(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2019년10월01일
(11) 등록번호 10-1990411
(24) 등록일자 2019년06월12일

(51) 국제특허분류(Int. Cl.)
G06F 9/50 (2018.01) *G06N 5/04* (2006.01)
G06N 99/00 (2019.01)
(52) CPC특허분류
G06F 9/5077 (2013.01)
G06F 9/5072 (2013.01)
(21) 출원번호 10-2017-0179097
(22) 출원일자 2017년12월26일
심사청구일자 2017년12월26일
(56) 선행기술조사문헌
KR101287448 B1*
(뒷면에 계속)

(73) 특허권자
경희대학교 산학협력단
경기도 용인시 기흥구 덕영대로 1732 (서천동, 경희대학교 국제캠퍼스내)
(72) 발명자
허의남
경기도 용인시 기흥구 이현로29번길 86-21 201동401호(보정동,e편한세상대림아파트)
손아영
경기도 용인시 기흥구 덕영대로 1732 경희대학교 국제캠퍼스 전자정보대학 331호
(뒷면에 계속)
(74) 대리인
김등용, 김홍석

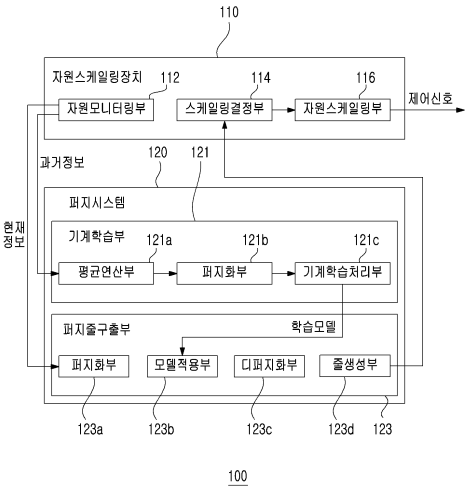
전체 청구항 수 : 총 9 항 심사관 : 유진태

(54) 발명의 명칭 **클라우드 시스템에서의 우선 순위 기반 자원 스케일링 시스템, 장치 및 방법**

(57) 요 약

클라우드 시스템에서의 우선 순위 기반 자원 스케일링 시스템, 장치 및 방법에 관한 것으로, 우선 순위 기반 자원 스케일링 시스템은, 자원 스케일링 시스템은, 가상 머신이 설치되는 물리 머신의 자원을 계속적으로 모니터링하는 자원 스케일링 장치 및 상기 물리 머신의 자원에 대한 계속적인 모니터링 결과를 기반으로 생성된 학습 모델을 생성하고, 상기 자원에 대한 현재의 모니터링 결과를 상기 학습 모델에 적용하여 퍼지 룰을 생성하는 퍼지 시스템을 포함하되, 상기 자원 스케일링 장치는, 상기 자원 모니터링부에 의한 자원에 대한 현재의 모니터링 결과를 기반으로 생성된 퍼지 룰에 따라 스케일링 여부 및 스케일링 방법을 결정하고, 상기 스케일링 여부 및 상기 스케일링 방법을 기반으로 상기 가상 머신 또는 새로 생성되는 가상 머신에 제공되는 자원에 대한 스케일링을 수행할 수 있다.

대 표 도 - 도3



(52) CPC특허분류
G06N 20/00 (2019.01)
G06N 5/048 (2013.01)

(72) 발명자
김윤곤
경기도 용인시 기흥구 덕영대로 1732 경희대학교 국제캠퍼스 전자정보대학 331호
이가원
경기도 용인시 기흥구 덕영대로 1732 경희대학교 국제캠퍼스 전자정보대학 331호
박준영
경기도 용인시 기흥구 덕영대로 1732 경희대학교 국제캠퍼스 전자정보대학 331호

이 발명을 지원한 국가연구개발사업
과제고유번호 1711055131
부처명 과학기술정보통신부
연구관리전문기관 정보통신기술진흥센터
연구사업명 대학ICT연구센터육성지원사업
연구과제명 실시간 모바일 클라우드 서비스 플랫폼 연구개발
기 여 율 1/1
주관기관 경희대학교 산학협력단
연구기간 2017.01.01 ~ 2017.12.31

(56) 선행기술조사문헌
KR1020160044623 A*
KR100954840 B1*
손아영 외 2명. '클라우드 컴퓨팅 환경에서 Multi-Metic 기반의 QoS를 위한 마이그레이션 연구'. 2017년도 한국통신학회 동계종합학술발표회 논문집, 2017.01., pp.1227-1228.
KR1020150007698 A
*는 심사관에 의하여 인용된 문헌

명 세 서

청구범위

청구항 1

가상 머신이 설치되는 물리 머신의 자원을 모니터링하는 자원 모니터링부;

상기 자원 모니터링부에서 주기적으로 또는 연속적으로 전달되는 모니터링 결과를 기반으로 기계학습을 수행하여 학습 모델을 생성하는 기계학습부;

상기 자원 모니터링부에서 주기적으로 또는 연속적으로 전달되는 모니터링 결과를 상기 학습 모델에 적용하여 퍼지 룰을 생성하는 퍼지룰 구축부;

상기 퍼지룰 구축부에서 생성된 퍼지 룰에 따라 스케일링 여부 및 스케일링 방법을 결정하는 스케일링 결정부; 및

상기 스케일링 여부 및 상기 스케일링 방법을 기반으로 상기 가상 머신 또는 새로 생성되는 가상 머신에 제공되는 자원에 대한 스케일링을 수행하는 자원 스케일링부;를 포함하고,

상기 자원 스케일링부는 퍼지-분석적 계층 프로세스(Fuzzy-AHP, Fuzzy-Analytic Hierarchy Process)를 기반으로 자원 스케일링을 수행하며, 자원 스케일링 대상이 되는 가상머신을 결정하며, 퍼지화 및 학습을 수행하는 퍼지 시스템의 복수의 메트릭의 비교결과를 기반으로 우선 순위에 따라 자원 스케일링 대상이 되는 가상머신을 결정하는, 자원 스케일링 장치.

청구항 2

삭제

청구항 3

삭제

청구항 4

삭제

청구항 5

삭제

청구항 6

제1항에 있어서,

상기 자원 스케일링부는, 상기 복수의 메트릭 중 적어도 둘 간의 쌍대 비교 행렬을 획득하고, 쌍대 비교 행렬을 기반으로 퍼지 가중치를 획득하고, 연산된 퍼지 가중치를 기반으로 최종 퍼지 가중치를 획득하고, 획득한 최종 퍼지 가중치를 기반으로 우선 순위를 결정하고, 결정된 우선 순위를 기반으로 자원 스케일링 대상이 되는 가상 머신을 결정하는 자원 스케일링 장치.

청구항 7

제1항에 있어서,

상기 새로 생성되는 가상 머신은, 기존의 가상 머신의 시스템이 마이그레이션되는 가상 머신을 포함하는 자원 스케일링 장치.

청구항 8

가상 머신이 설치되는 물리 머신의 자원을 계속적으로 모니터링하는 자원 스케일링 장치; 및

상기 물리 머신의 자원에 대한 계속적인 모니터링 결과를 기반으로 생성된 학습 모델을 생성하고, 상기 자원에 대한 현재의 모니터링 결과를 상기 학습 모델에 적용하여 퍼지 룰을 생성하는 퍼지 시스템;을 포함하되,

상기 자원 스케일링 장치는,

주기적으로 또는 연속적으로 전달되는 모니터링 결과를 기반으로 기계학습을 수행하여 학습 모델을 생성하고,

주기적으로 또는 연속적으로 전달되는 모니터링 결과를 상기 학습 모델에 적용하여 퍼지 룰을 생성하며,

생성된 퍼지 룰에 따라 스케일링 여부 및 스케일링 방법을 결정하고,

상기 스케일링 여부 및 상기 스케일링 방법을 기반으로 상기 가상 머신 또는 새로 생성되는 가상 머신에 제공되는 자원에 대한 스케일링을 수행하며,

상기 자원 스케일링 장치는 퍼지-분석적 계층 프로세스를 기반으로 자원 스케일링을 수행하고, 자원 스케일링 대상이 되는 가상머신을 결정하며, 퍼지화 및 학습을 수행하는 퍼지 시스템의 복수의 메트릭의 비교 결과를 기반으로 결정된 우선 순위에 따라서 자원 스케일링 대상이 되는 가상 머신을 결정하는 자원 스케일링 시스템.

청구항 9

삭제

청구항 10

삭제

청구항 11

삭제

청구항 12

제8항에 있어서,

상기 자원 스케일링 장치는, 상기 복수의 메트릭 중 적어도 둘 간의 쌍대 비교 행렬을 획득하고, 쌍대 비교 행렬을 기반으로 퍼지 가중치를 획득하고, 연산된 퍼지 가중치를 기반으로 최종 퍼지 가중치를 획득하고, 획득한 최종 퍼지 가중치를 기반으로 우선 순위를 결정하고, 결정된 우선 순위를 기반으로 자원 스케일링 대상이 되는 가상 머신을 결정하는 자원 스케일링 시스템.

청구항 13

제8항에 있어서,

상기 새로 생성되는 가상 머신은, 기존의 가상 머신의 시스템이 마이그레이션되는 가상 머신을 포함하는 자원 스케일링 시스템.

청구항 14

자원 스케일링 장치에 의해서 수행되는 자원 스케일링 방법으로서,

물리 머신의 자원을 모니터링하는 단계;

주기적으로 또는 연속적으로 전달되는 모니터링 결과를 기반으로 기계학습을 수행하여 학습 모델을 생성하는 단계;

주기적으로 또는 연속적으로 전달되는 모니터링 결과를 상기 학습 모델에 적용하여 퍼지 룰을 생성하는 단계;

자원 모니터링부에 의한 자원에 대한 현재의 모니터링 결과를 기반으로 생성된 퍼지 룰에 따라 스케일링 여부 및 스케일링 방법을 결정하는 단계; 및

상기 스케일링 여부 및 상기 스케일링 방법을 기반으로 상기 물리 머신에 설치된 가상 머신 또는 상기 물리 머신에 새로 생성되는 가상 머신에 제공되는 자원에 대한 스케일링을 수행하는 단계;를 포함하고,

상기 스케일링을 수행하는 단계는,

자원 스케일링 대상이 되는 가상 머신을 결정하는 단계;

퍼지-분석적 계층 프로세스를 기반으로 하는 우선 순위 리스트를 통해 자원 스케일링 대상이 되는 가상 머신을 결정하는 단계; 및

퍼지화 및 학습을 수행하는 퍼지 시스템의 복수의 메트릭의 비교 결과를 기반으로 결정된 우선 순위에 따라서 자원 스케일링 대상이 되는 가상 머신을 결정하는 단계를 포함하는, 자원 스케일링 방법.

청구항 15

삭제

청구항 16

삭제

청구항 17

삭제

청구항 18

삭제

청구항 19

제14항에 있어서,

상기 퍼지-분석적 계층 프로세스를 기반으로 하는 우선 순위 리스트를 통해 자원 스케일링 대상이 되는 가상 머신을 결정하는 단계는,

상기 복수의 메트릭 중 적어도 둘 간의 쌍대 비교 행렬을 획득하는 단계;

쌍대 비교 행렬을 기반으로 퍼지 가중치를 획득하는 단계;

연산된 퍼지 가중치를 기반으로 최종 퍼지 가중치를 획득하고, 획득한 최종 퍼지 가중치를 기반으로 우선 순위를 결정하는 단계; 및

결정된 우선 순위를 기반으로 자원 스케일링 대상이 되는 가상 머신을 결정하는 단계;를 포함하는 자원 스케일링 방법.

청구항 20

제14항에 있어서,

상기 새로 생성되는 가상 머신은, 기존의 가상 머신의 시스템이 마이그레이션 되는 가상 머신을 포함하는 자원 스케일링 방법.

발명의 설명

기 술 분 야

[0001] 클라우드 시스템에서의 우선 순위 기반 자원 스케일링 시스템, 장치 및 방법에 관한 것이다.

배 경 기 술

[0002] 가상 머신(VM, Virtual Machine)은 물리 머신(예를 들어, 데스크톱 컴퓨터나 서버 컴퓨터 등) 또는 가상 머신 내에 가상적으로 생성한 시스템을 의미한다. 가상 머신은 물리 머신의 자원(즉, 리소스)을 이용하여 물리 머신 내에 가상적으로 구축된다. 가상 머신은, 예를 들어, 물리 머신 내에 설치된 운영 체제 내에 동종 또는 이종의 운영 체제를 설치하는 등과 같이 다양한 방법을 통하여 구현될 수 있다.

[0003] 한편, 설치된 가상 머신이 요구하는 자원을 클라우드 시스템 내의 물리 머신이 충분하게 보유하지 못한 경우, 각각의 가상 머신은 요구하는 수준의 자원을 제공받을 수 없게 되고, 이에 따라 가상 머신의 성능은 저하 및 핫

스팟(Hot spot)의 문제가 발생하기도 한다. 따라서, 가상 머신을 이용함에 있어서 물리 머신 내의 각각의 가상 머신에게 적절하고 효율적으로 자원을 분배 및/또는 제공해야 할 필요성이 존재한다.

[0004] 클라우드 컴퓨팅을 이용하는 클라우드 시스템 역시 이와 같은 가상 머신은 생성 및 실행될 수 있다. 클라우드 컴퓨팅이란 인터넷 등과 같은 네트워크를 통하여 사용자가 언제 어디서든 단말기를 이용하여 소정의 작업(예를 들어, 소프트웨어의 이용이나 개발, 데이터의 저장, 백업 또는 복원 등)을 수행할 수 있도록 마련된 컴퓨팅 방법을 의미한다. 클라우드 컴퓨팅을 이용하는 경우, 사용자는 서로 상이한 위치에서 원하는 작업을 수행할 수 있게 된다. 클라우드 시스템은 클라우드 서비스를 제공하기 위해 동작하는 적어도 하나의 물리적 장치(예를 들어, 서버 장치)를 포함하는데, 이들 물리적 장치에도 소정의 운영 체제가 설치되어 있다. 클라우드 시스템은, 이와 같은 물리적 장치의 운영 체제 내에 동종 또는 이종의 운영 체제를 설치시킴으로써 가상 머신을 기반으로 클라우드 서비스를 제공할 수 있게 된다.

선행기술문헌

특허문헌

[0005] (특허문헌 0001) 대한민국 특허출원 제10-2014-0012932호(2014.02.05 출원)

(특허문헌 0002) 대한민국 특허출원 제10-2013-0162751호(2013.12.24 출원)

발명의 내용

해결하려는 과제

[0006] 가상 머신(VM, Virtual Machine)을 이용하는 경우에 있어서 자원 상태를 동적으로 판단하여 자원을 효율적으로 사용하면서도 그 성능을 향상시킬 수 있는 클라우드 시스템에서의 우선 순위 기반 자원 스케일링 시스템, 장치 및 방법을 제공하는 것을 해결하고자 하는 과제로 한다.

과제의 해결 수단

[0007] 상술한 과제를 해결하기 위하여 클라우드 시스템에서의 우선 순위 기반 자원 스케일링 시스템, 장치 및 방법이 제공된다.

[0008] 자원 스케일링 장치는, 가상 머신이 설치되는 물리 머신의 자원을 모니터링하는 자원 모니터링부; 상기 자원 모니터링부에 의한 자원에 대한 현재의 모니터링 결과를 기반으로 생성된 퍼지 룰에 따라 스케일링 여부 및 스케일링 방법을 결정하는 스케일링 결정부; 및 상기 스케일링 여부 및 상기 스케일링 방법을 기반으로 상기 가상 머신 또는 새로 생성되는 가상 머신에 제공되는 자원에 대한 스케일링을 수행하는 자원 스케일링부;를 포함할 수 있다.

[0009] 상기 퍼지 룰은, 상기 자원에 대한 현재의 모니터링 결과를, 상기 물리 머신의 자원에 대한 계속적인 모니터링 결과를 기반으로 생성된 학습 모델에 적용하여 획득되는 것일 수 있다.

[0010] 상기 자원 스케일링부는, 퍼지-분석적 계층 프로세스(Fuzzy-AHP, Fuzzy-Analytic Hierarchy Process)를 기반으로 자원 스케일링을 수행할 수 있다.

[0011] 상기 자원 스케일링부는, 자원 스케일링 대상이 되는 가상 머신을 더 결정할 수 있다.

[0012] 상기 자원 스케일링부는, 퍼지화 및 학습을 수행하는 퍼지 시스템의 복수의 메트릭의 비교 결과를 기반으로 결정된 우선 순위에 따라서 자원 스케일링 대상이 되는 가상 머신을 결정할 수 있다.

[0013] 상기 자원 스케일링부는, 상기 복수의 메트릭 중 적어도 둘 간의 쌍대 비교 행렬을 획득하고, 쌍대 비교 행렬을 기반으로 퍼지 가중치를 획득하고, 연산된 퍼지 가중치를 기반으로 최종 퍼지 가중치를 획득하고, 획득한 최종 퍼지 가중치를 기반으로 우선 순위를 결정하고, 결정된 우선 순위를 기반으로 자원 스케일링 대상이 되는 가상 머신을 결정할 수 있다.

[0014] 상기 새로 생성되는 가상 머신은, 기존의 가상 머신의 시스템이 마이그레이션되는 가상 머신을 포함할 수 있다.

[0015] 자원 스케일링 시스템은, 가상 머신이 설치되는 물리 머신의 자원을 계속적으로 모니터링하는 자원 스케일링 장

치; 및 상기 물리 머신의 자원에 대한 계속적인 모니터링 결과를 기반으로 생성된 학습 모델을 생성하고, 상기 자원에 대한 현재의 모니터링 결과를 상기 학습 모델에 적용하여 퍼지 룰을 생성하는 퍼지 시스템;을 포함하되, 상기 자원 스케일링 장치는, 상기 자원 모니터링부에 의한 자원에 대한 현재의 모니터링 결과를 기반으로 생성된 퍼지 룰에 따라 스케일링 여부 및 스케일링 방법을 결정하고, 상기 스케일링 여부 및 상기 스케일링 방법을 기반으로 상기 가상 머신 또는 새로 생성되는 가상 머신에 제공되는 자원에 대한 스케일링을 수행할 수 있다.

[0016] 상기 자원 스케일링 장치는, 퍼지-분석적 계층 프로세스를 기반으로 자원 스케일링을 수행할 수 있다.

[0017] 상기 자원 스케일링 장치는, 자원 스케일링 대상이 되는 가상 머신을 더 결정할 수 있다.

[0018] 상기 자원 스케일링 장치는, 퍼지화 및 학습을 수행하는 퍼지 시스템의 복수의 메트릭의 비교 결과를 기반으로 결정된 우선 순위에 따라서 자원 스케일링 대상이 되는 가상 머신을 결정할 수 있다.

[0019] 상기 자원 스케일링 장치는, 상기 복수의 메트릭 중 적어도 둘 간의 쌍대 비교 행렬을 획득하고, 쌍대 비교 행렬을 기반으로 퍼지 가중치를 획득하고, 연산된 퍼지 가중치를 기반으로 최종 퍼지 가중치를 획득하고, 획득한 최종 퍼지 가중치를 기반으로 우선 순위를 결정하고, 결정된 우선 순위를 기반으로 자원 스케일링 대상이 되는 가상 머신을 결정할 수 있다.

[0020] 상기 새로 생성되는 가상 머신은, 기존의 가상 머신의 시스템이 마이그레이션되는 가상 머신을 포함할 수 있다.

[0021] 자원 스케일링 방법은, 물리 머신의 자원을 모니터링하는 단계; 상기 자원 모니터링부에 의한 자원에 대한 현재의 모니터링 결과를 기반으로 생성된 퍼지 룰에 따라 스케일링 여부 및 스케일링 방법을 결정하는 단계; 및 상기 스케일링 여부 및 상기 스케일링 방법을 기반으로 상기 물리 머신에 설치된 가상 머신 또는 상기 물리 머신에 새로 생성되는 가상 머신에 제공되는 자원에 대한 스케일링을 수행하는 단계;를 포함할 수 있다.

[0022] 자원 스케일링 방법은, 상기 자원에 대한 현재의 모니터링 결과를, 상기 물리 머신의 자원에 대한 계속적인 모니터링 결과를 기반으로 생성된 학습 모델에 적용하여 상기 퍼지 룰을 획득하는 단계;를 더 포함할 수 있다.

[0023] 상기 스케일링 여부 및 상기 스케일링 방법을 기반으로 상기 물리 머신에 설치된 가상 머신 또는 상기 물리 머신에 새로 생성되는 가상 머신에 제공되는 자원에 대한 스케일링을 수행하는 단계는, 자원 스케일링 대상이 되는 가상 머신을 결정하는 단계;를 더 포함할 수 있다.

[0024] 상기 자원 스케일링 대상이 되는 가상 머신을 결정하는 단계는, 퍼지-분석적 계층 프로세스를 기반으로 하는 우선 순위 리스트를 통해 자원 스케일링 대상이 되는 가상 머신을 결정하는 단계;를 포함할 수 있다.

[0025] 상기 퍼지-분석적 계층 프로세스를 기반으로 하는 우선 순위 리스트를 통해 자원 스케일링 대상이 되는 가상 머신을 결정하는 단계는, 퍼지화 및 학습을 수행하는 퍼지 시스템의 복수의 메트릭의 비교 결과를 기반으로 결정된 우선 순위에 따라서 자원 스케일링 대상이 되는 가상 머신을 결정하는 단계;를 포함할 수 있다.

[0026] 상기 퍼지-분석적 계층 프로세스를 기반으로 하는 우선 순위 리스트를 통해 자원 스케일링 대상이 되는 가상 머신을 결정하는 단계는, 상기 복수의 메트릭 중 적어도 둘 간의 쌍대 비교 행렬을 획득하는 단계; 쌍대 비교 행렬을 기반으로 퍼지 가중치를 획득하는 단계; 연산된 퍼지 가중치를 기반으로 최종 퍼지 가중치를 획득하고, 획득한 최종 퍼지 가중치를 기반으로 우선 순위를 결정하는 단계; 및 결정된 우선 순위를 기반으로 자원 스케일링 대상이 되는 가상 머신을 결정하는 단계;를 포함할 수 있다.

[0027] 상기 새로 생성되는 가상 머신은, 기존의 가상 머신의 시스템이 마이그레이션되는 가상 머신을 포함할 수 있다.

발명의 효과

[0028] 상술한 클라우드 시스템에서의 우선 순위 기반 자원 스케일링 시스템, 장치 및 방법에 의하면, 자원 상태를 동적으로 판단하여 자원 스케일링을 수행함으로써 자원의 효율적 사용 및 가상 머신의 성능을 향상시킬 수 있게 되는 효과를 얻을 수 있다.

[0029] 상술한 클라우드 시스템에서의 우선 순위 기반 자원 스케일링 시스템, 장치 및 방법에 의하면, 분산 클라우드를 이용하는 경우, 특정 지역의 데이터 센터를 통해 클라우드 서비스를 제공받던 사용자가 다른 지역으로 이동하더라도 지리적으로 가까운 새로운 데이터센터로부터 기존의 클라우드 서비스를 끊김없이 계속해서 제공받을 수 있게 되는 효과도 얻을 수 있다.

[0030] 상술한 클라우드 시스템에서의 우선 순위 기반 자원 스케일링 시스템, 장치 및 방법에 의하면, 새로운 가상 머신으로 마이그레이션(migration, 이관 또는 이주)을 수행하는 경우 시스템이 마이그레이션될 가상 머신에 대해

적절한 자원 제공을 수행할 수 있게 되는 효과도 얻을 수 있다.

[0031] 상술한 클라우드 시스템에서의 우선 순위 기반 자원 스케일링 시스템, 장치 및 방법에 의하면, 기계 학습을 통하여 불안정한 자원의 상태를 판단하여 자원 스케일링을 결정할 수 있게 되는 효과도 얻을 수 있다.

[0032] 상술한 클라우드 시스템에서의 우선 순위 기반 자원 스케일링 시스템, 장치 및 방법에 의하면, 제공되는 클라우드 서비스의 목적이나 용도 등에 따라 적절하게 자원 스케일링을 수행할 수 있게 되고, 이에 따라 에너지 효율도 향상 및 개선할 수 있는 효과도 얻을 수 있다.

도면의 간단한 설명

[0033] 도 1은 클라우드 시스템의 일 실시예를 도시한 도면이다.

도 2는 클라우드 시스템 내의 어느 하나의 컴퓨팅 장치의 일 실시예를 도시한 블록도이다.

도 3은 자원 스케일링 장치 및 퍼지 시스템의 일 실시예를 도시한 도면이다.

도 4는 자원 스케일링 장치의 동작의 일례를 설명하기 위한 도면이다.

도 5는 목표 가상 머신을 선정하는 과정의 일례를 도시한 도면이다.

도 6은 목표 가상 머신의 우선 순위 결정과 관련하여 계층 구조의 일 실시예를 도시한 도면이다.

도 7은 상대적 중요도에 따라 정의된 스코어들의 일례를 도시한 도표이다.

도 8은 쌍대 비교 행렬의 일례를 도시한 도면이다.

도 9는 행렬의 변화에 따른 가중치를 도시한 도면이다.

도 10은 퍼지 시스템의 기계 학습부의 동작을 설명하기 위한 도면이다.

도 11은 종래 기술과의 에너지 소비량을 비교한 도면이다.

도 12는 우선 순위 기반 자원 스케일링 방법의 일 실시예를 도시한 흐름도이다.

발명을 실시하기 위한 구체적인 내용

[0034] 이하 명세서 전체에서 동일 참조 부호는 특별한 사정이 없는 한 동일 구성요소를 지칭한다. 이하에서 사용되는 '부'가 부가된 용어는, 소프트웨어 또는 하드웨어로 구현될 수 있으며, 실시예에 따라 '부'가 하나의 부품으로 구현되거나, 하나의 '부'가 복수의 부품들로 구현되는 것도 가능하다.

[0035] 명세서 전체에서 어떤 부분이 다른 부분과 연결되어 있다고 할 때, 이는 어떤 부분과 다른 부분에 따라서 물리적 연결을 의미할 수도 있고, 또는 전기적으로 연결된 것을 의미할 수도 있다.

[0036] 제 1 이나 제 2 등의 용어는 하나의 부분을 다른 부분으로부터 구별하기 위해 사용되는 것으로, 특별한 기재가 없는 이상 이들이 순차적인 표현을 의미하는 것은 아니다.

[0037] 단수의 표현은 문맥상 명백하게 예외가 있지 않는 한, 복수의 표현을 포함할 수 있다.

[0038] 이하 도 1 내지 도 11을 참조하여 클라우드 시스템에서의 우선 순위 기반 자원 스케일링 장치의 일 실시예에 대해서 설명한다. 이하 설명의 편의를 위하여 우선 순위 기반 자원 스케일링의 일 실시예로 분산 클라우드에서의 우선 순위 기반 자원 스케일링을 들어 설명하도록 하나, 후술하는 우선 순위 기반 자원 스케일링은 오직 도 1 등에 도시된 분산 클라우드에 한정하여 적용되는 것은 아니다. 후술하는 우선 순위 기반 자원 스케일링은 분산 클라우드 이외에도, 동일하게 또는 일부 변형을 거쳐, 중앙 집중형 클라우드나 가상 머신을 구축 가능한 통상적인 컴퓨팅 장치(예를 들어, 서버 장치)에도 적용 가능하다.

[0039] 도 1은 클라우드 시스템의 일 실시예를 도시한 도면이다.

[0040] 도 1에 도시된 바를 참조하면, 클라우드 시스템(1)은 복수의 클라우드 컴퓨팅 장치(11, 12)가 마련된 적어도 하나의 클라우드(10)를 포함할 수 있다.

[0041] 클라우드(10)는 분산 클라우드로 구현될 수 있다. 이 경우, 클라우드(10)는 적어도 하나의 클라우드 컴퓨팅 장치(11: 11-1 내지 11-6)를 포함할 수 있다.

[0042] 적어도 하나의 클라우드 컴퓨팅 장치(11: 11-1 내지 11-6)는 사용자의 단말기(9: 9-1 내지 9-5)가 유선 통신 네

트워크 및 무선 통신 네트워크 중 적어도 하나(이하 유무선 통신 네트워크)를 기반으로 접속 가능하게 마련되며, 접속한 사용자의 단말기(9: 9-1 내지 9-5)에 사용자가 요구하는 클라우드 서비스를 제공할 수 있다. 적어도 하나의 클라우드 컴퓨팅 장치(12: 12-1 내지 12-6)는 복수의 위치에 각각 설치될 수 있으며, 신속한 서비스의 제공을 위하여 상대적으로 사용자와 가까운 위치에 설치될 수 있다. 사용자의 단말기(9: 9-1 내지 9-5)는 인접한 클라우드 컴퓨팅 장치(12: 12-1 내지 12-6)에 접속하여 클라우드 서비스를 제공받을 수 있다.

[0043] 상황에 따라서 어느 하나의 클라우드 컴퓨팅 장치(12: 12-1 내지 12-6)에는 하나의 단말기(9: 9-1 내지 9-5)가 접속할 수도 있고, 또는 어느 하나의 클라우드 컴퓨팅 장치(12: 12-1 내지 12-6)에 복수의 단말기(9: 9-1 내지 9-5)가 접속할 수도 있다. 또한, 복수의 클라우드 컴퓨팅 장치(12: 12-1 내지 12-6)에 적어도 하나의 단말기(9: 9-1 내지 9-5)가 접속하는 것도 가능하다.

[0044] 일 실시예에 의하면, 적어도 하나의 클라우드 컴퓨팅 장치(11: 11-1 내지 11-5)는 자원 스케일링을 수행하도록 설계될 수도 있다. 다시 말해서, 적어도 하나의 클라우드 컴퓨팅 장치(11: 11-1 내지 11-5)가 후술하는 자원 스케일링 시스템(도 3의 100) 또는 자원 스케일링 장치(도 3의 110)의 일례가 될 수도 있다. 구체적으로 적어도 하나의 클라우드 컴퓨팅 장치(11: 11-1 내지 11-5)는 후술하는 바와 같이 자원 스케일링을 수행하여, 각각의 클라우드 컴퓨팅 장치(11, 12)가 가상 머신을 생성(또는 할당)하거나 실행할 때 생성 또는 실행되는 가상 머신에 제공되는 자원을 증가시키거나(스케일-업, scale-up), 감소시키거나(스케일-다운, scale-down), 적어도 하나의 가상 머신을 복수의 가상 머신으로 리 및 분산시키거나(스케일-아웃, scale- out) 또는 복수의 가상 머신을 더 적은 수의 가상 머신으로 통합시킬 수도 있다(스케일-인, scale-in). 적어도 하나의 클라우드 컴퓨팅 장치(11: 11-1 내지 11-5)는, 퍼지-분석적 계층 프로세스(Fuzzy-AHP, Fuzzy-Analytic Hierarchy Process)를 기반으로 자원 스케일링을 수행할 수도 있다.

[0045] 클라우드(10)는, 실시예에 따라서, 적어도 하나의 클라우드 컴퓨팅 장치(12: 12-1 내지 12-6)와 유무선 통신 네트워크를 통해 연결되고, 적어도 하나의 클라우드 컴퓨팅 장치(12: 12-1 내지 12-6)에 대한 소정의 제어 권한을 갖는 중앙 클라우드 컴퓨팅 장치(12)를 더 포함할 수도 있다. 중앙 클라우드 컴퓨팅 장치(12) 역시 필요에 따라 사용자의 단말기(9: 9-1 내지 9-5)에 사용자가 요구하는 클라우드 서비스를 제공할 수 있다.

[0046] 일 실시예에 의하면, 중앙 클라우드 컴퓨팅 장치(12)는 자원 스케일링을 수행할 수도 있다. 즉, 중앙 클라우드 컴퓨팅 장치(12)가 자원 스케일링 시스템(100) 또는 자원 스케일링 장치(110)로 동작하는 것도 가능하다. 마찬가지로 중앙 클라우드 컴퓨팅 장치(12) 역시 퍼지-분석적 계층 프로세스를 기반으로 자원 스케일링을 수행하도록 설계되는 것도 가능하다.

[0047] 도 2는 클라우드 시스템 내의 어느 하나의 컴퓨팅 장치의 일 실시예를 도시한 블록도이다.

[0048] 도 2에 도시된 바에 의하면, 적어도 하나의 클라우드 컴퓨팅 장치(11: 11-1 내지 11-6) 및/또는 중앙 클라우드 컴퓨팅 장치(12)는 물리 머신(16)을 포함할 수 있다. 물리 머신(16)은 예를 들어, 통신부(16), 프로세서(17) 및 저장부(18)를 포함할 수 있으며, 이외에도 그래픽 처리 장치(GPU, Graphic Processing Unit) 등과 같은 부가적인 장치를 더 포함할 수도 있다.

[0049] 통신부(16)는 적어도 하나의 클라우드 컴퓨팅 장치(11: 11-1 내지 11-6) 및/또는 중앙 클라우드 컴퓨팅 장치(12)가 네트워크에 접속할 수 있도록 한다. 통신부(16)는, 이용되는 통신 기술에 따라 다양한 부품을 포함할 수 있으며, 예를 들어, 외부 단자(예를 들어, 범용 직렬 버스(USB) 단자 등)나 네트워크 단자(RJ-45 단자 등) 등을 이용하여 구현되거나, 또는 안테나나 통신 칩 등을 이용하여 구현될 수 있다.

[0050] 프로세서(17)는 소정의 소프트웨어를 구동시켜, 미리 정의된 연산, 판단, 처리 및/또는 제어 동작 등을 수행할 수도 있다. 소정의 소프트웨어는 저장부(18)에 저장된 것일 수 있다. 저장부(18)에 저장된 소프트웨어는 설계자에 의해 미리 작성되어 저장된 것일 수도 있고, 또는 유선 또는 무선 통신 네트워크를 통해 접속 가능한 전자 소프트웨어 유통망을 통하여 획득 또는 갱신된 것일 수도 있다.

[0051] 일 실시예에 의하면, 프로세서(17)는 적어도 하나의 클라우드 컴퓨팅 장치(11: 11-1 내지 11-6) 및/또는 중앙 클라우드 컴퓨팅 장치(12)의 동작에 필요한 각종 연산 처리 및 제어를 수행할 수 있다. 예를 들어, 프로세서(17)는 클라우드 서비스의 제공에 필요한 각종 연산 및 제어 동작을 수행할 수도 있다. 또한, 프로세서(17)는 적어도 하나의 가상 머신(VM-1 내지 VM-k+1)의 생성 및 동작을 수행, 제어 또는 지원할 수도 있다. 또한, 프로세서(17)는 자원 스케일링을 수행할 수도 있으며, 이를 위해 적어도 하나의 클라우드 컴퓨팅 장치(11: 11-1 내지 11-6) 및/또는 중앙 클라우드 컴퓨팅 장치(12) 각각의 자원의 모니터링 결과를 획득하거나, 획득한 결과를 기반으로 퍼지 시스템(Fuzzy system)을 구동하거나, 퍼지 시스템의 구동 결과를 기반으로 각 가상 머신(VM-1 내

지 VM-k)의 스케일링 여부를 결정하거나, 스케일링을 수행할 가상 머신(VM-1 내지 VM-k+1)을 결정하거나, 및/또는 선택된 가상 머신(VM-1 내지 VM-k+1)의 스케일링을 수행할 수도 있다. 이들에 대한 자세한 설명은 후술한다.

[0052] 프로세서(17)는, 예를 들어, 중앙 처리 장치(17a, CPU, Central Processing Unit), 마이크로 컨트롤러 유닛(MCU, Micro Controller Unit), 마이컴(Micom, Micro Processor), 애플리케이션 프로세서(AP, Application Processor), 전자 제어 유닛(ECU, Electronic Controlling Unit) 및/또는 각종 연산 처리 및 제어 신호의 생성이 가능한 다른 전자 장치 등을 포함할 수 있다. 이들 장치는 예를 들어 하나 또는 둘 이상의 반도체 칩 및 관련 부품을 이용하여 구현 가능하다.

[0053] 저장부(18)는 적어도 하나의 클라우드 컴퓨팅 장치(11: 11-1 내지 11-6) 및/또는 중앙 클라우드 컴퓨팅 장치(12)의 동작에 필요한 소프트웨어(애플리케이션, 프로그램 또는 앱으로 지칭 가능하다) 데이터를 일시적 또는 비일시적으로 저장할 수 있다. 예를 들어, 저장부(18)는 자원 스케일링과 관련된 소프트웨어, 퍼지 시스템에 대한 소프트웨어, 퍼지 시스템 결과 획득된 학습된 모델(trained model) 및/또는 퍼지 룰 등을 저장할 수 있다.

[0054] 저장부(18)는, 예를 들어, 주기억장치 및 보조기억장치 중 적어도 하나를 포함할 수 있다. 주기억장치는 롬(ROM) 및/또는 램(RAM)과 같은 반도체 저장 매체를 이용하여 구현된 것일 수 있다. 롬은, 예를 들어, 통상적인 롬, 이프롬(EPROM), 이이프롬(EEPROM) 및/또는 마스크롬(MASK-ROM) 등을 포함할 수 있다. 램은 예를 들어, 디램(DRAM) 및/또는 에스램(SRAM) 등을 포함할 수 있다. 보조기억장치는, 플래시 메모리 장치, SD(Secure Digital) 카드, 솔리드 스테이트 드라이브(SSD, Solid State Drive), 하드 디스크 드라이브(HDD, Hard Disc Drive), 자기 드럼, 컴팩트 디스크(CD), 디브이디(DVD) 또는 레이저 디스크 등과 같은 광 기록 매체(optical media), 자기 테이프, 광자기 디스크 및/또는 플로피 디스크 등과 같이 데이터를 영구적 또는 반영구적으로 저장 가능한 적어도 하나의 저장 매체를 이용하여 구현될 수 있다.

[0055] 가상 머신(VM)은 물리 머신(15)의 자원을 기반으로 가상적으로 생성되어 동작하는 머신을 의미한다. 가상 머신(VM: VM-1 내지 VM-k+1)은, 프로세서(17)의 제어에 따라, 하나의 컴퓨팅 장치(11, 12)에 하나 또는 둘 이상 생성될 수 있다. 복수의 가상 머신(VM: VM-1 내지 VM-k+1) 각각은 가상적으로 서로 독립적인 머신(장치)으로 동작하게 된다. 예를 들어, 각각의 가상 머신(VM: VM-1 내지 VM-k+1)은 각각 별도로 운영체제(OS, Operating System)를 구동시켜 각각 분리된 컴퓨팅 장치인 것처럼 동작할 수 있다.

[0056] 사용자의 조작이나 미리 정의된 설정에 따라서, 어느 하나의 가상 머신(예를 들어, VM-k) 내의 시스템은 다른 가상 머신(VM-k+1)으로 마이그레이션(migration)될 수 있다. 마이그레이션은 특정한 환경(즉, 기존의 환경, 예를 들어, 하나의 가상 머신, VM-k)의 시스템, 애플리케이션 또는 모듈 등(이하 시스템 등)을 그대로 다른 환경(즉, 새로운 환경, 예를 들어, 다른 가상 머신 VM-k+1)으로 이주시키는 것을 의미한다. 이 경우, 시스템 등이 존재하던 기존의 환경과 새로운 환경은 서로 동일하지 않을 수 있다. 예를 들어, 기존의 가상 머신(VM-k)에 제공되던 시스템 자원(CPU, 메모리 등)과 새로운 가상 머신(VM-k+1)에 제공되는 시스템 자원은 서로 상이할 수 있다.

[0057] 가상 머신(VM-k)의 마이그레이션은 새로운 가상 머신(VM-k+1)의 생성과 동시에 또는 순차적으로 수행될 수도 있다. 즉, 자원 낭비나, 자원 부족이나, 또는 소프트웨어 업그레이드 등과 같은 이유로, 새로운 가상 머신(VM-k+1)을 생성하면서, 기존에 사용하던 가상 머신(VM-k)을 새로 생성된 가상 머신(VM-k+1)에 마이그레이션할 수도 있다. 이 경우, 새로운 가상 머신(VM-k+1)은 스케일링되면서 생성될 수 있다. 다시 말해서, 새로운 가상 머신(VM-k+1)은 기존의 가상 머신(VM-k)을 스케일-업, 스케일-다운, 스케일-인 또는 스케일-아웃하면서 생성될 수도 있다. 새로운 가상 머신(VM-k+1)의 생성 과정은, 예를 들어, 클라우드 컴퓨팅 장치(11, 12)의 프로세서(17)에 의해 수행될 수 있다.

[0058] 단말기(9: 9-1 내지 9-5)는 사용자가 조작하거나 또는 미리 정의된 설정에 따라서 클라우드(10)에 소정의 클라우드 서비스 제공을 요청하고, 제공되는 클라우드 서비스를 이용할 수 있도록 마련된 장치이다. 단말기(9: 9-1 내지 9-5)는 적어도 하나의 클라우드 컴퓨팅 장치(11: 11-1 내지 11-6)와 상호 데이터를 송수신하여 사용자가 소정의 클라우드 서비스를 제공받도록 할 수 있다. 또한, 필요에 따라서 단말기(9: 9-1 내지 9-5)는 중앙 클라우드 컴퓨팅 장치(12)와도 직접 또는 간접적으로 상호 데이터를 송수신하여 사용자가 중앙 클라우드 컴퓨팅 장치(12)로부터 클라우드 서비스를 제공받도록 할 수도 있다.

[0059] 또한, 사용자의 이동 등에 따라서 단말기(예를 들어, 9-1)는 그 위치가 변경될 수 있으며, 이 경우 단말 장치(9-1a, 9-1b)는 이동 전과 이동 후에 서로 상이한 클라우드 컴퓨팅 장치(11-6, 11-1)에 접속할 수도 있다. 다시 말해서, 이동 전의 단말 장치(9-1a)는 제6 클라우드 컴퓨팅 장치(11-6)에 접속하여 클라우드 서비스를

제공받고, 이동 후의 단말 장치(-1b)는 상대적으로 더 인접한 제1 클라우드(11-1)에 접속하여 클라우드 서비스를 제공받게 된다.

[0060] 단말기(9: 9-1 내지 9-5)는, 예를 들어, 셀룰러 폰, 스마트 폰, 태블릿 피씨, 두부 장착형 디스플레이(HMD, Head Mounted Display) 장치, 스마트 시계, 데스크톱 컴퓨터, 랩톱 컴퓨터, 디지털 텔레비전, 셋톱 박스, 내비게이션 장치, 개인용 디지털 보조기(PDA, Personal Digital Assistant), 휴대용 게임기, 인공 지능 음향 재생 장치 및/또는 이외 부호의 입력 및 수정이 가능한 다양한 장치 중 적어도 하나를 포함할 수 있다.

[0061] 일 실시예에 의하면, 클라우드(10)는 클라우드 컴퓨팅 장치(11, 12)와는 별도로 마련된 자원 스케일링 제어 장치(13)를 더 포함할 수도 있다. 자원 스케일링 제어 장치(13)는 후술하는 바와 같이 자원 스케일링을 수행하여, 각각의 클라우드 컴퓨팅 장치(11, 12)가 가상 머신을 생성하거나 실행할 때 가상 머신의 스케일-업, 스케일-다운, 스케일-인 및/또는 스케일-아웃을 제어할 수도 있다. 이 경우, 자원 스케일링 제어 장치(13)는, 클라우드 컴퓨팅 장치(11, 12)와 동일하게, 퍼지-분석적 계층 프로세스를 기반으로 자원 스케일링을 수행할 수도 있다.

[0062] 상술한 바와 같이, 자원 스케일링 제어 장치(13)는 적어도 하나의 클라우드 컴퓨팅 장치(11: 11-1 내지 11-5)에 의해 수행될 수도 있고, 및/또는 중앙 클라우드 컴퓨팅 장치(12)에 의해 수행될 수도 있다.

[0063] 적어도 하나의 클라우드 컴퓨팅 장치(11: 11-1 내지 11-6), 중앙 클라우드 컴퓨팅 장치(12), 자원 스케일링 제어 장치(13) 및 사용자의 단말기(9: 9-1 내지 9-5) 중 적어도 둘은, 상술한 바와 같이 유무선 통신 네트워크를 기반으로 통신을 수행할 수 있다. 여기서, 유선 통신 네트워크는, 통상 케이블을 이용하여 구축 가능하며, 케이블은, 예를 들어, 페어 케이블, 동축 케이블, 광섬유 케이블 또는 이더넷 케이블 등을 이용하여 구현된 것일 수 있다. 무선 통신 네트워크는 근거리 통신 네트워크 및 원거리 통신 네트워크 중 적어도 하나를 이용하여 구축될 수 있다. 근거리 통신 네트워크는, 예를 들어, 와이 파이(Wi-Fi), 지그비(zigbee), 블루투스(Bluetooth), 와이 파이 다이렉트(Wi-Fi Direct), 저전력 블루투스(Bluetooth Low Energy), 캔(CAN) 통신 또는 엔에프씨(NFC, Near Field Communication) 등을 이용하여 구현된 것일 수 있고, 원거리 통신 네트워크는, 예를 들어, 3GPP, 3GPP2 또는 와이맥스 계열 등의 이동 통신 표준을 기반으로 구현된 것일 수 있다.

[0064] 이하 자원 스케일링 장치(110) 및 퍼지 시스템(120)을 포함하는 자원 스케일링 시스템(100)에 대해 설명한다.

[0065] 도 3은 자원 스케일링 장치 및 퍼지 시스템의 일 실시예를 도시한 도면이고, 도 4는 자원 스케일링 장치의 동작의 일례를 설명하기 위한 도면이다.

[0066] 도 3에 도시된 바에 의하면 자원 스케일링 시스템(100)은 자원 스케일링 장치(110) 및 퍼지 시스템(120)을 포함할 수 있다.

[0067] 자원 스케일링 장치(110)는 적어도 하나의 클라우드 컴퓨팅 장치(11: 11-1 내지 11-6) 및/또는 중앙 클라우드 컴퓨팅 장치(12)의 자원을 모니터링하고, 기존의 또는 생성되는 가상 머신(VM-1 내지 VM-k+1)의 자원 스케일링을 수행한다.

[0068] 자원 스케일링 장치(110)는 자원 모니터링부(112), 스케일링 결정부(114) 및 자원 스케일링부(116)를 포함할 수 있다.

[0069] 도 4에 도시된 바를 참조하면, 자원 모니터링부(112)는, 가상 머신(VM-1 내지 VM-k+1)의 상태 및/또는 가상 머신(VM-1 내지 VM-k+1)에 제공되는 제공되는 자원을 모니터링할 수 있다(s1). 예를 들어, 자원 모니터링부(112)는 퍼지 시스템(120)의 기계 학습 결과를 기반으로 가상 머신(VM-1 내지 VM-k+1)의 워크 로드(work load)를 검사할 수 있다. 보다 구체적으로 예를 들어 자원 모니터링부(112)는 가상 머신(VM-1 내지 VM-k+1)의 중앙 처리 장치(CPU, 이는 애플리케이션 프로세서나 마이크로 컨트롤러 유닛 등과 같이 처리나 제어를 위해 이용되는 전자 장치일 수도 있다)의 사용량, 램 등의 메모리 사용량 및/또는 네트워크 자원 상태(예를 들어, 대역폭) 등을 모니터링함으로써, 가상 머신(VM-1 내지 VM-k+1)의 상태나 가상 머신(VM-1 내지 VM-k+1)에 제공되는 자원에 대한 정보를 획득할 수 있다. 또한, 자원 모니터링부(112)는 가상 머신(VM-1 내지 VM-k+1)이 대기한 시간(대기 시간)이나, 가상 머신(VM-1 내지 VM-k+1)이 동작한 시간(실행 시간)이나, 가상 머신(VM-1 내지 VM-k+1)의 동작이 종료된 기간이나, 및/또는 전력 공급의 중단의 빈도 등을 모니터링하는 것도 가능하다.

[0070] 자원 모니터링부(112)의 모니터링 결과는 퍼지 시스템(120)으로 전달될 수 있다. 이 경우, 자원 모니터링부(112)는 주기적으로 또는 연속적으로 퍼지 시스템(120)에 모니터링 결과를 전송할 수도 있다. 주기적으로 또는 연속적으로 전달된 모니터링 결과는 퍼지 시스템(120)의 기계 학습부(121)로 전달될 수 있다. 기계 학습부(121)는 주기적으로 또는 연속적으로 전달되는 모니터링 결과를 기반으로 기계 학습을 수행한다. 또한, 자원 모니

터링부(112)는 현재의 모니터링 결과를 퍼지 룰 구축부(123)로 전달할 수도 있다. 퍼지 룰 구축부(123)는 현재의 모니터링 결과를 기계 학습부(121)에서 학습 결과 생성된 학습 모델에 적용하여 퍼지 룰을 생성하고 생성한 퍼지 룰을 자원 스케일링 장치(110)의 스케일링 결정부(114)로 전달할 수 있다.

[0071] 스케일링 결정부(114)는 퍼지 시스템(120)에서 전달된 퍼지 룰을 이용하여 자원의 필요 정도를 평가하고, 가상 머신(VM-1 내지 VM-k+1)의 스케일링 여부 및 스케일링 방법을 결정할 수 있다. 예를 들어, 스케일링 결정부(114)는 퍼지 시스템(120)의 다양한 메트릭(metric)를 기반으로 룰에 따라서 스케일링 여부 및 방법을 결정할 수 있다. 예를 들어, 도 4에 도시된 바와 같이 스케일링 결정부(114)는 자원 낭비 여부나, 자원 활용 정도나, 에너지 소모 정도 및/또는 마이그레이션 퍼포먼스 수준 등을 결정할 수 있다(s2). 이 경우, 자원 낭비 여부는, 예를 들어, 중앙 처리 장치의 사용량, 메모리 사용량 및 네트워크 자원 상태 중 적어도 하나를 기반으로 결정될 수 있다. 보다 구체적으로 스케일링 결정부(114)는 중앙 처리 장치의 사용량, 메모리 사용량 및/또는 네트워크 자원 상태를 독립적으로 또는 종합하여 수치화하고 수치화된 결과가 미리 정의된 임계 값보다 크면 자원 낭비로 판단할 수 있다. 자원 활용 정도나 에너지 소모 정도는, 예를 들어, 중앙 처리 장치의 사용량, 메모리 사용량 및 네트워크 자원 상태 중 적어도 하나를 기반으로 판단될 수 있으며, 이는 퍼지 시스템(120)의 퍼지 룰을 통해 결정될 수 있다. 또한, 마이그레이션 퍼포먼스 수준은, 예를 들어, 자원 활용 정도와 대기 시간 및 실행 시간을 이용하여 판단될 수 있으며, 이를 위해 퍼지 룰이 이용될 수 있다.

[0072] 자원 스케일링부(116)는 스케일링 결정부(114)의 결정 결과를 기반으로 자원 스케일링을 수행할 수 있다(s3). 다시 말해서, 자원 스케일링부(116)는 퍼지-분석적 계층 프로세스 기반의 계산된 우선 순위 리스트를 통해 마이그레이션 등에 있어서 스케일링을 수행할 수 있다.

[0073] 일 실시예에 의하면, 자원 스케일링부(116)는 스케일-업, 스케일-다운, 스케일-인 및 스케일-아웃 중 적어도 하나에 대한 제어 신호를 생성하고, 이를 가상 머신(VM-1 내지 VM-k+1)에 전달하여 가상 머신(VM-1 내지 VM-k+1)의 스케일링을 수행할 수 있다. 자원 스케일링부(116)에 의해 스케일링되는 가상 머신(VM-1 내지 VM-k+1)은 기존에 생성된 가상 머신(VM-1 내지 VM-k)을 포함할 수도 있고, 및/또는 새로 생성되거나 생성되고 있는 가상 머신(VM-k+1)을 포함할 수도 있다. 수행되는 스케일링은, 예를 들어, 자원 활용 정도나, 에너지 소모 정도 및/또는 마이그레이션 퍼포먼스 수준 등을 기반으로 수행될 수 있다. 또한, 자원 스케일링부(116)는 자원 낭비가 발생되고 있다고 판단된 경우에는 컴퓨팅 장치(예를 들어, 서버 장치)의 동작을 종료시키거나 또는 자원을 낭비하고 있는 가상 머신(VM-1 내지 VM-k+1)을 제거하도록 설계될 수도 있다.

[0074] 도 5는 목표 가상 머신을 선정하는 과정의 일례를 도시한 도면이고, 도 6은 목표 가상 머신의 우선 순위 결정과 관련하여 계층 구조의 일 실시예를 도시한 도면이다. 도 7은 상대적 중요도에 따라 정의된 스코어들의 일례를 도시한 도표이고, 도 8은 쌍대 비교 행렬의 일례를 도시한 도면이다.

[0075] 일 실시예에 의하면, 자원 스케일링부(116)는, 자원 스케일링이 수행될 가상 머신(VM-1 내지 VM-k+1)을 결정할 수도 있다. 이 경우, 자원 스케일링부(116)는, 예를 들어, 도 5에 도시된 바와 같이, 먼저 퍼지 쌍대 비교 행렬(fuzzy pairwise comparison metric)을 생성하고(s12), 퍼지 쌍대 비교 행렬을 기반으로 최적 가중치를 도출하고(s14), 도출된 최적 가중치를 이용하여 최종 스코어를 결정한 후(s16), 스코어를 기반으로 도 6에 도시된 바와 같이 복수의 가상 머신(VM1 내지 VM6) 중 적어도 하나를 선정할 수 있다. 필요에 따라 퍼지 쌍대 비교 행렬의 생성 전에 결정 테이블(decision table)이 결정될 수도 있다.

[0076] 구체적으로 자원 스케일링부(116)는, 도 6에 도시된 바와 같이 사전에 미리 설정되거나 또는 자원 스케일링부(116)에 의해 생성 가능한 계층 구조를 기반으로 퍼지 시스템(120)의 메트릭을 비교하기 위한 퍼지 쌍대 비교 행렬을 생성한다(s12).

[0077] 계층 구조는 제공될 서비스(예를 들어 가상 머신이 제공할 클라우드 서비스)의 목적이나 기능 등에 따라 구성될 수 있다. 최상위(level 1)는 얻고자 하는 목표, 즉 목표 머신들(VM-1 내지 VM-7)의 순위로 결정된다. 다음 순위(level 2)는 목표 머신을 선정하기 위해 필요한 하위 요소들, 예를 들어 퍼포먼스와 에너지들로 결정된다. 제3 순위(level 3)는 제2 순위(level)에서 선정된 요소들에 대응하는 요소들로 결정된다. 예를 들어, 퍼포먼스의 하위 요소들은 시간 최소화, 자원 및 신뢰성을 포함하고, 에너지의 하위 요소들은 자원 및 비용을 포함할 수 있다. 또한, 제4 순위(level 4)는 제3 순위(level 3)의 요소들에 대응하는 하위 요소들을 포함할 수 있으며, 예를 들어, 시간 최소화는 총 마이그레이션 시간, 다운 타임(down time), 대기 시간 및 마이그레이션 개수 등의 하위 요소를, 자원은 중앙 처리 장치, 메모리 및 네트워크 등의 하위 요소를, 신뢰성은 실패율, 응답 시간 및 지원 비용(resource cost) 등의 하위 요소를 가질 수 있다. 또한, 자원은 중앙 처리 장치, 메모리 및 네트워크 등의 하위 요소를 가질 수 있다. 이들 순위(예를 들어, 제4 순위(level 4)의 요소들을 기반으로 목표 머신들

(VM-1 내지 VM-7)은 상호 비교될 수 있다. 예를 들어, 목표 머신들(VM-1 내지 VM-7) 중 적어도 두 개의 목표 머신들(VM-1 내지 VM-7)은 제4 순위의 하위 요소들을 기준으로 상호 비교될 수 있다. 도 6에 도시된 계층 구조는 예시적인 것으로, 계층 구조는 설계자의 선택 등에 따라서 다양하게 정의 가능함은 자명하다.

[0078] 우선 순위 리스트는 선정된 메트릭에 따라 변경될 수 있다. 다시 말해서, 메트릭의 선정에 따라서 선택될 가상 머신(VM-1 내지 VM-7) 각각에 대한 우선 순위가 상이하게 변경된다. 따라서, 메트릭의 선정은 가상 머신(VM-1 내지 VM-7)의 선택에 영향을 미친다.

[0079] 적어도 하나의 메트릭의 선정을 위하여 퍼지 시스템(120)의 메트릭 간에 비교가 수행된다. 메트릭 간의 비교는, 예를 들어, 도 7에 도시된 바와 같은 상대적 중요도에 따라 정의된 스코어들을 기반으로 수행될 수 있다. 이 경우, 스코어는 두 개의 메트릭(M1 내지 Mn) 사이의 중요도의 차이(이하 상대적 중요도)에 따라서 정의될 수 있다. 예를 들어, 스코어는 상대적 중요도가 증가할수록 더 크게 정의될 수 있다. 도 7에 도시된 바를 참고하면, 두 개의 선택지(즉, 메트릭(M1 내지 Mn)) 중 선택된 두 개의 메트릭(M1 내지 Mn)이 서로 동등하게 중요하거나 그 질(quality)이 동일하면, 두 개의 선택지 간의 상대적 중요도는 1로 정의되고, 두 개의 메트릭(M1 내지 Mn) 중 어느 하나가 다소 중요하거나 질이 다소 더 나으면, 두 개의 선택지 간의 상대적 중요도는 3으로 정의된다. 마찬가지로 두 개의 메트릭(M1 내지 Mn) 중 어느 하나가 꽤 중요하거나 또는 상당히 더 중요한 경우에는, 두 개의 선택지 간의 상대적 중요도는 각각 5 및 7로 정의될 수 있다. 만약 두 개의 메트릭(M1 내지 Mn) 중 어느 하나가 극히 더 중요한 경우에는 두 개의 선택지 간의 상대적 중요도는 9로 정의될 수도 있다. 도 7의 좌측 열의 상대적 중요도의 정의나, 또는 우측 열의 이에 대응하는 스코어 값은 예시적인 것으로 설계자의 선택에 따라서 이들은 다양하게 정의될 수 있다.

[0080] 상술한 메트릭(M1 내지 Mn)들 간의 비교는 도 8에 도시된 바와 같이 퍼지-쌍대 비교 행렬의 형태로 구성될 수 있다. 퍼지-쌍대 비교 행렬은 도 7의 상대적 중요도를 이용하여 구현될 수 있다. 도 8의 퍼지-쌍대 비교 행렬에서 M1 내지 Mn은 메트릭을 의미하고, i행과 j열이 서로 교차하는 지점의 값(즉, Vi/Vj)는 j번째 메트릭(Mj)에 대한 i번째 메트릭(Mi)의 중요도를 의미한다. Vi는 i번째 메트릭(Mi)의 중요도나 질 등을 설계자나 사용자 등에 의해 미리 정의된 기준에 따라서 수치화한 것이다. Vi/Vj는 i번째 메트릭(Mi)의 중요도나 질 등을 j번째 메트릭의 중요도나 질로 나눈 것으로, 상술한 상대적 중요도와 동일할 수도 있다. 실시예에 따라서, Vi/Vj는 상술한 중요도를 정규화하여 획득한 값일 수도 있다. 물론, 상술한 바와 같이 각각의 메트릭(M1 내지 Mn)들 마다 별도로 정의된 수치를 직접 나누어 획득된 것일 수도 있다. 쌍대 비교 행렬에서 i행과 j열이 서로 교차하는 지점의 값(즉, Vi/Vj)은 j행과 i열이 서로 교차하는 지점의 값(즉, Vj/Vi)의 역수가 된다. 또한, i행과 j열이 교차하는 지점(즉, 동일한 메트릭(Mi) 간의 비교 값)은 1로 정의된다.

[0081] 이와 같이 퍼지 시스템(120)에서 제공된 메트릭 간의 퍼지-쌍대 비교 행렬이 획득되면, 각 기준의 퍼지 가중치가 연산된다(s14).

[0082] 구체적으로 퍼지 가중치는 하기의 수학식 1 내지 수학식 3을 이용하여 연산된다.

수학식 1

[0083] $$W_m = \Sigma W_n$$

수학식 2

[0084] $$W_n = \frac{V_{n-1}}{V_n}$$

수학식 3

[0085] $$s.t\ \max\Sigma W_m$$

[0086] 수학식 1 내지 수학식 3에서 Wn은 메트릭 n(Mn)에 대한 가중치이고, Mm은 메트릭 n(Mn)에 대한 가중치의 합을 의미한다. 수학식 1 내지 수학식 3에 의하면 퍼지 가중치가 연산될 수 있다.

[0087] 도 9는 행렬의 변화에 따른 가중치를 도시한 도면이다.

[0088] 각 기준의 퍼지 가중치가 연산되면, 퍼지 가중치를 기반으로 최종 스코어, 즉 최종 퍼지 가중치가 계산될 수 있다(s16). 메트릭에 대한 최종 퍼지 가중치가 연산되면, 도 9에 도시된 바와 같이 메트릭에 따라서 각각의 요소, 예를 들어, 에너지 소비, 능력(capacity), 자원 활용 정도, 가상 머신 개수 및/또는 대기 시간 등 각각에 대한 다양한 가중치들의 조합(P1 내지 P5)이 미리 정의된 설정이나 데이터에 따라서 결정된다. 이에 따라 메트릭에 대한 퍼지-분석적 계층 프로세스 기반의 가중치가 획득될 수 있게 된다.

[0089] 최종 퍼지 가중치가 연산되면, 제공하는 클라우드 서비스 및 자원 사용량에 따라 우선 순위가 결정된다. 가상 머신(VM-1 내지 VM-k+1)은, 결정된 우선 순위에 따라 선정된다(s18).

[0090] 우선 순위에 따라 선정된 가상 머신(VM-1 내지 VM-k+1)은 상술한 바와 같이 스케일링되고, 이에 의하여 제공되는 클라우드 서비스의 목적이나 용도 등에 따라 적절하게 가상 머신(VM-1 내지 VM-k+1)의 선택 및 선택된 가상 머신(VM-1 내지 VM-k+1)에 대한 자원 스케일링이 수행될 수 있게 된다.

[0091] 도 10은 퍼지 시스템의 기계 학습부의 동작을 설명하기 위한 도면이다.

[0092] 퍼지 시스템(120)은 모니터링 결과(예를 들어, 자원 사용량 데이터)와 학습 결과(예를 들어, 학습된 자원 사용량에 대한 모델 등)을 기반으로 퍼지화를 수행하고, 퍼지화된 메트릭에 대해 퍼지-분석적 계층 프로세스 기반의 룰을 설정할 수 있다.

[0093] 도 3에 도시된 바를 참조하면, 퍼지 시스템(120)은 기계 학습부(121) 및 퍼지 룰 처리부(123)를 포함할 수 있다.

[0094] 기계 학습부(121)는 자원 모니터링부(112)에서 전달되는 모니터링 결과(예를 들어, 중앙 처리 장치의 점유율이나 메모리 사용량 등)을 기반으로 기계 학습을 수행한다.

[0095] 구체적으로 기계 학습부(121)는, 평균 연산부(121a), 퍼지화부(121b) 및 기계 학습 처리부(121c)를 포함할 수 있다.

[0096] 평균 연산부(121a)는 일정한 기간 동안 입력된 모니터링 결과들의 평균을 연산한다. 예를 들어, 중앙 처리 장치의 점유율이 연속적으로 입력되면, 미리 정의된 기간 동안의 점유율의 평균을 연산할 수 있다. 연산 결과는 퍼지화부(121b)로 전달된다.

[0097] 퍼지화부(121b)는 입력 변수(즉, 연산 결과)를 명령이 실현될 수 있는 소정의 수치로 변환시켜 퍼지화를 수행한다. 구체적으로 퍼지화부(121b)는 스케일 매핑(scale mapping)을 수행하여 입력 변수(즉, 연산 결과)의 값의 범위를 일치시키는 논의 영역(Universe of discourse)으로 변환시킬 수 있다. 이에 따라 퍼지화된 데이터 집합이 획득된다. 퍼지화된 데이터 집합은 기계 학습 처리부(121c)로 전달된다.

[0098] 기계 학습 처리부(121c)는 기계 학습을 수행한다. 기계 학습은 소정의 기계 학습 알고리즘을 이용하여 구현 가능하다. 기계 학습 알고리즘은, 예를 들어, 심층 신경망(DNN, Deep Neural Network), 콘볼루션 신경망(CNN, Convolutional Neural Network), 순환 신경망(RNN, Recurrent Neural Network), 심층 신뢰 신경망(DBN, Deep Belief Network) 및 심층 Q-네트워크(Deep Q-Networks) 중 적어도 하나를 이용하여 구현될 수 있으나, 구현 가능한 기계 학습 알고리즘은 이에 한정되는 것은 아니다.

[0099] 기계 학습 처리부(121c)는 도 10에 도시된 바와 같이, 입력층과, 퍼치 가중치들이 적용되어 획득된 데이터들로 이루어지는 은닉층과, 퍼지 가중치가 반영되어 데이터가 출력되는 출력층으로 이루어진 알고리즘을 기반으로 학습을 수행할 수 있다. 이에 따라 자원 상태를 반영하여 학습이 수행될 수 있게 된다. 학습에 따라 학습된 모델은 퍼지 룰 구축부(123), 구체적으로는 모델 적용부(123b)로 전달될 수 있다.

[0100] 퍼지 룰 구축부(123)는, 현재 또는 현재로부터 일정한 시간 범위에 해당하는 자원 모니터링 결과를 획득하고, 획득 결과를 기반으로 퍼지-분석적 계층 프로세스 기반의 룰을 설정할 수 있다.

[0101] 구체적으로 퍼지 룰 구축부(123)는, 도 3에 도시된 바와 같이, 퍼지화부(123a), 모델 적용부(123b), 디퍼지화부(123c) 및 룰 생성부(123d)를 포함할 수 있다.

[0102] 퍼지화부(123a)는 입력 변수(즉, 현재의 모니터링 결과)를 명령이 실현될 수 있는 소정의 수치로 변환시켜 퍼지화를 수행할 수 있다. 다시 말해서, 퍼지화부(123a)는 현재의 모니터링 결과에 대한 스케일 매핑 작업을 수행할 수 있다.

[0103] 모델 적용부(123b)는 퍼지화된 데이터를 기계 학습 처리부(121c)에서 획득된 학습 모델에 적용하여 결과 값을 획득할 수 있다. 학습 모델의 출력 값은 디퍼지화부(123c)로 전달된다.

[0104] 디퍼지화부(123c)는, 출력 변수(즉, 학습 모델의 출력 값)을 원하는 범위로 변환하여 논의 영역으로 전환하는 스케일 매핑 작업을 수행한다. 이 과정에서, 학습 모델의 출력 값은 모니터링 결과와 유사한 범위의 값을 가질 수 있게 된다. 스케일 매핑 작업에 따라 획득된 결과는 룰 생성부(123d)로 전달된다.

[0105] 룰 생성부(123d)는, 디퍼지화된 출력 값을 기반으로 퍼지 룰을 생성 및 구축한다. 룰 생성부(123d)는 학습 모델의 출력 값으로부터 룰을 찾아내고 찾아낸 룰을 기록 및 저장하여 퍼지 룰을 생성할 수 있다. 생성된 퍼지 룰은 스케일링 결정부(114)으로 전달되고, 스케일링 결정부(114)는 퍼지 룰을 기반으로 가상 머신(VM-1 내지 VM-k+1)의 스케일링 여부 및 스케일링 방법을 결정할 수 있다.

[0106] 상술한 자원 스케일링 장치(110)의 자원 모니터링부(112), 스케일링 결정부(114) 및 자원 스케일링부(116)와, 퍼지 시스템(120)의 기계 학습부(121), 평균 연산부(121a), 퍼지화부(121b) 및 기계 학습 처리부(121c) 및 퍼지 룰 구축부(123), 퍼지화부(123a), 모델 적용부(123b), 디퍼지화부(123c) 및 룰 생성부(123d)는 논리적으로 구분되는 것일 수도 있고 또는 물리적으로 구분되는 것일 수도 있다. 논리적으로 구분되는 경우 이들은 하나 또는 둘 이상의 반도체 칩 및 관련 부품 등에 의해 구현될 수도 있다. 또한, 물리적으로 구분되는 경우, 이들은 각각 별도의 반도체 칩 및 관련 부품 등에 의해 구현될 수도 있다.

[0107] 상술한 자원 스케일링 장치(110) 및 퍼지 시스템(120)은, 실시예에 따라서, 하나의 장치, 예를 들어 어느 하나의 클라우드 컴퓨팅 장치(11, 12) 또는 자원 스케일링 제어 장치(13)에 의해 수행될 수도 있고, 또는 복수의 장치, 예를 들어, 적어도 하나의 클라우드 컴퓨팅 장치(11, 12) 및 자원 스케일링 제어 장치(13) 중 적어도 둘에 의해 수행될 수도 있다. 후자의 경우, 예를 들어, 자원 스케일링 장치(110)는 어느 하나의 클라우드 컴퓨팅 장치(11, 12)에 의해 구현되고, 퍼지 시스템(120)은 자원 스케일링 제어 장치(13)에 의해 구현될 수도 있다. 또한, 실시예에 따라서, 퍼지 시스템(120)은 클라우드(10) 외부에 존재하는 다른 컴퓨팅 장치에 의해 구현될 수도 있다. 즉, 자원 스케일링 장치(110)는 적어도 하나의 클라우드 컴퓨팅 장치(11, 12) 또는 자원 스케일링 제어 장치(13)에 의해 구현되고, 퍼지 시스템(120)은 퍼지와 관련된 연산 처리를 위해 별도로 구축되고 하나 또는 둘 이상의 컴퓨팅 장치를 갖는 다른 시스템에 의해 구현될 수도 있다.

[0108] 도 11은 종래 기술과의 에너지 소비량을 비교한 도면이다. 도 11의 x축은 할당된 가상 머신(V-1 내지 V-k+1)의 개수를 의미하고, y축은 에너지 소비량의 상대값을 표시한 것이다. 또한 도 11의 복수의 선분의 집합으로 표현된 좌측의 블록은 종래 기술이 이용하는 에너지 소비량을 의미하고, 단색으로 표시된 우측의 블록은 상술한 자원 스케일링 장치의 에너지 소비량을 의미한다.

[0109] 상술한 바와 같이 자원 스케일링 장치(110)가 자원 스케일링을 수행하면, 도 11에 도시된 바와 같이 에너지 소비량이 종래의 클라우드 시스템에 비해서 상당히 감소한다. 예를 들어, 종래의 클라우드 시스템의 경우 가상 머신(V-1 내지 V-k+1)의 개수에 따라서 에너지 소비량의 변화가 크며 또한 그 값 역시 매우 크다. 반면에 상술한 자원 스케일링 장치(110)를 이용하는 경우 가상 머신(V-1 내지 V-k+1)의 개수에 따른 에너지 소비량의 변화가 대략 일정하며, 또한 에너지 소비량의 절대적 수치가 상대적으로 낮음을 알 수 있다.

[0110] 이하 도 12를 참조하여 우선 순위 기반 자원 스케일링 방법의 일 실시예에 대해서 설명하도록 한다.

[0111] 도 12는 우선 순위 기반 자원 스케일링 방법의 일 실시예를 도시한 흐름도이다.

[0112] 도 12에 도시된 바에 의하면, 자원 스케일링을 수행하기 위한 장치나 시스템(예를 들어, 분산형 클라우드 시스템의 클라우드 컴퓨팅 장치나 자원 스케일링 제어 장치 등)이 구동을 개시한다(310).

[0113] 자원 스케일링을 수행하기 위한 장치나 시스템은, 자원 스케일링 장치의 자원이나 또는 다른 장치(예를 들어,

클라우드 컴퓨팅 장치 등)의 자원을 모니터링하고 자원 모니터링 결과를 획득한다(312). 이 경우, 자원 모니터링은 지속적으로, 연속적으로 수행될 수 있다.

[0114] 퍼지 시스템은 자원 모니터링 획득 결과를 기 획득한 학습 모델에 적용하여, 퍼지 룰을 결정한다(314). 구체적으로 퍼지 시스템은 현재의 자원 모니터링 획득 결과에 대한 퍼지화를 수행하고, 퍼지화된 데이터를 기 획득한 학습 모델에 적용하여 출력 값을 얻고, 출력 값에 대한 디퍼지화를 수행한 후, 디퍼지화된 결과를 기반으로 퍼지 룰을 생성할 수 있다. 여기서, 기 획득한 학습 모델은, 지속적으로, 연속적으로 입력되는 자원 모니터링 결과에 대한 기계 학습 결과 획득된 모델일 수 있다. 상세하게는 퍼지 시스템은 일정 기간 동안 연속적으로 입력되는 자원 모니터링 결과의 평균을 획득하고, 획득한 평균에 대한 퍼지화를 수행한 후 퍼지화된 결과에 대해 기계 학습을 수행하여 학습 모델을 획득할 수 있다.

[0115] 자원 스케일링 장치는 퍼지 시스템에서 획득한 퍼지 룰을 기반으로 스케일링 여부를 결정할 수 있다(316). 예를 들어, 자원 스케일링 장치는 전달된 퍼지 룰을 이용하여 자원의 필요 정도를 평가하고, 이를 기반으로 자원 스케일링 대상(예를 들어, 가상 머신)의 스케일링 여부 및 스케일링 방법을 결정할 수 있다. 가상 머신의 스케일링 여부 및 방법의 결정은 마이그레이션 등이 수행될 때 함께 또는 순차적으로 수행될 수도 있다.

[0116] 한편, 스케일링 여부나 방법의 결정과 동시에 또는 이시에, 자원 스케일링이 수행될 가상 머신이 결정될 수 있다(318). 자원 스케일링이 수행될 가능 머신은, 퍼지 쌍대 비교 행렬을 생성하고, 이를 기반으로 최적 가중치를 도출하고, 최적 가중치를 기반으로 최종적인 가중치를 결정한 후, 결정된 최종 가중치를 기반으로 복수의 가상 머신 중 적어도 하나의 가상 머신이 선택될 수 있다.

[0117] 가상 머신이 선택되면, 결정된 스케일링 방법에 따라서 가상 머신의 스케일링이 수행될 수 있다(320). 예를 들어, 기존의 가상 머신 또는 마이그레이션을 위해 새로 생성되는 가상 머신의 스케일-업, 스케일-다운, 스케일-인 또는 스케일-아웃이 수행될 수도 있다.

[0118] 상술한 실시예에 따른 우선 순위 기반 자원 방법은, 컴퓨터 장치에 의해 구동될 수 있는 프로그램의 형태로 구현될 수 있다. 여기서 프로그램은, 프로그램 명령, 데이터 파일 및 데이터 구조 등을 단독으로 또는 조합하여 포함할 수 있다. 프로그램은 기계어 코드나 고급 언어 코드를 이용하여 설계 및 제작된 것일 수 있다. 프로그램은 상술한 우선 순위 기반 자원 방법을 구현하기 위하여 특별히 설계된 것일 수도 있고, 컴퓨터 소프트웨어 분야에서 통상의 기술자에게 기 공지되어 사용 가능한 각종 함수나 정의를 이용하여 구현된 것일 수도 있다. 또한, 여기서, 컴퓨터 장치는, 프로그램의 기능을 실현 가능하게 하는 프로세서나 메모리 등을 포함하여 구현된 것일 수 있으며, 필요에 따라 통신 장치를 더 포함할 수도 있다.

[0119] 상술한 방법을 구현하기 위한 우선 순위 기반 자원 프로그램은, 컴퓨터에 의해 판독 가능한 기록 매체에 기록될 수 있다. 컴퓨터에 의해 판독 가능한 기록 매체는, 예를 들어, 하드 디스크나 플로피 디스크와 같은 자기 디스크 저장 매체, 자기 테이프, 콤팩트 디스크나 디브이디와 같은 광 기록 매체, 플롭티컬 디스크와 같은 자기-광 기록 매체 및 롬, 램 또는 플래시 메모리 등과 같은 반도체 저장 장치 등 컴퓨터 등의 호출에 따라 실행되는 특정 프로그램을 저장 가능한 다양한 종류의 하드웨어 장치를 포함할 수 있다.

[0120] 이상 우선 순위 기반 자원 장치 및 방법의 여러 실시예에 대해 설명하였으나, 우선 순위 기반 자원 장치 및 방법은 오직 상술한 실시예에 한정되는 것은 아니다. 해당 기술 분야에서 통상의 지식을 가진 자가 상술한 실시예를 기초로 수정 및 변형하여 구현 가능한 다양한 장치나 방법 역시 상술한 우선 순위 기반 자원 장치 및 방법의 일례가 될 수 있다. 예를 들어, 설명된 기술들이 설명된 방법과 다른 순서로 수행되거나, 및/또는 설명된 시스템, 구조, 장치, 회로 등의 구성 요소들이 설명된 방법과 다른 형태로 결합 또는 조합되거나 다른 구성 요소 또는 균등물에 의하여 대치되거나 또는 치환되더라도 상술한 우선 순위 기반 자원 장치 및 방법의 일 실시예가 될 수 있다.

부호의 설명

[0121]

1: 클라우드 시스템	10: 클라우드
11: 클라우드 컴퓨팅 장치	12: 중앙 클라우드 컴퓨팅 장치
13: 자원 스케일링 제어 장치	16: 통신부
17: 프로세서	18: 저장부
100: 자원 스케일링 시스템	110: 자원 스케일링 장치

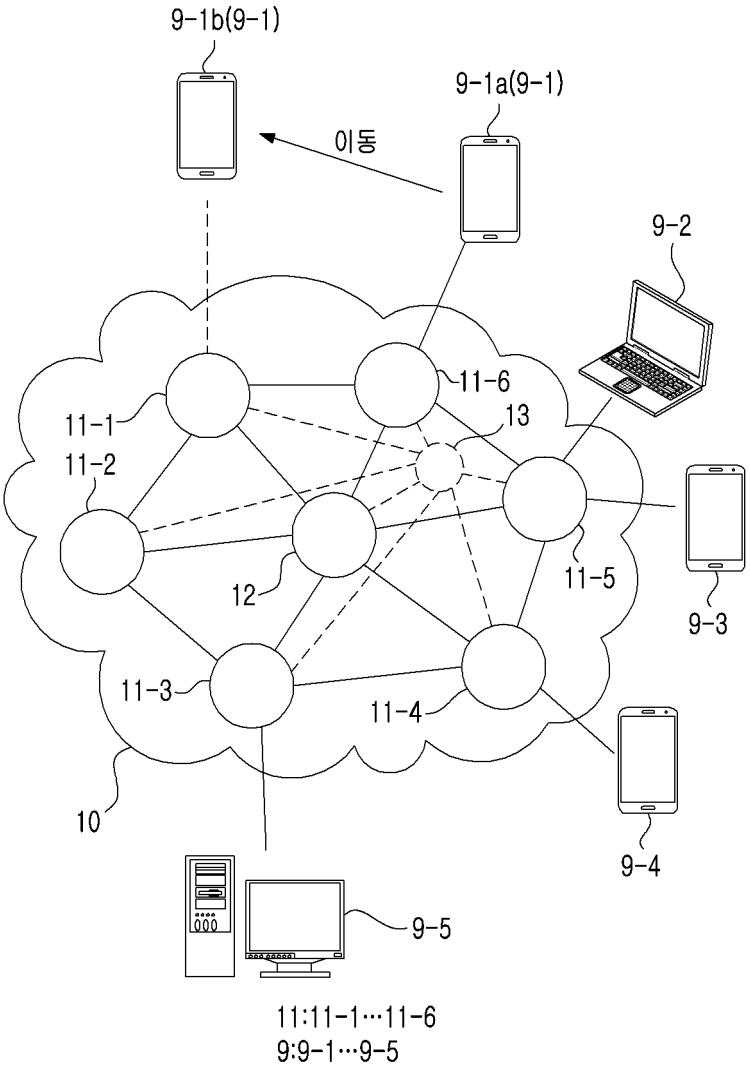
112: 자원 모니터링부 114: 스케일링 결정부

116: 자원 스케일링부 120: 퍼지 시스템

121: 기계 학습부 123: 퍼지 룰 구축부

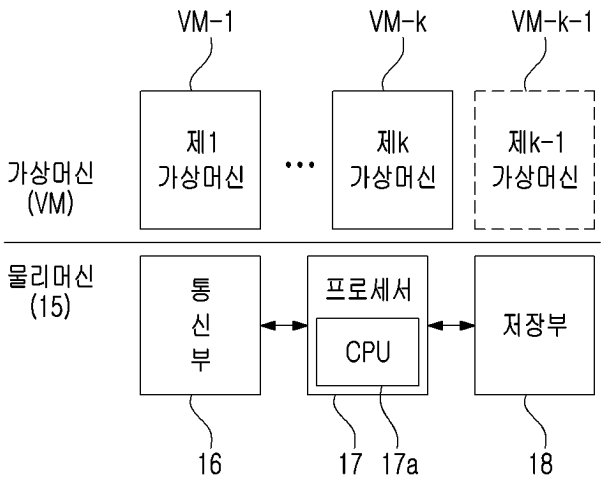
도면

도면1



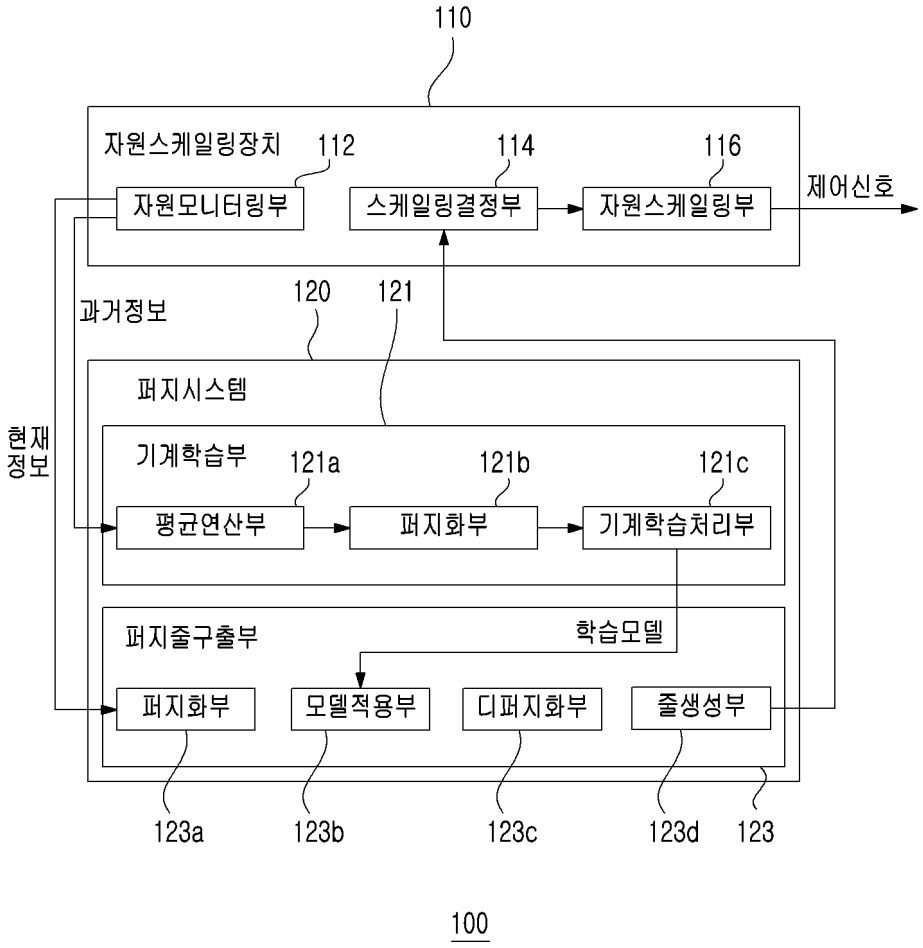
1

도면2


VM-1
VM-k
VM-k-1
제1
가상머신
제k
가상머신
제k-1
가상머신
가상머신
(VM)
물리머신
(15)
통
신
부
프로세서
CPU
저장부
16
17
17a
18

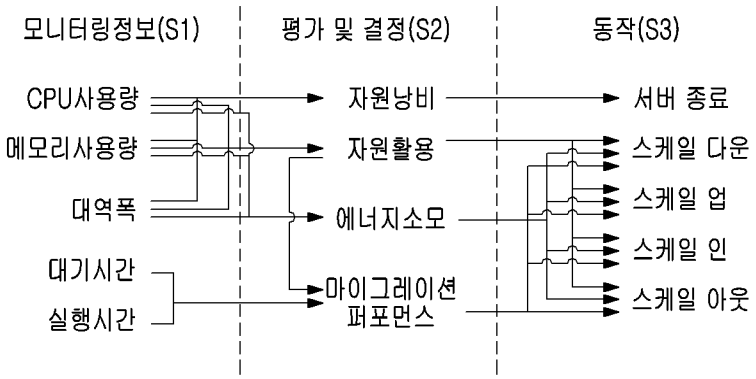
11,12

도면3

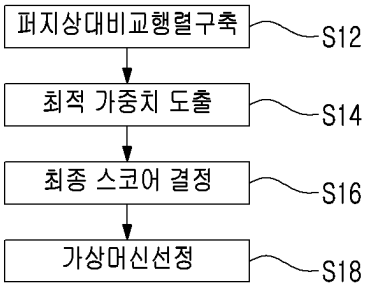

110
자원스케일링장치
112
114
116
자원모니터링부
스케일링결정부
자원스케일링부
제어신호
120
121
과거정보
퍼지시스템
현재
정보
기계학습부
121a
121b
121c
평균연산부
퍼지화부
기계학습처리부
퍼지출구출부
학습모델
퍼지화부
모델적용부
디퍼지화부
줄생성부
123a
123b
123c
123d
123

100

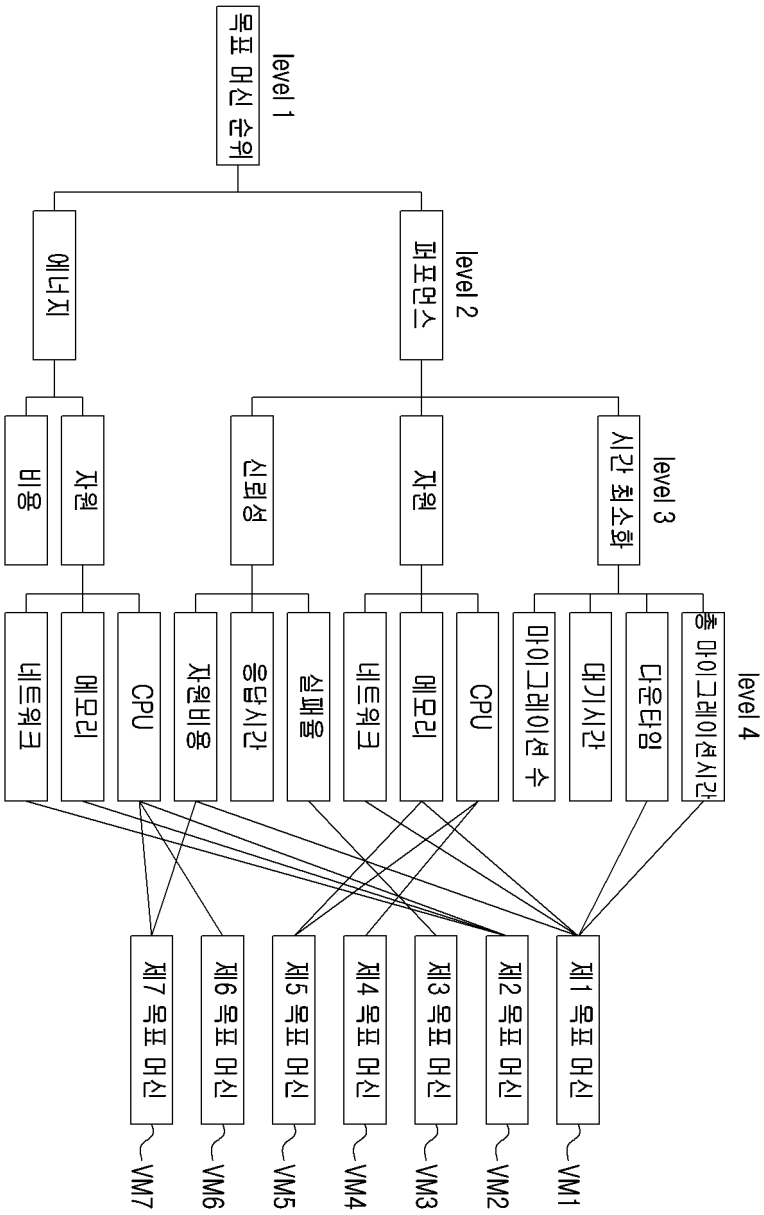
도면4


모니터링정보(S1)
평가 및 결정(S2)
동작(S3)
CPU사용량
메모리사용량
대역폭
대기시간
실행시간
자원낭비
자원활용
에너지소모
마이그레이션
퍼포먼스
서버 종료
스케일 다운
스케일 업
스케일 인
스케일 아웃

도면5


퍼지상대비교행렬구축
S12
최적 가중치 도출
S14
최종 스코어 결정
S16
가상머신선정
S18

도면6


level 1
목표 머신 순위
level 2
에너지
퍼포먼스
level 3
비용
자원
신뢰성
자원
시간 최소화
level 4
네트워크
메모리
CPU
자원비용
응답시간
실패율
네트워크
메모리
CPU
마이그레이션 수
대기시간
다운타임
총 마이그레이션시간
제7 목표 머신
제6 목표 머신
제5 목표 머신
제4 목표 머신
제3 목표 머신
제2 목표 머신
제1 목표 머신
VM7
VM6
VM5
VM4
VM3
VM2
VM1

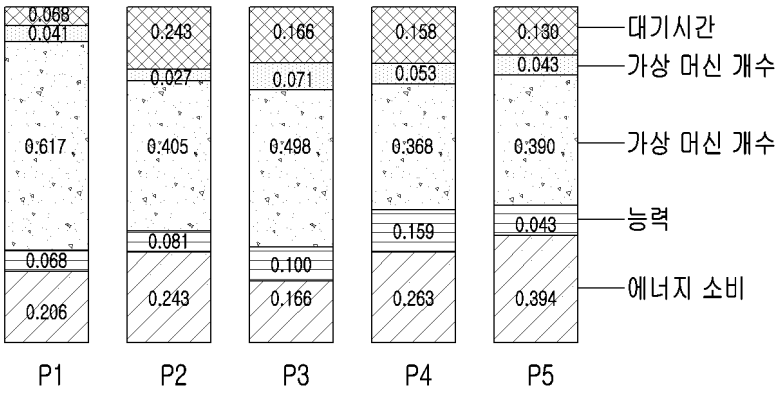
도면7

상대적 중요도	값
동등하게 중요함	1
다소 중요함	3
꽤 중요함	5
상당히 더 중요함	7
극히 더 중요함	9

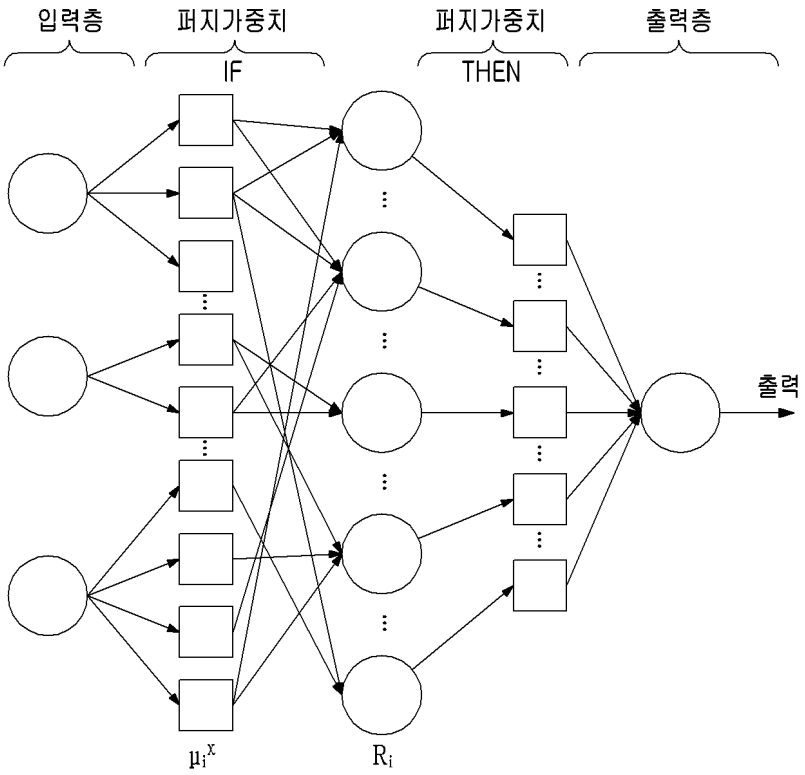
도면8

	M1	M2	…	Mn
M1	V1/V1	V1/V2	…	V1/Vn
M2	V2/V1	V2/V2	…	V2/Vn
…	…	…	…	…
Mn	Vn/V1	Vn/V2	…	Vn/Vn

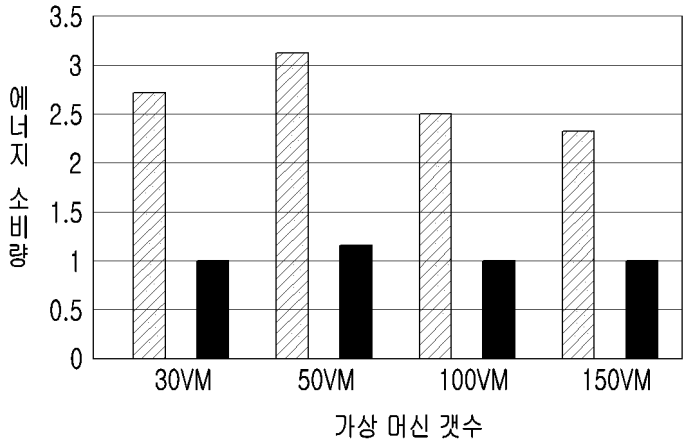
도면9



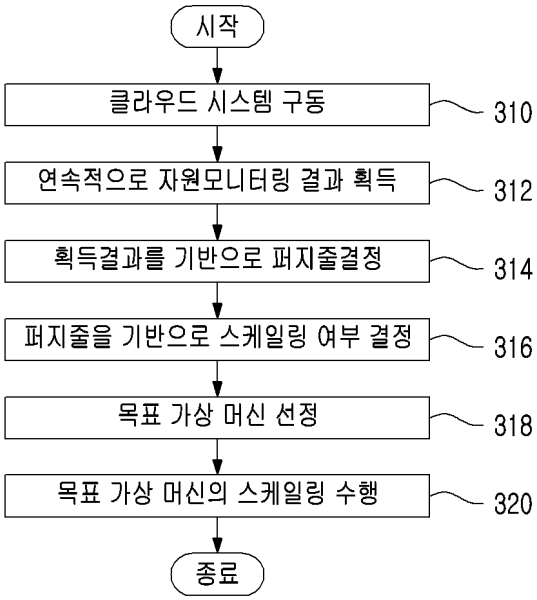
도면10


입력층
퍼지가중치
IF
퍼지가중치
THEN
출력층
출력
μ_i^x
R_i

도면11


3.5
3
2.5
2
1.5
1
0.5
0
에너지 소비량
30VM
50VM
100VM
150VM
가상 머신 갯수

도면12



【심사관 직권보정사항】

【직권보정 1】

【보정항목】 발명(고안)의 설명

【보정세부항목】 문단번호 [0006], [0028]

【변경전】

자원 상태를 동작으로 판단하여

【변경후】

자원 상태를 동적으로 판단하여

【직권보정 2】

【보정항목】 청구범위

【보정세부항목】 청구항 14

【변경전】

상기 자원 모니터링부에 의한

【변경후】

자원 모니터링부에 의한