



(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2018년08월23일
(11) 등록번호 10-1891365
(24) 등록일자 2018년08월17일

(51) 국제특허분류(Int. Cl.)
H04L 12/16 (2006.01) H04L 29/06 (2006.01)
(21) 출원번호 10-2013-7025774
(22) 출원일자(국제) 2012년03월06일
심사청구일자 2017년02월09일
(85) 번역문제출일자 2013년09월30일
(65) 공개번호 10-2014-0016309
(43) 공개일자 2014년02월07일
(86) 국제출원번호 PCT/US2012/027796
(87) 국제공개번호 WO 2012/134729
국제공개일자 2012년10월04일
(30) 우선권주장
13/074,920 2011년03월29일 미국(US)
(56) 선행기술조사문헌
US20100146085 A1*
US07913105 B1*
US07664125 B1*
*는 심사관에 의하여 인용된 문헌

(73) 특허권자
마이크로소프트 테크놀로지 라이선싱, 엘엘씨
미국 워싱턴주 (우편번호 : 98052) 레드몬드 원
마이크로소프트 웨이
(72) 발명자
프라샨스 프라할라드
미국 워싱턴주 98052-6399 레드몬드 원 마이크로
소프트 웨이 엘씨에이 - 인터내셔널 페이턴즈 마
이크로소프트 코포레이션
조지 매튜
미국 워싱턴주 98052-6399 레드몬드 원 마이크로
소프트 웨이 엘씨에이 - 인터내셔널 페이턴즈 마
이크로소프트 코포레이션
(뒷면에 계속)
(74) 대리인
제일특허법인(유)

전체 청구항 수 : 총 20 항

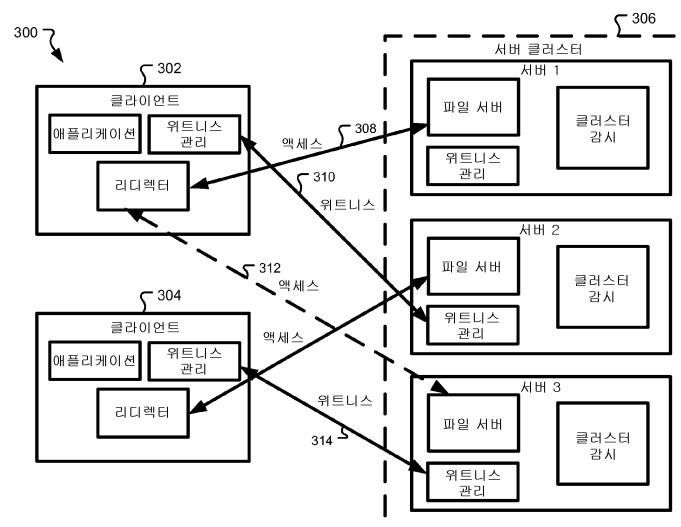
심사관 : 박보미

(54) 발명의 명칭 위트니스 서비스의 제공

(57) 요약

리소스 상태에 대한 통지를 클라이언트에게 송신하는 위트니스 서비스를 제공하는 것에 관한 실시예들이 설명된다. 실시예는 리소스의 상태에 관한 통지를 등록하고 수신하기 위한 다양한 메시지를 포함하는 프로토콜을 제공한다. 프로토콜은 클러스터 내의 제1 노드로부터 노드 정보를 요청하기 위한 메시지를 포함할 수도 있다. 노드 정보는 클러스터 내에서 리소스를 감시하는 위트니스 서비스를 제공하는 노드를 식별한다. 프로토콜은 네트워크나 클러스터 리소스의 상태 또는 상태 변화에 관한 통지를 위한 위트니스 서비스를 등록하는 데 사용되는 메시지를 포함한다. 프로토콜은 리소스의 상태 정보에 대한 통지를 송신하기 위한 메시지도 포함한다.

대표도 - 도3



(72) 발명자

크루즈 데이비드 앤

미국 워싱턴주 98052-6399 레드몬드 원 마이크로소프트 웨이 엘씨에이 - 인터내셔널 패이턴츠 마이크로소프트 코포레이션

펑커턴 제임스 티

미국 워싱턴주 98052-6399 레드몬드 원 마이크로소프트 웨이 엘씨에이 - 인터내셔널 패이턴츠 마이크로소프트 코포레이션

줄리 토마스 이

미국 워싱턴주 98052-6399 레드몬드 원 마이크로소프트 웨이 엘씨에이 - 인터내셔널 패이턴츠 마이크로소프트 코포레이션

명세서

청구범위

청구항 1

상태 통지를 클라이언트에게 제공하기 위한 컴퓨터 시스템으로서,

적어도 하나의 프로세서와,

상기 적어도 하나의 프로세서에 동작가능하게 접속되고, 상기 적어도 하나의 프로세서에 의해 실행될 때, 상기 적어도 하나의 프로세서로 하여금 방법을 수행하게 하는 명령어가 저장된 메모리를 포함하되,

상기 방법은,

상기 컴퓨터 시스템에서, 복수의 노드에 관한 노드 정보에 대한 제1 클라이언트로부터의 요청을 수신하는 단계와,

상기 요청에 응답하여, 상기 노드 정보를 갖는 응답을 상기 제1 클라이언트로 송신하는 단계 - 상기 노드 정보는 리소스를 모니터링하기 위한 위트니스 노드(witness nodes)로서의 역할을 하도록 상기 제1 클라이언트에 의하여 선택되기 위한 상기 복수의 노드 중 적어도 2개의 노드를 식별함 - 와,

리소스에 관한 통지를 등록하기 위한 제2 클라이언트로부터의 등록 요청을 상기 복수의 노드 중 제1 노드에서 수신하는 단계 - 상기 제1 노드는 제1 위트니스 노드로서의 역할을 하고 상기 리소스를 모니터링함 - 와,

상기 적어도 하나의 리소스의 상태에 대한 통지를 상기 제1 노드로부터 상기 제2 클라이언트로 송신하는 단계를 포함하는

컴퓨터 시스템.

청구항 2

제1항에 있어서,

상기 리소스는 네트워크 리소스이고 상기 제1 노드는 상기 네트워크 리소스로부터 이벤트를 수신하는

컴퓨터 시스템.

청구항 3

제2항에 있어서,

상기 리소스는 클러스터 리소스이고 상기 제1 노드는 상기 클러스터에서 실행되는 클러스터 서비스로부터 이벤트를 수신하는

컴퓨터 시스템.

청구항 4

제1항에 있어서,

상기 통지는 상기 리소스의 실패를 나타내는

컴퓨터 시스템.

청구항 5

제1항에 있어서,

상기 제1 노드로 접속하기 위한 상기 제1 클라이언트로부터의 요청을 수신하는 단계와,

상기 제1 노드로부터 상기 제1 클라이언트로 응답을 송신하는 단계 - 상기 응답은 상기 제1 노드 상의 정보에 액세스하기 위한 세션을 개설함 - 를 더 포함하되,

상기 제1 노드로 접속하기 위한 요청 및 상기 제1 노드 상의 정보에 액세스하기 위한 세션을 개설하는 응답은 파일 액세스 프로토콜에 따라 포맷되는

컴퓨터 시스템.

청구항 6

제5항에 있어서,

노드 정보에 대한 상기 요청, 상기 노드 정보를 갖는 응답, 상기 등록 요청, 및 상기 통지는 상기 파일 액세스 프로토콜과 상이한 위트니스 프로토콜에 따라 포맷되는

컴퓨터 시스템.

청구항 7

제6항에 있어서,

상기 위트니스 프로토콜은 전송 제어 프로토콜(transmission control protocol: TCP) 전송 수단을 통해 원격 절차 호출(remote procedure call: RPC)을 사용하는

컴퓨터 시스템.

청구항 8

제1항에 있어서,

상기 노드 정보는 위트니스로서의 역할을 하는 복수의 노드를 식별하는

컴퓨터 시스템.

청구항 9

제1항에 있어서,

상기 방법은 상기 제 1 노드에서 상기 통지에 대한 등록을 해제(unregister)하기 위한 등록 해제 요청을 수신하는 단계를 더 포함하는

컴퓨터 시스템.

청구항 10

상대 통지를 수신하기 위한 컴퓨터 시스템으로서,

적어도 하나의 프로세서와,

상기 적어도 하나의 프로세서에 동작가능하게 접속되고, 상기 적어도 하나의 프로세서에 의해 실행될 때, 상기

적어도 하나의 프로세서로 하여금 방법을 수행하게 하는 명령어가 저장된 메모리를 포함하되,
 상기 방법은,
 제1 노드 상의 정보에 액세스하도록 상기 제1 노드로 접속하기 위한 요청을 송신하는 단계 - 상기 제1 노드는 클러스터 내의 복수의 노드 중 하나임 - 와,
 상기 제1 노드로부터 상기 제1 노드 상의 정보에 액세스하기 위한 세션을 개설하는 응답을 수신하는 단계와,
 상기 복수의 노드에 관한 노드 정보에 대한 제2 요청을 상기 제1 노드로 송신하는 단계와,
 상기 제2 요청에 응답하여, 상기 노드 정보를 갖는 제2 응답을 수신하는 단계 - 상기 노드 정보는 상기 복수의 노드 중에서 상기 제1 노드와 다르게 위트니스 노드로서의 역할을 하는 적어도 2개의 노드를 식별함 - 와,
 상기 노드 정보에 기초하여, 식별된 상기 노드 중에서 제2 노드를 선택하고, 적어도 하나의 리소스에 관한 통지를 등록하기 위한 등록 요청을 상기 제2 노드에 송신하는 단계와,
 상기 적어도 하나의 리소스의 상태에 대한 통지를 수신하는 단계를 포함하는
 컴퓨터 시스템.

청구항 11

제10항에 있어서,
 상기 리소스는 상기 제1 노드의 리소스인
 컴퓨터 시스템.

청구항 12

제11항에 있어서,
 상기 통지는 상기 제1 노드의 실패를 나타내는
 컴퓨터 시스템.

청구항 13

제12항에 있어서,
 상기 방법은,
 상기 제1 노드로부터 응답을 수신하면, 상기 제1 노드 상에 저장된 정보에 액세스하기 위한 액세스 요청을 상기 제1 노드에 송신하는 단계를 더 포함하는
 컴퓨터 시스템.

청구항 14

제10항에 있어서,
 상기 복수의 노드는 리소스를 모니터링하기 위한 위트니스 노드로서의 역할을 하는
 컴퓨터 시스템.

청구항 15

제14항에 있어서,
상기 노드 정보는 상기 복수의 노드에 액세스하기 위한 액세스 순서를 식별하는
컴퓨터 시스템.

청구항 16

제15항에 있어서,
상기 리소스는 네트워크 리소스 또는 클러스터 리소스인
컴퓨터 시스템.

청구항 17

제15항에 있어서,
상기 제2 요청은 상기 제1 노드 및 상기 제1 노드 상의 리소스 중 적어도 하나의 상태에 관한 통지를 수신하기
위한 요청을 포함하는
컴퓨터 시스템.

청구항 18

클라이언트 통신 서비스를 제공하기 위한 컴퓨터 시스템으로서,
클러스터 내의 복수의 서버를 포함하되,
상기 복수의 서버의 적어도 일부는,
서버에 저장된 파일 정보에 대한 액세스를 제공하기 위한 제1 클라이언트와의 세션을 개설하고,
위트니스 서비스를 제공하도록 상기 제1 클라이언트에 의하여 선택되기 위한 복수의 서버의 상기 적어도 일부
내의 기타 서버들을 식별하는 위트니스 정보를 저장하고,
상기 위트니스 정보를 상기 제1 클라이언트에게 제공하고,
적어도 하나의 리소스에 관한 이벤트를 모니터링하고,
상기 적어도 하나의 리소스에 관한 통지를 등록하기 위한 제2 클라이언트로부터의 등록 요청을 수신하고,
상기 적어도 하나의 리소스의 상태에 관한 통지를 상기 제2 클라이언트로 송신하도록 구성되는
컴퓨터 시스템.

청구항 19

제18항에 있어서,
상기 리소스는 클러스터 리소스이고 상기 컴퓨터 시스템은 상기 클러스터 상에서 실행되는 클러스터 서비스로부
터 이벤트를 수신하는
컴퓨터 시스템.

청구항 20

제19항에 있어서,
상기 통지는 상기 적어도 하나의 리소스의 실패를 나타내는
컴퓨터 시스템.

발명의 설명

기술 분야

배경 기술

[0001] 서버 클러스터는 페일오버(failover) 및 정보의고가용성을 클라이언트에게 제공하기 위해서 흔히 사용된다. 클러스터는 통상적으로 클러스터 페일오버 및 부하 균형을 관리하기 위한 기능을 수행하는 클러스터 서비스를 갖는다. 클러스터 서비스는 통상적으로 최대 서비스 가용성을 제공하기 위해 빠른 실패 탐지(failure detection)를 제공한다. 그러나 클라이언트의 관점에서 대부분의 실패 탐지는 네트워크 타임아웃을 통해 수행된다. 만약 클라이언트가 전송 제어 프로토콜(TCP)을 사용하는 파일 액세스 요청 또는 데이터베이스 액세스 요청과 같은 클러스터로의 요청을 개시하면, 타임아웃에 도달할 때까지 실패는 탐지되지 않는다. 클라이언트는 소정 시간 기간 동안 서버로부터의 확인응답(acknowledgement)을 기다리기도 하고, 또는 만약 클라이언트가 동작을 수행하고 확인응답을 수신했다면 소정 시간 기간 동안 서버로부터의 응답을 기다리기도 한다. 따라서 서버 실패를 탐지하는데 필요한 시간은 상대적으로 빠른 30초에서부터 더 긴 시간까지 다양해진다. 클라이언트가 실패를 더 빠르게 탐지하기 위해서는, 서버가 실패했다는 것을 판별할 수 있도록 활성화(예컨대, 요청을 주기적으로 송신함)되어야 한다. 빠른 탐지를 보장하는 한가지 방법은 클라이언트가 킵-얼라이브 패킷(keep-alive packet)을 끊임없이 송신하여, 서버가 실패하면 유틸 클라이언트가 네트워크 연결해제(network disconnection)를 수신하지 않는 것을 방지하도록 하는 것이다. 그러나, 킵-얼라이브 패킷은 네트워크 대역폭을 소비한다.

[0002] 이러한 사항들 및 기타 고려사항에 관하여 실시예들이 안출되었다. 또한, 비록 문제점들이 비교적 구체적으로 논의되었지만, 실시예들은 배경기술 내에서 확인된 구체적인 문제점들을 해결하는 것으로 한정되는 것이 아님을 이해하여야 한다.

발명의 내용

해결하려는 과제

과제의 해결 수단

[0003] 본 요약은 이하 상세한 설명에서 자세히 설명될 개념들 중 선택된 것을 단순화된 형태로 소개하기 위해 제공된다. 본 요약은 청구된 발명의 주요 특징 또는 필수적 특징을 나타내는 것을 의도하지 않을 뿐만 아니라, 청구된 발명의 범위를 결정하는 데에 사용되는 것을 의도하지도 않는다.

[0004] 등록된 클라이언트에게 리소스 상태의 통지를 송신하는 위트니스 서비스를 제공하기 위한 실시예들이 설명된다. 실시예들은 예컨대 클러스터 또는 네트워크 내의 리소스와 같은 리소스의 상태에 관한 통지를 수신하고 등록하기 위한 다양한 메시지를 포함하는 프로토콜을 제공한다. 일 실시예에서 프로토콜은 클라이언트가 클러스터 내의 노드로부터 위트니스 정보를 요청하는 메시지를 포함한다. 위트니스 정보는 위트니스 서비스를 제공하는 클러스터 내의 노드를 식별한다. 프로토콜은 네트워크 또는 클러스터 리소스의 상태나 상태 변화에 관한 통지를 위한 위트니스 서비스에 등록하는 데 사용되는 메시지를 더 포함한다. 일 실시예에서 메시지는 노드 실패의 통지를 등록하는 데 사용된다. 프로토콜은 리소스의 상태 정보를 갖는 통지를 전송하는 위트니스 서비스에 대한 메시지도 포함한다. 실시예들은 위트니스 서비스로부터 등록을 해제하기 위한 메시지도 제공한다.

[0005] 실시예들은 컴퓨터 프로세스, 컴퓨팅 시스템 또는 컴퓨터 프로그램 제품 또는 컴퓨터 판독 가능 매체와 같은 제조품으로 구현된다. 컴퓨터 프로그램 제품은 컴퓨터 시스템에 의해 판독 가능하고 컴퓨터 프로세스를 실행하기 위한 컴퓨터 프로그램 명령어로 인코딩된 컴퓨터 저장 매체가 될 수도 있다. 컴퓨터 프로그램 제품은 컴퓨

팅 시스템에 의해 판독 가능하고 컴퓨터 프로세스를 실행하기 위한 컴퓨터 프로그램 명령어로 인코딩된 반송파상의 전파 신호가 될 수도 있다.

도면의 간단한 설명

[0006]

아래의 도면들을 참고하여 실시예들이 설명되는데, 실시예들이 이에 한정되는 것은 아니며 빠뜨림 없이 설명되는 것도 아니다.

도 1은 실시예들을 구현하는 데 사용되는 시스템을 도시한다.

도 2는 일부 실시예에 따른 위트니스 프로토콜을 사용하여 메시지를 교환하는 클라이언트 및 노드 클러스터의 블록도를 도시한다.

도 3은 일부 실시예에 따른 위트니스 프로토콜 및 파일 액세스 프로토콜을 사용하여 클라이언트와 파일 서버 클러스터가 통신하는 블록도를 도시한다.

도 4는 도 3에 도시된 실시예 내에서 교환될 수 있는 메시지의 순서를 보여주는 순서도를 도시한다.

도 5는 일부 실시예에 따른 리소스 상태 정보의 상태 통지를 송수신하기 위한 동작 순서를 도시한다.

도 6은 일부 실시예에 따른 클러스터 리소스의 상태 통지를 송수신하기 위한 동작 순서를 도시한다.

도 7은 실시예들을 구현하는 데 적합한 컴퓨팅 환경의 블록도를 도시한다.

발명을 실시하기 위한 구체적인 내용

[0007]

다양한 실시예들이 본 명세서의 일부로서 구체적이고 예시적인 실시예들을 보여주는 첨부된 도면을 참조하여 아래에서 더욱 상세하게 설명된다. 그러나 실시예들은 여러 다양한 형태로 구현될 수 있고 본 명세서에서 제시된 실시예들로 한정되는 것으로 해석되지 않아야 하며, 오히려 본 명세서에서 제시된 실시예들은 당업자에게 실시예들의 범주를 빈틈없이 완벽하고 충분하게 전달하기 위해 제공되는 것이다. 실시예들은 방법, 시스템 또는 장치로 실시될 수 있다. 따라서 실시예들은 하드웨어 구현, 완전한 소프트웨어 구현 또는 소프트웨어와 하드웨어 측면의 결합된 구현의 형태를 취할 수 있다. 그러므로 이하 상세한 설명에서 취하는 방식으로 한정되는 것은 아니다.

[0008]

도 1은 일부 실시예를 구현하는 데 사용되는 시스템(100)을 도시한다. 시스템(100)은 클라이언트(102, 104) 및 서버 클러스터(106)를 포함한다. 클라이언트(102, 104)는 서버 클러스터(106)와 네트워크(108)를 통해 통신한다. 서버 클러스터(106)는 클라이언트(102, 104) 상의 애플리케이션에 의해 액세스되는 정보를 저장한다. 클라이언트(102, 104)는 클러스터(106) 상의 정보에 액세스하기 위해 클러스터(106)와의 세션을 개설한다. 도 1에서는 비록 클라이언트(102, 104)만이 클러스터(106)와 통신하는 것으로 도시되어 있으나, 다른 실시예에서는 서버 클러스터(106)로부터의 정보에 액세스하는 클라이언트가 2개 이상이 될 수도 있다.

[0009]

도 1에 도시된 것처럼, 서버 클러스터(106)는 클러스터(106) 상에 저장되는 정보를 위한 여분 및 고가용성을 모두 제공하는 서버들(106A, 106B, 106C)을 포함한다. 실시예에서, 클러스터(106)는 파일 시스템, 데이터베이스, 또는 클라이언트(102, 104)에 의해 액세스되는 다른 정보를 가질 수 있다. 비록 3개의 서버가 도 1에 도시되어 있으나, 다른 실시예에서는 클러스터(106)가 3개 이상의 서버를 포함할 수도 있고, 또는 3개보다 적은 서버를 포함할 수도 있다.

[0010]

일 실시예에 따르면, 클라이언트(102, 104)에 의해 액세스되는 정보를 저장하는 것 이외에, 클러스터(106)는 위트니스 서비스도 제공한다. 위트니스 서비스는 클라이언트(102, 104)가 클러스터(106)에 의해 감시되는 리소스의 상태에 관한 통지를 수신하도록 해준다. 리소스는 클러스터 리소스 또는 네트워크 리소스일 수 있다. 일 실시예에서, 각각의 서버(106A, 106B, 106C)는 위트니스 서비스를 제공하는 것이 가능하다. 즉, 클라이언트(102, 104)는 그 클라이언트가 클러스터 리소스에 액세스하는 데 사용하는 서버가 아닌, 서버들(106A, 106B, 106C) 중 임의의 것을 위트니스 서비스용으로 등록할 수 있다. 다른 실시예에서는, 서버들 중 일부만이 위트니스 서비스를 제공할 수도 있다. 예를 들어, 이러한 실시예에서는 오직 서버(106B, 106C)만 위트니스 서비스를 실행할 수도 있다. 그러나 다른 실시예에서는, 클러스터(106)가 위트니스 서비스 제공 전용의 서버를 포함할 수도 있다. 비록 도시되지는 않았으나, 이러한 실시예에서 클러스터(106)는 위트니스 서비스를 실행하도록 특

별히 구성되며 클라이언트(102, 104)에게 액세스를 제공하지 않는 서버를 포함할 수 있다.

- [0011] 위트니스 서비스를 제공하는 클러스터(106)로부터의 서버는 클러스터(106) 내의 리소스 및/또는 네트워크 리소스로부터의 이벤트를 수신한다. 서버는 이러한 이벤트로부터의 정보를 처리하고 저장하여 이들 리소스의 상태를 추적 파악한다. 이러한 이벤트에 응답하여, 서버는 리소스에 관한 상태 정보를 수신하도록 위트니스 서비스에 등록되어 있는 클라이언트에게 통지를 전송할 것이다. 일 실시예로서, 이벤트는 클러스터(106) 상에서 실행되는 클러스터 서비스에 의해 부하 균형 및 실패 탐지를 제공하기 위해 생성될 수 있다.
- [0012] 전술된 것처럼, 클라이언트(102, 104)는 위트니스 서비스로부터의 통지를 요청하기 위해 클러스터(106)와 통신한다. 실시예에서는 위트니스 서비스 및 클라이언트(102, 104)에 의해 사용되는 위트니스 프로토콜이 제공된다. 프로토콜은 클라이언트(102, 104)가 클러스터(106) 내의 서버로부터 위트니스 서비스에 대한 정보를 요청하게 해주는 메시지를 포함한다. 정보는 클러스터(106) 내의 어떤 서버가 위트니스 서비스를 제공하는지 나타낼 수 있다. 프로토콜은 위트니스 서비스로부터의 통지를 등록하기 위한 메시지를 포함할 수도 있다. 위트니스 서비스에 의해 클라이언트(102, 104)에게 송신되는 리소스의 상태에 관한 메시지는 위트니스 프로토콜에 따라 포맷(formatted)된다.
- [0013] 일 실시예를 설명하기 위해, 클라이언트(102)는 클러스터(106)의 서버와 세션을 개설한다. 예를 들어, 클라이언트(102)는 서버(106A) 상에 저장된 데이터베이스에 액세스하기 위해 서버(106A)와의 세션을 개설할 수 있다. 클라이언트(102)는 서버(106A)로부터의 정보에 액세스하기 때문에, 서버(106A) 상의 임의의 실패에 대해 빨리 지시받기를 원할 수도 있다. 실패의 빠른 탐지는 빠른 복구를 가능케 하여, 클라이언트(102) 상의 애플리케이션이 대기해야 하는 총 시간을 줄여준다. 빠른 탐지 및 복구는 클라이언트(102)가 재접속하는 데 예약된 리소스가 다른 클라이언트에게 이용 불가능한 시간의 윈도우도 줄여준다.
- [0014] 위트니스 서비스에 등록하도록, 클라이언트(102)는 먼저 클러스터(106) 상에서 어떤 서버 또는 서버들이 위트니스 서비스를 실행하는지에 관한 정보를 위해 위트니스 프로토콜에 따라 포맷된 메시지를 서버(106A)로 송신한다. 서버(106A)는 서버(106B, 106C)가 위트니스 서비스를 실행한다는 식별을 포함하는 정보와 이들 서버 중 어느 하나에 접속하기 위해 필요한 임의의 정보로 응답한다. 이러한 예시에서, 서버(106A)로부터 수신된 정보에 기초하여, 클라이언트(102)는 서버(106A) 상의 리소스에 대한 상태나 상태 변화의 통지를 수신하기 위해 등록 요청을 서버(106B)로 송신한다. 클라이언트는 서버(106A) 상의 리소스가 실패한 때, 그리고 그 후 서버에 재접속 가능할 때에 통지를 받게 된다. 상태는 서버(106A)가 온라인인지 오프라인인지 여부와 같이 단순한 것일 수도 있다. 다른 실시예에서, 통지는 서버(106A) 상의 디스크 또는 네트워크 어댑터가 실패하거나 오프라인인지 여부와 같은 세부적인 사항을 포함할 수도 있다.
- [0015] 일부 실시예에서 클라이언트(102)는 위트니스 서비스에 등록하기 위해 접속하는 특정 서버를 선택하기 위한 소정의 방법을 가질 수도 있다. 예를 들어, 클라이언트는 서버를 무작위 방식, 라운드-로빈 방식, 또는 클러스터로부터 수신하는 정보에 기초하여 선택할 수 있다.
- [0016] 클라이언트(102)가 서버(106A) 상의 리소스에 관한 통지를 위해 서버(106B)에 등록된 후에, 서버(106B)에 의해 탐지되는 리소스의 상태 변화는 그 리소스의 상태 변화를 나타내는 정보를 갖는 통지가 클라이언트(102)에게 송신되도록 한다. 예를 들어, 통지는 서버(106A)가 오프라인이라는 것을 나타낼 수 있다. 통지에 응답하여, 클라이언트(102)는 서버(106A)로 정보 요청을 송신하는 것을 멈추고 클러스터(106) 내의 다른 서버에 접속을 시도할 수 있다. 그렇지 않으면, 클라이언트(102)는 서버(106A)가 온라인으로 돌아왔다는 후속 통지를 기다려서 재접속하고 서버(106A)로부터의 정보에 액세스하는 것을 재개할 수도 있다.
- [0017] 이하의 설명은 도 1에 도시된 실시예가 동작할 수 있는 방식의 단지 한가지 예시일 뿐이다. 더욱 자세하게 후술되는 바와 같이, 실시예는 상이한 단계 또는 동작을 포함할 수도 있다. 이들은 임의의 적합한 소프트웨어나 하드웨어 컴포넌트 또는 모듈을 사용해서 구현될 수도 있다.
- [0018] 이제 도 2로 가면, 클라이언트(202, 204)와 노드 클러스터(206)를 보여주는 소프트웨어 환경(200)의 블록도가 도 2에 도시된다. 클라이언트 및 노드 클러스터는 일부 실시예에 따른 위트니스 프로토콜을 사용하여 메시지를 교환한다. 도 2는 위트니스 프로토콜을 사용하기 위한 환경의 일반적인 예시를 설명하기 위해 도시된다. 도 2는 노드 클러스터(206) 상에 저장된 파일 시스템에 액세스하는 클라이언트(202, 204)와 같이 여러 상황에 적용 가능한 일반적인 것을 의도한 것인데, 여기서는 노드 클러스터(206)가 분산 파일 시스템을 갖춘 서버 클러스터이다. 다른 예시로서, 환경(200)은 클라이언트(202, 204)가 노드 클러스터(206) 상에 저장된 데이터베이스에 액세스하는 상황에서 사용될 수도 있다.

- [0019] 도 2에 도시된 것처럼, 클라이언트(202, 204)는 노드 클러스터(206)에 의해 제공되는 정보에 액세스한다. 노드 클러스터(206)는 3개의 노드, 즉 노드 1, 노드 2, 및 노드 3을 포함한다. 도 2에 도시된 바와 같이, 노드 1, 2, 3 각각은 클라이언트(202, 204)와 통신하는 다양한 컴포넌트를 포함한다. 예를 들어, 각각의 노드는 클라이언트로 하여금 노드 클러스터 상에 저장된 정보에 액세스하게 해주는 세션을 개설하기 위해 클라이언트(202, 204)와 통신하는 클라이언트 액세스 모듈을 포함한다.
- [0020] 도 2에 도시된 실시예에서는, 노드 1, 2, 3이 노드 클러스터(206) 상에서 제공되는 위트니스 서비스의 일부로서 클라이언트(202, 204)와 통신하는 위트니스 관리 컴포넌트도 포함한다. 위트니스 관리 컴포넌트는 노드 클러스터(206) 내에서 어떤 노드가 위트니스 서비스를 실행하는지에 대한 정보를 제공하여 클라이언트(202, 204)가 위트니스 서비스에 등록할 수 있도록 한다. 또한, 위트니스 관리 컴포넌트는 리소스 상태의 통지를 위해 클라이언트를 등록하고, 도 2에 도시된 바와 같이 노드 및 네트워크 리소스로부터의 이벤트를 수신하고 처리하는 감시 컴포넌트와 통신한다. 감시 컴포넌트는 이벤트를 처리하고 상태 및 상태 변화 정보를 위트니스 관리 컴포넌트로 제공하는데, 위트니스 컴포넌트는 결국 리소스의 상태나 상태 변화의 통지를 위해 등록된 임의의 클라이언트에게 통지하게 된다. 일부 실시예에서는, 감시 컴포넌트가 클러스터 내의 리소스의 상태를 감시하도록 클러스터 서비스에 의해 제공되는 API를 사용할 수도 있다.
- [0021] 도 2에 도시된 실시예에서, 클라이언트(202)는 노드 1과의 세션을 개설하고, 채널(208)을 사용하여 노드 1의 클라이언트 액세스 컴포넌트로 액세스 요청을 송신한다. 클라이언트(202)는 서버가 위트니스 서비스에게 제공하는 것과 같은, 위트니스 서비스에 관한 정보를 수신하기 위해 노드 1과 분리된 채널(미도시)도 개시한다. 노드 1에 대해 개설된 분리된 채널은 노드 1이 클라이언트(202)에 의해 송신되는 액세스 요청에 간섭받지 않고 위트니스 서비스에 관한 정보를 통신하게 한다.
- [0022] 도 2에서, 클라이언트(202)는 노드 2의 위트니스 관리 컴포넌트에 의해 제공된 위트니스 서비스에 등록한다. 도 2에 도시된 것처럼, 위트니스 관리 컴포넌트와의 통신은 분리된 채널을 통해 발생한다. 따라서, 상태 통지를 포함한 임의의 유형의 위트니스 정보는 채널(208)을 통한 액세스 요청에 간섭을 받지 않는다. 파선으로 도시된 채널(216)은 노드 1의 리소스에 대해 실패한 경우 액세스 요청을 송신하기 위하여 클라이언트(202)가 개설하는 분리된 채널을 도시한다. 일부 실시예에서, 위트니스 관리 컴포넌트는 클라이언트(202)가 실패할 경우 액세스 요청을 송신할 새로운 리소스나, 새로운 리소스에 접속하기 위한 추천 네트워크 경로 또는 다수의 추천 네트워크 경로들을 가리키는 정보를 제공한다.
- [0023] 마찬가지로, 클라이언트(204)는 노드 2와의 세션을 개설하고 채널(212)을 통해 노드 2의 클라이언트 액세스 컴포넌트로 액세스 요청을 송신한다. 클라이언트(204)는 위트니스 서비스를 등록하고 채널(214)을 사용하여 노드 3의 위트니스 관리 컴포넌트와 통신한다.
- [0024] 일 실시예에서, 위트니스 관리 컴포넌트는 추가적인 기능을 제공한다. 일부 실시예에서, 컴포넌트는 부하 균형 기능을 제공한다. 다시 말해, 리소스의 상태에 관한 통지를 송신하는 것 이외에, 위트니스 관리 컴포넌트는 클라이언트(202, 204)가 노드(206) 상의 부하 균형을 위한 알고리즘에 기초하여 다른 리소스에 접속하라고 요청하는 통지도 송신할 수 있다. 예를 들어, 도 2에 도시된 바와 같이, 클라이언트(202)는 노드 1의 클라이언트 액세스 컴포넌트와 통신한다. 게다가, 클라이언트(202)는 노드 2의 위트니스 관리 컴포넌트와 통신하고, 이는 클라이언트(202)에게 노드 1 내의 리소스에 대한 실패와 같은 리소스의 임의의 상태 변화에 대한 통지를 제공한다. 이러한 실시예에서, 노드 2의 위트니스 관리 컴포넌트는 노드 1에 의해 서비스 중인 액세스 요청의 과도한 총량으로 인해 노드 1이 저조하게(예컨대, 높은 대기 시간) 실행 중임을 판별할 수 있다. 다른 실시예에 따르면, 판별을 통해 노드 2로부터 클라이언트를 없애서, 노드 2가 전력 절약 또는 유지보수 등의 기타 이유에 따라 셧다운될 수 있도록 한다. 알고리즘에 따르면, 노드 2가 클라이언트(202)로 하여금 향후의 액세스 요청은 활용도가 낮은 노드 3으로 송신하도록 지시하는 통지를 송신할 수도 있다. 그 후 클라이언트(202)는 향후의 요청을 송신하기 위해 파선으로 도시된 노드 3과의 채널(216)로 세션을 개설한다. 다른 실시예에서, 위트니스 관리 컴포넌트는, 클라이언트(202)가 액세스 요청을 보낼 새로운 리소스뿐만 아니라, 추천 네트워크 경로나 다수의 추천 네트워크 경로들을 가리키는 정보도 제공한다.
- [0025] 도 3은 도 2에서 설명된 환경(200)과 유사한 환경(300)을 도시한다. 그러나 환경(300)은 서버 클러스터(306) 상에 있는 파일 서버로 판독/기록 액세스를 요청하는 클라이언트(302, 304)를 도시한다. 서버 클러스터(306)는 서버 1, 서버 2, 서버 3을 포함하고, 이들은 각각 파일 서버, 위트니스 관리 컴포넌트, 및 클러스터 감시 컴포넌트를 포함한다. 도 3에 도시된 실시예에서, 클라이언트(302, 304)는 애플리케이션, 위트니스 관리 컴포넌트, 및 리더렉터를 포함한다. 도 4는 실시예에 따른 환경(300) 내에서 교환될 수 있는 메시지의 한가지 예시를 보

여주는 순서도를 도시한다.

- [0026] 도 3 및 4에 대한 이하의 설명은 파일 액세스 프로토콜로서 서버 메시지 블록(SMB) 프로토콜을 사용하여 설명된다. 그러나 실시예들이 이것에 제한되는 것은 아니다. 네트워크 파일 시스템(NFS) 또는 SMB의 다른 버전을 포함하는 임의의 파일 액세스 프로토콜이 실시예 내의 파일 액세스 프로토콜로 사용될 수도 있다. SMB는 단지 설명의 용이함과 편의를 위해 사용된 것일 뿐이다.
- [0027] 도 4를 참조하면, 순서도는 일 실시예에 따라 클라이언트(도 3의 302), 네트워크 DNS(400), 및 서버 클러스터(도 3의 306) 사이에서 교환되는 메시지를 도시한다. 먼저 클라이언트(302)는 SMB 버전을 사용하여 액세스될 수 있는 파일 시스템을 저장하는 서버 클러스터(306)의 IP 주소를 얻기(resolve) 위해 네트워크 DNS(400)로 요청(402)을 송신한다. 요청(402)에 응답하여, 네트워크 DNS(400)는 서버 클러스터(306)의 IP 주소를 포함하는 메시지(404)로 응답한다.
- [0028] 응답 메시지(404)를 수신한 후, 클라이언트(302)는 액세스 요청 메시지(406)를 서버 클러스터(306)로 송신한다. 액세스 요청은 클라이언트(302) 상의 리더렉터에 의해 송신될 수 있다. 도 3 및 4에 도시된 바와 같이, 서버 1은 클라이언트(302)로부터의 요청을 처리한다. 도 3에 도시된 것처럼, 서버 1은 파일 액세스 세션을 개설하기 위해 클라이언트(302) 상의 리더렉터와 통신하는 파일 서버 컴포넌트를 포함한다. 실시예에서, 요청(406)은 SMB 세션을 교섭(negotiate)하는 제1 요청일 수 있다. 예를 들어, 요청은 SMB2 교섭 패킷(negotiate packet)일 수 있다. 도 4에 도시된 바와 같이, 서버 1은 클라이언트(302)와의 세션을 개설하는 액세스 응답(408)을 송신한다. 응답은 예를 들어 SMB2 세션 셋업 응답 또는 SMB2 트리 접속 응답일 수 있다.
- [0029] 비록 도 4 및 관련 설명은 단일 액세스 요청(406)과 단일 액세스 응답(408)을 제공하고 있으나, 실시예에서 이것은 세션을 개설하기 위해 송신되는 여러 메시지 중 단지 2개의 메시지에 불과한 것임을 주목해야 할 것이다. 이해할 수 있는 바와 같이, SMB2 프로토콜을 사용하여 세션을 개설하기 위해 많은 메시지가 클라이언트(302) 상의 리더렉터 및 서버 1 상의 파일 서버 사이에서 교환될 수 있다. 메시지는 하나 이상의 SMB2 교섭 메시지, SMB2 세션 셋업 메시지, 및/또는 SMB2 트리 접속 메시지를 포함할 수 있다. 세션이 개설된 이후, 클라이언트(302)는 예컨대 파일 및/또는 디렉토리의 판독/기록과 같은 액세스 요청(미도시)을 개설된 채널(도 3의 308)을 통해 클러스터(306)의 서버로 송신하기 시작한다.
- [0030] 클라이언트(302) 상의 위트니스 관리 컴포넌트는 서버 1이 클러스터(306)의 일부라는 것을 탐지한다. 이러한 탐지는 응답 메시지(408) 처리의 일부로서 또는 기타 그 밖의 다른 수단에 의해 실시예 내에서 일어난다. 서버 1이 클러스터(306)의 일부라는 것을 판별한 결과로서, 클라이언트(302) 상의 위트니스 관리 컴포넌트는 클러스터(306)에 의해 제공되는 위트니스 서비스를 등록하도록 실시예에서 구성된다. 그 결과, 클라이언트(302) 상의 위트니스 관리 컴포넌트는 위트니스 정보에 대한 요청(410)을 송신하기 위해 새로운 채널(도 3의 310)을 사용한다.
- [0031] 실시예에서, 요청은 위트니스 서비스용으로 구체적으로 구성되는 위트니스 프로토콜에 따라 포맷된다. 위트니스 프로토콜은 위트니스 서비스와 관련된 많은 메시지들을 포함할 수 있다. 예를 들어, 실시예에서 위트니스 프로토콜은 적어도 위트니스 정보를 요청하는 위트니스 요청 메시지, 클러스터 내의 리소스에 대한 통지를 등록하기 위한 위트니스 등록 메시지, 및 리소스의 상태 변화(예컨대, 리소스의 온라인 또는 오프라인 상태)를 클라이언트에게 통지하기 위한 위트니스 리소스 변화 메시지를 포함한다. 이들은 단지 일부 예시에 불과하고, 실시예에서는 위트니스 서비스가 다른 메시지를 포함할 수도 있다. 일 실시예에서, 위트니스 프로토콜은 원격 절차 호출(RPC)을 전송 제어 프로토콜(TCP)과 함께 전송 수단으로 사용한다.
- [0032] 도 4를 다시 참조하면, 서버 1의 위트니스 관리 컴포넌트는 요청(410)에 대해 위트니스 정보를 포함하는 응답(412)으로서 응답한다. 응답(412)은 위트니스 프로토콜에 따라 포맷되고 위트니스 서비스를 제공하는 클러스터(306) 내의 다른 서버를 식별하는 정보를 포함하게 된다. 실시예에서 응답(412)은 클라이언트(302)가 클러스터(306)의 다른 서버들 중 하나 상의 위트니스 서비스에 접속하기 위해 필요한 기타 정보도 포함한다. 본 예시에서, 응답(412) 내의 위트니스 정보는 위트니스 서비스를 제공하는 것으로 적어도 서버 2를 식별하는 정보를 포함한다.
- [0033] 일부 실시예에서, 위트니스 정보는 특수한 포맷으로 구조화된다. 예를 들어, 위트니스 정보는 랭크(ranked)될 수 있다. 즉, 위트니스 서비스를 제공하는 서버는 특수한 순서에 따라 리스팅될 수 있다. 그러면 클라이언트(302)는 리스트 내의 첫 번째 서버로 시작할 수 있고, 만약 그 서버가 위트니스 서비스를 제공하는 데 이용할 수 없다면, 클라이언트는 리스트 상의 다음 서버로 시도할 수 있다. 랭킹은 파라미터를 고려하는 여러 다양한

알고리즘에 기초할 수 있는데, 일부 비-한정적인 예시로서, 대량의 클라이언트를 처리하는 능력을 나타낼 수 있는, 특정 서버와 네트워크 접속의 하드웨어 및 소프트웨어 특성을 비롯하여, 위트니스 서비스를 제공하는 서버, 위트니스 서비스를 등록한 클라이언트의 수를 포함한다.

[0034] 서버 1로부터 수신된 위트니스 정보에 응답하여, 클라이언트(302)는 서버 2로 등록 요청(414)을 송신하기 위해 채널(도 3의 310)을 사용할 것이다. 등록 요청(414)도 위트니스 프로토콜에 따라 포맷되고 클라이언트(302)가 어떤 리소스에 대한 통지를 수신하기를 바라는지 식별하는 정보를 포함한다. 일 예시에서, 클라이언트(302)는 서버 1의 실패에 대한 조기 지시(early indication)를 갖도록 서버 1이 언제 오프라인되는지 알고 싶을 수도 있다. 실패 탐지는 빠른 복구를 가능하게 하여, 클라이언트(302) 상의 애플리케이션이 동작(예컨대, 관독/기록)을 처리하기 위해 대기하는 시간의 총량을 줄이게 된다. 본 예시에서는, 클라이언트(302)가 서버 1의 온라인/오프라인 상태에 관한 통지를 위해 서버 2의 위트니스 관리 컴포넌트에 등록한다. 서버 2는 등록 요청을 승인(acknowledge)하는 상태 응답(416)을 송신한다. 실시예에서, 응답(416)은 클라이언트(302)가 통지를 수신하기 위해 등록한 리소스의 현재 상태도 포함하게 된다.

[0035] 일부 실시예에서, 동작(416) 이후, 위트니스 서버(서버 2)는 파선으로 도시된 서버 실패(417)와 같은 실패를 겪을 수도 있다. 이 경우, 클라이언트(302)는 일부 실시예에서 메시지(410)와 유사한 다른 메시지를 송신하는 것과 (서버 1 또는 상이한 서버로부터의) 메시지(412)와 유사한 다른 메시지를 수신하는 것을 포함하여 대체 서버를 재선택해야 할 수도 있다. 클라이언트(302)는 대체 서버로부터 위트니스 서비스를 등록하기 위해 요청(414)과 유사한 위트니스 등록 요청을 송신할 수도 있다.

[0036] 다른 실시예에서, 클라이언트는 복수의 위트니스 서버 접속을 개설할 수도 있다. 즉, 클라이언트는 위트니스 서비스를 위해 클러스터(306) 내의 하나 이상의 서버에 등록할 수 있다. 복수의 접속은 클라이언트로 하여금 서버들 중 어느 하나가 실패한 경우에도 위트니스 서비스에 계속해서 액세스할 수 있도록 한다. 클라이언트는 실시예 내의 접속들 중 하나를 기본 접속으로 지정하고 나머지 접속을 기본 접속이 실패한 경우에만 보조적으로 사용하게 할 수도 있다.

[0037] 만약 어느 순간 서버 1의 서버 실패(418)가 있다면, 서버 2는 그 실패를 탐지할 것이다. 전술된 바와 같이, 서버 2는 클러스터(306) 상에서 실행되는 클러스터 서비스에 의해 생성된 이벤트를 수신함으로써 실패를 탐지한다. 서버 2가 서버 실패(418)를 탐지하면, 서버 2는 그 실패에 대한 통지(420)를 클라이언트(302)에게 송신할 것이다. 실패에 응답하여, 본 실시예에서 클라이언트(302)는 추후의 파일 액세스 요청(422)을 중지하도록 구성된다. 이해되는 바와 같이, 클라이언트(302)가 추가적인 액세스 요청을 보낸다고 하더라도, 이들은 서버 1의 실패 때문에 처리될 수가 없다. 실시예에서 서버 2의 위트니스 서비스는 클라이언트(302)로부터의 액세스 요청을 처리하기 위해 이용가능한 리소스가 서버 3에 있다는 것을 나타내는 메시지(424)를 송신하도록 구성된다. 메시지(424)의 수신 이후, 클라이언트(302)는 파일 액세스 요청(426)을 서버 3으로 송신하는 것을 재개하게 된다. 클라이언트(302) 상의 리더렉터는 채널(도 3의 312)을 통해 파일 액세스 요청(426)을 서버 3의 파일 서버 컴포넌트로 송신한다.

[0038] 다른 실시예에서, 클라이언트(302)는 서버 1이 온라인으로 돌아올 때까지 파일 액세스 요청을 중지하도록 구성된다. 서버 2의 위트니스 서비스는 서버 1이 이용가능하게 될 때, 서버 1이 클라이언트(302)로부터의 액세스 요청 처리가 가능하다는 것을 나타내는 메시지(424)를 클라이언트(302)에게 송신하도록 구성된다.

[0039] 비록 도시되지는 않았지만, 일부 실시예에서는, 클라이언트(302)가 서버 3으로 파일 액세스 요청을 송신하기 시작할 때, 위트니스 서비스를 수신하는 서버를 변경할 수도 있다. 예를 들어, 서버 3으로 파일 액세스 요청을 송신한 이후, 클라이언트(302)는 위트니스 정보에 대한 요청을 서버 3으로 송신하는 프로세스를 개시할 수 있다. 서버 3은 어떤 서버가 위트니스 서비스를 제공하는지에 관한 정보를 포함하여 서버 1에 의해 송신된 응답(412)과 유사한 정보로 응답할 수 있다. 정보는 달라질 수도 있는데, 예를 들어 그것은 상이한 서버를 식별할 수도 있다. 위트니스 서비스를 제공하는 것으로 이전에 식별되었던 서버는 실패되고 더 이상 가용될 수 없거나, 다른 서버가 온라인이 되기도 한다. 또한, 일부 실시예에서, 위트니스 관리 컴포넌트는 위트니스 서비스를 제공하는 것이 가능한 서버의 랭킹을 위트니스 정보에게 제공하도록 구성된다. 서버 3에 의해 제공되는 위트니스 정보는 서버 2에 의해 이전에 제공된 위트니스 정보와 상이할 수 있다. 서버 3에 의해 송신된 위트니스 정보에 응답하여, 클라이언트(302)는 상이한 서버로부터 위트니스 서비스에 대한 등록을 요청할지 여부를 결정할 수 있다.

[0040] 도 5 및 6은 실시예에 따른 동작 순서(500, 550, 600, 650)를 도시한다. 동작 순서(500, 550, 600, 650)는 임의의 적합한 컴퓨팅 환경에서 실행될 수 있다. 예를 들어, 동작 순서는 도 1 내지 3에서 도시된 것과 같은 시

시스템 및 환경에 의해 실행될 수 있다. 따라서 동작 순서(500, 550, 600, 650)에 대한 설명은 도 1 내지 3의 컴포넌트들 중 적어도 하나를 참조할 수도 있다. 그러나, 도 1 내지 3의 컴포넌트에 대한 임의의 참조는 오직 설명을 위한 것에 불과할 뿐이고, 도 1 내지 3의 구현이 동작 순서(500, 550, 600, 650)에 대한 환경을 한정하는 것이 아님을 이해할 수 있을 것이다.

- [0041] 또한, 비록 동작 순서(500, 550, 600, 650)가 특정 순서에 따라 순차적으로 설명되고 도시되지만, 다른 실시예에서는 이 동작들이 상이한 순서로, 복수의 시간에서, 및/또는 병렬적으로 실행될 수 있다. 게다가, 일부 실시예에서는 하나 이상의 동작이 생략되거나 결합될 수도 있다.
- [0042] 동작 순서(500) 내의 단계들이 동작 순서(550) 내의 단계들에 대응할 수 있음을 보여주기 위해 동작 순서(500, 550)는 도 5 내에서 함께 도시된다. 동작 순서(500)는 리소스에 대한 상태 통지를 수신하기 위한 단계를 도시한다. 동작 순서(550)는 리소스에 대한 상태 통지를 제공하기 위한 단계를 도시한다. 실시예에서, 앞서 설명된 클라이언트(도 1의 102, 104) 또는 클라이언트(도 2의 202, 204)와 같은 클라이언트가 동작 순서(500)를 구현할 수도 있다. 동작 순서(500)를 구현하는 클라이언트는 랩톱 컴퓨터, 데스크톱 컴퓨터, 스마트폰 장치, 또는 태블릿 컴퓨터를 포함하는 임의의 유형의 클라이언트가 될 수 있다. 실시예에서, 도 2에 도시된 노트 1, 노트 2, 노트 3과 같은 노트가 동작 순서(550)를 구현한다. 노트는 클러스터 서비스를 실행시키는 노트 클러스터의 일부일 수도 있다.
- [0043] 순서(500)는 제1 노트로 접속하라는 요청이 송신되는 동작(502)에서 개시된다. 제1 노트는 액세스되는 정보를 저장하는 노트 클러스터의 일부이다. 일부 실시예에 따르면, 동작(502)에서 송신되는 요청은 노트 클러스터 상에 저장된 정보에 액세스하기 위하여 제1 노트와의 세션을 개설하라는 요청이다. 동작(502) 이후, 순서(500)는 제1 노트와의 세션이 개설되었다는 것을 나타내는 응답이 수신되는 동작(504)으로 넘어간다. 비록 순서(500)에는 도시되지 않았으나, 실시예에 따르면 동작(504) 이후에 액세스 요청이 제1 노트 상에 저장된 정보에 액세스하기 위해 제1 노트로 송신된다.
- [0044] 도 5에 도시된 실시예처럼, 동작 순서는 동작(504)으로부터 제2 요청이 위트니스 정보를 위해 제1 노트로 송신되는 동작(506)으로 이어진다. 위트니스 정보는 클러스터 내의 어떤 노트가 리소스에 대한 통지를 제공하기 위해 등록될 수 있는 위트니스 서비스를 제공하는지 식별한다. 리소스는 예컨대 네트워크 리소스 또는 클러스터 리소스가 될 수 있다.
- [0045] 동작(508)에서는 위트니스 정보가 수신된다. 그 후 위트니스 정보는 클러스터 내에서 어떤 노트가 위트니스 서비스를 제공하는지 판별하는 데 사용된다. 위트니스 정보는 동작(510)에서 리소스에 관한 통지를 등록하는 등록 요청을 송신하는 데 사용된다. 일 실시예에서, 리소스는 정보가 액세스되는 제1 노트일 수 있다. 요청은 제1 노트 또는 제1 노트 상의 리소스에 대한 상태에 관한 통지를 수신할 수 있다. 예를 들어, 통지는 제1 노트가 온라인인지 오프라인인지 여부를 나타낼 수 있다.
- [0046] 순서(500)는 동작(510)으로부터 리소스에 관해 통지가 수신되는 동작(512)으로 이어진다. 그 후 순서(500)는 동작(514)에서 종료한다. 일부 실시예에서, 단계(512)에서의 통지 수신에 응답하여 실행되는 추가적인 단계가 있을 수도 있는데, 예컨대 액세스 요청의 중지, 클러스터 내의 다른 노트로의 재접속, 및 클러스터 내의 다른 노트로 액세스 요청의 재개 등이 될 수 있다.
- [0047] 실시예에 따르면, 동작(512)에서 수신되는 통지는 상태 통지가 아니라 부하 균형 통지일 수 있다. 부하 균형 통지는 향후의 액세스 요청이 상이한 서버로 송신되어야 한다는 것을 지시할 수 있다. 일부 실시예에서, 통지는 사용되어야 하는 상이한 네트워크, 네트워크 리소스, 또는 클러스터 리소스를 구체적으로 식별할 수 있다. 이러한 실시예에서, 많은 동작들이 동작(512)에서의 통지 수신에 응답하여 실행되기도 한다.
- [0048] 순서(550)는 순서(500)에 관해 전술된 단계들에 대응하는 단계들을 가진다. 전술된 바와 같이, 실시예에 따르면, 도 2에 도시된 노트 1, 노트 2, 및 노트 3이 순서(550)를 구현한다. 순서(550)는 노트 클러스터 내의 노트로 접속하라는 요청이 수신되는 동작(552)에서 개시된다. 요청은 노트 클러스터 상에 저장된 정보에 액세스하기 위한 세션을 개설하라는 요청일 수 있다. 순서는 동작(552)으로부터 노트 클러스터 내의 제1 노트 상에 저장된 정보에 액세스하도록 하는 세션이 개설되었음을 나타내는 응답이 송신되는 동작(554)으로 이어진다. 일부 실시예에서는 순서(550)에 도시되지 않은 추가적인 단계가 동작(554) 이후에 이어질 수도 있다. 예를 들어, 동작(554)에서 송신한 응답에서 지시된 바에 따라 세션이 개설되면, 노트 클러스터의 제1 노트에 저장된 정보에 액세스하기 위하여 다수의 액세스 요청이 수신될 수도 있다.
- [0049] 동작(554) 이후, 위트니스 정보에 대한 제2 요청이 동작(556)에서 수신된다. 위트니스 정보는 노트 클러스터

내에서 어떤 노드가 위트니스 서비스를 제공하는지 식별하는 정보를 포함할 수 있다. 순서(550)는 위트니스 정보를 포함하는 응답이 송신되는 동작(558)으로 이어진다. 실시예에 따르면, 위트니스 정보가 구조화되어 위트니스 서비스를 제공하는 노드 클러스터 내의 노드에 대한 지시만 포함하는 것이 아니라 일부 노드에 대한 선호도를 나타내도록 랭크(ranked)될 수도 있다.

- [0050] 동작(560)에서는 리소스에 대한 상태 통지를 등록하라는 등록 요청이 수신된다. 실시예에서 요청은 네트워크 리소스에 대한 상태 통지를 위한 것이다. 다른 실시예에서, 요청은 클러스터 리소스에 대한 상태 통지를 위한 것으로서, 예컨대 클러스터 내의 노드가 온라인인지 오프라인인지 여부와 같은 것이다. 동작(560)에서 실시예에 따라 수신되는 요청은 제1 요청(동작(552)) 및 제2 요청(동작(556))을 송신하는 클라이언트와 상이한 클라이언트로부터 수신된다는 것을 이해할 수 있을 것이다. 비록 순서(550) 동안 수신되는 요청들은 상이한 클라이언트들로부터 송신된 것일 수 있지만, 동작들(552 내지 564)은 실시예 내에서 단일 노드에서 실행된다.
- [0051] 동작(560) 이후, 순서는 리소스가 감시되는 동작(562)으로 이어진다. 동작(562)은 클러스터 리소스와 네트워크 리소스를 포함하는 다수의 리소스들을 감시하는 것을 포함할 수 있다. 리소스는 비록 그 리소스에 관한 통지를 수신하라는 요청이 현재 없다고 하더라도 감시될 수 있다. 동작(562)은 네트워크 리소스로부터 또는 클러스터 서비스로부터 이벤트를 수신하고 처리하는 것을 포함할 수 있다.
- [0052] 동작(562) 이후, 순서는 리소스에 관한 통지가 송신되는 동작(564)으로 이어진다. 통지는 동작(560)에서 수신되는 등록 요청 내에 지시된 리소스의 상태에 관한 정보를 제공한다. 일 실시예에서, 통지는 제1 노드가 오프라인이라는 것을 나타낼 수 있다. 동작(564) 이후, 순서는 동작(514)에서 종료한다. 실시예에서 순서(550)는 동작(564) 이후에 실행되는 추가적인 단계를 더 포함할 수도 있다. 일 예시로서, 이전에 오프라인이었던 리소스가 지금은 온라인이며 액세스 가능하다는 것이나, 다른 리소스가 이용가능하다는 것을 나타내는 추가적인 메시지가 송신될 수 있다.
- [0053] 일부 실시예에서, 동작(552 내지 558)은 한 장소에서 실행되고 동작(560 내지 564)은 상이한 장소에서 실행되는 것을 주목해야 한다. 예를 들어, 동작(552 내지 558)은 노드 클러스터의 노드 중 동작(560 내지 564)을 실행하는 노드와 상이한 노드 상에서 실행될 수 있다. 다른 실시예에서는, 단일 노드가 모든 동작(552 내지 564)을 실행할 수도 있으나, 동작(552 내지 558)은 제1 클라이언트에 관해 실행되고 동작(560 내지 564)은 제2 클라이언트에 관해 실행될 수 있다. 다시 말해, 노드가 제1 클라이언트로부터는 정보에 액세스하라는 요청을 서비스하면서 제2 클라이언트에게는 위트니스 서비스를 제공할 수도 있다.
- [0054] 동작 순서(600)는 클라이언트에게 분산 파일 시스템을 제공하는 클러스터로부터의 클러스터 리소스에 대한 상태 통지를 수신하기 위한 단계를 도시한다. 동작 순서(650)는 파일 시스템을 제공하는 클러스터 상의 클러스터 리소스에 대한 상태 통지를 제공하기 위한 단계를 도시한다. 실시예에서, 전술된 클라이언트(도 1의 102, 104)나 클라이언트(도 3의 302, 304)와 같은 클라이언트가 동작 순서(600)를 구현할 수도 있다. 실시예에서, 도 3에 도시된 서버 1, 서버 2, 및 서버 3과 같은 서버 클러스터(306)의 서버가 동작 순서(650)를 구현한다.
- [0055] 순서(600)는 서버 클러스터 상의 파일 시스템에 접속하라는 요청이 송신되는 동작(602)에서 개시된다. 서버 클러스터는 하나 이상의 서버를 포함하고 클라이언트에 의해 액세스되는 파일 정보를 저장한다. 일부 실시예에서, 동작(602)에서 송신되는 요청은 클러스터 상의 파일에 액세스하기 위해 서버 클러스터 내의 서버와의 세션을 개설하라는 요청이다. 실시예에 따른 세션은 SMB 또는 NFS의 버전과 같은 파일 액세스 프로토콜로 개설된다. 동작(602) 이후, 순서는 클러스터 내의 서버와의 세션이 개설되었다는 것을 나타내는 응답을 수신하는 동작(604)으로 이어진다. 동작(602)과 동작(604)은 일련의 동작들 중에서 세션을 교섭하기 위해 실행되는 단지 2개의 동작일 수 있다. 다시 말해, 다른 실시예에서는 동작(602)과 동작(604) 사이에서 실행되는 다수의 동작들이 있을 수도 있다.
- [0056] 동작(604) 이후, 액세스 요청은 서버 클러스터 내에 저장된 파일 시스템으로부터의 정보에 액세스하기 위해 동작(606)에서 서버로 송신된다. 액세스 요청은 SMB의 버전 또는 NFS의 버전과 같은 특정 파일 액세스 프로토콜에 따라 포맷된 판독/기록 요청과 같은 것을 포함할 수 있다.
- [0057] 도 6에서의 실시예에 도시된 바와 같이, 동작 순서는 동작(606)으로부터 제2 요청이 위트니스 서비스 정보를 위해 서버로 송신되는 동작(608)으로 이어진다. 다른 것들 중에서도 위트니스 서비스 정보는 클러스터 내에서 어떤 서버가 클러스터 리소스에 대한 통지를 제공하기 위해 등록될 수 있는지 식별한다. 동작(608)에서 송신되는 요청은 동작(602 내지 606)에서 사용되는 파일 액세스 프로토콜과 상이한 위트니스 프로토콜에 따라 포맷될 수도 있다.

- [0058] 동작(610)에서는 위트니스 서비스 정보가 수신된다. 그 후 위트니스 서비스 정보는 클러스터 내의 어떤 서버가 위트니스 서비스를 제공하는지 판별하는 데 사용된다. 동작(612)에서 위트니스 서비스 정보는 클러스터 리소스에 관한 통지를 등록하기 위한 등록 요청을 송신하는 데 사용된다. 일 실시예에서, 리소스는 동작(604)에서 세션이 개설되는 서버이면서, 동작(606)에서 송신되는 요청을 사용하여 정보가 액세스되는 서버가 될 수도 있다. 등록 요청은 클라이언트가 서버의 상태에 관한 통지를 수신하기 바란다는 것을 나타낼 수 있다.
- [0059] 순서(600)는 동작(612)으로부터 리소스에 관한 통지가 수신되는 동작(614)으로 이어진다. 일부 실시예에서는, 단계(614)에서의 통지의 수신에 응답하여 실행되는 추가적인 단계가 더 있을 수 있는데, 예를 들어 액세스 요청을 중지하는 것, 클러스터 내의 다른 서버로의 재접속, 및 클러스터 내의 다른 서버로의 액세스 요청의 재개 등이 있다. 동작(614) 이후에는, 동작(616)에서 등록을 해제하기 위한 요청이 송신된다. 순서(600)는 그 후 동작(618)에서 종료한다.
- [0060] 순서(650)는 순서(600)에 관해 전술된 단계들에 대응하는 단계들을 가진다. 순서(650)는 서버 클러스터로 접속하기 위한 요청을 수신하는 동작(652)에서 개시된다. 요청은 실시예에서 서버 클러스터 상에 저장된 파일 정보에 액세스하기 위한 세션을 개설하라는 요청이다. 동작은 서버 클러스터의 파일 서버 상에서 일어난다. 순서는 동작(652)으로부터 서버 클러스터 내의 파일 서버 상에 저장된 파일에 액세스하도록 하기 위해 세션이 개설되었다는 것을 나타내는 응답이 송신되는 동작(654)으로 이어진다. 전술된 것처럼, 동작(652)과 동작(654)은 세션을 개설하기 위해 실행되는 단지 2개의 동작일 수 있다. 일부 실시예에서는, 세션을 개설하기 위한 동작(652)과 동작(654) 사이에서 실행되는 다수의 다른 동작들이 있을 수도 있다.
- [0061] 동작(654) 다음에는 파일 액세스 요청이 수신되는 동작(656)이 뒤따른다. 파일 액세스 요청은 서버 상에 저장되는 파일 정보를 판독 또는 기록하기 위한 요청일 수 있다. 실시예에서, 동작(652 내지 656)에서 송수신되는 메시지는 SMB의 버전 또는 NFS의 버전과 같은 파일 액세스 프로토콜에 따라 포맷된다.
- [0062] 동작(656) 이후, 위트니스 정보에 대한 요청이 동작(658)에서 수신된다. 요청은 서버 클러스터 내의 어떤 서버가 위트니스 서비스를 제공하는지에 관한 위트니스 정보를 위한 것이다. 순서(650)는 위트니스 서비스를 제공하는 서버 클러스터 내의 서버에 대한 지시를 갖는 위트니스 정보를 포함한 응답이 송신되는 동작(660)으로 이어진다. 동작(658)에서 실시예에 따라 수신되는 요청은 제1 요청(동작(652)) 및 제2 요청(동작(658))을 송신하는 클라이언트와 상이한 클라이언트로부터 수신된다는 것을 이해할 수 있을 것이다. 비록 순서(650) 동안 수신되는 요청들은 상이한 클라이언트들로부터 송신된 것일 수 있지만, 동작들(652 내지 668)은 실시예 내에서 단일 노드에서 실행된다.
- [0063] 동작(662)에서는, 클러스터 리소스에 대한 상태 통지를 등록하기 위한 등록 요청이 수신된다. 실시예에 따르면, 요청은 예컨대 서버와의 세션이 개설되고 파일 액세스 요청을 서비스하고 있는지 아닌지 여부, 서버가 온라인인지 오프라인인지 여부와 같은 클러스터 리소스에 대한 상태 통지에 대한 것이다. 동작(662) 이후, 순서는 리소스가 감시되는 동작(664)으로 이어진다. 동작(664)은 클러스터 상에서 실행되는 클러스터 서비스에 의해 제공되는 API를 이용하는 것을 포함할 수 있다. 동작(664)은 클러스터 서비스로부터 수신된 이벤트를 수신하고 처리하는 것을 포함할 수도 있다.
- [0064] 동작(664) 이후, 순서는 리소스에 관한 통지가 송신되는 동작(666)으로 이어진다. 통지는 동작(662)에서 수신된 등록 요청 내에 지시된 리소스의 상태에 관한 정보를 제공한다. 일 실시예에서, 통지는 서버가 지금 오프라인이라는 것을 나타낼 수 있다. 동작(666) 이후, 순서는 등록을 해제하라는 요청이 수신되는 동작(668)으로 이어진다. 동작(668)에서 수신된 요청은 이전의 등록, 예컨대 동작(662)에서 수신된 등록 요청이 등록 해제되어야 함을 지시한다. 순서(650)는 동작(618)에서 종료한다.
- [0065] 일부 실시예에서, 동작(652 내지 658)은 한 서버 상에서 실행되고 동작(660 내지 664)은 상이한 서버 상에서 실행된다는 것을 주목해야 한다. 예를 들어, 동작(652 내지 658)은 서버 클러스터의 제1 서버 상에서, 동작(660 내지 664)은 서버 클러스터의 제2 서버 상에서 실행될 수 있다. 다른 실시예에서는, 단일 서버가 모든 동작(652 내지 664)을 실행할 수도 있으나, 동작(652 내지 658)은 제1 클라이언트에 관해 실행되고 동작(660 내지 664)은 제2 클라이언트에 관해 실행될 수 있다. 다시 말해, 노드가 제1 클라이언트로부터는 정보에 액세스하라는 요청을 서비스하면서 제2 클라이언트에게는 위트니스 서비스를 제공할 수도 있다.
- [0066] 도 7은 본 명세서에서 설명된 실시예들을 구현하는 데 사용될 수 있는 일반적인 컴퓨터 시스템(700)을 도시한다. 컴퓨터 시스템(700)은 컴퓨팅 환경에 대한 단지 하나의 예시일 뿐이고, 컴퓨터 및 네트워크 아키텍처의 기능이나 사용 범위에 관한 어떤 한정사항을 제안하기 위한 것은 아니다. 예시적인 컴퓨터 시스템(700)

내에 도시된 컴포넌트들 중 임의의 하나 또는 조합에 관하여 컴퓨터 시스템(700)이 종속되거나 필수적인 것으로 해석되어서는 안 된다. 실시예에서 시스템(700)은 도 1에 관하여 전술된 클라이언트 및/또는 서버로서 사용될 수 있다.

[0067] 가장 기본적인 구성에 있어서, 시스템(700)은 통상적으로 적어도 하나의 프로세싱 유닛(702)과 메모리(704)를 포함한다. 엄밀한 구성 및 컴퓨팅 장치의 타입에 따라, 메모리(704)는 휘발성(예컨대 RAM), 비-휘발성(예컨대 ROM, 플래시 메모리 등) 또는 이들의 어떤 조합이 될 수 있다. 이러한 가장 기본적인 구성은 도 7에서 파선(706)으로 도시된다. 시스템 메모리(704)는 시스템(700) 상에서 실행되는 애플리케이션을 저장한다. 예를 들어, 메모리(704)는 감시되고 있는 리소스에 관한 정보(720)를 저장할 수 있다. 메모리(704)는 통지를 위한 클라이언트 등록(722)도 포함할 수 있다.

[0068] 본 명세서에서 컴퓨터 판독 가능 매체라는 용어는 컴퓨터 저장 매체를 포함할 수 있다. 컴퓨터 저장 매체는 예컨대 컴퓨터 판독 가능 명령어, 데이터 구조, 프로그램 모듈 또는 기타 데이터와 같은 정보의 저장을 위한 임의의 방법 또는 기법으로 구현되는 휘발성 및 비휘발성, 이동형 및 고정형 매체를 포함할 수 있다. 시스템 메모리(704), 이동형 스토리지, 및 고정형 스토리지(708)이 모두 컴퓨터 저장 매체 예시들(즉, 메모리 스토리지)이다. 컴퓨터 저장 매체는 RAM, ROM, EEPROM, 플래시 메모리 또는 기타 메모리 기법, CD-ROM, DVD 또는 기타 광학 스토리지, 자기 카세트, 자기 테이프, 자기 디스크 스토리지 EH는 기타 자기 저장 장치, 또는 정보를 저장하는 데 사용되고 컴퓨팅 장치(700)에 의해 액세스될 수 있는 임의의 다른 매체를 포함할 수 있다. 임의의 이러한 컴퓨터 저장 매체는 장치(700)의 일부가 될 수 있다. 컴퓨팅 장치(700)는 키보드, 마우스, 펜, 사운드 입력 장치, 터치 입력 장치 등과 같은 입력 장치(들)(714)도 포함할 수 있다. 디스플레이, 스피커, 프린터 등과 같은 출력 장치(들)(716)도 포함될 수 있다. 전술된 장치들은 예시일 뿐이고 다른 것들이 사용될 수도 있다.

[0069] 본 명세서에서 컴퓨터 판독 가능 매체라는 용어는 통신 매체를 포함할 수도 있다. 통신 매체는 컴퓨터 판독 가능 명령어, 데이터 구조, 프로그램 모듈, 또는 변조 데이터 신호 내의 기타 데이터에 의해 구현되는 반송파나 기타 전송 메커니즘과 같은 것이 될 수 있고, 임의의 정보 전달 매체를 포함한다. "변조 데이터 신호"라는 용어는 신호 내의 부호화 정보에 관한 방식에 따라 변경되거나 설정된 하나 이상의 특성을 갖는 신호를 가리킨다. 예시로서, 통신 매체는 유선 네트워크나 직결 접속과 같은 유선 매체, 및 음향, 고주파(RF), 적외선 등과 같은 무선 매체를 포함할 수 있으나, 이에 한정되는 것은 아니다.

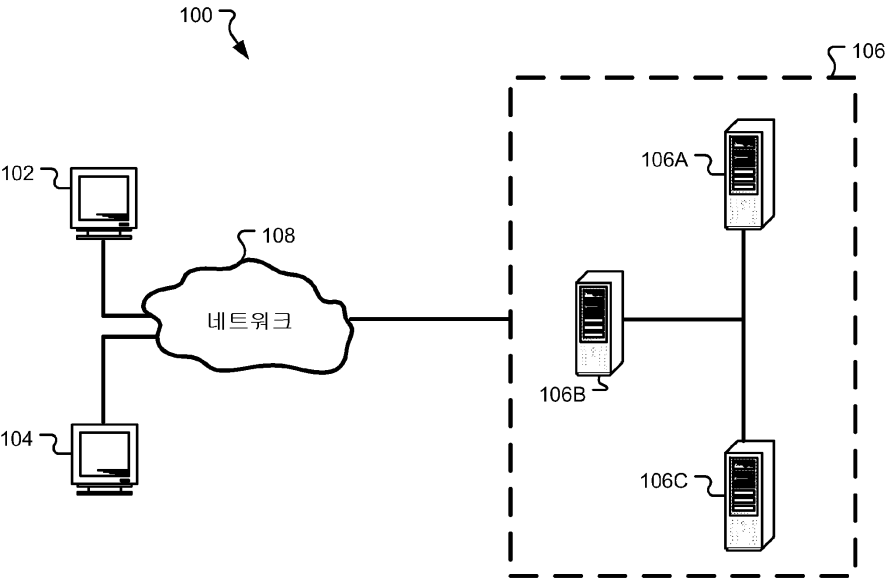
[0070] 본 명세서를 통해 일 실시예 또는 실시예로 언급된 것은 상세하게 설명된 특징, 구조, 또는 특성이 적어도 하나의 실시예 내에 포함된다는 것을 의미한다. 따라서, 이러한 구절의 사용은 하나 이상의 실시예를 언급하는 것일 수도 있다. 또한, 설명된 특징, 구조, 또는 특성은 하나 이상의 실시예에서 임의의 적합한 방식에 따라 결합될 수도 있다.

[0071] 관련 기술분야의 당업자는 실시예들이 하나 이상의 구체적인 세부 설명 없이도 실시될 수 있으며, 또는 다른 방법, 리소스, 재료 등과 함께 실시될 수도 있다는 것을 인식할 수 있을 것이다. 주지의 구조, 리소스, 또는 동작을 다른 예시들에서 자세하게 설명하거나 도시하지 않은 것은 실시예에 대한 명료하지 않은 측면을 피하기 위한 것일 뿐이다.

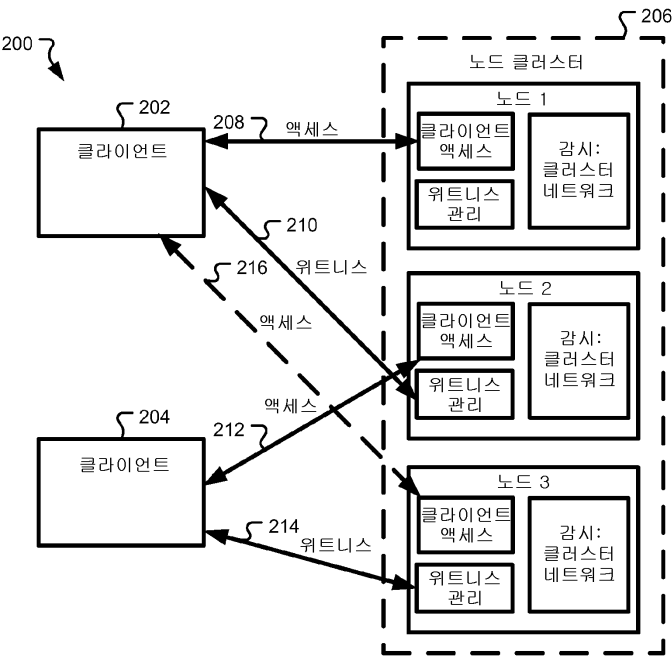
[0072] 예시적인 실시예들과 응용예들이 도시되고 설명되었으나, 실시예들이 전술된 정확한 구성 및 리소스로 한정되는 것은 아님을 이해할 수 있을 것이다. 당업자에게 명백한 다양한 수정, 변경, 및 변화는 청구된 실시예들의 범주를 벗어나지 않는 한 본 명세서에서 개시된 방법 및 시스템의 배치, 동작, 및 세부 사항을 구성할 수도 있다.

도면

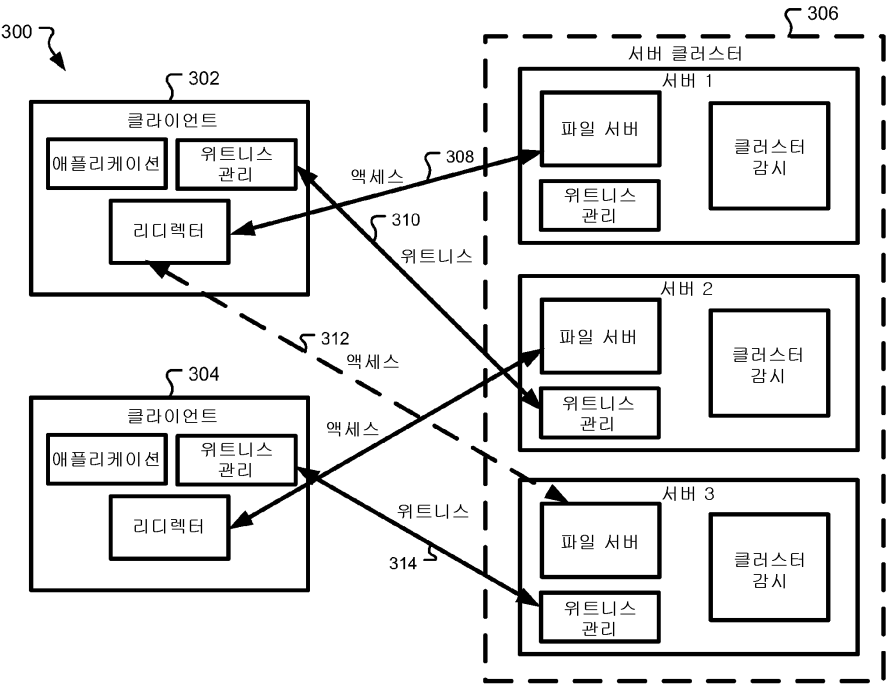
도면1



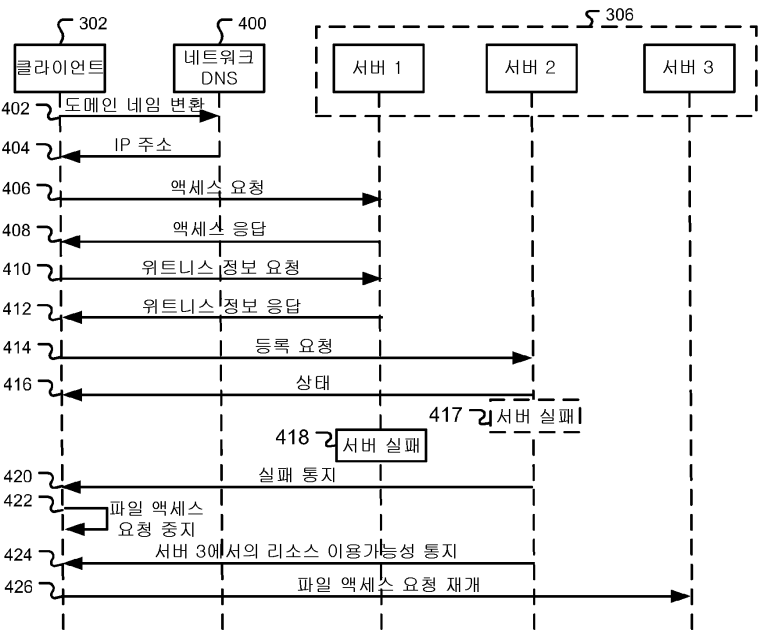
도면2



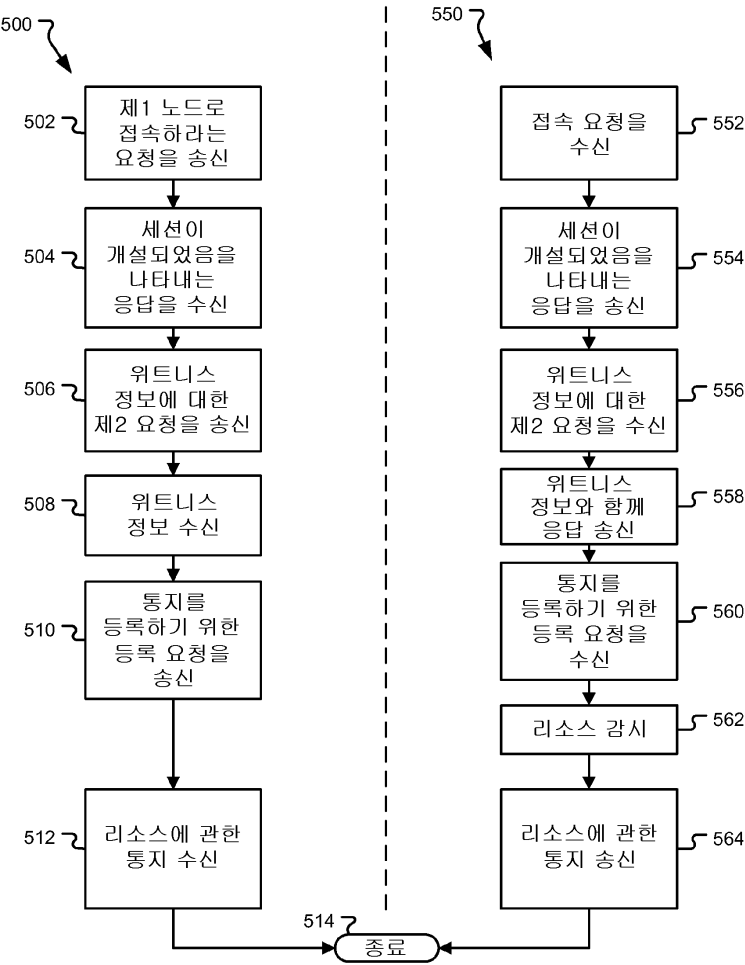
도면3



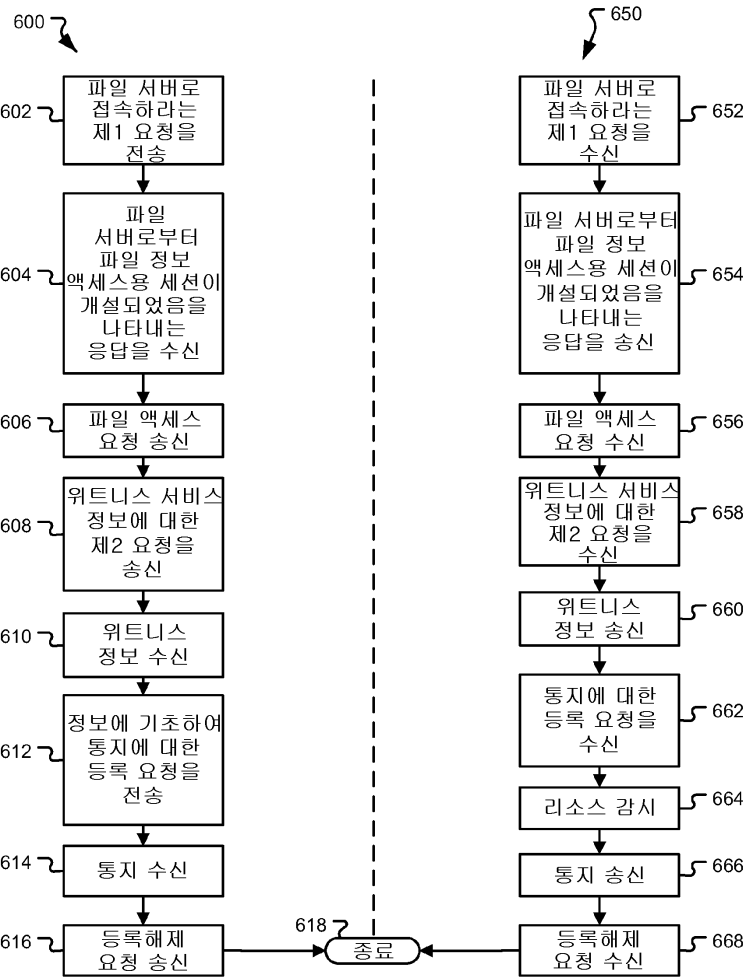
도면4



도면5



도면6



도면7

