

협업 클라우드 환경에서의 요청 패턴 기반 자원 공급 기법

강동기*, 김병상*, 윤찬현*

*한국과학기술원 전기 및 전자 공학과

e-mail:dkkang@kaist.ac.kr, bs.kim@kaist.ac.kr, chyou@kaist.ac.kr

Request Pattern based Resource Provisioning Method in Cloud Collaboration Environment

Dong-Ki Kang*, Byung-Sang Kim*, Chan-Hyun Youn*

*Dept of Electrical Engineering, KAIST

요 약

고용량 멀티미디어 및 대용량 콘텐츠와 같은 트래픽이 급속도로 증가하고 사용자들의 다양한 요구 사항이 발생함에 따라 기존의 단일 클라우드 서비스 환경의 자원 부족 문제가 대두되고 있다. 이를 극복하기 위해 협업 클라우드 환경이 제안되었는데, 이는 이종의 클라우드 서비스 환경을 통합하여 부족한 자원 문제를 해결하고, 클라우드 서비스 제공자들의 이익을 최대화시키는 것에 목적을 두고 있다. 본 논문에서는 협업 클라우드 환경에서 각 데이터 센터내의 자원을 최적으로 분배하고 입력되는 사용자 요청 패턴을 반영할 수 있는 수익 함수 모델을 제안하고 이에 따라 각 서비스 제공자들의 이익이 증가함을 보인다.

1. 서론

컴퓨팅 기술의 발전속도가 지속적으로 증가하고, 다양한 어플리케이션들이 개발되며, 이를 이용하는 사용자들의 숫자와 관련 요구사항이 크게 증가함에 따라, 기업은 서비스를 제공하기 위한 자원 (서버, 네트워크 장치, 소프트웨어 등) 을 준비하고 이를 지속적으로 관리하며 성능을 수시로 개선시켜야 했다. 이를 위하여 발생하는 운용 비용이나 장소 등 문제의 심각성이 증가함에 따라 이를 해결하기 위한 효과적인 방법의 필요성이 대두되었다.

클라우드 컴퓨팅 서비스는 기업 또는 일반 사용자가 자신이 직접 하드웨어 및 소프트웨어 자원 환경을 구축할 필요 없이 클라우드 컴퓨팅 서비스 제공자에게 요청을 하면 요청한 만큼 서비스를 제공받고 그에 대한 요금을 지불하는 방식이다. 클라우드 기술을 통하여 기업 또는 일반 사용자는 자원을 구입하거나 확장 및 운용하는 것에 대해 신경 쓸 필요가 없으며 이를 위해 소비하는 비용도 크게 줄일 수 있다.

그러나 클라우드 컴퓨팅 자원을 요청하는 사용자가 지속적으로 증가하고 있고 대용량의 콘텐츠 및 복잡한 어플리케이션의 증가로 인하여, 기존의 단일 클라우드 서비스 제공자들이 가진 클라우드 자원으로는 요구사항을 전부 수용할 수 없게 되었다.

클라우드 자원의 부족은 사용자들의 요청량을 제대로 처리하지 못하게 하므로, QoS 가 만족되지 못하여 신뢰성이 떨어지고 서비스 제공자의 이익을 감소시킨다. 반대로

사용자들의 요청량이 적어서 유휴 (Idle) 자원이 많이 발생하는 클라우드 서비스 제공자의 경우에는 자원의 운용 비용은 지속적으로 발생하는데 반해 이익의 발생 크기가 작으므로 이를 해결하기 위한 연구가 활발히 진행되고 있다.

최근의 연구에서 가장 많이 고려되는 방법 중 하나로 협업 클라우드 환경을 고려할 수 있는데, 이는 이종의 클라우드 서비스 환경이 하나의 가상적인 클라우드 서비스 환경으로 통합되어서 자원 부족이 발생할 경우 다른 클라우드 서비스 제공자로부터 남은 자원을 할당받으며, 또한 자신도 유휴 자원을 다른 서비스 제공자에게 제공함으로써 서비스 사용자의 요구사항을 만족시키며 제공자의 이익을 최대화시킬 수 있도록 하는 것이다.

본 논문에서는 협업 클라우드 환경에서 각 데이터 센터 내의 유휴 자원을 효율적으로 아웃소싱 (Outsourcing) 및 인소싱 (Insourcing) 할 수 있도록 하기 위한 수익 함수 (Profit Function) 를 제안하며 이동 평균을 적용한 간단한 요구량 예측 모델 기법을 사용하여 비용에 최적인 자원 공급 기법을 제안한다. 제안한 기법을 통해 클라우드 서비스 제공자는 현재 사용자의 요청 패턴에 따른 최적의 결정을 내릴 수 있으며 이에 따라 이익이 증가될 수 있음을 보인다.

2. 협업 클라우드에서의 수익 함수 모델

그림 1 에서는 이종의 클라우드 자원으로 구성된 협업

클라우드 아키텍처를 보인다.

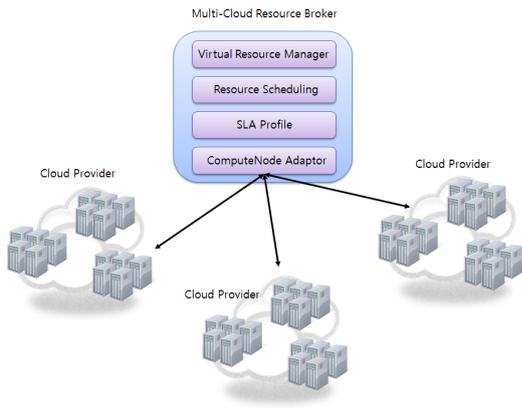


그림 1. 협업 클라우드 환경 아키텍처

그림 1 에서 다중 클라우드 자원 브로커 (Multi-Cloud Resource Broker : MCRB) 는 분산되어 있는 클라우드 서비스 제공자들을 한데 묶어서 가상 클라우드 (Virtual Cloud) 환경을 구성한다. 각 클라우드 서비스 제공자들은 ComputeNode Adaptor 모듈을 통해 MCRB 와의 연동을 수행하며 Virtual Resource Manager 모듈은 각 클라우드 서비스 제공자의 요청사항을 확인하고 정의된 수익 함수에 기반하여 최적의 자원 공급 기능을 수행한다.

각 클라우드 서비스 제공자가 가지는 이익의 크기를 계산하기 위한 수익 함수는 다음과 같다.

$$r_t = r_a + r_i + r_o + r_e + r_l \quad (1)$$

$$s_t = s_a + s_o + s_e, \quad (s_a \geq r_a + r_i, s_o \geq r_o, s_e \geq r_e) \quad (2)$$

$$s_a \leq s_c \quad (3)$$

$$revenue(\Delta t) = profit(\Delta t) - cost(\Delta t) \quad (4)$$

$$profit(\Delta t) = (r_a(\Delta t) + r_o(\Delta t) + r_e(\Delta t))p_r + r_i(\Delta t)p_i \quad (5)$$

$$cost(\Delta t) = c_i(s_a(\Delta t) - (r_a(\Delta t) + r_i(\Delta t))) + c_e(s_e(\Delta t) - r_e(\Delta t)) + c_b(r_a(\Delta t) + r_i(\Delta t) + r_e(\Delta t)) + c_o r_o(\Delta t) \quad (6)$$

r 은 사용자의 요청량을 의미하며 r_a, r_i, r_o, r_l, r_e 는 각각 로컬 자원에서 처리하는 요청량, 인소싱으로 처리하는 요청량, 아웃소싱으로 처리하는 요청량, 손실이 발생한 요청량, 최소 유지 자원에서 처리하는 요청량을 의미한다.

s 는 현재 클라우드 서비스 제공자가 가지고 있는 자원량을 의미하며, s_a, s_o, s_e 는 각각 로컬에서의 자원 가용량, 아웃소싱에서의 자원량 및 최소 유지 자원량을 의미한다. 최소 유지 자원량 s_e 는 클라우드 서비스 제공자가

항상 여유분으로 가지고 있어야 하는 최소 자원량의 크기를 나타낸다. 만약 이 값이 감소하면 사용자의 요청 손실의 확률이 증가하며 이 값이 증가하면 불필요한 자원 유지 비용이 증가한다. s_c 는 로컬에서의 최대 자원량을 의미한다. c_i, c_b, c_o 는 각각 공급된 자원이 유휴 상태인 경우의 유지비용과 유휴상태가 아닌 경우의 유지비용 및 자원을 아웃소싱할 때 지불하는 사용료를 나타낸다. p_r 은 사용자가 클라우드 서비스 제공자에게 지불하는 사용료를 나타내며 p_i 는 다른 클라우드 서비스 제공자에게 자원을 인소싱할때의 이익을 나타낸다. 식 (5) 와 (6) 은 각각 시간 Δt 동안 클라우드 서비스 제공자가 얻을 수 있는 순수 수익을 계산하기 위한 이익 및 비용 함수를 나타낸다.

각 값은 $p_r > c_o > c_b > c_i, p_i > c_b$ 이며 서비스 제공자가 이익을 얻기 위해서는 $profit(\Delta t) > cost(\Delta t)$ 이어야 한다. 만약 $r_t > s_t$ 인 경우에는 요청 손실이 발생하는데 그 크기는 $r_t(\Delta t) - s_t(\Delta t) = r_l(\Delta t)$ 이며 발생하는 손실 비용은 $r_l(\Delta t)p_r$ 이다. 요청 손실이 발생할 경우 사용자의 요청 사항을 처리하지 못함으로써 발생하는 이익 감소 문제 외에도 평판치가 감소되어 해당 클라우드 서비스를 이용하는 사용자의 수가 줄어들 확률이 높아진다.

r 값은 사용자의 요청에 의해 결정되며 c 와 p 역시 고정된 값으로 가정할 수 있다. 그러므로 클라우드 서비스 제공자가 조정할 수 있는 값은 s 값이며, 이익을 최대화하고 손실률을 최소화하기 위해서는 사용자의 요청을 고려한 최적의 s 값을 찾아야 한다.

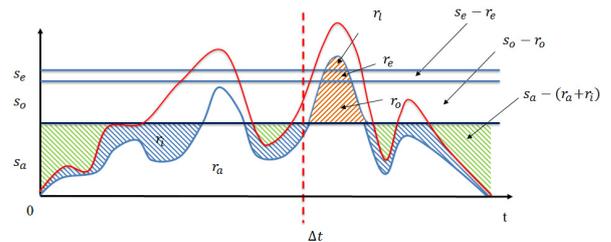


그림 2. 시간에 따른 자원 공급 그래프

그림 2 에서는 s 와 r 의 관계 그래프가 시간을 가로축으로 하여 나타나고 있다. 그림 2 를 기반으로 고려할 때 클라우드 서비스 제공자의 이익이 최대화 될 때는 입력되는 $r_a + r_i$ 에 대해서 녹색으로 빗금친 $s_a - (r_a + r_i)$, $s_o - r_o$, r_l 지역의 크기가 0 이 되면서 s_o 가 최소의 크기를 가지는 값이 되는 것이다. 이를 위해서 s_a 와 s_o 의 값은 들어오는 사용자의 요청 패턴을 고려하여 동적으로 조절될 수 있다.

사용자의 요청 패턴을 고려하기 위하여 상수 모델, 경향 모델 및 경향-분기 모델을 사용할 수 있는데 먼저 상

수 모델은 다음과 같다.

$$x_t = a + \epsilon_t \quad (7)$$

이 때, x_t 는 현재 주기 t 동안의 사용자의 요청량을 나타내며, a 는 평균적인 고정량이고 ϵ_t 0 을 평균으로 가지는 독립적인 임의의 편차를 의미한다.

경향 모델은 규칙적으로 증가 또는 감소하는 경향을 나타내는데 이는 선형적인 경향 패턴으로 표현할 수 있다. b 를 주기동안 일정하게 증가 혹은 감소하는 경향 계수라고 하면 경향 모델은 다음과 같이 표현된다.

$$x_t = a + bt + \epsilon_t \quad (8)$$

경향-분기 모델은 연속 기간 T 가 존재할 때 주기 t 마다의 증가 혹은 감소 경향을 나타내는 모델로서 각 주기의 경향 가중치 값을 F_t 라고 정의하면 $\sum_{k=1}^T F_{t+k} = T$ 이며, 모델 식은

$$x_t = (a + bt)F_t + \epsilon_t \quad (9)$$

로 표현될 수 있다.

사용자의 요청량은 분기마다 증가 혹은 감소하는 추세를 가진다고 고려할 수 있으므로 요청량 예측을 위해 식 (9) 를 선택할 수 있으며, 예측값의 계산을 위해 단순 이동 평균 (Simple Moving Average : SMA) 을 고려하면 t 주기에서의 이동 평균 값 x'_t 는

$$x'_t = (x_t + x_{t-1} + \dots + x_{t-N+1})/N \quad (10)$$

와 같이 계산될 수 있고 이 때, a 와 b 는

$$a = \frac{1}{k} \sum_{t=1}^k x'_{k-t+1} \quad (11)$$

$$b = \left[\frac{2}{k} \sum_{t=\frac{k}{2}}^k x'_{k-t+1} - \frac{2}{k} \sum_{t=1}^{\frac{k}{2}} x'_{\frac{k}{2}-t+1} \right] \quad (12)$$

와 같이 간단한 방법으로 예측할 수 있다.

식 (9) - (12) 를 사용하여, 주기 $t+1$ 에서의 요청량 x_{t+1} 를 계산하면 이는 $x_{t+1} = r_{total}$ 이며, 주어진 r_t 값을 통하여 다음과 같이 s_a 와 s_o 를 결정할 수 있다.

$$(s_a, s_o) = \begin{cases} (s_c - r_t, 0), & \text{if } s_c \geq r_t \\ (s_c, r_t - s_c), & \text{otherwise} \end{cases} \quad (13)$$

즉 사용자의 요청량이 현재 클라우드 서비스 제공자의 로컬 자원량 보다 적은 경우에는 아웃소싱을 수행하지 않으며, 요청량이 로컬 자원량 보다 많은 경우에는 그것의 차이 만큼 아웃소싱을 수행한다.

그러므로 본 논문에서 제안한 기법을 사용하면 사용자의 요청사항이 많지 않을 경우 불필요한 아웃소싱의 발생을 막음으로써 발생 비용을 줄이고 사용자의 요청 사항이 큰 경우에는 아웃소싱을 통해 이익을 증가시킴으로써 클라우드 서비스 제공자의 이익을 최대화시킬 수 있다.

3. 결 론

본 논문에서는 협업 클라우드 환경에서의 클라우드 서비스 제공자의 이익을 최대화시키기 위한 수익함수모델을 제안하고 사용자의 요청 패턴을 고려하여 적시에 적절한 자원 공급을 수행하기 위해 경향-분기 모델 및 단순 이동 평균 기법을 사용하여 서비스 제공자의 로컬 자원 공급 및 아웃소싱 공급량을 결정하는 방법을 보였다. 향후 연구에서는 본 연구에 대한 검증을 수행하고, 사용자 요청사항의 버스트 특성 (Burst Characteristic) 을 반영할 수 있는 실질적인 모델을 수립하고 이에 대한 예측 기법에 초점을 맞추고자 한다.

Acknowledgment

이 논문은 2012 년도 정부(교육과학기술부)의 재원으로 한국연구재단-클라우드 Collaboration 기술 사업과 BK21 사업의 지원을 받아 수행된 연구임(No. 2012-0020522)

참고문헌

- [1] A.N. Toosi, R.N. Calheiros, R.K. Thulasiram and R. Buyya "Resource Provisioning Policies to Increase IaaS Provider's Profit in a Federated Cloud Environment" in Proceedings of 13th High Performance Computing and Communications (HPCC), pp.279-287, 2011
- [2] I. Goiri, J. Guitart and J. Torres "Characterizing Cloud Federation for Enhancing Providers' Profit" in Proceedings of 3rd International Conference on Cloud Computing, pp.123-130, 2010
- [3] Y. LAILI, F. TAO, L. ZHANG, L. REN "The Optimal Allocation Model of Computing Resources in Cloud Manufacturing System" in Proceedings of 7th International Conference on Natural Computation, pp.2322-2326, 2011