



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2013-0049720
(43) 공개일자 2013년05월14일

- (51) 국제특허분류(Int. Cl.)
G06F 13/14 (2006.01)
- (21) 출원번호 10-2012-0115293
(22) 출원일자 2012년10월17일
심사청구일자 없음
- (30) 우선권주장
13/289,617 2011년11월04일 미국(US)
- (71) 출원인
엘에스아이 코퍼레이션
미국 캘리포니아주 95131, 새너제이, 라이더 파크 드라이브 1320
- (72) 발명자
바차브트치크 루이즈 디
미국 켄사스주 67220 위치타 이스트 팔콘 코트 4609
언레인 제이슨 에이
미국 켄사스주 67220 위치타 이스트 팔콘 코트 4617
카우프맨 레이드 에이
미국 켄사스주 67002 앤도버 웨스트 2번 스트리트 610
- (74) 대리인
제일특허법인

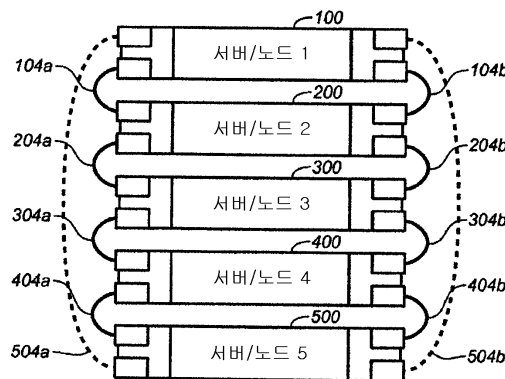
전체 청구항 수 : 총 19 항

(54) 발명의 명칭 **가상 SAS 익스팬더를 통해 공유된 서버 직접 연결 저장장치**

(57) 요약

데이터 저장 시스템은 제 1서버로서, 데이터를 저장하도록 구성되는 제 1 복수의 저장 디스크들과 1 가상 익스팬더 및 제 1 로직 컴포넌트를 제공하도록 구성되는 제 1 프로세서를 포함하는 제 1 호스트 버스 어댑터를 포함하는, 제 1 서버와; 제 2 서버로서, 데이터를 저장하도록 구성되는 제 2 복수의 저장 디스크들과, 제 2 가상 익스팬더 및 제 2 로직 컴포넌트를 제공하도록 구성되는 제 2 프로세서를 포함하는 제 2 호스트 버스 어댑터를 포함하는, 제 2 서버를 포함하고, 상기 제 1 서버의 상기 제 1 호스트 버스 어댑터는 SAS 접속부를 통하여 상기 제 2 서버의 상기 제 2 호스트 버스 어댑터에 결합되고, 상기 제 1 복수의 저장 디스크들 및 상기 제 2 복수의 저장 디스크들 각각은 상기 제 1 서버 및 상기 제 2 서버 각각에 의해 액세스 가능하다.

대표도 - 도3a



특허청구의 범위

청구항 1

제 1 서버와 제 2 서버를 포함하는 데이터 저장 시스템으로서,

상기 제 1 서버는

데이터를 저장하도록 구성되는 제 1 복수의 저장 디스크들과,

제 1 가상 익스팬더(expander) 및 제 1 로직 컴포넌트를 제공하도록 구성되는 제 1 프로세서를 포함하는 제 1 호스트 버스 어댑터를 포함하며,

상기 제 2 서버는

데이터를 저장하도록 구성되는 제 2 복수의 저장 디스크들과,

제 2 가상 익스팬더 및 제 2 로직 컴포넌트를 제공하도록 구성되는 제 2 프로세서를 포함하는 제 2 호스트 버스 어댑터를 포함하며,

상기 제 1 서버의 상기 제 1 호스트 버스 어댑터는 직렬 연결 소형 컴퓨터 시스템 인터페이스(SAS) 접속부를 통하여 상기 제 2 서버의 상기 제 2 호스트 버스 어댑터에 결합되고, 상기 제 1 복수의 저장 디스크들 및 상기 제 2 복수의 저장 디스크들 각각은 상기 제 1 서버 및 상기 제 2 서버 각각에 의해 액세스 가능한

데이터 저장 시스템.

청구항 2

제1항에 있어서,

상기 SAS 접속부는 SAS 케이블인

데이터 저장 시스템.

청구항 3

제1항에 있어서,

상기 제 1 서버의 상기 제 1 호스트 버스 어댑터의 상기 제 1 가상 익스팬더는 상기 SAS 접속부를 통하여 상기 제 2 서버의 상기 제 2 호스트 버스 어댑터의 상기 제 2 가상 익스팬더에 결합되는

데이터 저장 시스템.

청구항 4

제1항에 있어서,

상기 제 1 서버 및 상기 제 2 서버 각각은 제 2(secondary) 호스트 버스 어댑터를 더 포함하는

데이터 저장 시스템.

청구항 5

제4항에 있어서,

제 1 서버의 상기 제 2 호스트 버스 어댑터는 상기 제 2 서버의 상기 제 2 호스트 버스 어댑터에 결합되는

데이터 저장 시스템.

청구항 6

제4항에 있어서,

상기 제 1 서버의 상기 제 1 호스트 버스 어댑터 및 상기 제 2 호스트 버스 어댑터 각각에 결합되는 버스를 더 포함하는

데이터 저장 시스템.

청구항 7

제1항에 있어서,

제 3 서버를 더 포함하고,

상기 제 3 서버는,

데이터를 저장하도록 구성되는 제 3 복수의 저장 디스크들과,

제 3 가상 익스팬더 및 제 3 로직 컴포넌트를 제공하도록 구성되는 제 3 프로세서를 포함하는 제 3 호스트 버스 어댑터를 포함하고,

상기 제 1 서버는 상기 제 3 서버에 결합되고, 상기 제 2 서버는 상기 제 3 서버에 결합되는

데이터 저장 시스템.

청구항 8

제7항에 있어서,

상기 제 1 서버와 상기 제 2 서버, 상기 제 1 서버와 상기 제 3 서버, 또는 상기 제 2 서버와 상기 제 3 서버 사이에 페일오버 접속부(failover connection)를 포함하는

데이터 저장 시스템.

청구항 9

제1항에 있어서,

상기 제 1 복수의 저장 디스크들 및 상기 제 2 복수의 저장 디스크들은 RAID(redundant array of independent disks) 구성으로 구성되는

데이터 저장 시스템.

청구항 10

제 1 서버와 제 2 서버를 포함하는 데이터 저장 시스템으로서,

상기 제 1 서버는

데이터를 저장하도록 구성되는 제 1 복수의 저장 디스크들과,

1개의 코어가 제 1 가상 익스팬더를 제공하도록 구성되는 제 1 멀티-코어 프로세서를 포함하는 제 1 호스트 버스 어댑터를 포함하며,

상기 제 2 서버는

데이터를 저장하도록 구성되는 제 2 복수의 저장 디스크들과,

1개의 코어가 제 2 가상 익스팬더를 제공하도록 구성되는 제 2 멀티-코어 프로세서를 포함하는 제 2 호스트 버스 어댑터를 포함하며,

상기 제 1 서버의 상기 제 1 호스트 버스 어댑터는 직렬 연결 소형 컴퓨터 시스템 인터페이스(SAS) 접속부를 통하여 상기 제 2 서버의 상기 제 2 호스트 버스 어댑터에 결합되고, 상기 제 1 복수의 저장 디스크들 및 상기 제 2 복수의 저장 디스크들 각각은 상기 제 1 서버 및 상기 제 2 서버 각각에 의해 액세스 가능한

데이터 저장 시스템.

청구항 11

제10항에 있어서,

상기 SAS 접속부는 SAS 케이블인

데이터 저장 시스템.

청구항 12

제10항에 있어서,

상기 제 1 서버의 상기 제 1 호스트 버스 어댑터의 상기 제 1 가상 익스팬더는 상기 SAS 접속부를 통하여 상기 제 2 서버의 상기 제 2 호스트 버스 어댑터의 상기 제 2 가상 익스팬더에 결합되는

데이터 저장 시스템.

청구항 13

제10항에 있어서,

상기 제 1 서버 및 상기 제 2 서버 각각은 제 2 호스트 버스 어댑터를 더 포함하는

데이터 저장 시스템.

청구항 14

제13항에 있어서,

제 1 서버의 상기 제 2 호스트 버스 어댑터는 상기 제 2 서버의 상기 제 2 호스트 버스 어댑터에 결합되는

데이터 저장 시스템.

청구항 15

제13항에 있어서,

상기 제 1 서버의 상기 제 1 호스트 버스 어댑터 및 상기 제 1 서버의 상기 제 2 호스트 버스 어댑터 각각에 결합되는 버스를 더 포함하는

데이터 저장 시스템.

청구항 16

제10항에 있어서,
제3 서버를 더 포함하고,
상기 제3 서버는,
데이터를 저장하도록 구성되는 제 3 복수의 저장 디스크들과,
1개의 코어가 제 3 가상 익스팬더를 제공하도록 구성되는 제 3 멀티-코어 프로세서를 포함하는 제 3 호스트 버스 어댑터를 포함하고,
상기 제 1 서버는 상기 제 3 서버에 결합되고, 상기 제 2 서버는 상기 제 3 서버에 결합되는
데이터 저장 시스템.

청구항 17

제16항에 있어서,
상기 제 1 서버와 상기 제 2 서버, 상기 1 서버와 상기 제 3 서버, 또는 상기 제 2 서버와 상기 제 3 서버 사이에 페일오버 접속부를 포함하는
데이터 저장 시스템.

청구항 18

제10항에 있어서,
상기 제 1 복수의 저장 디스크들 및 상기 제 2 복수의 저장 디스크들은 RAID(redundant array of independent disks) 구성으로 구성되는
데이터 저장 시스템.

청구항 19

적어도 4개의 서버들을 포함하고, 상기 적어도 4개의 서버들 각각은:
데이터를 저장하도록 구성되는 복수의 저장 디스크들과,
제 1 가상 익스팬더를 제공하도록 구성되는 제 1 프로세서를 포함하는 제 1 호스트 버스 어댑터와,
제 2 가상 익스팬더를 제공하도록 구성되는 제 2 프로세서를 포함하는 제 2 호스트 버스 어댑터를 포함
하고,
상기 적어도 4개의 서버들 각각은 상기 적어도 4개의 서버들 중 하나의 상기 제 1 가상 익스팬더를 상기 적어도 4개의 서버들 중 2개의 다른 서버들의 상이한 제 1 가상 익스팬더에 접속시키는 제 1 접속 구성을 포함하고, 상기 적어도 4개의 서버들 각각은 상기 적어도 4개의 서버들 중 하나의 상기 제 2 가상 익스팬더를 상기 적어도 4개의 서버들 중 2개의 다른 서버들의 상이한 제 2 가상 익스팬더에 접속시키는 제 2 접속 구성을 포함하며, 상기 적어도 4개의 서버들 중 적어도 하나의 서버의 상기 제 1 접속 구성은 상기 제 1 접속 구성 및 상기 제 2 접속 구성이 어느 서버들에 관련되는지에 따라 상기 적어도 4개의 서버들 중 상기 적어도 하나의 서버의 상기 제 2 접속 구성과 상이한
데이터 저장 시스템.

명세서

기술 분야

- [0001] 본 개시물은 일반적으로 데이터 저장 시스템 분야에 관한 것이며, 더 상세하게는, SAS(serial attached SCSI, 직렬 연결 소형 컴퓨터 시스템 인터페이스(small computer system interface: SCSI) 익스팬더(expander)들을 통해 공유되는 서버 직접 연결 저장장치에 관한 것이다.

배경 기술

- [0002] 클라우드 컴퓨팅(cloud computing)은 최소의 관리 노력 또는 서비스 제공자 인터랙션(service provide interaction)으로 고속으로 제공 및 해제될 수 있는 구성 가능한 컴퓨팅 자원들(예를 들어, 네트워크(network)들, 서버(server)들, 저장장치, 애플리케이션(application)들, 및 서비스(service)들)의 공유된 풀(pool)로의 온-디맨드 네트워크 액세스(on-demand network access) 인에이블(enable)하기 위한 모델(model)을 제공함으로써 부상하고 있다. 클라우드 컴퓨팅은 일반적으로 리던던시(redundancy)를 위한 클러스터링(clustering)을 이용하는데, 이는 여기서 4개가 제공되고 다음의 문제가 되는 특징들을 각각 포함하는 다양한 저장장치 구성들을 통해 성취될 수 있다:
- [0003] (1) 각 노드는 저장장치에 저 레이턴시 블록 인터페이스(lower latency block interface)를 제공할 수 있는 공통 저장장치 영역 네트워크(storage area network: SNA) 패브릭(fabric)에 접속될 수 있다; (2) 각 노드는 공유된 저장장치로의 파일 액세스(file access)를 이용할 수 있는 이더넷 네트워크(Ethernet network)에 접속될 수 있다; (3) 외부 JBOD("just a bunch of disk")들; 및 (4) 직접 연결 드라이브들(내부).
- [0004] 구성들 (1) 및 (2)는 클러스터(cluster)를 형성하기 위한 노드(node)를 공통 저장장치에 접속하기 위하여 파이버 또는 이더넷 스위치(Fiber or Ethernet switch)와 같은 추가적인 외부 소자를 필요로 할 수 있다. 이와 같은 외부 소자는 단일 고장점(single point of failure)을 제공하기 때문에 바람직하지 않을 수 있다. 결과적으로, 상기 구성들의 높은 유용성을 제공하기 위하여 리던던시 컴포넌트(redundancy component)가 필요할 수 있는데, 이는 시스템들에 추가적인 비용을 발생시킬 것이다.
- [0005] 구성 (3)은 일반적으로 비용 효율적이지만, 클러스터 내의 노드들의 양을 JBOD 상의 커넥터(connector)들의 수로 제한하는데, 이는 상당히 제한적일 수 있고 확장성을 제한할 수 있다. 더구나, 구성들 (1) 내지 (3)은 일반적으로 외장 인클로저(external enclosure) 내에 위치될 저장 시스템을 필요로 하는데, 이는 추가적인 전력, 공간, 및 유지보수 비용들을 부과한다.
- [0006] 구성 (4)는 일반적으로 경제적이지만, 연결된 드라이브들에 대해 공유된 저장장치가 존재하지 않기 때문에, 고가용성 클러스터링을 제공하지 못한다. 이와 같이, 이러한 구성들은 문제가 되는 비용들 및 복잡성을 제공하며, 고가용성 클러스터링을 위한 저장장치 요건들(예를 들어, 리던던시 및 공통 액세스)에 대한 바람직한 솔루션(solution)들을 제공하지 못한다.

발명의 내용

과제의 해결 수단

- [0007] 본 개시물의 하나의 구현예에서, 데이터 저장 시스템은 제 1 서버로서, 데이터를 저장하도록 구성되는 제 1 복수의 저장 디스크들과, 1 가상 익스팬더(expander) 및 제 1 로직 컴포넌트를 제공하도록 구성되는 제 1 프로세서를 포함하는 제 1 호스트 버스 어댑터를 포함하는, 상기 제 1 서버와; 제 2 서버로서, 데이터를 저장하도록 구성되는 제 2 복수의 저장 디스크들과, 제 2 가상 익스팬더 및 제 2 로직 컴포넌트를 제공하도록 구성되는 제 2 프로세서를 포함하는 제 2 호스트 버스 어댑터를 포함하는, 상기 제 2 서버를 포함하고, 상기 제 1 서버의 상기 제 1 호스트 버스 어댑터는 SAS 접속부를 통하여 상기 제 2 서버의 상기 제 2 호스트 버스 어댑터에 결합되고, 상기 제 1 복수의 저장 디스크들 및 상기 제 2 복수의 저장 디스크들 각각은 상기 제 1 서버 및 상기 제 2 서버 각각에 의해 액세스 가능하다.
- [0008] 본 개시물의 또 다른 구현예에서, 데이터 저장 시스템은 제 1 서버로서, 데이터를 저장하도록 구성되는 제 1 복수의 저장 디스크들과, 1개의 코어가 제 1 가상 익스팬더를 제공하도록 구성되는 제 1 멀티-코어 프로세서를 포함하는 제 1 호스트 버스 어댑터를 포함하는, 상기 제 1 서버와; 제 2 서버로서, 데이터를 저장하도록 구성되는

제 2 복수의 저장 디스크들과, 1개의 코어가 제 2 가상 익스팬더를 제공하도록 구성되는 제 2 멀티-코어 프로세서를 포함하는 제 2 호스트 버스 어댑터를 포함하는, 상기 제 2 서버를 포함하고, 상기 제 1 서버의 상기 제 1 호스트 버스 어댑터는 SAS 접속부를 통하여 상기 제 2 서버의 상기 제 2 호스트 버스 어댑터에 결합되고, 상기 제 1 복수의 저장 디스크들 및 상기 제 2 복수의 저장 디스크들 각각은 상기 제 1 서버 및 상기 제 2 서버 각각에 의해 액세스 가능하다.

- [0009] 본 개시물의 부가적인 구현예에서, 데이터 저장 시스템은 적어도 4개의 서버들을 포함하고, 상기 적어도 4개의 서버들 각각은: 데이터를 저장하도록 구성되는 복수의 저장 디스크들과, 제 1 가상 익스팬더를 제공하도록 구성되는 제 1 프로세서를 포함하는 제 1 호스트 버스 어댑터와, 제 2 가상 익스팬더를 제공하도록 구성되는 제 2 프로세서를 포함하는 제 2 호스트 버스 어댑터를 포함하고, 상기 적어도 4개의 서버들 각각은 상기 적어도 4개의 서버들 중 하나의 상기 제 1 가상 익스팬더를 상기 적어도 4개의 서버들 중 2개의 다른 서버들의 상이한 제 1 가상 익스팬더에 접속시키는 제 1 접속 구성을 포함하고, 상기 적어도 4개의 서버들 각각은 상기 적어도 4개의 서버들 중 하나의 상기 제 2 가상 익스팬더를 상기 적어도 4개의 서버들 중 2개의 다른 서버들의 상이한 제 2 가상 익스팬더에 접속시키는 제 2 접속 구성을 포함하며, 상기 적어도 4개의 서버들 중 적어도 하나의 서버의 상기 제 1 접속 구성은 상기 제 1 접속 구성 및 상기 제 2 접속 구성이 어느 서버들에 관련되는지에 따라 상기 적어도 4개의 서버들 중 상기 적어도 하나의 서버의 상기 제 2 접속 구성과 상이하다.
- [0010] 상술된 일반적인 설명 및 다음의 상세한 설명 둘 모두가 단지 예시적이고 설명을 위한 것이며, 반드시 청구항 바와 같이 본 개시물을 제한하는 것은 아니라는 점이 이해되어야 한다. 본 개시물에 포함되고 본 개시물의 일부를 구성하는 첨부 도면들은 본 개시물의 실시예를 도시하고, 일반적인 설명과 함께 본 명세서의 원리들을 설명하는 역할을 한다.

도면의 간단한 설명

- [0011] 본 개시물의 다수의 장점들은 첨부 도면을 참조하여 당업자들에 의해 더 양호하게 이해될 수 있다:

도 1은 서버의 내부 레이아웃(internal layout)의 개략도.
 도 2는 호스트 버스 어댑터(host bus adaptor)의 개략도.
 도 3a는 캐스케이드(cascade)된 직접-연결 저장장치(DAS) 클러스터에 대한 구성의 개략도.
 도 3b는 캐스케이드된 DAS 클러스터의 또 다른 구성의 개략도.
 도 4는 도 3a의 캐스케이드된 DAS의 일부의 개략도.
 도 5는 캐스케이드된 DAS 클러스터의 일 구현예의 개략도.

발명을 실시하기 위한 구체적인 내용

- [0012] 예들이 첨부 도면들에 도시되어 있는 본 개시물의 바람직한 실시예들이 상세히 참조될 것이다.
- [0013] 본 개시물은 서버들이 스위치 또는 외부 저장장치인 외부 소자를 사용하지 않고 공통으로 저장장치를 공유하는 노드들(예를 들어, 서버들)의 클러스터에 속하는 구현예들을 제공한다. 일반적으로, SAS 기술은 각 노드에서의 직접 연결 디스크들, 각 노드 간의 접속부들과 함께 사용됨으로써, 캐스케이드된 SAS 토폴로지(cascaded SAS topology)를 통해 SAN 환경을 에뮬레이팅(emulating)한다. 현대의 컴퓨팅 서버들은 SAS를 통해 임베딩(embedding)된 디스크들을 포함할 수 있는데, 여기서 1개의 서버의 내부 저장장치는 다른 접속된 서버들 사이에서 공유될 수 있다. 내부 저장장치가 공유되는 경우에, 중요 데이터 액세스를 위한 외부 저장장치는 필요하지 않을 수 있다. 익스팬더들을 에뮬레이팅할 수 있는 SAS 호스트 버스 어댑터(host bus adapter: HBA)들이 모든 다른 노드들 및 대응하는 연결 디스크들로의 양방향성 트래픽(bi-directional traffic)을 인에이블하는데 사용될 수 있다.
- [0014] 도 1은 노드들의 클러스터들 내로 통합될 수 있는 서버(100)의 개략도를 도시한다. 서버(100)는 1개 이상의 HBA들(예를 들어, SAS HBA들)을 포함할 수 있는데, 도 1에서는 2개의 HBA들(102a 및 102b)이 도시되어 있다. 도 2는 HBA(102a)의 컴포넌트들의 개략도를 도시한다. 도시된 바와 같이, HAS(102a)는 총 16개의 phys에 대하여 한 쌍의 4개의 외부 커넥터들(104a 및 104b) 및 한 쌍의 4개의 내부 커넥터들(106a 및 106b)을 포함한다. HBA(102)는 또한 HBA(102a)의 동작을 관리하기 위한 2-코어 CPU(2-core CPU)(108)와 같은 프로세서를

포함한다. 도 1에 도시된 바와 같이, 한 쌍의 4개의 내부 커넥터들(106a 및 106b)은 HBA(102a)를 서버(100) 상에서 저장장치로서 이용 가능한 복수의 디스크들(110)에 접속시킨다. 유사하게, HBA(102b)들은 상기 HBA(102b)를 서버(100) 상에서 복수의 디스크들(110)에 접속시키는 커넥터들(112a 및 112b)를 포함한다.

[0015] HBA 102a 및 HBA 102b의 외부 커넥터들(예를 들어, 104a 및 104b)은 상기 서버(100)를 클러스터의 부분인 다른 서버들에 접속시키는 역할을 한다. 각 서버는 클러스터 내의 다른 서버들로 접속될 적어도 1개의 HBA를 포함하는데, 여기서 서버당 1개 이상의 HBA는 리던던시를 허용한다. 예를 들어, 각 서버/노드는 리던던시를 위해 (각 서버/노드의 HBA들을 통한) 2개의 다른 노드들로의 SAS 접속부들을 포함할 수 있다. 도 3a에 도시된 바와 같이, 캐스케이드된 DAS 클러스터에 대한 구성의 개략도가 제공된다. 상기 구성은 5개의 서버들/노드들(100, 200, 300, 400, 및 500)를 포함하는데, 여기서 서버/노드(100)는 최초 노드라고 특징지을 수 있고, 서버/노드(500)는 최종 노드라고 특징지을 수 있다. 서버/노드(100)는 커넥터들(104a 및 104b)를 통해 서버/노드(200)에 링크(link)된다. 서버/노드(200)는 커넥터들(204a 및 204b)를 통해 서버/노드(300)에 링크된다. 서버/노드(300)는 커넥터들(304a 및 304b)를 통해 서버/노드(400)에 링크된다. 서버/노드(400)는 커넥터들(404a 및 404b)를 통해 서버/노드(500)에 링크된다. 최초 노드 및 최종 노드는 서로 접속될 수도 있지만, 상기 접속부는 루프(loop)(예를 들어, 무효한 SAS 토폴로지)를 방지하기 위하여 디스에이블(disable)될 수 있다. 도 3a에 도시된 바와 같이, 서버/노드(100)는 클러스터 내의 노드가 이용 불가능해질 때까지 디스에이블된 상태인 커넥터들(504a 및 504b)를 통해 서버/노드(500)에 접속된다. 노드 또는 접속부가 더 이상 작동하지 않는 경우에(예를 들어, 노드 고장), 최초 노드 및 최종 노드 간의 디스에이블된 접속부들(예를 들어, 커넥터들(504a 및 504b))은 모든 이용 가능한 노드들로의 중단없는 액세스를 보증하기 위하여 펌웨어(firmware)에 의해 즉시 인에이블될 수 있다. 시스템의 각 서버/노드는 복수의 디스크들(110)과 같이, 모든 노드들에 액세스 가능한 로컬 SAS(또는 직렬 AT 어태치먼트(serial AT attachment: SATA)) 저장장치를 포함할 수 있다. 본원에 제공된 바와 같이, 각 노드는 2개의 다른 노드들로의 리던던트 접속부(redundant connection)들을 포함할 수 있는데, 즉, 모든 엔드 디바이스(end device)들로의 이중 경로들이 리던던시를 위해 사용될 수 있지만, 상기 리던던트 접속부들은 본 개시물의 모든 구현예들에서 필요하지는 않을 수 있다.

[0016] 도 3b를 참조하면, 클러스터가 2개의 상이한 케이블링 패턴(cabling pattern)들을 포함하는 캐스케이드된 DAS 클러스터의 또 다른 구성이 도시되어 있다. 예를 들어, 커넥터들(104a, 204a, 304a, 404a, 및 504a)의 구성은 도 3a를 참조하여 설명된 바와 같은 구성과 동일한 반면, 도 3b에서의 커넥터들(104b, 204b, 304b, 404b, 및 504b)의 구성은 도 3a에서의 커넥터들(104b, 204b, 304b, 404b, 및 504b)의 구성과 상이하다. 도 3b에 도시된 구성에서 상이한 케이블링 패턴들을 포함함으로써, 상기 케이블링 패턴들은 서버/노드의 각 HBA가 동일한 서버/노드에 접속되는 경우보다 레이턴시를 감소시킬 수 있고 시스템/드라이브 가용성을 증가시킬 수 있다. 도 3b의 커넥터들(104b 및 504a)은 클러스터의 각 서버/노드가 동작하고 있을 때 디스에이블되지만, 클러스터 내의 노드 또는 접속부가 더 이상 동작하고 있지 않을 때(예를 들어, 노드 고장) 활성화되는 페일오버 접속부(failover connection)들일 수 있다. 펌웨어는 클러스터 내의 모든 이용 가능한 노드들로의 중단없는 액세스를 제공하기 위하여 클러스터 내의 노드들 또는 접속부들이 더 이상 동작하고 있지 않을 때 커넥터(104b 및 504a)를 즉시 활성화시킬 수 있다.

[0017] 도 4를 참조하면, 도 3a의 캐스케이드된 DAS의 일부의 개략도가 도시되어 있다. 도 4에 도시된 바와 같이, 각 서버/노드의 각 HBA는 2개의 메인 컴포넌트들을 포함할 수 있다: (1) HBA의 동작 및 시스템(100) 상의 다수의 HBA들 간의 통신을 제공하기 위한 PCI(peripheral component interconnect) 로직 및 HBA 로직 및 (2) 드라이브들 및 HBA 로직 컴포넌트 간 및 HBA 로직 컴포넌트 및 외부 phys 간의 트래픽의 라우팅(routing)을 핸들링(handling)하기 위한 가상 익스팬더. 예를 들어, 서버/노드(100)의 HBA(102a)는 PCI/HBA 로직 컴포넌트(114a) 및 가상 익스팬더(116a)를 포함하는 반면, 서버/노드(100)의 HBA(102b)는 PCI/HBA 로직 컴포넌트(114b) 및 가상 익스팬더(116b)를 포함한다. 커넥터들(106a 및 106b)은 복수의 드라이브들(110)을 HBA(102a)의 가상 익스팬더(116a)에 결합시킬 수 있고, 커넥터들(112a 및 112b)은 상기 복수의 드라이브들(110)을 HBA(102b)의 가상 익스팬더(116b)에 결합시킬 수 있다. 클러스터의 부분인 다른 서버들/노드들에서도 유사한 구성들이 존재할 수 있는데, 예를 들어, 서버/노드(200)의 HBA(202a)는 PCI/HBA 로직 컴포넌트(214a) 및 가상 익스팬더(216a)를 포함하는 반면, 서버/노드(200)의 HBA(202b)는 PCI/HBA 로직 컴포넌트(214b) 및 가상 익스팬더(216b)를 포함하며, 접속들은 복수의 드라이브들(210) 및 가상 익스팬더들(216a 및 216b) 사이에서 이루어진다.

[0018] 각 서버/노드는 또한 상기 서버/노드의 컴포넌트들 사이의 통신을 제공하기 위한 버스를 포함할 수 있다. 예를 들어, 서버/노드(100)는 HBA(102a) 및 HBA(102b) 각각에 결합될 수 있는 PCI 버스(118)를 포함할 수 있는 반면, 서버 노드(200)는 HBA(202a) 및 HBA(202b) 각각에 결합되는 PCI 버스(218)를 포함할 수 있다. 또한, 각

서버/노드는 도 3a 및 3b를 참조하여 설명된 바와 같이, 2개의 다른 서버/노드들에 접속될 수 있다. 각 서버/노드 간의 접속부들은 각 서버/노드 간의 외부 결합을 제공하는 SAS 케이블들(406)과 같은 SAS 커넥터를 포함할 수 있다. 도 4에 도시된 바와 같이, 서버/노드(100)는 클러스터 내의 최종적인 기계(예를 들어, 엔드 노드(end node))와 결합되도록 루프(loop)되는 2개의 외부 SAS 케이블들(406)을 포함한다. 본원에 설명된 바와 같이, SAS 케이블들 중 1개 이상은 패일오버 케이블들의 역할을 함으로써 무효한 SAS 토폴로지들을 방지하기 위하여 디스에이블될 수 있다.

[0019] 도 5를 참조하면, 캐스케이드된 DAS 클러스터의 일 구현예의 개략도가 도시되어 있다. 일반적으로, 도 5의 캐스케이드된 DAS 클러스터의 구현예는 서버들/노드들 간의 접속들을 기반으로 한 도 4의 구현예와 상이하다. 도 5의 구현예는 서버/노드 시스템의 타측 상의 대응하는 외부 phys와 상이한 결합 패턴을 갖는 외부 phys를 서버/노드 시스템의 일측 상에 표시한다. 예를 들어, 도시된 바와 같이, 서버/노드(100)의 HBA(102a)는 커넥터(502a)를 통해 서버/노드(400)의 HBA(402a)에 결합되고 커넥터(504a)를 통해 서버/노드(200)의 HBA(202a)에 결합되는 반면, 서버/노드(100)의 HBA(102b)는 커넥터(502b)를 통해 서버/노드(400)의 HBA(402b)에 결합되고 커넥터(504b)를 통해 서버/노드(300)의 HBA(302b)에 결합되며; 서버/노드(200)의 HBA(202a)는 커넥터(504a)를 통해 서버/노드(100)의 HBA(102a)에 결합되고 커넥터(506a)를 통해 서버/노드(300)의 HBA(302a)에 결합되는 반면, 서버/노드(200)의 HBA(202b)는 커넥터(506b)를 통해 서버/노드(400)의 HBA(402b)에 결합되고 커넥터(508b)를 통해 서버/노드(300)의 HBA(302b)에 결합되며; HBA(302a)는 커넥터(506a)를 통해 서버/노드(200)의 HBA(202a)에 결합되고 커넥터(508a)를 통해 서버/노드(400)의 HBA(402a)에 결합되는 반면, HBA(302b)는 커넥터(508b)를 통해 서버/노드(200)의 HBA(202b)에 결합되고 커넥터(504b)를 통해 서버/노드(100)의 HBA(102b)에 결합된다. 이와 같은 결합 패턴은 서버/노드의 각 HBA가 동일한 서버/노드에 접속되는 경우보다 레이턴시를 감소시킬 수 있고 시스템/드라이브 가용성을 증가시킬 수 있다.

[0020] 커넥터들(502a 및 508b)은 클러스터의 각 서버/노드가 동작하고 있을 때 디스에이블되지만, 클러스터 내의 노드 또는 접속부가 더 이상 동작하고 있지 않을 때(예를 들어, 노드 고장) 활성화되는 패일오버 접속부들일 수 있다. 펌웨어는 클러스터 내의 모든 이용 가능한 노드들로의 중단없는 액세스를 제공하기 위하여 클러스터 내의 노드 또는 접속부가 더 이상 동작하고 있지 않을 때 커넥터(502a 및/또는 508b)를 즉시 활성화시킬 수 있다.

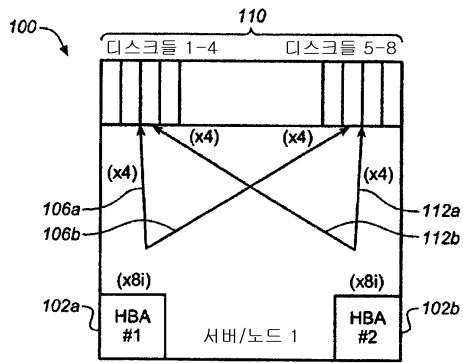
[0021] 데이터 액세스/핸들링을 촉진하기 위하여, 인커밍 IO(incoming IO)(입력/출력)는 도 2에 도시된 2-코어 CPU(108)과 같은 멀티-코어 프로세서를 HBA 상에서 사용하는 효율적인 라우팅 알고리즘(routing algorithm)에 의해 프로세싱될 수 있다. 이와 같은 사용은 HBA의 가상 익스팬더(예를 들어, 가상 익스팬더(116a))의 레이턴시를 감소시킬 수 있다. 예를 들어, HBA가 2-코어 프로세서를 포함하는 경우에, 제 2 코어는 가상 익스팬더에 전용될 수 있다.

[0022] 도 5에 도시된 캐스케이드된 DAS 클러스터 구현예는 RAID(redundant array of independent disks) 동작성을 위해 구성될 수도 있다. 예를 들어, 클러스터의 서버/노드들의 복수의 드라이브들(110, 210, 310, 410)은 드라이브 고장, 시스템 고장, HBA 고장, 또는 케이블 고장 중 하나 이상을 완화함으로써 클러스터에 증가된 가용성을 제공하기 위하여 (도 5에 도시된 것들과 같은) RADI 구성들 내에 배치될 수 있다.

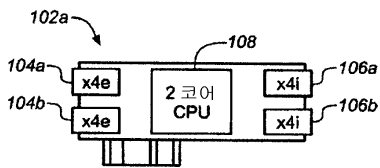
[0023] 본 개시물 및 본 개시물에 수반되는 다수의 장점들이 상술된 설명에 의해 이해될 것이라고 여겨지며, 본 명세서의 사상 및 범위를 벗어나지 않거나 본 명세서의 모든 본질적인 장점들을 희생시키지 않고 컴포넌트들의 형태, 구성, 및 배열에서 다양한 변화들이 행해질 수 있다는 점이 명백할 것이다. 본원에 상술된 형태가 단지 설명을 위한 실시예이기 때문에, 다음의 청구범위들은 이와 같은 변화들을 포함하게 된다.

도면

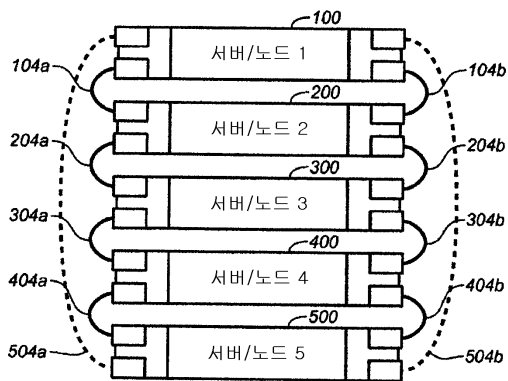
도면1



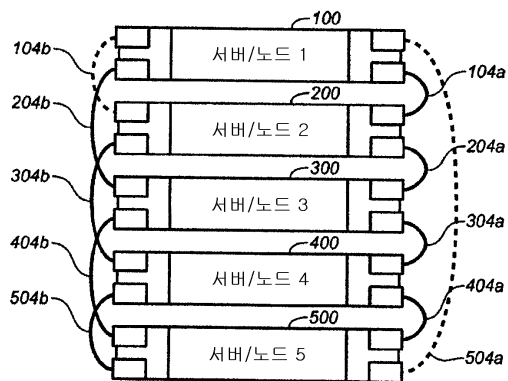
도면2



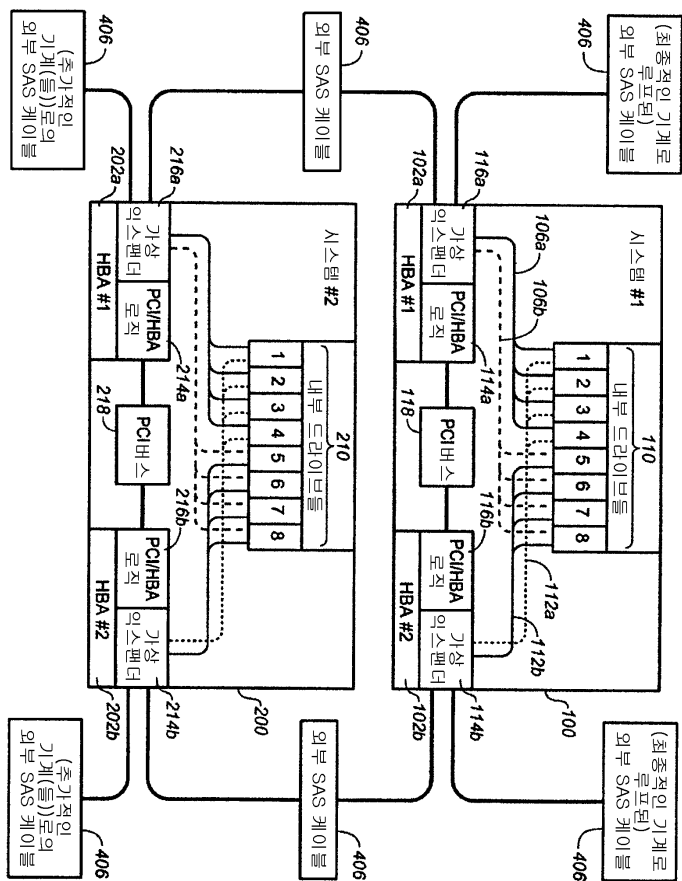
도면3a



도면3b



도면4



도면5

