



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2017-0094899
(43) 공개일자 2017년08월22일

(51) 국제특허분류(Int. Cl.)
G06F 9/50 (2006.01) G06F 15/16 (2006.01)
(52) CPC특허분류
G06F 9/5055 (2013.01)
G06F 15/16 (2013.01)
(21) 출원번호 10-2016-0016199
(22) 출원일자 2016년02월12일
심사청구일자 없음

(71) 출원인
삼성에스디에스 주식회사
서울특별시 송파구 올림픽로35길 125 (신천동)
(72) 발명자
김성일
서울특별시 송파구 올림픽로35길 125 (신천동, 삼성SDS West Campus)
김병수
서울특별시 송파구 올림픽로35길 125 (신천동, 삼성SDS West Campus)
김재정
서울특별시 송파구 올림픽로35길 125 (신천동, 삼성SDS West Campus)
(74) 대리인
두호특허법인

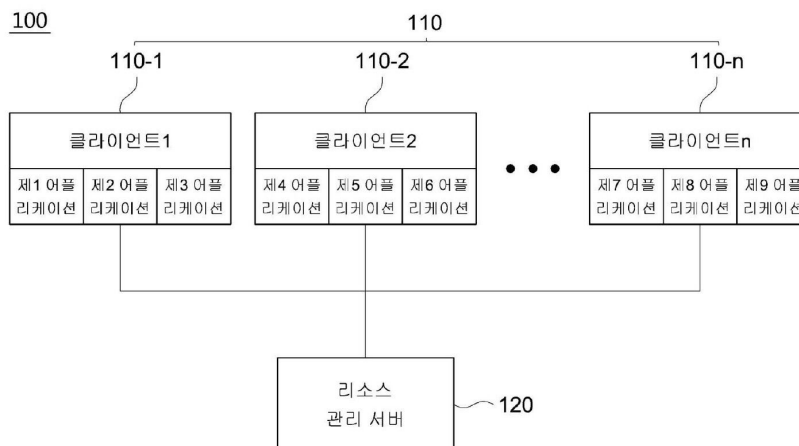
전체 청구항 수 : 총 10 항

(54) 발명의 명칭 컴퓨팅 리소스 관리 시스템 및 방법

(57) 요약

컴퓨팅 리소스 관리 시스템 및 방법이 개시된다. 본 발명의 일 실시예에 따른 컴퓨팅 리소스 관리 시스템은, 하나 이상의 어플리케이션을 구비하고, 설정된 리소스 할당 정책에 따라 씨그룹(cgroups: control groups)를 이용하여 컴퓨팅 리소스를 각 어플리케이션별로 할당하며, 상기 각 어플리케이션별 컴퓨팅 리소스의 사용량에 관한 정보를 수집하는 클라이언트; 및 상기 클라이언트로부터 상기 사용량에 관한 정보를 수신하고, 상기 사용량에 따라 상기 리소스 할당 정책을 변경하는 리소스 관리 서버를 포함하며, 상기 클라이언트는, 변경된 상기 리소스 할당 정책을 상기 리소스 관리 서버로부터 수신하고, 변경된 상기 리소스 할당 정책에 따라 상기 씨그룹스를 이용하여 상기 컴퓨팅 리소스를 상기 각 어플리케이션별로 재할당한다.

대표도



(52) CPC특허분류
G06F 9/5061 (2013.01)

명세서

청구범위

청구항 1

하나 이상의 어플리케이션을 구비하고, 설정된 리소스 할당 정책에 따라 씨그룹스(cgroups: control groups)를 이용하여 컴퓨팅 리소스를 각 어플리케이션별로 할당하며, 상기 각 어플리케이션별 컴퓨팅 리소스의 사용량에 관한 정보를 수집하는 클라이언트; 및

상기 클라이언트로부터 상기 사용량에 관한 정보를 수신하고, 상기 사용량에 따라 상기 리소스 할당 정책을 변경하는 리소스 관리 서버를 포함하며,

상기 클라이언트는, 변경된 상기 리소스 할당 정책을 상기 리소스 관리 서버로부터 수신하고, 변경된 상기 리소스 할당 정책에 따라 상기 씨그룹스를 이용하여 상기 컴퓨팅 리소스를 상기 각 어플리케이션별로 재할당하는, 컴퓨팅 리소스 관리 시스템.

청구항 2

청구항 1에 있어서,

상기 사용량에 관한 정보는, 설정된 시간 동안 수집된 상기 각 어플리케이션별 컴퓨팅 리소스의 실시간 사용량, 최대 사용량, 최소 사용량 및 평균 사용량 중 하나 이상을 포함하는, 컴퓨팅 리소스 관리 시스템.

청구항 3

청구항 1에 있어서,

상기 리소스 관리 서버는, 상기 리소스 할당 정책에 따라 상기 어플리케이션에 할당된 컴퓨팅 리소스의 할당량과 수집된 상기 어플리케이션의 컴퓨팅 리소스의 사용량을 비교하고, 상기 비교 결과에 따라 상기 리소스 할당 정책을 변경하는, 컴퓨팅 리소스 관리 시스템.

청구항 4

청구항 3에 있어서,

상기 리소스 관리 서버는, 상기 사용량이 상기 할당량 이상인 경우 상기 어플리케이션에 할당되는 컴퓨팅 리소스를 증가시키도록 상기 리소스 할당 정책을 변경하며, 상기 사용량이 상기 할당량 미만인 경우 상기 어플리케이션에 할당되는 컴퓨팅 리소스를 감소시키도록 상기 리소스 할당 정책을 변경하는, 컴퓨팅 리소스 관리 시스템.

청구항 5

청구항 1에 있어서,

상기 리소스 관리 서버는, 수집된 상기 어플리케이션의 컴퓨팅 리소스의 사용량이 상기 사용량을 수집하는 시간대에 따라 변화하는 경우 상기 어플리케이션에 할당되는 컴퓨팅 리소스를 상기 시간대에 따라 증가시키거나 감소시키도록 상기 리소스 할당 정책을 변경하는, 컴퓨팅 리소스 관리 시스템.

청구항 6

하나 이상의 어플리케이션을 구비하는 클라이언트에서, 설정된 리소스 할당 정책에 따라 씨그룹스(cgroups: control groups)를 이용하여 컴퓨팅 리소스를 각 어플리케이션별로 할당하는 단계;

상기 클라이언트에서, 상기 각 어플리케이션별 컴퓨팅 리소스의 사용량에 관한 정보를 수집하는 단계;

리소스 관리 서버에서, 상기 클라이언트로부터 상기 사용량에 관한 정보를 수신하는 단계;

상기 리소스 관리 서버에서, 상기 사용량에 따라 상기 리소스 할당 정책을 변경하는 단계;

상기 클라이언트에서, 변경된 상기 리소스 할당 정책을 상기 리소스 관리 서버로부터 수신하는 단계; 및

상기 클라이언트에서, 변경된 상기 리소스 할당 정책에 따라 상기 씨그룹스를 이용하여 상기 컴퓨팅 리소스를 상기 각 어플리케이션별로 재할당하는 단계를 포함하는, 컴퓨팅 리소스 관리 방법.

청구항 7

청구항 6에 있어서,

상기 사용량에 관한 정보는, 설정된 시간 동안 수집된 상기 각 어플리케이션별 컴퓨팅 리소스의 실시간 사용량, 최대 사용량, 최소 사용량 및 평균 사용량 중 하나 이상을 포함하는, 컴퓨팅 리소스 관리 방법.

청구항 8

청구항 6에 있어서,

상기 리소스 할당 정책을 변경하는 단계는, 상기 리소스 할당 정책에 따라 상기 어플리케이션에 할당된 컴퓨팅 리소스의 할당량과 수집된 상기 어플리케이션의 컴퓨팅 리소스의 사용량을 비교하고, 상기 비교 결과에 따라 상기 리소스 할당 정책을 변경하는, 컴퓨팅 리소스 관리 방법.

청구항 9

청구항 8에 있어서,

상기 리소스 할당 정책을 변경하는 단계는, 상기 사용량이 상기 할당량 이상인 경우 상기 어플리케이션에 할당되는 컴퓨팅 리소스를 증가시키도록 상기 리소스 할당 정책을 변경하며, 상기 사용량이 상기 할당량 미만인 경우 상기 어플리케이션에 할당되는 컴퓨팅 리소스를 감소시키도록 상기 리소스 할당 정책을 변경하는, 컴퓨팅 리소스 관리 방법.

청구항 10

청구항 6에 있어서,

상기 리소스 할당 정책을 변경하는 단계는, 수집된 상기 어플리케이션의 컴퓨팅 리소스의 사용량이 상기 사용량을 수집하는 시간대에 따라 변화하는 경우 상기 어플리케이션에 할당되는 컴퓨팅 리소스를 상기 시간대에 따라 증가시키거나 감소시키도록 상기 리소스 할당 정책을 변경하는, 컴퓨팅 리소스 관리 방법.

발명의 설명

기술 분야

본 발명의 실시예들은 컴퓨팅 리소스를 효율적으로 관리하는 기술과 관련된다.

[0001]

배경 기술

- [0002] 컴퓨팅 장치의 하드웨어의 성능이 향상됨에 따라 단일의 하드웨어 및 단일의 운영체제 상에서 복수의 어플리케이션을 실행시킬 수 있는 환경이 구축되었다. 그러나, 하나의 운영체제 상에서 복수의 어플리케이션을 동시에 실행시키는 경우, 어느 하나의 어플리케이션이 컴퓨팅 리소스를 독점할 수 있다. 이에 따라, 상기 운영체제에 설치된 타 어플리케이션을 실행시키기 위한 컴퓨팅 장치의 성능이 저하되는 문제가 발생한다. 이러한 문제를 극복하기 위해, 컴퓨팅 리소스(CPU, 메모리, 네트워크 등)의 활용률을 높이기 위한 기술들이 개발되고 있다.
- [0003] 일 예시로서, 가상화된 하드웨어 리소스(가상 머신)를 생성하는 가상화 기술이 개발되었다. 그러나, 이 경우 가상화된 컴퓨터 환경에 독립된 운영체제를 설치하여야 하며, 각 운영 체제에서는 다른 운영체제를 별도의 호스트로 인식한다. 또한, 각 가상 머신이 독립적으로 실행되므로 가상 머신의 운영체제 간의 캐쉬(cache)의 공유가 불가능하다. 따라서, 가상화 기술에 따르면, 절차가 복잡해지고 메모리가 낭비되는 문제점이 있다.
- [0004] 다른 예시로서, 리눅스 컨테이너 기술은 가상 머신을 위해 별도의 계층이 존재하지 않고 운영체제 수준에서 자원의 분할 및 제어를 담당한다. 따라서, 리눅스 컨테이너 기술에 따르면, 성능 저하가 완화되고 운영체제를 별도로 설치하지 않아도 되는 장점이 있다. 그러나, 리눅스 컨테이너 기술에 따르면, 초기에 설정한 컴퓨팅 리소스를 모니터링하고 각 어플리케이션별 리소스 할당량을 동적으로 변경하기가 어려운 문제점이 있다.

선행기술문헌

특허문헌

- [0005] (특허문헌 0001) 한국공개특허공보 제10-2012-0071979호(2012.07.03)

발명의 내용

해결하려는 과제

- [0006] 본 발명의 실시예들은 어플리케이션별 컴퓨팅 리소스의 사용량 정보에 기초하여 컴퓨팅 리소스를 어플리케이션별로 재할당함으로써 컴퓨팅 리소스 활용률을 향상시키기 위한 것이다.

과제의 해결 수단

- [0007] 본 발명의 예시적인 실시예에 따르면, 하나 이상의 어플리케이션을 구비하고, 설정된 리소스 할당 정책에 따라 씨그룹스(cgroups: control groups)를 이용하여 컴퓨팅 리소스를 각 어플리케이션별로 할당하며, 상기 각 어플리케이션별 컴퓨팅 리소스의 사용량에 관한 정보를 수집하는 클라이언트; 및 상기 클라이언트로부터 상기 사용량에 관한 정보를 수신하고, 상기 사용량에 따라 상기 리소스 할당 정책을 변경하는 리소스 관리 서버를 포함하며, 상기 클라이언트는, 변경된 상기 리소스 할당 정책을 상기 리소스 관리 서버로부터 수신하고, 변경된 상기 리소스 할당 정책에 따라 상기 씨그룹스를 이용하여 상기 컴퓨팅 리소스를 상기 각 어플리케이션별로 재할당하는, 컴퓨팅 리소스 관리 시스템이 제공된다.
- [0008] 상기 사용량에 관한 정보는, 설정된 시간 동안 수집된 상기 각 어플리케이션별 컴퓨팅 리소스의 실시간 사용량, 최대 사용량, 최소 사용량 및 평균 사용량 중 하나 이상을 포함할 수 있다.
- [0009] 상기 리소스 관리 서버는, 상기 리소스 할당 정책에 따라 상기 어플리케이션에 할당된 컴퓨팅 리소스의 할당량과 수집된 상기 어플리케이션의 컴퓨팅 리소스의 사용량을 비교하고, 상기 비교 결과에 따라 상기 리소스 할당 정책을 변경할 수 있다.
- [0010] 상기 리소스 관리 서버는, 상기 사용량이 상기 할당량 이상인 경우 상기 어플리케이션에 할당되는 컴퓨팅 리소스를 증가시키도록 상기 리소스 할당 정책을 변경하며, 상기 사용량이 상기 할당량 미만인 경우 상기 어플리케이션에 할당되는 컴퓨팅 리소스를 감소시키도록 상기 리소스 할당 정책을 변경할 수 있다.
- [0011] 상기 리소스 관리 서버는, 수집된 상기 어플리케이션의 컴퓨팅 리소스의 사용량이 상기 사용량을 수집하는 시간

대에 따라 변화하는 경우 상기 어플리케이션에 할당되는 컴퓨팅 리소스를 상기 시간대에 따라 증가시키거나 감소시키도록 상기 리소스 할당 정책을 변경할 수 있다.

[0012] 본 발명의 다른 예시적인 실시예에 따르면, 클라이언트에서, 하나 이상의 어플리케이션을 구비하고, 설정된 리소스 할당 정책에 따라 씨그룹스(cgroups: control groups)를 이용하여 컴퓨팅 리소스를 각 어플리케이션별로 할당하는 단계; 상기 클라이언트에서, 상기 각 어플리케이션별 컴퓨팅 리소스의 사용량에 관한 정보를 수집하는 단계; 리소스 관리 서버에서, 상기 클라이언트로부터 상기 사용량에 관한 정보를 수신하는 단계; 상기 리소스 관리 서버에서, 상기 사용량에 따라 상기 리소스 할당 정책을 변경하는 단계; 상기 클라이언트에서, 변경된 상기 리소스 할당 정책을 상기 리소스 관리 서버로부터 수신하는 단계; 및 상기 클라이언트에서, 변경된 상기 리소스 할당 정책에 따라 상기 씨그룹스를 이용하여 상기 컴퓨팅 리소스를 상기 각 어플리케이션별로 재할당하는 단계를 포함하는, 컴퓨팅 리소스 관리 방법이 제공된다.

[0013] 상기 사용량에 관한 정보는, 설정된 시간 동안 수집된 상기 각 어플리케이션별 컴퓨팅 리소스의 실시간 사용량, 최대 사용량, 최소 사용량 및 평균 사용량 중 하나 이상을 포함할 수 있다.

[0014] 상기 리소스 할당 정책을 변경하는 단계는, 상기 리소스 할당 정책에 따라 상기 어플리케이션에 할당된 컴퓨팅 리소스의 할당량과 수집된 상기 어플리케이션의 컴퓨팅 리소스의 사용량을 비교하고, 상기 비교 결과에 따라 상기 리소스 할당 정책을 변경할 수 있다.

[0015] 상기 리소스 할당 정책을 변경하는 단계는, 상기 사용량이 상기 할당량 이상인 경우 상기 어플리케이션에 할당되는 컴퓨팅 리소스를 증가시키도록 상기 리소스 할당 정책을 변경하며, 상기 사용량이 상기 할당량 미만인 경우 상기 어플리케이션에 할당되는 컴퓨팅 리소스를 감소시키도록 상기 리소스 할당 정책을 변경할 수 있다.

[0016] 상기 리소스 할당 정책을 변경하는 단계는, 수집된 상기 어플리케이션의 컴퓨팅 리소스의 사용량이 상기 사용량을 수집하는 시간대에 따라 변화하는 경우 상기 어플리케이션에 할당되는 컴퓨팅 리소스를 상기 시간대에 따라 증가시키거나 감소시키도록 상기 리소스 할당 정책을 변경할 수 있다.

발명의 효과

[0017] 본 발명의 실시예들에 따르면, 클라이언트가 각 어플리케이션별로 할당된 범위 내의 컴퓨팅 리소스만 사용하도록 제한함으로써, 타 어플리케이션이 컴퓨팅 리소스를 선점하여 전체적인 컴퓨터의 성능이 저하되는 것을 방지할 수 있다.

[0018] 또한, 본 발명의 실시예들에 따르면, 리소스 관리 서버가 설정된 시간 동안 각 어플리케이션별로 수집된 컴퓨팅 리소스 사용량을 분석하고, 상기 분석 결과에 따라 리소스 할당 정책을 변경함으로써, 상기 어플리케이션별로 최적 크기의 컴퓨팅 리소스를 할당할 수 있다.

[0019] 또한, 본 발명의 실시예들에 따르면, 특정 시간대에 따라 어플리케이션의 컴퓨팅 리소스의 사용량이 변경되는 경우, 상기 시간대에 따라 상기 어플리케이션에 할당되는 컴퓨팅 리소스를 동적으로 변화시킴으로써, 한정된 컴퓨팅 리소스를 효율적으로 활용할 수 있다.

도면의 간단한 설명

[0020] 도 1은 본 발명의 일 실시예에 따른 컴퓨팅 리소스 관리 시스템의 상세 구성을 나타낸 블록도

도 2는 본 발명의 일 실시예에 따른 클라이언트의 상세 구성을 나타낸 블록도

도 3은 본 발명의 일 실시예에 따른 리소스 관리 서버가 할당 정책을 수정하는 과정을 설명하기 위한 예시도

도 4는 본 발명의 일 실시예에 따른 컴퓨팅 리소스 관리 방법을 설명하기 위한 흐름도

발명을 실시하기 위한 구체적인 내용

[0021] 이하, 도면을 참조하여 본 발명의 구체적인 실시형태를 설명하기로 한다. 이하의 상세한 설명은 본 명세서에서 기술된 방법, 장치 및/또는 시스템에 대한 포괄적인 이해를 돕기 위해 제공된다. 그러나 이는 예시에 불과하며 본 발명은 이에 제한되지 않는다.

- [0022] 본 발명의 실시예들을 설명함에 있어서, 본 발명과 관련된 공지기술에 대한 구체적인 설명이 본 발명의 요지를 불필요하게 흐릴 수 있다고 판단되는 경우에는 그 상세한 설명을 생략하기로 한다. 그리고, 후술되는 용어들은 본 발명에서의 기능을 고려하여 정의된 용어들로서 이는 사용자, 운용자의 의도 또는 관례 등에 따라 달라질 수 있다. 그러므로 그 정의는 본 명세서 전반에 걸친 내용을 토대로 내려져야 할 것이다. 상세한 설명에서 사용되는 용어는 단지 본 발명의 실시예들을 기술하기 위한 것이며, 결코 제한적이어서는 안 된다. 명확하게 달리 사용되지 않는 한, 단수 형태의 표현은 복수 형태의 의미를 포함한다. 본 설명에서, "포함" 또는 "구비"와 같은 표현은 어떤 특성들, 숫자들, 단계들, 동작들, 요소들, 이들의 일부 또는 조합을 가리키기 위한 것이며, 기술된 것 이외에 하나 또는 그 이상의 다른 특성, 숫자, 단계, 동작, 요소, 이들의 일부 또는 조합의 존재 또는 가능성을 배제하도록 해석되어서는 안 된다.
- [0023] 도 1은 본 발명의 일 실시예에 따른 컴퓨팅 리소스 관리 시스템(100)의 상세 구성을 나타낸 블록도이다. 도 1에 도시된 바와 같이, 본 발명의 일 실시예에 따른 컴퓨팅 리소스 관리 시스템(100)은 하나 이상의 클라이언트(110-1, 110-2, ..., 110-n) 및 리소스 관리 서버(120)를 포함한다.
- [0024] 본 실시예들에 있어서 클라이언트(110)는 하드웨어, 소프트웨어 및 상기 하드웨어와 소프트웨어를 관리 및 제어하는 운영체제(OS: operating system)를 포함하는 사용자 단말로서, 예를 들어 데스크탑, 노트북, PDA 등이 될 수 있다. 여기서, 하드웨어는 예를 들어, 중앙처리장치(CPU: central processing unit), 메모리(램), 디스크 입출력 장치(하드디스크), 네트워크 장치 등이 될 수 있다. 또한, CPU, 메모리, 디스크 입출력 장치, 네트워크 장치는 컴퓨팅 리소스를 구성하는 각각의 요소일 수 있다. 운영체제는 예를 들어, 리눅스(Linux)일 수 있다.
- [0025] 클라이언트(110)는 하나 이상의 어플리케이션을 구비할 수 있다. 어플리케이션은 클라이언트(110)에서 실행되는 소프트웨어로서, 예를 들어 로그 분석기일 수 있다. 또한, 하나 이상의 어플리케이션은 클라이언트(110)에 구비된 운영체제 상에서 실행될 수 있다. 이에 따라, 동일한 운영체제 상에서 실행된 어플리케이션들은 각각의 실행에 의해 발생한 캐시(cache)를 공유할 수 있다.
- [0026] 클라이언트(110)는 상기 클라이언트(110)에 구비된 어플리케이션별로 컴퓨팅 리소스를 각각 할당할 수 있다. 일 실시예에 따르면, 컴퓨팅 리소스는 예를 들어, CPU, 메모리, 디스크 입출력 장치 및 네트워크 장치 중 하나 이상을 포함할 수 있다.
- [0027] 클라이언트(110)는 설정된 리소스 할당 정책에 따라 씨그룹스(cgroups: control groups)를 이용하여 컴퓨팅 리소스를 각 어플리케이션별로 할당할 수 있다. 씨그룹스는 리눅스에 구비되는 컴퓨팅 리소스 제어 기능으로서, 컴퓨팅 리소스를 어플리케이션별로 할당하도록 상기 컴퓨팅 리소스를 제어할 수 있다. 한편, 리소스 할당 정책은 클라이언트(110)의 각 어플리케이션별 컴퓨팅 리소스의 할당량에 관한 정보를 포함할 수 있다. 예를 들어, 리소스 할당 정책은 제1 어플리케이션에 클라이언트1(110-1)의 CPU 전체 크기의 30% 및 클라이언트1(110-1)의 메모리 전체 크기의 40%가 할당되는 것을 나타내는 정보를 포함할 수 있으며, 제2 어플리케이션에 클라이언트1(110-1)의 CPU 전체 크기의 40% 및 클라이언트1(110-1)의 메모리 전체 크기의 50%가 할당되는 것을 나타내는 정보를 포함할 수 있다. 클라이언트(110)는 어플리케이션별로 할당된 컴퓨팅 리소스의 할당량 중 적어도 일부를 사용하여 상기 어플리케이션을 각각 실행할 수 있다. 예를 들어, 제1 어플리케이션에 상기 클라이언트(110)의 CPU 전체 크기의 50%가 할당되는 경우, 제1 어플리케이션은 상기 할당된 컴퓨팅 리소스 범위 내에서 실행될 수 있다.
- [0028] 이에 따라, 클라이언트(110)는 각 어플리케이션을 실행함에 있어서 타 어플리케이션으로 할당된 컴퓨팅 리소스에는 접근할 수 없다. 즉, 본 발명의 실시예들에 따르면, 클라이언트(110)가 각 어플리케이션별로 할당된 범위 내의 컴퓨팅 리소스를 사용하도록 제한함으로써, 타 어플리케이션이 컴퓨팅 리소스를 선점하여 컴퓨터의 성능이 저하되는 것을 방지할 수 있다.
- [0029] 또한, 클라이언트(110)는 설정된 시간(예를 들어, 하루) 동안 어플리케이션별로 사용되는 컴퓨팅 리소스의 사용량에 관한 정보를 수집할 수 있다. 사용량에 관한 정보는 설정된 시간 동안 수집된 상기 각 어플리케이션별 컴퓨팅 리소스의 실시간 사용량, 최대 사용량, 최소 사용량 및 평균 사용량 중 하나 이상을 포함할 수 있다. 이와 같이, 클라이언트(110)는 각 어플리케이션별 컴퓨터 리소스의 사용량에 관한 정보를 수집함으로써, 어플리케이션별 컴퓨팅 리소스의 사용 패턴을 획득할 수 있다.
- [0030] 클라이언트(110)는 수집된 각 어플리케이션별 컴퓨팅 리소스의 사용량에 관한 정보를 리소스 관리 서버(120)로 전송하고, 리소스 관리 서버(120)로부터 변경된 리소스 할당 정책을 수신할 수 있다. 후술할 바와 같이, 리소스

관리 서버(120)는 상기 각 어플리케이션별 컴퓨팅 리소스의 사용량에 따라 리소스 할당 정책을 변경하고, 클라이언트(110)로 상기 리소스 할당 정책을 전송할 수 있다. 상술한 바와 같이, 리소스 할당 정책은 각 어플리케이션별 컴퓨팅 리소스의 할당량에 관한 정보를 포함할 수 있다.

[0031] 클라이언트(110)는 변경된 리소스 할당 정책에 따라 씨그먼트를 이용하여 각 어플리케이션별로 상기 컴퓨팅 리소스를 재할당할 수 있다. 예를 들어, 클라이언트(110)는 리소스 할당 정책에 따라 제1 어플리케이션에 할당되는 CPU의 할당량을 증가시키거나, 감소시킬 수 있다.

[0032] 리소스 관리 서버(120)는 리소스 할당 정책을 관리하는 장치로서, 리소스 할당 정책을 각 클라이언트(110)에 전송하고, 각 클라이언트(110)로부터 수신된 각 어플리케이션별 컴퓨팅 리소스의 사용량에 관한 정보로부터 상기 리소스 할당 정책을 변경할 수 있다. 일 예시로서, 리소스 관리 서버(120)는 상기 리소스 할당 정책에 따라 상기 어플리케이션에 할당된 컴퓨팅 리소스의 할당량과 수집된 상기 어플리케이션의 컴퓨팅 리소스의 사용량을 비교하고, 상기 비교 결과에 따라 상기 리소스 할당 정책을 변경할 수 있다. 구체적으로, 리소스 관리 서버(120)는 상기 사용량이 상기 할당량 이상인 경우 상기 어플리케이션에 할당되는 컴퓨팅 리소스를 증가시키도록 상기 리소스 할당 정책을 변경하며, 상기 사용량이 상기 할당량 미만인 경우 상기 어플리케이션에 할당되는 컴퓨팅 리소스를 감소시키도록 상기 리소스 할당 정책을 변경할 수 있다. 예를 들어, 클라이언트(110)가 제1 어플리케이션을 실행하는 데 사용하는 컴퓨팅 리소스의 최대 사용량이 제1 어플리케이션에 할당된 컴퓨팅 리소스의 할당량보다 설정된 값(예를 들어, CPU 전체 크기의 5%) 이상 작은 경우, 리소스 관리 서버(120)는 제1 어플리케이션에 할당되는 컴퓨팅 리소스의 할당량을 감소시키도록 상기 리소스 할당 정책을 변경할 수 있다. 또한, 클라이언트(110)가 어플리케이션의 실행 과정에서 어플리케이션에 할당된 컴퓨팅 리소스를 모두 사용하는 경우, 즉 컴퓨팅 리소스의 사용량이 할당량과 같아지는 경우 상기 어플리케이션에는 더 많은 컴퓨팅 리소스가 필요할 수 있다. 예를 들어, 제1 어플리케이션 및 제2 어플리케이션에 할당되는 CPU의 할당량이 각각 50% 및 30%이며, 제1 어플리케이션 및 제2 어플리케이션의 CPU의 사용량이 각각 40% 및 30%인 경우, 리소스 관리 서버(120)는 제1 어플리케이션에 할당되는 CPU의 할당량을 감소시키고 제2 어플리케이션에 할당된 CPU의 할당량을 증가시키도록 상기 리소스 할당 정책을 변경할 수 있다.

[0033] 다른 예시로서, 리소스 관리 서버(120)는 수집된 상기 어플리케이션의 컴퓨팅 리소스의 사용량이 상기 사용량을 수집하는 시간대에 따라 변화하는 경우 상기 어플리케이션에 할당되는 컴퓨팅 리소스가 상기 시간대에 따라 수집된 컴퓨팅 리소스의 사용량이 되도록 상기 리소스 할당 정책을 변경할 수 있다. 즉, 리소스 관리 서버(120)는 어플리케이션별로 할당된 컴퓨팅 리소스의 할당량이 특정 시간대에 따라 동적으로 변경되도록 리소스 할당 정책을 설정 및 변경할 수 있다. 이를 위해, 리소스 관리 서버(120)는 설정된 시간(예를 들어, 24시간)동안 수집된 각 어플리케이션별 컴퓨팅 리소스의 실시간 사용량을 분석할 수 있다. 다음으로, 리소스 관리 서버(120)는 각 어플리케이션별 컴퓨팅 리소스의 사용량이 시간대에 따라 설정된 값 이상 변경되는지 또는 주기적으로 변경되는지 여부를 판단할 수 있다. 예를 들어, 주간에는 제1 어플리케이션(예를 들어, 웹 서비스를 수반하는 어플리케이션)의 컴퓨팅 리소스 사용량이 많고, 야간에는 제2 어플리케이션(예를 들어, 로그 분석기)의 컴퓨팅 리소스의 사용량이 많을 수 있다. 이 경우, 리소스 관리 서버(120)는 주간에는 웹 서비스를 수반하는 상기 어플리케이션에 할당되는 컴퓨팅 리소스를 증가시키고, 야간에는 로그 분석기에 할당되는 컴퓨팅 리소스를 증가시키도록 리소스 할당 정책을 변경할 수 있다. 이에 따라, 각 어플리케이션에 할당되는 컴퓨팅 리소스의 할당량이 특정 시간대에 따라 동적으로 변경될 수 있다.

[0034] 또한, 리소스 관리 서버(120)는 관리자의 입력에 따라 어플리케이션별 컴퓨팅 리소스의 할당량을 설정할 수도 있다. 구체적으로, 관리자는 어플리케이션별 컴퓨팅 리소스의 할당량을 직접 입력할 수 있고, 리소스 관리 서버(120)는 관리자의 입력 값에 기반하여 리소스 할당 정책을 수립할 수 있다. 이때, 관리자는 각 어플리케이션별, 컴퓨팅 리소스별 가중치를 부여할 수 있다.

[0035] 한편, 리소스 관리 서버(120)는 리소스 할당 정책의 수립 및 변경시 어플리케이션의 아이디 또는 상기 아이디가 속한 그룹별로, 어플리케이션의 프로세스명(예를 들어, 아파치 웹서버의 httpd)별로, 또는 어플리케이션의 프로세스 아이디(PID)별로 리소스 할당 정책을 수립할 수 있다. 예를 들어, 리소스 관리 서버(120)는 아이디 A 및 아이디 B에 대해서는 각각 CPU 전체 크기의 50% 및 30%를 할당하도록 리소스 할당 정책을 수립할 수 있다. 또한, 리소스 관리 서버(120)는 아이디 C 및 아이디 D가 그룹 a에 속하고 아이디 E 및 아이디 F가 그룹 b에 속할 때 그룹 a와 그룹 b에 다른 리소스 할당 정책을 설정할 수 있다.

[0036] 도 2는 본 발명의 일 실시예에 따른 클라이언트(110)의 상세 구성을 나타낸 블록도이다. 도 2에 도시된 바와 같

이, 본 발명의 일 실시예에 따른 클라이언트(110)는 하드웨어(202), 운영체제(204), 복수의 어플리케이션(206-1, 206-2, 206-3) 및 컴퓨팅 리소스 에이전트(208)를 구비할 수 있다.

- [0037] 하드웨어(202)는 중앙처리장치(CPU: central processing unit), 메모리, 디스크 입출력 장치 및 네트워크 장치 등이 될 수 있다. 여기서, CPU는 총 8개의 코어(0번 내지 7번)로 구성될 수 있다. 또한, 메모리의 크기는 8GB일 수 있다. 또한, 디스크 입출력 장치는 200Mbps의 입출력 속도를 가질 수 있으며, 네트워크 장치는 1Gbps의 속도로 데이터를 전송할 수 있다. 이러한, CPU, 메모리, 디스크 입출력 장치 및 네트워크 장치는 각각 컴퓨팅 리소스의 일 요소가 될 수 있다.
- [0038] 운영체제(204, OS: operating system)는 하드웨어(202)를 제어할 수 있다. 상술한 바와 같이, 복수의 어플리케이션(206)은 상기 운영체제(204) 상에 설치될 수 있다. 운영체제는 예를 들어 리눅스일 수 있으며, 씨그룹스(cgroups: control groups)을 구비할 수 있다. 이에 따라, 클라이언트(110)는 특정 어플리케이션(206)을 실행하는 경우 씨그룹스를 통해 상기 어플리케이션(206)에 할당된 컴퓨팅 리소스를 사용하도록 제어할 수 있다.
- [0039] 어플리케이션(206)은 각각 할당된 범위 내의 컴퓨팅 리소스를 사용하여 실행될 수 있다. 예를 들어, 제1 어플리케이션(206-1)에 할당된 컴퓨팅 리소스는 4개의 CPU 코어(0번 내지 3번), 4GB의 메모리, 100Mbps의 입출력 속도를 가지는 디스크 입출력 장치 및 400Mbps의 데이터 전송 속도를 가지는 네트워크 장치일 수 있다. 또한, 제2 어플리케이션(206-2)에 할당된 컴퓨팅 리소스는 2개의 CPU 코어(4번 내지 5번), 2GB의 메모리, 20Mbps의 입출력 속도를 가지는 디스크 입출력 장치 및 100Mbps의 데이터 전송 속도를 가지는 네트워크 장치일 수 있다. 또한, 제3 어플리케이션(206-3)에 할당된 컴퓨팅 리소스는 2개의 CPU 코어(6번 내지 7번), 2GB의 메모리, 80Mbps의 입출력 속도를 가지는 디스크 입출력 장치 및 500Mbps의 데이터 전송 속도를 가지는 네트워크 장치일 수 있다.
- [0040] 컴퓨팅 리소스 에이전트(208)는 어플리케이션(206)별 컴퓨팅 리소스의 사용량에 관한 정보를 수집하고 이를 리소스 관리 서버(120)로 전송할 수 있다. 상술한 바와 같이, 사용량에 관한 정보는 설정된 시간 동안 수집된 상기 각 어플리케이션별 컴퓨팅 리소스의 실시간 사용량, 최대 사용량, 최소 사용량 및 평균 사용량 중 하나 이상을 포함할 수 있다.
- [0041] 또한, 컴퓨팅 리소스 에이전트(208)는 리소스 관리 서버(120)로부터 리소스 할당 정책을 수신하고, 수신된 리소스 할당 정책에 따라 어플리케이션(206)별로 컴퓨팅 리소스를 할당할 수 있다. 컴퓨팅 리소스 에이전트(208)는 예를 들어, 씨그룹스를 통해 각 어플리케이션별로 컴퓨팅 리소스를 할당할 수 있다.
- [0042] 도 3은 본 발명의 일 실시예에 따른 리소스 관리 서버(120)가 할당 정책을 변경하는 과정을 설명하기 위한 예시 도이다.
- [0043] 도 3에 도시된 바와 같이, 리소스 관리 서버(120)는 리소스 할당 정책에 포함된 각 어플리케이션(206)별 컴퓨팅 리소스의 할당량을 변경할 수 있다. 예를 들어, 제2 어플리케이션(206-2)에 할당되는 CPU의 할당량을 증가시키고자 하는 경우, 리소스 할당 정책에 포함된 제2 어플리케이션(206-2)의 CPU 할당량을 증가시킬 수 있다.
- [0044] 이와 같이, 리소스 관리 서버(120)는 컴퓨팅 리소스를 구성하는 요소 각각(예를 들어, CPU, 메모리 등)에 대한 사용량을 수집하고, 컴퓨팅 리소스를 구성하는 요소 각각에 대한 할당량을 개별적으로 설정할 수 있다.
- [0045] 도 4는 본 발명의 일 실시예에 따른 컴퓨팅 리소스 관리 방법(400)을 나타내기 위한 흐름도이다. 도시된 흐름도에서는 상기 방법을 복수 개의 단계로 나누어 기재하였으나, 적어도 일부의 단계들은 순서를 바꾸어 수행되거나, 다른 단계와 결합되어 함께 수행되거나, 생략되거나, 세부 단계들로 나뉘어 수행되거나, 또는 도시되지 않은 하나 이상의 단계가 부가되어 수행될 수 있다.
- [0046] 하나 이상의 어플리케이션을 구비하는 클라이언트(110)는, 설정된 리소스 할당 정책에 따라 씨그룹스(cgroups: control groups)를 이용하여 컴퓨팅 리소스를 각 어플리케이션별로 할당한다(S402). 이때, 컴퓨팅 리소스는, 중앙처리장치(central processing unit), 메모리(memory), 하드 디스크 및 네트워크 장치 중 하나 이상을 포함할 수 있다.
- [0047] 클라이언트(110)는, 각 어플리케이션(206)별 컴퓨팅 리소스의 사용량을 수집한다(S404). 상기 사용량에 관한 정보는, 설정된 시간 동안 수집된 상기 각 어플리케이션별 컴퓨팅 리소스의 실시간 사용량, 최대 사용량, 최소 사용량 및 평균 사용량 중 하나 이상을 포함할 수 있다.

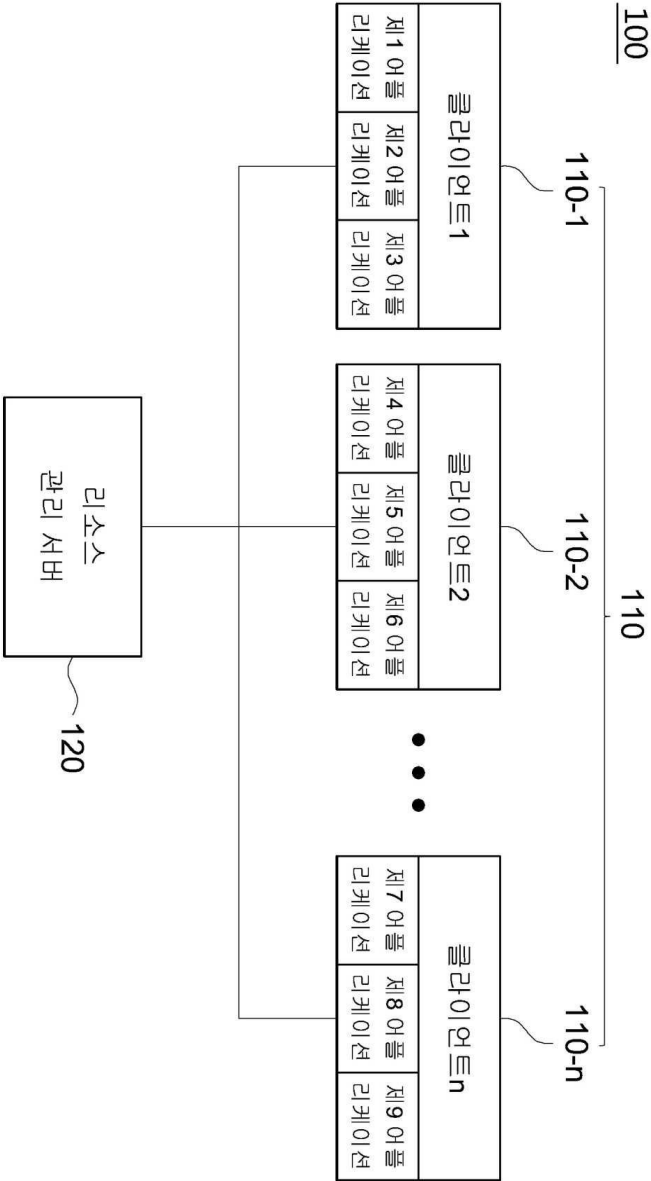
- [0048] 리소스 관리 서버(120)는, 상기 클라이언트로부터 상기 사용량에 관한 정보를 수신하고, 상기 사용량에 따라 상기 리소스 할당 정책을 변경할 수 있다(S406). 구체적으로, 상기 리소스 할당 정책에 따라 상기 어플리케이션에 할당된 컴퓨팅 리소스의 할당량과 수집된 상기 어플리케이션의 컴퓨팅 리소스의 사용량을 비교하고, 상기 비교 결과에 따라 상기 리소스 할당 정책을 변경할 수 있다. 구체적으로, 리소스 관리 서버(120)는, 상기 사용량이 상기 할당량 이상인 경우 상기 어플리케이션에 할당되는 컴퓨팅 리소스를 증가시키도록 상기 리소스 할당 정책을 변경하며, 상기 사용량이 상기 할당량 미만인 경우 상기 어플리케이션에 할당되는 컴퓨팅 리소스를 감소시키도록 상기 리소스 할당 정책을 변경할 수 있다. 또한, 리소스 관리 서버(120)는, 수집된 상기 어플리케이션의 컴퓨팅 리소스의 사용량이 상기 사용량을 수집하는 시간대에 따라 변화하는 경우 상기 어플리케이션에 할당되는 컴퓨팅 리소스를 상기 시간대에서 증가시키거나 감소시키도록 상기 리소스 할당 정책을 변경할 수 있다.
- [0049] 클라이언트(110)는, 상기 리소스 관리 서버(120)로부터 변경된 상기 리소스 할당 정책을 수신한다(S408).
- [0050] 클라이언트(110)는, 변경된 상기 리소스 할당 정책에 따라 상기 씨그룹스를 이용하여 상기 어플리케이션(206)별로 상기 컴퓨팅 리소스를 재할당한다(S410).
- [0051] 이후, S404 단계 내지 S410 단계가 반복적으로 수행될 수 있다.
- [0052] 이상에서 본 발명의 대표적인 실시예들을 상세하게 설명하였으나, 본 발명이 속하는 기술분야에서 통상의 지식을 가진 자는 상술한 실시예에 대하여 본 발명의 범주에서 벗어나지 않는 한도 내에서 다양한 변형이 가능함을 이해할 것이다. 그러므로 본 발명의 권리범위는 설명된 실시예에 국한되어 정해져서는 안 되며, 후술하는 특허 청구범위뿐만 아니라 이 특허청구범위와 균등한 것들에 의해 정해져야 한다.

부호의 설명

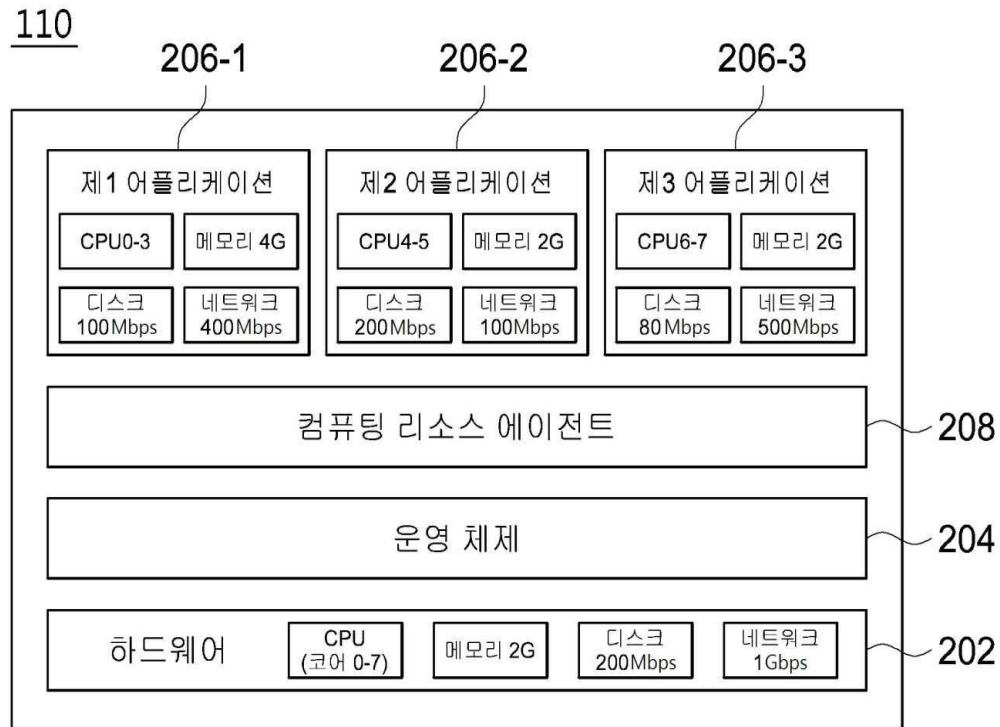
- [0053] 100: 컴퓨팅 리소스 관리 시스템
- 110: 클라이언트
- 120: 리소스 관리 서버
- 202: 하드웨어
- 204: 운영체제
- 206: 어플리케이션
- 208: 컴퓨팅 리소스 에이전트

도면

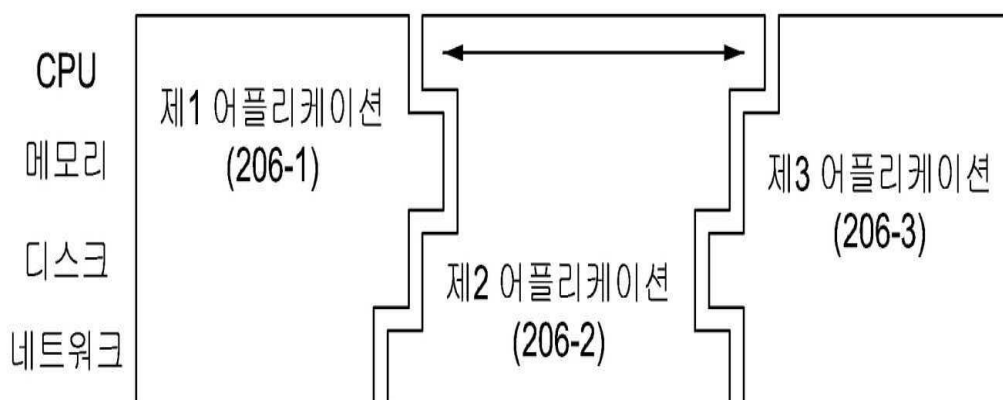
도면1



도면2



도면3



도면4

