

프로그램 특성을 고려한 이중 자원 탐색 기술

김유식, 윤찬현
한국과학기술원 전기및전자공학과
e-mail: {yusiky326, chyoun}@kaist.ac.kr,

A Study on Methods of Searching Multi-Cloud Resources in Cloud Collaboration Based on Program Similarity

Yusik Kim, Chan-Hyun Youn
Dept of Electrical Engineering, KAIST

요 약

사용자 지향형 클라우드 협업 시스템이 가지는 장점은 다양한 클라우드 서비스 제공자들로부터 컴퓨팅 자원을 제공받기 때문에, 넓은 폭에서 선택이 가능한 점이다. 이것이 가능하려면, 다양한 컴퓨팅 자원들의 기대 성능값을 사전에 알아서 사용자가 사용하는 애플리케이션의 특성을 고려하여 자원들을 평가하고 가장 높은 성능 기댓값을 가지는 자원을 찾을 수 있어야 한다.

1. 서론

클라우드 협업(Cloud Collaboration) 시스템이란 브로커를 통해 다수의 클라우드 서비스 제공자(CSP, Cloud Service Providers)들로부터 자원을 제공받아 사용자에게 전달해주는 시스템이다. 이 브로커를 활용한 시스템에는 한 가지 어려움이 있는데 클라우드 서비스 제공자들이 독립적으로 갖고 있는 가상 머신 배치(Virtual Machine Placement)에 사용하는 운영 정책이나 실시간 배치 상황 등은 그들이 직접 관련 정보들을 공개하지 않는 한 브로커 단에서 정확하게 알 수 없다는 점이다. 하나의 물리 머신(PM, Physical Machine)에 몇 개의 가상 머신이 할당되느냐에 따라, 가상 머신 간의 간섭 효과의 정도가 달라지기 때문에 성능이 저하되면서 성능에 대한 예측 또한 어려워지게 된다. 사용자에게 같은 비용 중 최고의 성능을 내는 자원을 탐색하여 전달하기 위해서는, 브로커 단에서 각 클라우드 서비스 제공자들이 제공하는 가상 머신이 어느 정도의 성능이 낼지 알 수 있어야 하며, 또한 실시간으로 변화하는 간섭(Interference) 효과의 정도에 따라 저하될 성능의 크기 또한 알 수 있어야 한다. 이에 본 논문에서는 예약형 가상 머신(Reserved Virtual Machine)과 벤치마킹 도구를 활용하여 사전에 평균적인 성능 및 간섭 효과를 측정하여 이를 바탕으로한 자원 탐색 기술을 제안하고자 한다.

2. 가상 머신간의 간섭을 고려한 탐색 기법

하나의 물리 머신에 여러 개의 가상 머신이 실행하고 있을 때 이상적인 경우라면, 하나의 가상 머신이 다른 가

상 머신들의 성능에 영향을 주어서는 안 된다. 하지만, 현존하는 가상화 기술로는 간섭 효과를 완전히 없애는 것이 불가능하다. Xen 하이퍼바이저를 사용할 경우, CPU는 비교적 간섭 효과가 적지만, 캐쉬(Cache)나 Disk I/O, Network I/O같은 경우에는 간섭에 의한 성능 저하가 발생한다. 그 뿐만 아니라, 같은 클라우드 서비스 제공자라고 하더라도, 시간에 지남에 따라, 새로운 CPU 아키텍처를 사용하는 물리 머신을 추가하기도 하기 때문에, 다양한 CPU 아키텍처가 공존하며 메모리 대역폭의 차이나 장기간 사용에 의한 디스크 성능 저하 상태 등 다양한 원인으로 인해 가상 머신의 성능은 일정하지 않다[1].

클라우드 협업 시스템의 장점은 사용자에게 다양한 클라우드 서비스 제공자들 중 하나를 선택할 수 있겠끔 선택권을 보장해준다는 점이다. 특정 클라우드 서비스 제공자가 부족한 물리 머신 또는, 통합(Consolidation) 정책으로 간섭으로 인한 성능 저하가 커진 상황이라면, 그렇지 않은 다른 클라우드 서비스 제공자를 선택할 수 있도록 하는 것이다. 이것이 가능하려면, 예약형 가상 머신을 미리 구입하여 사전에 벤치마킹 도구를 활용하여 성능 등을 측정된 데이터가 있어야 한다. 평균적인 성능과, 성능의 분산도에 따라, 해당 가상 머신을 사용할 경우 예상되는 성능 기대치를 예측하여, 사용자에게 가장 높은 성능 기대치를 가지는 가상 머신을 제공하도록 한다.

3. 프로그램 특성을 고려한 탐색 기법

7zip과 같이 압축하는 애플리케이션들은 CPU를 주요 자원으로 사용하는가 하면, Cat과 같은 애플리케이션들은 Disk I/O를 주로 사용한다. 이와 같이 애플리케이션마다

사용하는 자원의 정도가 다 다르기 때문에 이를 고려한 자원 탐색 기술이 필요하다. 예를 들면, 사용자가 실행하고자 하는 애플리케이션이 주로 CPU 연산에 의존하는 것이라면, 다른 자원들에 발생하고 있는 간섭으로 인한 성능 저하나 평균적으로 낮은 성능을 갖고 있더라도 사용자에게 중요하지 않기 때문이다. 그렇다면, 이를 알기 위해서는 사용자가 사용할 애플리케이션에 대해서 알고 있어야 한다. 사용자 지향형 브로커 시스템이기 때문에, 사전에 해당 애플리케이션을 실행해보아 CPU, 캐쉬, 디스크, 네트워크 등에 여러 자원에 대한 사용 정보를 확인하여 한다. CPU의 경우 사용률(Utilization), 캐쉬의 경우, 캐쉬 히트(Cache Hit)과 캐쉬 미스(Cache Miss) 등으로 확인할 수 있으며, 디스크의 경우 읽기(Read), 쓰기(Write) 횟수 등으로 알 수 있다. 뿐만 아니라, 각 자원별 간섭으로 미치는 성능 저하의 경우는, 백그라운드 애플리케이션으로 각 자원별로 부하를 주는 벤치마킹툴과의 영향도를 체크함으로써 간섭 정도를 파악할 수 있다.

사전에 사용자의 애플리케이션이 가지는 프로그램 특성을 파악하고 이를 계수(Coefficient)화 하여 자원 탐색시 성능 점수에 곱하여 합함으로써 자원별 성능 점수를 애플리케이션에 해당되게 계산하여 탐색을 하는 것이다. 이와 같은 방법이라면, 단순히 가상 자원의 사이즈에 따라 제공받는 가상 자원이 아니라, 다양한 클라우드 서비스 제공자들로부터 다양한 성능을 가지는 이종 자원들 중 평균적으로 최고의 성능을 내면서도 간섭을 적게 받고 있으며, 애플리케이션이 주로 사용하는 자원들 중심으로 상대적으로 높은 성능을 가지는 자원을 제공할 수 있게 된다.

4. 예약형 가상 머신을 활용한 성능 측정

아마존의 경우 예약형 가상 머신은 온 디맨드(On-demand) 가상 머신에 비해 상당히 가격이 저렴하다. 클라우드 협업 시스템의 브로커는 각 클라우드 서비스 제공자들로부터 저렴한 가격으로 장기간 대여가 가능한 가상 머신들을 구입해와 이를 약간의 이윤을 덧붙여 사용자에게 다시 빌려주는 방식으로 하는 비즈니스 모델이 가능하다. 이 때, 브로커는 각 가상 머신에 대한 성능 평가를 하여야 한다. 똑같은 벤치마킹툴을 사용하여, 각 자원 별 평균적인 성능을 상대적으로 비교할 수 있도록 정리하여야 한다. 뿐만 아니라, 평균적인 성능외에 성능의 분포도를 조사하여, 실시간으로 변화하는 간섭 효과에 대해서도 조사를 하여야 한다. 예를 들면, 평균적으로 1의 성능을 가지던 가상 머신이, 어느 순간 1.2의 성능을 내기 시작한다면 간섭 효과가 있는 것으로 파악하여야 한다는 것이다. 현존하는 가상화 기술로는 간섭으로 인한 성능 저하가 불가피하기 때문에 감수하여야 하며, 이 두 가지 모두 브로커 단에서 알 수 있는 방법은 직접 브로커에서 테스트하는 수뿐이 없다. 클라우드 서비스 제공자들이 본인들에게 불리한 정보를 알려줄 리가 없기 때문이다.

5. 결론

본 논문에서는 프로그램의 특성을 사전에 조사하여 사용하는 주요 자원들에 대해 조사하고 이를 바탕으로 해당 자원들에 대해 간섭이 적으며 평균적인 성능 기댓값이 높은 자원을 탐색하는 방법을 제안하였다.

이를 위해서는, 브로커 단에서 사전에 가상 머신에 대한 조사가 이루어져야 하는데, 저렴하지만 장기간 대여를 해야만하는 예약형 가상 머신을 활용함으로써, 이를 여러 사용자에게 여러 번 판매하는 방식으로 하고자 한다.

참고문헌

[1] B. Farley, V. Varadarajan, K. Bowers, A. Juels, T. Ristenpart, and M. Swift, More for your money: Exploiting performance heterogeneity in public clouds. In SOCC, 2012