



(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2019년01월24일
 (11) 등록번호 10-1916847
 (24) 등록일자 2018년11월02일

- (51) 국제특허분류(Int. Cl.)
G06F 15/16 (2018.01) *G06F 9/06* (2018.01)
- (21) 출원번호 10-2013-7030613
- (22) 출원일자(국제) 2012년05월18일
 심사청구일자 2017년04월14일
- (85) 번역문제출일자 2013년11월19일
- (65) 공개번호 10-2014-0026503
- (43) 공개일자 2014년03월05일
- (86) 국제출원번호 PCT/US2012/038647
- (87) 국제공개번호 WO 2012/162171
 국제공개일자 2012년11월29일
- (30) 우선권주장
 13/111,956 2011년05월20일 미국(US)
- (56) 선행기술조사문헌
 US20090300149 A1
 US20020103886 A1
 US20090300423 A1
 US20110055378 A1

- (73) 특허권자
마이크로소프트 테크놀로지 라이선싱, 엘엘씨
 미국 워싱턴주 (우편번호 : 98052) 레드몬드 원
 마이크로소프트 웨이
- (72) 발명자
아이어 카난 씨
 미국 워싱턴주 98052-6399 레드몬드 원 마이크로
 소프트 웨이 엘씨에이 - 인터내셔널 페이턴츠 마
 이크로소프트 코포레이션
왓슨 에릭 비
 미국 워싱턴주 98052-6399 레드몬드 원 마이크로
 소프트 웨이 엘씨에이 - 인터내셔널 페이턴츠 마
 이크로소프트 코포레이션
- (74) 대리인
제일특허법인(유)

전체 청구항 수 : 총 19 항

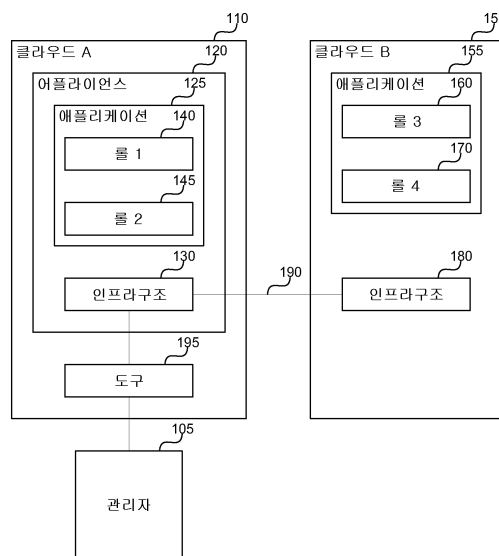
심사관 : 윤혜숙

(54) 발명의 명칭 **크로스-클라우드식 관리 및 고장수리 기법**

(57) 요약

본 명세서에는 애플리케이션이 단일 애플리케이션에서 같이 분산된 애플리케이션의 동작, 관리 및 고장수리를 할 수 있게 하는 동시에, 둘 이상의 클라우드에 걸쳐도록 하는 능력을 제공하는 클라우드 관리 시스템이 기술되었다. 본 명세서에 기술된 클라우드 관리 시스템은 애플리케이션이 둘 이상의 클라우드에 걸쳐도록 하는

(뒷면에 계속)
대표도 - 도1



능력을 제공하는 한편, 분산된 애플리케이션의 동작, 관리 및 고장수리가 단일 애플리케이션과 같게 할 수 있다. 시스템은 서로 다른 위치에서 실행 중인 애플리케이션의 인스턴스의 실행 및 그에 대한 지식의 집중화를 위해 데이터센터를 가로질러 통신하는 인프라구조를 제공한다. 인프라구조는 문제가 애플리케이션 내에 있는지 또는 클라우드 사이에서의 접속으로 인한 것인지를 알기 위한 지능을 이용하여, 클라우드들 사이의 접속 및 애플리케이션을 모두 모니터링하는 시스템에 의해 제공된다. 시스템은 복수의 클라우드 플랫폼/위치에 걸쳐 관리 기능을 조직화한다. 따라서, 복수의 클라우드를 가로질러 퍼져있는 애플리케이션들이 더욱 쉽게 모니터링, 관리 및 디버깅될 수 있도록, 클라우드 관리 시스템은 단일 모니터링 및 고장수리 인터페이스와 지식 및 실행 패브릭을 생성한다.

명세서

청구범위

청구항 1

분산된 애플리케이션 인스턴스(distributed application instances)에서 애플리케이션 관리 데이터에 액세스하도록 하는 관리 도구로부터의 요청을 처리하는 컴퓨터로 구현되는 방법으로서,

하나 이상의 데이터센터 내의 애플리케이션 실행 인스턴스와 관련된 관리 데이터에 액세스하도록 하는 관리 도구 요청을 애플리케이션 관리 도구로부터 수신하는 단계 - 상기 관리 도구 요청을 수신하는 단계는 하나 이상의 애플리케이션 인스턴스의 동작을 기술하는 상태 정보에 액세스하도록 하는 요청을 수행 모니터링 도구로부터 수신하는 단계를 포함함 - 와,

수신된 상기 요청을 만족시키는 하나 이상의 관리 데이터 유형을 식별하는 단계와,

상기 애플리케이션의 둘 이상의 인스턴스를 포함하여 상기 애플리케이션의 분산화를 결정하는 단계와,

각 분산된 애플리케이션 인스턴스로부터의 상기 요청을 만족시키도록 관리 데이터를 수집하는 단계와,

상기 수신된 관리 도구의 요청에 대해 균일한 응답을 제공하도록 상기 수집된 관리 데이터를 통합하는 단계와,

상기 수신된 관리 도구의 요청에 응답하여 상기 수집되고 통합된 관리 데이터를 보고하는 단계

를 포함하되,

상기 단계들은 적어도 하나의 프로세서에 의해 수행되는

컴퓨터로 구현되는 방법.

청구항 2

제 1 항에 있어서,

상기 관리 도구 요청을 수신하는 단계는, 관리 데이터를 요청하기 위한 도구에 노출된 프로그램의 애플리케이션 프로그래밍 인터페이스(API)를 통해 상기 관리 도구 요청을 수신하는 단계를 포함하는

컴퓨터로 구현되는 방법.

청구항 3

제 2 항에 있어서,

상기 API는 어디에서, 또는 얼마나 많은 장소에서 애플리케이션 인스턴스가 실행중인지 이해하기 위해 도구를 요구하지 않고도 관리 데이터에 액세스할 수 있게 하는 균일한 인터페이스를 포함하는

컴퓨터로 구현되는 방법.

청구항 4

제 1 항에 있어서,

상기 하나 이상의 관리 데이터 유형을 식별하는 단계는, 상기 관리 도구 요청이 상기 애플리케이션의 각 인스턴스에 의해 생성되는 정보를 요청한다고 판정하는 단계를 포함하는

컴퓨터로 구현되는 방법.

청구항 5

제 1 항에 있어서,

상기 하나 이상의 관리 데이터의 유형을 식별하는 단계는, 각 애플리케이션 인스턴스로부터 어떤 정보를 수집할 지 및 각 애플리케이션 인스턴스에 의해 중앙 위치로 이동된(pushed) 데이터로부터 상기 하나 이상의 관리 데이터의 유형이 이미 로컬에서 수집되었는지 여부를 결정하는 단계를 포함하는

컴퓨터로 구현되는 방법.

청구항 6

제 1 항에 있어서,

상기 애플리케이션의 분산화를 결정하는 단계는, 상기 애플리케이션이 실행중인 위치 및 컴퓨터 시스템이 상기 요청을 만족시키는 관리 데이터를 찾을 위치를 결정하는 단계를 포함하는

컴퓨터로 구현되는 방법.

청구항 7

제 1 항에 있어서,

상기 관리 데이터를 수집하는 단계는, 사설 데이터센터 내의 적어도 하나의 애플리케이션 인스턴스 및 클라우드 컴퓨팅 시설 내의 적어도 하나의 애플리케이션 인스턴스에 액세스하는 단계를 포함하는

컴퓨터로 구현되는 방법.

청구항 8

제 1 항에 있어서,

상기 관리 데이터를 수집하는 단계는, 상기 수신된 관리 도구 요청을 만족시키기 위해 상기 애플리케이션의 각 인스턴스에 접촉하는 단계를 포함하는

컴퓨터로 구현되는 방법.

청구항 9

제 1 항에 있어서,

하나 이상의 원격 애플리케이션 인스턴스에 하나 이상의 고장수리 커맨드(troubleshooting command)를 전송하는 단계를 더 포함하되,

상기 원격 애플리케이션 인스턴스는 상기 고장수리 커맨드를 수행하고, 요청된 데이터를 중앙 위치로 다시 보고하여 상기 관리 도구가 상기 보고된 요청 데이터와 연관된 정보에 액세스할 수 있도록 하는

컴퓨터로 구현되는 방법.

청구항 10

제 1 항에 있어서,

상기 수집된 데이터를 통합하는 단계는, 상기 관리 도구에 의해 관리되는 애플리케이션의 다양한 가능한 분산을 이해하는 것을 포함하기 위해 상기 관리 도구가 기록될 필요가 없도록 데이터를 포맷(format)하는 단계를 포함하는

컴퓨터로 구현되는 방법.

청구항 11

제 1 항에 있어서,

상기 수집된 데이터를 통합하는 단계는, 컴퓨터 시스템이 애플리케이션 부하를 처리해야 하는 필요에 따라 애플리케이션을 하나의 위치에서 다른 위치로 또는 복수의 위치들로 자유롭게 이동시킬 수 있도록 하면서, 관리자에게 일정한 관리 및 고장수리 경험을 제공하도록 데이터를 포맷하는 단계를 포함하는

컴퓨터로 구현되는 방법.

청구항 12

제 1 항에 있어서,

상기 데이터를 보고하는 단계는, 상기 관리 도구 요청이 수신되는 인터페이스를 통해서 상기 데이터를 상기 관리 도구로 전송하는 단계를 포함하는

컴퓨터로 구현되는 방법.

청구항 13

크로스-클라우드(cross-cloud)식 관리 및 고장수리를 위한 크로스-클라우드 컴퓨팅용 컴퓨터 시스템으로서,

프로세서 및 메모리와,

애플리케이션의 인스턴스가 실행중인 복수의 데이터센터 위치에 대한 정보를 관리하는 위치 관리 구성요소와,

상기 애플리케이션의 인스턴스가 실행중인 위치를 기술하는 정보를 저장하는 위치 데이터 스토어와,

상기 컴퓨터 시스템에 인터페이스를 제공하는 도구 인터페이스 구성요소 - 상기 인터페이스를 통해 하나 이상의 도구가 상기 애플리케이션에 대한 관리 및 고장수리 정보에 액세스할 수 있음 - 와,

관리 정보에 액세스하거나 애플리케이션 고장수리를 수행하기 위해 상기 도구 인터페이스 구성요소에 접속하는 하나 이상의 관리 도구와,

상기 애플리케이션이 실행중인 하나 이상의 원거리 위치에서의 관리 정보를 상기 애플리케이션의 홈 위치로 다시 이송하는 데이터 이송 구성요소와,

하나 이상의 위치에서 상기 애플리케이션에 대한 하나 이상의 고장수리 업무를 수행하는 고장수리 구성요소를 포함하되,

상기 프로세서 및 메모리는 상기 위치 관리 구성요소, 상기 도구 인터페이스 구성요소, 상기 데이터 이송 구성요소 또는 상기 고장수리 구성요소 내에 포함되는 소프트웨어 명령어를 실행하도록 구성되는,

컴퓨터 시스템.

청구항 14

제 13 항에 있어서,

상기 위치 관리 구성요소는, 각 위치에 어떻게 도달하는지, 관리 정보를 검색하기 위해 이용가능한 접속, 각 접속에 사용하기 위한 사용자 계정에 대한 자격, 및 고장수리 정보를 수집하고 고장수리 커맨드를 전송하는 하나 이상의 애플리케이션 및 데이터센터 구성요소를 기술하는 정보를 포함하는 컴퓨터 시스템.

청구항 15

제 13 항에 있어서,
 상기 위치 관리 구성요소는,
 하나의 데이터센터 또는 클라우드로부터 다른 데이터센터 또는 클라우드로의 애플리케이션 부하의 동적 이송을 기술하는 정보를 수신하고, 상기 위치 관리 구성요소가 상기 애플리케이션이 실행중인 모든 데이터센터 위치에 대한 완전한 그림을 갖도록 업데이트하는 컴퓨터 시스템.

청구항 16

제 13 항에 있어서,
 상기 도구 인터페이스 구성요소는, 사업체의 사설 데이터센터 내에 위치한 클라우드 컴퓨팅 어플라이언스 (appliance)에서 상기 애플리케이션과 관련된 정보에 액세스하기 위해 도구에 대한 초기 접속 포인트를 제공하는 컴퓨터 시스템.

청구항 17

제 13 항에 있어서,
 적어도 하나의 관리 도구는 로컬 애플리케이션과 작업하도록 설계되고, 상기 시스템은 상기 도구에 상기 적어도 하나의 관리 도구에 대한 인식 없이 복수의 위치에서 실행중인 분산된 애플리케이션을 기술하는 정보를 제공하는 컴퓨터 시스템.

청구항 18

제 13 항에 있어서,
 상기 애플리케이션이 실행중인 하나 이상의 위치와 관련된 발부(billing) 정보를 보고하는 발부 구성요소를 더 포함하는 컴퓨터 시스템.

청구항 19

원거리 애플리케이션 인스턴스의 위치로부터의 데이터를 보고하고 고장수리 요청을 처리하도록 컴퓨터 시스템을 제어하기 위한 명령어를 포함하는 컴퓨터 판독가능 저장 매체로서,
 상기 명령어는 실행 시 프로세서로 하여금,
 애플리케이션의 클라이언트로부터의 요청에 의해 생성된 부하의 일부분을 처리하는 원거리 애플리케이션 인스턴

스에서의 관리 데이터를 수신하는 것과,

관리자가 하나 이상의 분산된 원거리 위치에서 실행중인 상기 애플리케이션의 복수의 인스턴스에 의해 보고되는 관리 데이터에 액세스할 수 있는 상기 애플리케이션의 홈 위치를 결정하는 것 - 상기 홈 위치는 상기 애플리케이션이 정상적으로 실행되는 위치임 - 과,

상기 원거리 애플리케이션 인스턴스로부터 상기 애플리케이션의 결정된 홈 위치로 수신된 상기 관리 데이터를 전송하는 것과,

상기 원거리 애플리케이션 인스턴스를 고장수리하기 위해 상기 홈 위치에서 실행하는 관리 도구로부터 고장수리 요청을 수신하는 것과,

수신된 상기 고장수리 요청에 응답하여 하나 이상의 고장수리 동작을 수행하는 것과,

상기 수신된 고장수리 요청에 응답하여 상기 홈 위치에서 고장수리 결과를 전송하는 것

을 포함하는 동작을 수행하게 하는

컴퓨터 판독가능 저장 매체.

발명의 설명

기술 분야

배경 기술

[0001]

데이터센터는 큰 규모의 애플리케이션을 실행하기 위한 서버를 제공한다. 기업은 종종 판매, 홍보, 인적자원, 청구서 발부(billing), 제품 카탈로그 등과 같은 핵심 사업 기능을 수행하기 위해 데이터센터를 사용한다. 데이터센터는 또한 웹사이트, 웹서비스, 이메일 호스트, 데이터베이스 및 다수의 그외의 애플리케이션과 같은 고객-응대 애플리케이션도 실행할 수 있다. 데이터센터는 전형적으로 예상 피크 부하(peak load)를 판단하여 피크 부하 레벨을 조절하도록 서버, 네트워크 인프라구조, 냉각 및 그외의 리소스를 제공함으로써 설계된다. 데이터센터는 매우 비용이 높으며 피크가 아닌 시간에는 충분히 활용되지 않는 것으로 알려져 있다. 또한 데이터센터를 모니터링하고 데이터센터에 대한 유지보수를 수행하기 위한 장비와 인력 모두의 측면에서 상대적으로 높은 관리 비용을 갖는다. 거의 모든 사업체가 일종의 데이터센터를 이용하기 때문에, 전세계의 조직(organization)에 의해 수행되는 다수의 중복적인 기능이 존재한다.

[0002]

클라우드 컴퓨팅은 종래의 데이터센터의 최적화의 하나로서 부상하고 있다. 클라우드는 사업체에 대한 종래의 데이터센터 기능을 적어도 일부분 수행하는, 네트워크를 통해 입수가능한 리소스들의 세트(예로서, 프로세싱, 스토리지, 또는 그외의 리소스)로서 정의된다. 클라우드는 종종 추상화 계층(layer of abstraction)을 포함하여, 애플리케이션과 클라우드의 사용자가 애플리케이션이 실행중인 명확한 하드웨어와, 그 하드웨어가 위치되어 있는 곳 등을 알지 못하게 할 수 있다. 이는 클라우드 운영자에게 서비스에 투입되고 나오는 리소스의 회전, 유지보수 등의 측면에서 일부 추가적인 자유를 허용한다. 클라우드는 MICROSOFT™ Azure, Amazon Web Services 등과 같은 공용 클라우드뿐 아니라 Eucalyptus System, MICROSOFT™ 등에 의해 제공되는 것과 같은 개인용 클라우드를 포함할 수 있다. 기업들은 사업체가 데이터센터를 다양한 레벨의 클라우드 기능들과 접속시키도록 자체의 데이터센터를 둘 수 있는 어플라이언스(appliance)를 제공하기 시작하였다.

[0003]

클라우드 기반의 리소스가 레버리지된(leveraged) 경우에도, 데이터센터를 갖는 사업체는 큰 규모의 데이터센터를 증축하는 데에 상당한 비용을 겪는다. 사업체는 종종 여전히 "최악의 경우의" 피크 시나리오에 대한 계획을 세우며, 따라서 상당한 양의 하드웨어를 포함하고, 이러한 하드웨어 중 적어도 일부는 추가 프로세싱 용량, 추가 저장 공간 등의 측면에서 거의 사용되지 않거나 충분히 활용되지 않는다. 이러한 추가 리소스는 적은 양의 반환에 대해 높은 비용을 발생시킨다. 전체(premise)에 클라우드 기반 컴퓨팅을 사용하는 고객들은 피크 용량 시간에, 재해 복구 시나리오에서, 또는 단지 용량 관리를 위해 다른 호환가능 클라우드(예로서, 다른 위치에 있는 자체의 제 2 인스턴스, Microsoft 사의 공용 클라우드 등) 내의 용량을 사용할 수 있을 것으로 기대한다. 이렇게 하는 것은 최악의 경우에 대한 시나리오를 위한 증축으로 인한 중복(redundancy)을 배가시키는 것보다 비용이 훨씬 낮다. 또한, 고객들은 복수의 클라우드에 걸쳐 나누어진 애플리케이션을 관리(예로서 고장수리(troubleshooting), 운영) 할 수 있을 것으로 기대한다. 오늘날, 애플리케이션, 클라우드 관리 및 고장수리는

클라우드 또는 다른 데이터센터에 걸쳐 동작하지 않는다.

발명의 내용

해결하려는 과제

과제의 해결 수단

[0004] 본 명세서에는 애플리케이션이 단일 애플리케이션에서 같이 분산된 애플리케이션의 동작, 관리 및 고장수리 (troubleshooting)를 할 수 있게 하는 동시에, (매우 먼 거리만큼 떨어져 있는) 둘 이상의 클라우드에 걸쳐도록 하는 능력을 제공하는 클라우드 관리 시스템이 기술되었다. 이러한 시스템은 서로 다른 위치에서 실행중인 애플리케이션의 인스턴스의 실행 및 그에 대한 지식의 집중화를 위해 데이터센터를 가로질러 통신하는 인프라구조를 제공한다. 일부 경우에서, 시스템은 사업체가 자체의 사설 데이터센터를 둘 수 있게 하는 컴퓨팅 어플라이언스를 제공하며, 이는 관리자가 적어도 일부 애플리케이션 부하를 공용 클라우드 또는 다른 개별 위치로 분산시킬 수 있게 하는 동시에 컴퓨팅 어플라이언스를 통해 통합된 관리를 제공할 수 있게 한다. 인프라구조는 문제가 애플리케이션 내에 있는지 또는 클라우드 사이에서의 접속으로 인한 것인지를 알기 위한 지능을 이용하여, 클라우드들 사이의 접속 및 애플리케이션을 모두 모니터링하는 시스템에 의해 제공된다. 시스템은 복수의 클라우드 플랫폼/위치에 걸쳐 관리 기능을 조직화한다. 만약 관리자가 애플리케이션을 디버깅(debugging)하기 원한다면, 시스템은 매끄러운 통합 인터페이스를 통해 올바른 위치에서 제대로 작동하는 디버깅을 가능하게 한다. 따라서, 복수의 클라우드를 가로질러 퍼져있는 애플리케이션들이 더욱 쉽게 모니터링, 관리 및 디버깅될 수 있도록, 클라우드 관리 시스템은 단일 모니터링 및 고장수리 인터페이스와 지식 및 실행 "팩트릭(fabric)"을 생성한다.

[0005] 본 요약부는 아래의 상세한 설명에서 추가로 설명되는 개념들의 선택을 단순화된 형태로 도입하도록 제공되었다. 본 요약부는 본 발명의 청구사항의 중요 특성 또는 기본 특성을 식별하고자 하는 것이 아니며, 청구사항의 범주를 한정하도록 사용하고자 하는 것 또한 아니다.

도면의 간단한 설명

[0006] 도 1은 일 실시예에서, 연관된 관리 인프라구조를 갖는 두 개의 클라우드에서 실행중인 애플리케이션을 도시한다.

도 2는 일 실시예에서, 클라우드 관리 시스템의 구성요소를 도시한 블록도이다.

도 3은 일 실시예에서, 분산된 애플리케이션 인스턴스로부터 데이터를 액세스하기 위해 관리 도구로부터의 요청을 처리하기 위한 클라우드 관리 시스템의 프로세싱을 도시한 순서도이다.

도 4는 일 실시예에서, 원거리 애플리케이션 인스턴스의 위치로부터의 데이터를 보고하고 고장수리 요청을 처리하기 위한 클라우드 관리 시스템의 프로세싱을 도시한 순서도이다.

발명을 실시하기 위한 구체적인 내용

[0007] 본 명세서에는 애플리케이션이 단일 애플리케이션에서 같이 분산된 애플리케이션의 동작, 관리 및 고장수리 (troubleshooting)를 할 수 있게 하는 동시에, (매우 먼 거리만큼 떨어져 있을 수 있는) 둘 이상의 클라우드에 걸쳐도록 하는 능력을 제공하는 클라우드 관리 시스템이 기술되었다. 이러한 시스템은 서로 다른 위치에서 실행중인 애플리케이션의 인스턴스에 대한 인식을 실행 및 집중화하기 위해 데이터센터를 가로질러 통신하는 인프라구조를 제공한다. 예를 들어, 시스템은 애플리케이션이 어디에서 실행중인지와 무관하게, 로그 기록 (logging), 수행 추적 및 그 외의 관리 기능을 집중화할 수 있다. 일부 경우에서, 시스템은 사업체가 자체의 사설 데이터센터를 둘 수 있게 하는 컴퓨팅 어플라이언스를 제공하며, 이는 관리자가 적어도 일부 애플리케이션 부하를 공용 클라우드 또는 다른 개별 위치로 분산시킬 수 있게 하는 동시에 컴퓨팅 어플라이언스를 통해 통합된 관리를 제공한다.

- [0008] 클라우드 관리 시스템에 의해 제공되는 인프라구조는, 문제가 애플리케이션 내에 있는지 또는 클라우드 사이의 접속으로 인한 것인지를 알기 위한 지능을 이용하여, 클라우드들 사이의 접속 및 애플리케이션을 모두 모니터링한다. 이러한 시스템은 복수의 클라우드 플랫폼/위치를 가로지르는 관리 기능을 조직화한다(하나의 클라우드의 인프라구조로부터, 태스크가 둘 이상의 클라우드를 가로질러 실행하도록 조직화된다). 만약 관리자가 애플리케이션을 디버깅(debugging)하기 원한다면, 시스템은 매끄러운 통합 인터페이스를 통해 올바른 위치에서의 제대로 작동하는 디버깅을 가능하게 한다. 따라서, 복수의 클라우드 전체에 퍼져있는 애플리케이션이 보다 쉽게 모니터링, 관리 및 디버깅될 수 있도록, 클라우드 관리 시스템은 복수의 클라우드를 가로지르는 단일 모니터링 및 고장수리 인터페이스와 인식 및 실행 "팩트릭(fabric)"을 생성한다.
- [0009] 도 1은 일 실시예에서 연관된 관리 인프라구조를 갖는 두 개의 클라우드에서 실행중인 애플리케이션을 도시한다. 일부 실시예에서, 클라우드 관리 시스템은 애플리케이션을 완전히 모니터링/고장수리 할 수 있도록 데이터를 구비/모든 위치에서의 데이터에 대한 액세스를 갖는 하나의 클라우드 내의 인프라구조를 이용하여 애플리케이션(및/또는 관리자)을 포함한다. 예시로서, 도 1에 도시된 클라우드(110, 150)와 같은 두 클라우드에서 실행중인 인스턴스를 갖는 애플리케이션을 고려하자. 클라우드(110)는 인프라구조(130)를 포함하는 MICROSOFT™ Azure 어플라이언스 인스턴스(120)를 포함한다. 어플라이언스 인스턴스(120)는 롤(role)(140, 145)을 실행중인 애플리케이션 인스턴스(125)를 포함한다. 제 2 클라우드(150)는 롤(160, 170)을 실행중인 애플리케이션 인스턴스(155)를 포함한다. 제 2 클라우드(150)는 또한 인프라구조(180)를 포함한다. 어플라이언스 인스턴스(120)는 각각의 롤에 대해 알고 있으며 이들은 동일한 애플리케이션의 일부임을 알고 있다. 각 위치에서 플럼빙하는(plumbing) 인프라구조는 어플라이언스 인스턴스(120)가 제 2 클라우드(150)에서 실행중인 롤(160, 170)에 대한 정보를 검색하게 할 수 있다. 시스템은 개별적인 롤, 전체 애플리케이션, 또는 둘 모두를 분산시킬 수 있다. (예로서, 애플리케이션, 장치 및 인프라구조로부터의 로그와 같은) 모든 관리 데이터를 이용하여, 시스템은 모든 롤이 사전정의된 건강 롤을 적용함으로써 로컬화되는 것과 같이 애플리케이션의 건강을 평가할 수 있다. 시스템은 만약 애플리케이션 또는 인프라구조/네트워크에서 문제가 발생하였을 경우의 평가를 위해서 양 위치 사이의 접속(190)뿐 아니라 양 위치를 가로지른 인프라구조의 건강도 볼 수 있다.
- [0010] 유사하게, 자동화 또는 수동 고장수리 또는 개선 단계가 필요할 때, 클라우드(110) 내의 인프라구조(130)는 고장수리 및 디버깅 지원을 제공하기 위해서 클라우드(150) 내의 인프라구조(180)와 조직화할 수 있다. 예를 들어, 시스템 팩트릭은 애플리케이션 광범위 업데이트, 셋다운 등을 실행하기 위해 위치를 가로질러 도달할 수 있다. 당업자는 위치를 가로지르는 제어를 수행하는 다수의 방식을 인지하고 있을 것이다. 예를 들어, 인프라구조(130)는 인프라구조(180)를 직접 제어할 수 있고, 인프라구조(130)는 인프라구조(180)에 인프라구조(130) 대신 실행할 것 등을 요청할 수 있다. 유사하게, 운영자/관리자 고장수리 도구(예컨대, 모니터링 시각화, 경고, 로그 기록 및 구성 데이터 열람 등)를 이용하여 애플리케이션 및 인프라구조의 위치가 입수가능하며 논리적으로 디스플레이되지만, 이들을 합하기 위한 관리자로부터의 개별적인 도구 및 정신적 훈련을 포함하지는 않는다. 예를 들어 모든 롤에 대한 고장수리 및 데이터 열람시에, 만약 관리자(105)의 다음 단계가 애플리케이션의 로그를 열람하거나 롤 인스턴스에 대한 원거리 세션을 시작하기 위해 하나 이상의 도구(195)를 이용하는 것이라면, 롤이 존재하는 위치와 무관하게 시스템은 관리자(105)를 직접 접속시킨다.
- [0011] 클라우드 관리 시스템의 설계는 복수의 클라우드/위치를 가로지르는 서비스의 단순화되고 일관된 실행을 제공한다. 시스템은 "컴퓨팅 리소스"의 정의를 서버로부터 데이터센터를 넘어 인터넷의 일부로 이동시킨다(데이터센터들과 그 사이의 접속). 이것은 SLA(service level agreement)가 서비스 소유자가 종종 가장 신경 쓰는 서비스 레벨에서 정의, 모니터링 및 관리되는 것을 가능하게 한다.
- [0012] 일부 실시예에서, 클라우드 관리 시스템은 버스팅(bursting)이라 불리는, 필요에 따라 애플리케이션을 하나의 위치로부터 다른 위치로 균일하게 이동시키는 클라우드 이송 시스템(cloud migration system)과 협력하여 동작한다. 클라우드 이송 시스템은 피크 부하 컨디션을 검출하여 자동으로 컴퓨팅을 다른 소스로 (그리고 다시 원래대로) 이동시킴으로써, 그리고 둘 이상의 클라우드를 가로질러 컴퓨팅을 제공하고 하나의 사이트에서 재해가 발생한 경우에 하나로 완전히 이동시킴으로써, 용량 관리 및 재해 복구를 제공한다. 이는 사업체가 일관된 레벨의 부하에 대한 로컬 리소스를 계획하고 피크 또는 다른 일상적이지 않은 부하에 대해 클라우드 기반의 리소스를 레버리지하는 것을 가능하게 한다. 다수의 경우에, 사업체의 사업은 한 해의 특정 시기에 더욱 바쁘며, 추가 리소스가 해당 시기 동안에만 필요할 수 있다. 예를 들어, 세무전략 사업체는 4월 중반에 특히 바쁘고, 전자 상거래 사이트는 추수감사절과 크리스마스 등의 기간에 연휴 폭주를 경험한다. 클라우드 이송 시스템은 데이터센터 내에서의 부하를 모니터하고 현재 부하가 데이터센터의 용량 근처에 있음을 나타내는 문턱값을 검출한다. 예를 들어, 시스템은 데이터센터가 현재 부하를 얼마나 잘 처리하고 있는지를 결정하기 위해 중앙 처리

장치(CPU) 사용도, 메모리 사용도, 스토리지 사용도, 네트워크 대역폭 및 그외의 메트릭을 모니터링할 수 있다. 시스템은 또한 문턱값에 도달하였는지 또는 곧 도달할 것인지 여부를 결정하기 위해 추세(예컨대, 리소스 사용의 가속도)를 관찰할 수도 있다.

[0013] 문턱값에 도달할 것임을 검출하면, 클라우드 이송 시스템은 데이터센터 부하의 적어도 일부를 다른 데이터센터 또는 클라우드 기반의 리소스로 질서있게 이동시키는 것을 용이하게 한다. 예를 들어, 시스템은 일부 피크 부하를 공용 클라우드로 이동시킬 수 있다. 클라우드 가격 모델이 다양할 수 있기 때문에, 시스템은 결정에 있어서 비용을 고려할 수 있다. 예를 들어, 시스템은 비용을 감소시키기 위해서 사업체 데이터센터에서 가능한 한 많은 부하를 호스팅할 것을 선호할 수 있는 반면, 클라이언트 요청을 만족시키기 위해 요구되는 정도까지의 클라우드 리소스만을 레버리징할 수 있다. 시스템은 또한 특정한 부하가 실행중인 장소와 무관하게 (예로서, 사업체 내에서 로컬로 또는 클라우드를 이용해 공개적으로) IT 인력에 일관적인 경험을 제공하는 관리 및 모니터링 도구를 제공할 수 있다. 시스템은 또한 높은 부하를 겪는 동안 다른 리소스로 이동하기 위한 적절한 작업부하 또는 애플리케이션을 결정하는 것을 돕도록 계획 도구(planning tool)를 제공할 수 있다. 예를 들어, 애플리케이션은 이송에 더욱 적합하게 하거나 덜 적합하게 하는 다양한 규정/규제 또는 네트워크/설계 제한을 가질 수 있다. 시스템은 또한 재해 시에 빠른 작업부하 전이를 관리하기 위해서 데이터센터/네트워크 레벨에서 재해 복구 아키텍처로서 사용될 수도 있다. 만약 데이터센터 리소스가 영구적으로 고장 난다면, 시스템은 재빠르고 효율적으로 추가적인 부하를 클라우드 또는 다른 리소스로 이송하여 데이터센터의 클라이언트가 고장에 의해 영향을 받지 않거나 영향을 덜 받도록 할 수 있다. 따라서, 클라우드 이송 시스템은 사업체가 드물게 발생하는 추가 부하를 위해 다른 리소스를 레버리징하는 더 작고 보다 효율적인 데이터센터를 건설할 수 있게 한다.

[0014] 클라우드 관리 시스템은 애플리케이션이 하나의 위치에서 다른 위치로 이송될 때 매끄러운 관리 및 고장수리를 제공하도록 클라우드 이송 시스템과 함께 작동한다. 전술된 바와 같이, 클라우드 이송 시스템은 데이터센터와 클라우드 사이에서 일시적으로(즉, 버스팅) 또는 영구적으로(즉, 재해 복구) 리소스를 이동시킬 수 있다. 일시적인 이동은 피크 또는 데이터센터의 용량을 초과하는 그 외의 높은 부하를 처리하도록 짧은 시간 동안 애플리케이션 또는 다른 부하를 버스팅하는 것을 포함한다. 일시적인 이동은 전체 애플리케이션을 버스팅하거나 또는 애플리케이션의 부하를 둘 이상의 위치를 가로질러 나누는 것을 포함할 수 있다. 영구적인 이동은 데이터센터 내의 하드웨어 고장으로 인한 부하의 장기간 이동, 용량 수요의 보다 지속적인 증가, 동적 부하 균형(dynamic load balancing)을 갖는 애플리케이션의 전세계적 분산에 대한 요구 등을 포함한다. 아래에 시스템이 사업체에 의해 사용될 수 있는 몇몇 예시적인 시나리오가 기술되었다.

[0015] 제 1 예시에서, 사업체는 용량을 관리하기 위해 애플리케이션 부하를 공용 클라우드에 버스팅한다. 사업상 결정자(즉, CEO, CFO, 또는 VIP 홍보/판매) 및 데이터센터 시스템 관리자는 이것이 보다 비용 효율적일 것이라고 판단하고, 일 년 중 사용/트래픽의 피크 레벨이 가장 높은 3일에 일부 작업을 공용 클라우드로 버스팅하도록 보다 나은 고객 경험을 제공하며, 그들의 매월 사용 레벨 피크에는 자체의 데이터센터를 유지한다(잠재적으로 클라우드 어플라이언스와 함께). 이들은 클라우드에 작업을 버스팅하고 언제 얼마나 많은 작업량이 될 것인지에 대한 추정을 예측하도록 클라우드 제공자와의 사업 협약에 사인한다. 이들의 계정이 셋업되면 정보가 클라우드 어플라이언스로 입력된다. 계획 기간 동안에, 관리자는 접속이 적절하게 작동하는지를 보장하도록 클라우드 제공자로부터의 테스트 애플리케이션으로 테스트를 실행한다. 그 다음 관리자는 용량 관리 도구에서 특정한 레벨에 용량을 유지시키는 애플리케이션의 버스팅을 시작하기 위한 용량값(예컨대, 문턱값)을 설정한다. 관리자는 이러한 상황에서 이동할 자격이 있는 애플리케이션을 추가로 명시하기 위한 도구로 들어간다(예컨대, 우수한 기술적 피트(fit), 일시적인 움직임에 규제를 가하지 않는다).

[0016] 사용이 한계를 초과하면 시스템은 자동으로 애플리케이션을 공용 클라우드로 이동시킨다. 언제 용량이 시작된 버스팅의 5% 내에 들어오는지, 시스템이 언제 버스팅하는지, 시스템이 무엇을 버스팅하는지, 그리고 언제 시스템이 애플리케이션을 다시 가져오는지에 대한 정보가 모니터링/사용도 시스템에서 발생한다. 명확한 로그가 이동된 모든 컴퓨팅 리소스 및/또는 스토리지를 보관하며, 관리자는 청구서 발부(billing)를 위해 자신의 공용 클라우드 계정으로 가도록 알림을 받는다. 이동가능한 것으로 태그된 애플리케이션 및 버스팅 파라미터의 리뷰는 사업체 데이터센터 그룹 및 관리 내에서의 정기적인 용량 계획 미팅에서 리뷰된다.

[0017] 제 2 예시에서, 사업체는 용량을 관리하기 위해서 애플리케이션을 클라우드들에 걸쳐 나눈다. 이 시나리오는 이동되는 애플리케이션의 유형이 보다 복잡하여 서로 다르게 우선순위를 매기도록 나눈다는 점을 제외하면 전술된 시나리오와 유사하다. 기업체는 애플리케이션을 클라우드로 나누기 위해(버스팅 형식), 클라우드 제공자와 관계를 갖도록 결정한다. 이러한 경우에, 큰 애플리케이션은 버스팅 후보로서 사전-식별된다. 용량이 문턱값에 도달하면, 100 중 50의 작업자 인스턴스가 자동으로 공용 클라우드로 이동된다. 애플리케이션은 이제 두 개

의 어플라이언스 인스턴스 또는 클라우드 인스턴스로 나누어지며, 모든 모니터링 및 청구서 발부 데이터는 중점적으로 관리될 수 있도록 시작 인스턴스로 전송된다. 사업체 자체의 데이터센터 내의 클라우드 어플라이언스는 나누어진 애플리케이션의 가능한 문제점을 디버깅하는 것을 돕기 위한 고장수리 도구(troubleshooting tool)를 구비한다(예컨대, 네트워킹 문제, 네트워크 대역폭/지연 문제, 패브릭 통신 문제 등). 용량 상황이 어플라이언스 상에서 진정되었다면, 50의 작업자 인스턴스가 다시 어플라이언스로 이동되어 다시 정상적으로 기능 하는 애플리케이션이 된다.

[0018] 다른 예시에서, 클라우드 제공자는 하나의 클러스터를 다른 클러스터로 버스팅하도록 결정한다. 공용 클라우드 용량 계획 팀은 Chicago 지역 데이터센터 내의 클러스터가 아슬아슬하게 가득 채워졌지만 높은 활용도를 유지하길 원한다고 판단한다. 이들은 활용도가 90%에 도달할 때 West coast 지역 데이터센터의 충분히 활용되고 있지 않은 클러스터로의 버스팅을 셋업한다. 관리자는 용량 관리 도구로 들어가서 이동 후보가 될 적절한 고객/애플리케이션(예컨대, 낮은 데이터 사용률을 가짐)을 선택한다. Chicago 지역 클러스터의 이용도가 문턱값에 도달하면, 시스템을 그날 하루 동안 선택된 애플리케이션(예컨대, 클러스터의 애플리케이션의 10%)을 West coast 지역 데이터센터로 자동으로 이동시킨다. 이용도가 문턱값 아래로 복귀하면, 시스템은 애플리케이션을 다시 Chicago 지역으로 이동시킨다. 시스템은 버스팅에 대한 사전대책을 강구하는 지정된 모니터링 팀이 고객 질문에 대응할 수 있음을 통지한다.

[0019] 다른 예시에서, 시스템은 크로스-클라우드(cross-cloud) 포트폴리오 관리에 사용된다. 사업체는 자체의 클라우드 어플라이언스 상에서 용량을 효율적으로 관리하기 위해서, 모든 가변적인 수요의 애플리케이션을 공용 클라우드에 두고 일정한 수요의 애플리케이션을 어플라이언스 또는 로컬 데이터센터 리소스에 두기를 (그리고 그에 따라 어플라이언스를 보다 높은 활용도로 구동할 수 있을 것을) 원한다고 결정한다. 사업체는 자신의 컴퓨팅 리소스를 나누기를 원하는 동시에, 자신의 애플리케이션 개발자들이 동일한 방식으로 애플리케이션을 관리하기 위해, 그리고 둘을 가로지르는 부서별 청구서 발부의 단일 시야를 유지하기 위해서(예를 들어, 고객 판매 그룹, 내부 IT, B2B 세일 등에 어느 비용이 할당되는지) 여전히 자신의 모든 애플리케이션의 건강을 가로지르는 세계적인 시야를 원한다. 사업체는 어플라이언스와 동일한 그룹을 갖는 공용 클라우드와의 통합 계정을 셋업하고 사업체 측에서 통합하기 위한 청구서 발부 데이터를 획득할 수 있다. 유사하게, 애플리케이션 레벨 모니터링 뿐만 아니라 애플리케이션이 실행중인 플랫폼에 대한 데이터를 모니터링하는 공용 클라우드에 대한 애플리케이션-프로그래밍 인터페이스(API) 액세스를 획득할 수 있으며, 따라서 네트워크 운영 센터(NOC)는 사업체의 컴퓨팅 활동의 상태에 대한 완전하고 일정한 시야를 갖는다.

[0020] 다른 예시에서, 사업체는 동적 부하 균형을 갖는 전 세계적으로 분산된 애플리케이션을 셋업한다. 사업체 고객은 둘 이상의 클라우드 인스턴스에 걸쳐 용량을 관리하길 원하고 상당한 양의 부하를 독립적으로, 그러나 지리적으로 분산된 인스턴스로 갖는다(예를 들어, Bing은 독일어 쿼리를 수행하는 US 및 UK 데이터센터를 검색한다). 일반적인 환경 하에서, 글로벌 트래픽 매니저는 50%의 트래픽을 각 위치에 전송한다. 주요한 위치에서 부하가 높아지면, 시스템은 부하 밸런서(load balancer)에게 트래픽의 75%를 UK 시스템으로 전송하도록 지시하며, 그에 따라 US 클라우드 인스턴스로부터의 용량을 자유롭게 하고, 허용가능한 레벨로 만든다. 용량이 정상치로 복귀하면, 시스템은 부하 밸런서에게 50/50 스플릿(split)으로 복귀하도록 지시한다. 이러한 변화는 공용 클라우드가 제 2의 데이터센터로 사용되기 위한 것이다(1%의 부하로 말하면, 어플라이언스를 이용하는 고객의 사이트는 다른 99%가 된다). 재해 또는 고객의 사이트로부터 부하를 이동해야 할 다른 이유가 존재할 경우에, 100%의 트래픽이 공용 클라우드로 이동된다.

[0021] 다른 예시에서, 사업체는 자신의 데이터센터의 용량에 도달하여 추가 컴퓨팅 리소스를 필요로 하지만, 아직 데이터센터를 확장하기 위해 지불하기 위한 이용가능한 자금을 가지고 있지 않다. 이러한 경우에, 기업은 그들이 하드웨어 구매를 완료할 수 있을 때까지 넘치는 부하에 대해 공용 클라우드를 사용할 수 있다.

[0022] 도 2는 일 실시예에서 클라우드 관리 시스템의 구성요소들을 도시한 블록도이다. 시스템(200)은 위치 관리 구성요소(210), 위치 데이터 스토어(220), 도구 인터페이스 구성요소(230), 하나 이상의 관리 도구(240), 데이터 이송 구성요소(250), 고장수리 구성요소(260) 및 청구서 발부 구성요소(270)를 포함한다. 이들 각각의 구성요소는 아래에서 보다 자세하게 기술된다.

[0023] 위치 관리 구성요소(210)는 애플리케이션의 인스턴스가 실행중인 복수의 데이터센터 위치에 대한 정보를 관리한다. 구성요소(210)는 각 위치에 어떻게 도달하는지, 관리 정보를 검색하기 위해 이용가능한 접속, 연관된 보안 자격증을 갖는 각 위치에 대해 사용하기 위한 사용자 계정, 고장수리 정보를 수집하고 고장수리 커맨드를 전송하는 애플리케이션 및 데이터센터 구성요소 등을 기술하는 정보를 포함한다. 위치 관리 구성요소(210)는 애플

리케이션 부하의 이송 또는 하나의 데이터센터/클라우드로부터 다른 데이터센터/클라우드로의 버스팅을 기술하는 정보를 수신하여, 구성요소(210)가 애플리케이션이 실행중인 모든 위치에 대한 완전한 그림을 가질 수 있도록 관리되는 정보를 업데이트한다. 이는 애플리케이션이 어디에서 실행중인지 또는 얼마나 많은 위치에서 애플리케이션이 실행중인지와 무관하게, 시스템(200)이 완전한 그림을 제시하고 애플리케이션 관리를 균일하게 할 수 있게 한다. 조건이 변화하고 애플리케이션이 분산됨에 따라, 위치 관리 구성요소(210)는 관리 데이터의 포괄적인 세트를 갖는 관리 도구를 제시할 수 있다.

[0024] 위치 데이터 스토어(220)는 애플리케이션의 인스턴스가 실행중인 위치를 기술하는 정보를 저장한다. 데이터 스토어(220)는 하나 이상의 파일, 파일 시스템, 하드 드라이브, 데이터베이스, 클라우드 기반의 저장 서비스, 또는 시스템(200)의 세션들 사이의 정보를 지속시키기 위한 다른 기구를 포함할 수 있다. 저장된 정보는 접속 정보, 사용자 룰, 관리 데이터의 소스, 이용가능한 로그 파일 및 복수의 위치로 분산된 애플리케이션의 관리 또는 고장수리와 관련된 임의의 다른 정보를 포함할 수 있다.

[0025] 도구 인터페이스 구성요소(230)는 시스템(200)에 대한 인터페이스를 제공하며, 도구 인터페이스 구성요소(230)를 통해 하나 이상의 도구가 애플리케이션에 대한 관리 및 고장수리 정보에 액세스할 수 있다. 인터페이스는 하나 이상의 웹페이지, 웹서비스, API, 또는 관리자나 도구가 시스템(200)의 관리 및 고장수리 정보에 직접적인 또는 프로그램적인 액세스를 할 수 있게 하는 다른 인터페이스를 포함할 수 있다. 일부 실시예에서, 도구 인터페이스 구성요소(230)는 사업체의 사설 데이터센터 내에 위치한 클라우드-컴퓨팅 어플라이언스에서 애플리케이션과 관련된 정보에 액세스하기 위한 도구에 대해 초기 접속 포인트를 제공한다. 어플라이언스는 공용 클라우드 또는 다른 데이터센터로의 애플리케이션 부하의 이송 및 분산을 관리할 수 있으며, 관리 정보를 수집하는 도구를 위한 중심 접속 포인트를 제공하거나 또는 애플리케이션 고장수리를 제공할 수 있다.

[0026] 하나 이상의 관리 도구(240)는 관리 정보 액세스 또는 애플리케이션 고장수리 수행을 위해 도구 인터페이스 구성요소(230)에 접속한다. 하나 이상의 관리 도구(240)는 로그 뷰어, 보고 도구, 디버깅 도구, 또는 실행중인 애플리케이션의 문제의 해결과 관련된 또는 해결을 돕는 정보를 디스플레이하는 다른 도구를 포함할 수 있다. 관리 도구(240)는 로컬 애플리케이션과 작업하도록 설계된 도구를 포함할 수 있으며, 시스템(200)은 도구에 대해 도구에 대한 인식 없이 복수의 위치에서 실행중인 분산된 애플리케이션을 기술하는 정보를 제공한다. 이것은 관리자가 의존하는 현존하는 도구가 자동 애플리케이션 부하 이송이 데이터센터 또는 클라우드에 도입되고 있을 때에도 사용되도록 할 수 있다. 다른 경우에, 도구는 분산된 애플리케이션을 이해하고 복수의 위치와 관련된 특정한 관리 정보 또는 고장수리를 제공하도록 명확하게 기록될 수 있다. 도구 인터페이스 구성요소(230)는 복수의 인터페이스를 제공할 수 있으며, 이러한 복수의 인터페이스를 통해서 관리 도구(240)가 각 도구에 의해 이해되는 패러다임을 이용하는 시스템(200)에 접속한다.

[0027] 데이터 이송 구성요소(250)는 애플리케이션이 실행중인 하나 이상의 원거리 위치에 있는 관리 정보를 다시 애플리케이션의 홈 위치로 이송한다. 홈 위치는 사설 데이터센터, 위치 또는 클라우드 컴퓨팅 어플라이언스, 또는 애플리케이션이 고정적인 조건 하에서 일상적으로 실행하는 다른 위치를 포함할 수 있다. 소정의 레벨의 부하에 도달하면(예컨대, 피크 또는 주기적 버스트), 애플리케이션은 클라이언트 요청을 만족시키는 것을 돕기 위해 일부 부하를 하나 이상의 다른 데이터센터 또는 공용 클라우드로 이동시킬 수 있다. 이러한 다른 위치는 홈 위치와 같이 로그 파일, 거래 데이터 등과 같은 관리 데이터를 생성하며, 데이터 이송 구성요소(250)는 이러한 데이터를 다시 홈 위치로 이동시키거나 또는 홈 위치로부터의 데이터에 대한 액세스를 제공하며, 관리 도구(240)가 애플리케이션의 활동의 포괄적인 그림을 관리자에게 제공할 수 있게 한다.

[0028] 고장수리 구성요소(260)는 하나 이상의 위치에서의 애플리케이션에 대한 고장수리 업무를 수행한다. 고장수리는 디버깅, 테스트 데이터 프로세싱, 또는 애플리케이션이 올바르게 동작하는지 여부를 결정하는 다른 형태를 포함할 수 있다. 고장수리는 일반적으로 홈 위치에서 잘 이해되지만, 애플리케이션이 복수의 데이터센터 또는 클라우드에 걸치기 시작하면서 보다 복잡해진다. 클라우드 관리 시스템(200)은 균일한 인터페이스를 제공함으로써 이러한 복잡도로부터 관리 도구(240) 및 관리자를 분리시킴, 이러한 인터페이스를 통해 도구 및 관리자가 관리 정보에 액세스하고 복수의 위치에서 고장수리를 수행한다. 따라서, 만약 관리 도구가 관리자로부터 하여금 홈 위치에서의 애플리케이션 코드의 특정한 조각으로부터 트레이스 정보를 수신하거나 중단점(breakpoint)을 위치시킬 수 있게 한다면, 고장수리 구성요소(260)는 애플리케이션의 원거리 클라우드 기반 인스턴스에서도 이를 용이하게 만들 수 있다. 도구 및 관리자는 애플리케이션이 실행중인 모든 위치를 인식하지조차 못할 수 있지만, 마치 애플리케이션이 홈 위치에서만 실행중인 것처럼 관리 업무를 수행할 수 있다.

[0029] 청구서 발부 구성요소(270)는 애플리케이션이 실행중인 하나 이상의 위치와 관련된 청구서 발부 정보를 보고한

다. 하나의 공통적인 관리 업무는 컴퓨팅 비용을 관리하는 것이며, 공용 클라우드는 종종 작업부하와 관련된 메트릭에 기초하여 (예컨대, 컴퓨팅 시간, 사용된 저장 공간, 등) 청구한다. 다양한 위치에서 애플리케이션 인스턴스가 발생시키는 비용의 그림을 모으는 것이 관리자에게 유용할 수 있으며, 클라우드 관리 시스템(200)은 정보가 관리 도구 및 리포트를 통해 기록될 수 있도록 이러한 유형의 정보를 수집하기 위해 청구서 발부 구성요소(270)를 선택적으로 제공할 수 있다.

[0030] 클라우드 관리 시스템이 구현되는 컴퓨팅 디바이스는 중앙 처리 장치, 메모리, 입력 디바이스(예컨대, 키보드 및 포인팅 디바이스), 출력 디바이스(예컨대, 디스플레이 디바이스) 및 저장 디바이스(예컨대, 디스크 드라이브 또는 다른 비휘발성 저장 매체)를 포함할 수 있다. 메모리 및 저장 디바이스는 시스템을 구현하거나 가능하게 하는 컴퓨터 실행가능한 명령(예컨대, 소프트웨어)으로 인코딩될 수 있는 컴퓨터 판독가능한 저장 매체이다. 또한, 데이터 구조 및 메시지 구조가 통신 링크 상의 신호와 같이 데이터 전송 매체를 통해 저장 또는 전송될 수 있다. 인터넷, 로컬 영역 네트워크, 광역 네트워크, 포인트-투-포인트 다이얼-업 접속(point-to-point dial-up connection), 셀 폰 네트워크 등과 같은 다양한 통신 링크가 사용될 수 있다.

[0031] 시스템의 실시예는 다양한 운영 환경에서 구현될 수 있으며, 이러한 운영 환경은 개인 컴퓨터, 서버 컴퓨터, 휴대용 또는 랩탑 디바이스, 멀티프로세서 시스템, 마이크로프로세서 기반의 시스템, 프로그램 가능한 소비자 전자기기, 디지털 카메라, 네트워크 PC, 미니컴퓨터, 메인프레임 컴퓨터, 전술된 임의의 시스템 또는 디바이스를 포함하는 분산 컴퓨팅 환경, 셋톱 박스, SOC(system on chip), 등을 포함한다. 컴퓨터 시스템은 셀 폰, PDA, 스마트폰, 개인 컴퓨터, 프로그램 가능한 소비자 전자기기, 디지털 카메라 등일 수 있다.

[0032] 시스템은 프로그램 모듈과 같은, 하나 이상의 컴퓨터 또는 다른 디바이스에 의해 실행되는 컴퓨터 실행가능한 명령의 일반적인 맥락에서 기술될 수 있다. 일반적으로, 프로그램 모듈은 루틴, 프로그램, 객체, 구성요소, 데이터 구조 등 특정한 업무를 수행하거나 특정한 추출 데이터 유형을 구현하는 것을 포함한다. 전형적으로 프로그램 모듈의 기능은 다양한 실시예에서 필요에 따라 결합되거나 분산될 수 있다.

[0033] 도 3은 일 실시예에서, 분산된 애플리케이션 인스턴스로부터 데이터를 액세스하도록 관리 도구로부터의 요청을 처리하기 위한 클라우드 관리 시스템의 프로세싱을 도시한 순서도이다. 블록(310)에서 시작하면, 관리 도구로부터 하나 이상의 데이터센터 내의 애플리케이션 실행 인스턴스와 관련된 관리 데이터에 액세스하라는 요청을 수신한다. 예를 들어, 수행-모니터링 도구는 애플리케이션이 얼마나 많은 클라이언트 요청을 처리하는지, 애플리케이션의 리소스 이용도를 기술하는 상태 정보, 또는 애플리케이션으로부터의 다른 정보를 요청할 수 있다. 시스템은 관리 데이터를 요청하기 위해 시스템이 도구에 노출시키는 API를 통해 도구 요청을 수신할 수 있다. 애플리케이션 인스턴스가 어디에서 실행중인지 또는 얼마나 많은 위치에서 애플리케이션 인스턴스가 실행중인지와 무관하게 API는 관리 데이터에 액세스하기 위해 균일한 인터페이스를 포함할 수 있다.

[0034] 계속해서 블록(320)을 참조하면, 시스템은 수신된 요청을 만족시키는 하나 이상의 관리 데이터의 유형을 식별한다. 예를 들어, 시스템은 요청이 애플리케이션의 각 인스턴스에 의해 생산되는 로그 정보를 요청할 것을 결정할 수 있다. 요청된 데이터를 식별하는 것은, 시스템이 각 애플리케이션 인스턴스로부터 어느 정보를 수집해야 하는지 또는 각 애플리케이션 인스턴스에 의해 이미 중앙 위치로 이동된 데이터로부터 로컬에서 수집되었는지 여부를 결정하는 것을 가능케 한다.

[0035] 계속해서 블록(330)을 참조하면, 시스템은 애플리케이션의 둘 이상의 인스턴스를 포함하는 애플리케이션의 분산화를 결정한다. 분산화는 애플리케이션이 어디에서 실행중인지 그리고 시스템이 어디에서 요청을 만족시키기 위한 관리 데이터를 찾을 것인지를 결정한다. 시스템은 다른 데이터센터로/로부터의 애플리케이션 부하의 버스트 또는 다른 이송 각각을 기술하는 정보를 추적하는 데이터 스토어를 포함할 수 있으며, 그에 따라 시스템이 애플리케이션 인스턴스가 실행중인 각각의 위치를 인식하도록 한다. 관리 도구 요청을 수신하면, 이러한 정보는 시스템이 어디에서 관리 데이터를 수집해야 할지를 결정할 수 있게 한다.

[0036] 계속해서 블록(340)을 참조하면, 시스템은 각각의 분산된 애플리케이션 인스턴스로부터의 요청을 만족시키기 위해 관리 데이터를 수집한다. 인스턴스는 로컬 사설 데이터센터, 원거리 사설 데이터센터, 사설 클라우드 컴퓨팅 시설, 공용 클라우드 컴퓨팅 시설, 다른 사설 데이터센터에 의해 제공되는 여분의 리소스 등을 포함할 수 있다. 시스템은 각 인스턴스로부터 수신된 관리 도구 요청을 만족시키기 위해서 정보(예컨대, 수행 데이터, 고장 등)를 포함하는 앞서 전송된 정보에 액세스하거나 애플리케이션의 각 인스턴스에 접촉한다.

[0037] 계속해서 블록(350)을 참조하면, 하나 이상의 고장수리 커맨드를 하나 이상의 원거리 애플리케이션 인스턴스로 선택적으로 전송한다. 예를 들어, 만약 하나의 위치가 고장을 겪으면, 관리자는 추가적인 트레이스 정보를 요

청하기 위해, 하나 이상의 테스트 요청을 전송하기 위해, 또는 다른 유형의 디버깅을 수행하기 위해 관리 도구를 사용할 수 있다. 원격리 애플리케이션 인스턴스는 고장수리 커맨드를 수행하고, 관리 도구가 정보에 액세스할 수 있는 중앙 위치로 요청된 데이터를 보고한다.

- [0038] 계속해서 블록(360)을 참조하면, 시스템은 수신된 관리 도구 요청에 대한 균일한 응답을 제공하기 위해서 수집된 데이터를 통합한다. 이러한 방식으로, 관리 도구는 도구에 의해 관리되는 애플리케이션의 다양한 가능한 분산에 대한 이해를 포함하기 위해 기록될 필요가 없다. 따라서 시스템은 애플리케이션 부하를 처리해야 하는 필요에 따라 애플리케이션을 한 위치에서 다른 위치로 또는 복수의 위치로 자유롭게 이동시킬 수 있는 한편, 관리자에게 간단한 관리 및 고장수리 경험을 제공할 수 있다.
- [0039] 계속해서 블록(370)을 참조하면, 수신된 관리 도구 요청에 응답하여 수집되어 통합된 관리 데이터를 보고한다. 시스템은 요청이 수신되는 인터페이스, 또는 통지 인터페이스, 또는 데이터를 도구에 제공하기 위한 다른 시설을 통해 데이터를 전송할 수 있다. 블록(370) 후에, 이들 단계들이 종료된다.
- [0040] 도 4 는 일 실시예에서, 원격리 애플리케이션 인스턴스의 위치로부터의 데이터를 보고하고 원격리 애플리케이션 인스턴스의 위치에서의 고장수리 요청을 처리하기 위한 클라우드 관리 시스템의 프로세싱을 도시한 순서도이다. 블록(410)에서 시작하면, 시스템은 애플리케이션의 클라이언트로부터의 요청에 의해 생성된 부하의 일부분을 처리하는 원격 애플리케이션 인스턴스에서의 관리 데이터를 수신한다. 관리 데이터는 수행 데이터, 로그 정보, 오류 세부사항, 통계적 정보, 판매 히스토리, 또는 애플리케이션을 관리하기 위해 유용한 애플리케이션 동작의 표시를 포함할 수 있다.
- [0041] 계속해서 블록(420)을 참조하면, 시스템은 관리자가 분산된 원격리 위치에서 실행중인 애플리케이션의 복수의 인스턴스에 의해 보고되는 관리 데이터에 액세스할 수 있는 애플리케이션의 홈 위치를 결정한다. 애플리케이션 인스턴스는 홈 위치가 접촉될 수 있는 장소를 명시하고 애플리케이션 인스턴스가 애플리케이션의 원격리 인스턴스인 인스턴스의 생성에 따라 홈 위치로부터 구성 정보를 수신할 수 있다. 시스템은 피크 부하를 처리하기 위해서 애플리케이션을 복수의 위치로 이동시킬 수 있고, 프로세싱이 피크가 아닌 위치에서 낮은 우선권의 업무를 수행할 수 있으며, 따라서 보다 저렴하거나 또는 다른 이유로 관리자에 의해 결정된다. 애플리케이션은 애플리케이션이 정상적으로 실행하는 홈 위치를 가질 수 있으며, 하나 이상의 분산된 원격리 위치에서 피크 또는 다른 부하를 처리할 수 있다.
- [0042] 계속해서 블록(430)을 참조하면, 시스템은 원격리 애플리케이션 인스턴스로부터 애플리케이션의 결정된 홈 위치로 수신된 관리 데이터를 전송한다. 시스템은 관리 데이터가 관리자 및 관리 도구의 편의를 위해 홈 위치에 있는 한 장소에서 이용가능하도록 분산된 인스턴스에서 생성된 데이터를 홈 위치에 다시 주기적으로 이동시킬 수 있다. 시스템은 또한 수요에 따라 또는 다양한 도구(예로서, 도 3을 참조)에 의한 요청으로서 데이터를 이동시킬 수 있다. 일부 경우에서, 시스템은 짧은 기간 동안 애플리케이션 부하를 원격리 위치로 버스팅할 수 있으며, 부하가 다시 홈 위치로 이동되어 원격리 인스턴스가 종료되었을 때 애플리케이션의 실행과 관련된 정보를 수집할 수 있다.
- [0043] 계속해서 블록(440)을 참조하면, 시스템은 원격 애플리케이션 인스턴스를 고장수리하기 위해 홈 위치에서 실행하는 관리 도구로부터 고장수리 요청을 선택적으로 수신한다. 고장수리 요청은 디버깅 중단점, 상세한 트레이스 정보에 대한 요청, 또는 고장수리 동작을 수행하라는 다른 커맨드 또는 요청을 포함할 수 있다.
- [0044] 계속해서 블록(450)을 참조하면, 시스템은 수신된 고장수리 요청에 응답하여 하나 이상의 고장수리 동작을 수행한다. 이 동작은 디버깅 중단점 설정, 로그 기록 레벨 터닝 업, 애플리케이션으로 테스트 데이터 전송, 또는 애플리케이션이 적절하게 동작하는지 여부를 결정하기 위한 요청에 의해 명시되는 임의의 다른 동작의 수행을 포함할 수 있다.
- [0045] 계속해서 블록(460)을 참조하면, 시스템은 수신된 고장수리 요청에 응답하여 홈 위치에서 고장수리 결과를 전송한다. 고장수리 커맨드를 원격으로 실행하기 위한 시설을 제공함으로써, 시스템은 홈 위치에서 동작중인 고장수리 도구가 인스턴스가 실행중인 장소와 무관하게 애플리케이션 인스턴스를 수리할 수 있게 하며, 애플리케이션을 관리 및 고장수리하는 관리자의 능력을 방해하지 않고 애플리케이션의 인스턴스를 다양한 위치로 매끄럽게 이송할 수 있게 한다. 블록(460) 후에, 이들 단계들은 종료된다.
- [0046] 일부 실시예에서, 클라우드 관리 시스템은 도메인 명칭 서비스(DNS) 기록을 수정함으로써 애플리케이션 부하를 이송한다. 이 시스템은 DNS 서버를 하나 이상의 새로운 목적지 인터넷 프로토콜(IP) 어드레스에 부하를 소스 데이터센터로부터 멀리 떨어진 타겟 데이터센터/클라우드로 다이렉팅하게 하라고 요청하는 인커밍 클라이언트를

포인팅하도록 수정할 수 있다. 글로벌 트래픽 매니저(GTM)는 종종 클라이언트의 요청을 처리하기 위해 가장 가까운 서버로 클라이언트를 포인팅하며, 이러한 솔루션은 부하 또는 다른 조건에 기초하여 트래픽을 재다이렉팅(redirect)하도록 수정될 수 있다. 따라서, 하나의 데이터센터가 과부하되거나 용량에 근접하면, 시스템은 GTM에게 적어도 일부 클라이언트 요청을 초과 부하를 처리할 수 있는 새로운 위치로 다이렉팅하라고 알릴 수 있다. 유사하게, 시스템은 관리 도구가 관리 요청을 해결할 수 있고 어디에 존재하는지와 무관하게 애플리케이션 인스턴스에 접속될 수 있는 다른 어드레스 또는 DNS를 제공할 수 있다.

[0047] 일부 실시예에서, 클라우드 관리 시스템은 이송 상태가 용이해진 후에 타겟 컴퓨팅 리소스로부터 로그 및 다른 데이터를 다시 이송한다. 예를 들어, 피크 부하 기간 이후에, 시스템은 모든 부하를 다시 원래의 데이터센터로 이동시킬 수 있고, 애플리케이션 로그와 같이 타겟 데이터에서 생성된 정보를 마지막 분석을 위해 다시 원래의 데이터센터로 가져올 수 있다. 일부 애플리케이션에 있어서, 클라이언트 요청을 추적하는 것은 규정 준수의 문제일 수 있거나 또는 단순히 디버깅 및 보고에 유용한 것일 수 있다. 어떠한 경우에도, 소스 위치에서 로그를 통합하는 것은 소스 위치로의 성공적인 재이송의 일부일 수 있다.

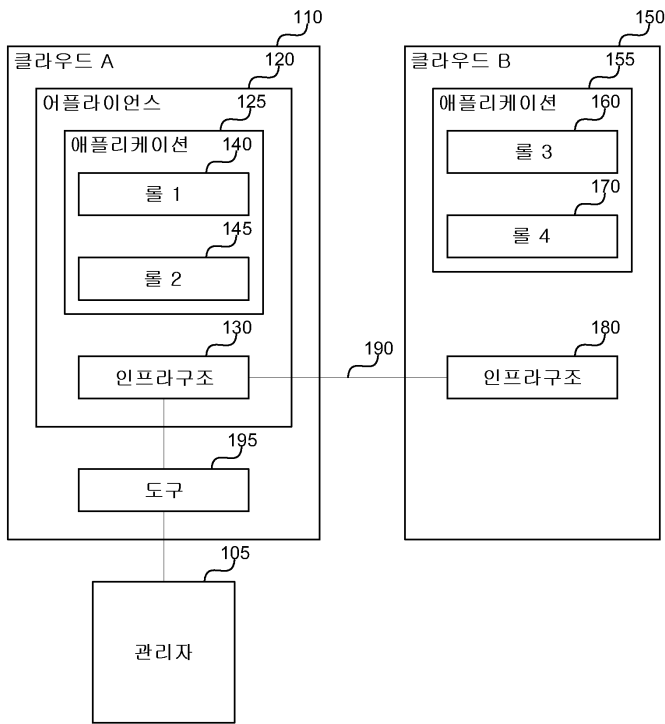
[0048] 일부 실시예에서, 클라우드 관리 시스템은 소스 컴퓨팅 리소스와 하나 이상의 타겟 컴퓨팅 리소스 사이에서 동적으로 변환할 수 있는 애플리케이션 부하의 양을 할당한다. 예를 들어, 시스템은 소스 컴퓨팅 리소스를 용량이 완전히 차거나 용량에 거의 근접한 상태로 유지하라는 요청을 동적으로 라우팅하고 단지 요청을 소스 컴퓨팅 리소스가 성공적으로 처리될 수 없는 외부 컴퓨팅 리소스에 전송할 수 있다. 이러한 결정은 비용, 데이터 안전, 또는 필요한 만큼의 가능한 한 적은 양의 애플리케이션 부하를 이동시키거나 또는 애플리케이션 부하를 가장 저렴하게 수행될 수 있거나 또는 가장 효율적으로 수행될 수 있는 곳에 배치하기 위한 다른 고려의 문제일 수 있다. 일부 경우에서, 이러한 결정은 애플리케이션의 규정 요구사항에 기초할 수 있다. 예를 들어, 건강관리 또는 다른 문서보존 관련 법률 하에 있는 애플리케이션들은 이들이 동작할 수 있는 데이터센터/클라우드에 대한 제한을 가질 수 있다.

[0049] 일부 실시예에서, 클라우드 관리 시스템은 재해 복구를 위한 다양한 옵션을 제공한다. 일부 경우에, 시스템은 메인 데이터센터의 고장정지(outage)를 모니터링하기 위해 외부 데이터센터에서의 리소스를 요청할 수 있다. 만약 외부 데이터센터가 메인 데이터센터에 도달할 수 없게 되면, 외부 데이터센터는 재해가 발생하였으며 애플리케이션 부하를 외부 데이터센터에 이동시키도록 결정할 수 있다. 과거 시스템에서, 재해를 성공적으로 처리하기 위해서 조직이 (상당한 비용으로) 필요한 용량의 200%를 유지하는 것이 일반적이었다. 클라우드 관리 시스템에서, 조직은 제 2 위치에서 더 적은 양의 이용가능한 용량을 유지할 수 있으며(예컨대, 10%), 고정시에는 신속하게 필요한 만큼의 용량을 더 요청할 수 있다. 보험과 매우 유사하게, 클라우드 제공자의 모든 클라이언트에게 동시에 고장이 발생하여 높은 여분의 용량을 요청할 가능성은 낮으며, 복수의 클라이언트는 제 1 리소스의 고장 시에 사용되기 위한 중복되는 제 2 리소스를 공유할 수 있다. 시스템은 또한 재해 복구 이후에 관리 도구 및 고장수리 리소스를 새로운 위치에 포인팅하도록 다시 홈에 위치시켜, 방해받지 않고 관리가 계속되도록 할 수 있다.

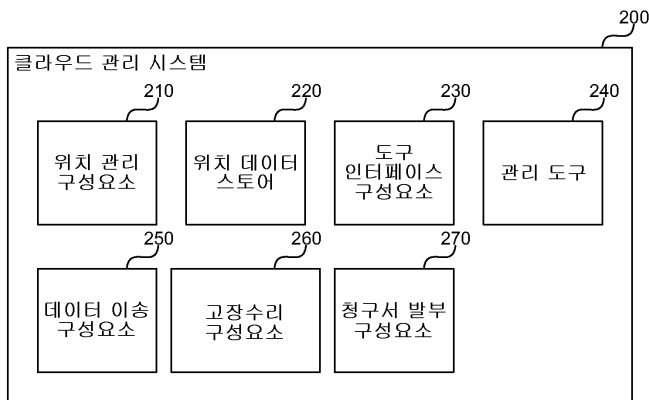
[0050] 전술된 내용으로부터, 클라우드 관리 시스템의 특정한 실시예가 설명을 위해서 본 명세서에 기술되었지만, 본 발명의 범주 및 사상으로부터 벗어나지 않고 다양한 수정이 이루어질 수 있음을 이해할 것이다. 따라서, 본 발명은 첨부된 특허청구범위에 의해서만 한정된다.

도면

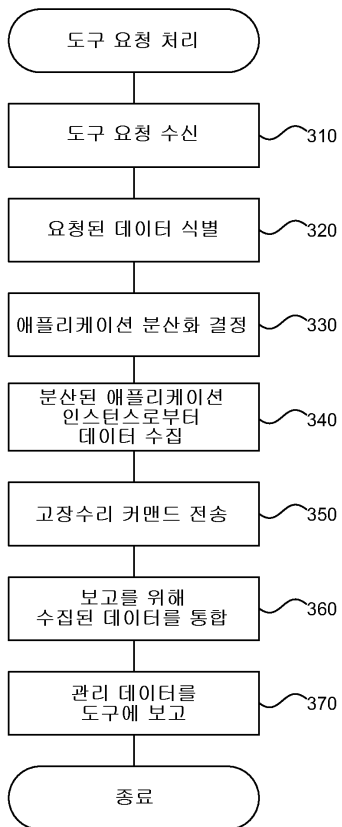
도면1



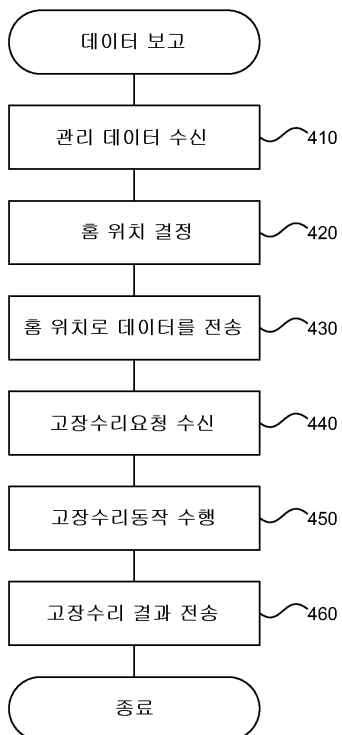
도면2



도면3



도면4



【심사관 직권보정사항】

【직권보정 1】

【보정항목】 청구범위

【보정세부항목】 제19항

【변경전】

컴퓨터 판독가능 저장 장치.

【변경후】

컴퓨터 판독가능 저장 매체.