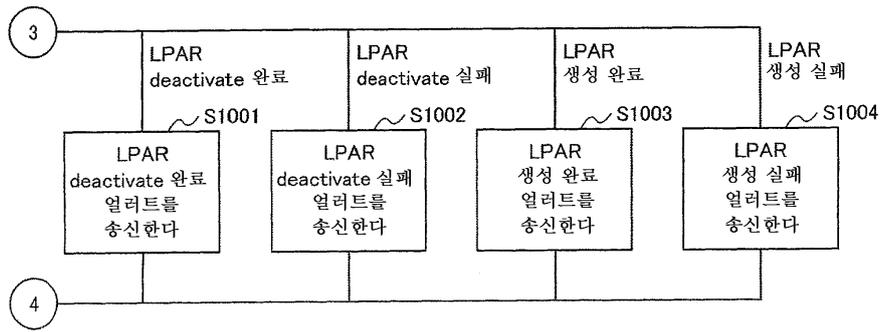
도면8



도면9

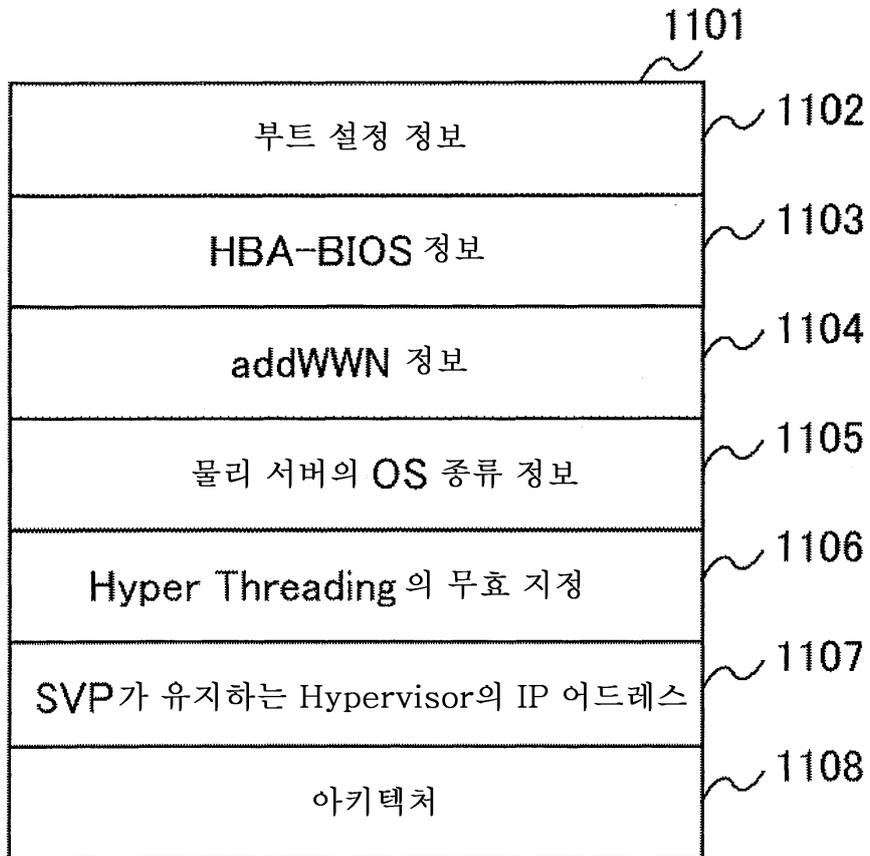


도면10

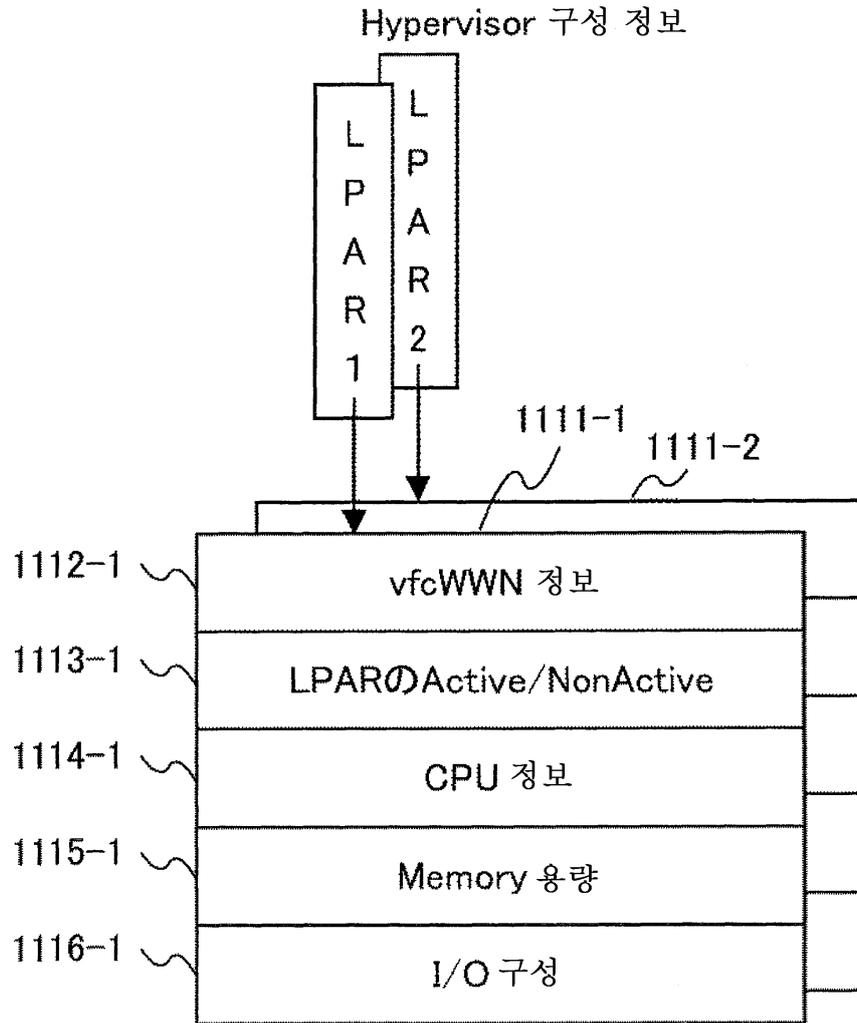


도면11

하드웨어 구성 정보



도면12



도면13

1201	1202	1203	1204	1205	1206	1207	1208	1209	1210
서버 모듈 번호	아키텍처 종별	실장 메모리 용량	가동 중 LPAR의 합계 메모리 사용량	빈 메모리 용량	실장 CPU 성능	할당 완료 CPU 성능의 합계	빈 CPU 성능	빈 NIC수	빈 HBA수

107