

# 클라우드 컴퓨팅을 위한 예약 기반 가상 자원 관리 기법

강동기\*, 김성환\*, 윤찬현\*

\*한국과학기술원 전기 및 전자 공학과

e-mail:dkkang@kaist.ac.kr, s.h\_kim@kaist.ac.kr, chyoun@kaist.ac.kr

## Reserved Virtual Resource Management for Cloud Computing

Dong-Ki Kang\*, Seong-Hwan Kim\*, Chan-Hyun Youn\*

\*Dept of Electrical Engineering, KAIST

### 요 약

클라우드 컴퓨팅 환경에서, 모든 IT 자원은 가상화된 자원(Virtualized Resource)의 형태로 클라우드 서비스 사용자에게 제공되고 사용자는 자신이 점유한 시간만큼만 자원 사용 비용을 클라우드 서비스 제공자에게 지불하면 된다. 서비스 제공자는 서비스 사용자가 처리하고자 하는 어플리케이션의 특성을 고려하여 온디맨드(On-demand) 및 예약형(Reserved) 가상 자원 가격 정책을 제공한다. 따라서 서비스 사용자는 서비스 제공자의 제안 가격 정책을 기반으로 어플리케이션의 처리 예상 시간 및 지불 비용을 고려하여 효율적인 자원 요청 계획을 수립해야 한다. 본 논문에서는 클라우드 브로커링 시스템(Cloud Brokering System)을 통해 자원 할당 정책을 수행하여 연동된 다중 사용자에게 비용 효율적인 자원을 제공하는 구조 및 프로세스를 소개한다.

### 1. 서론

기존의 컴퓨팅 시스템과 달리 클라우드 컴퓨팅 환경에서는 클라우드 서비스 제공자(Cloud Service Provider)가 보유한 가상 자원 인스턴스(Virtual Resource Instance)를 통해 클라우드 서비스 사용자(Cloud Service Customer)는 자신이 원하는 명세를 가진 컴퓨팅 자원을 할당받고 자원을 점유한 시간만큼만 제공자에게 비용을 지불할 수 있다. 지불 비용에 관한 정책은 각 클라우드 서비스 제공자마다 조금씩 다르지만, 아마존 EC2 및 GoGrid 와 같은 대형 벤더(Vendor) 들은 일반적으로 온디맨드, 예약형 및 스팟 인스턴스의 3 가지 형태로 가격 정책을 달리하여 사용자에게 제공한다[1,2]. 온디맨드는 사용자가 자원을 요청하면 즉각적으로 할당될 수 있으며 가격이 상대적으로 비싸다. 이는 일반적으로 무계획적이고 간헐적인 응용을 처리하는데 적합하다. 예약형 인스턴스는 응용 처리 시간이 상대적으로 긴 경우(일반적으로 한달에서 1년 이상)에 적합하며 시간당 자원 사용 비용이 온디맨드에 비하여 저렴하다. 스팟 인스턴스는 세 가지 종류의 인스턴스 중 평균적으로 가격이 가장 저렴하지만 주기에 따라서 가격이 동적으로 바뀌며 입찰 가격(bid price)을 사용자가 협의하지 못한다면 할당된 자원이 자동적으로 해제되기 때문에 신뢰성이 떨어진다. 그러므로 사용자가 가장 최적의 비용으로 응용을 처리하기 위해서는 이와 같은 세 종류의 가격 정책 및 응용을 특성을 고려하여 효율적인 자원 할당 요청 정책을 수립해야 한다.

본 논문에서는 클라우드 브로커링 시스템(Cloud Brokering System)의 아키텍처를 제안하고 이를 통하여 직접 온디맨드 및 예약형 가상 자원 할당 정책을 수행함으로써 사용자가 별다른 최적화 기법을 수행하지 않고서도 비용 최적의 서비스를 제공받을 수 있도록 한다. 제안하는 구조 및 프로세스를 통하여 클라우드 서비스 사용자는 처리 응용의 서비스 수준 협약(Service Level Agreement)을 만족하면서도 비용은 효과적으로 감소시킬 수 있다.

### 2. 클라우드 브로커링 시스템

본 논문에서는 스팟 인스턴스(Spot Instance)를 제외한 온디맨드 및 예약형 인스턴스만을 고려한 자원 할당 기법이 적용된 클라우드 브로커링 시스템을 제안한다. 온디맨드 인스턴스의 단위 시간당 사용 비용이  $C_{on}$  이고, 예약형 인스턴스의 단위 시간당 사용 비용이  $C_{resv}$  일 때, 자원 총 사용 비용은 다음과 같다.

$$C_{total} = \sum_{i=0}^k \left( \sum_{j=0}^{n_i} (t_{on}^{ij} \times C_{on}^i) + \sum_{j=0}^{n_{resv}^i} (t_{resv}^{ij} \times C_{resv}^i) \right) \quad (1)$$

이 때  $i$  는 할당받은 인스턴스의 명세(Specification) 을 의미하며  $t^{ij}$  는  $i$  타입을 가지는  $j$  번째 인스턴스의 할당 시간을 나타내고,  $n$  은 할당받은 해당 인스턴스의 개수를

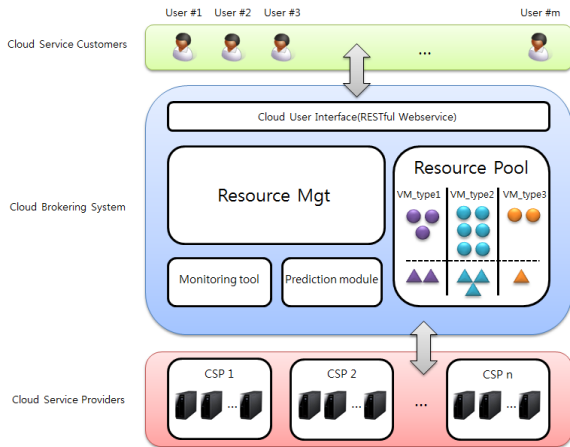


그림 1. 클라우드 브로커링 시스템

나타낸다. 직관적으로  $C_{on} > C_{resv}$  이지만  $t_{resv}^{ij} > t_{on}^{ij}$  이므로 시간과 비용을 같이 고려하여 적절한 온디맨드 및 예약형 인스턴스의 개수를 결정하는 것이 중요하다.

그림 1에서는 자원 풀을 관리하는 클라우드 브로커링 시스템의 구조를 간단히 나타내고 있다. 클라우드 브로커링 시스템은 연동된 다수의 클라우드 서비스 제공자의 서비스 품질을 주기적으로 모니터링 하여 CSP 프로파일링을 수행하며 이를 고려하여 자원 풀에 예약형 가상 자원 인스턴스에 대한 프로비저닝을 수행한다. 즉 단위 시간당 자원 사용 비용이 저렴한 예약형 인스턴스를 미리 확보하여 온디맨드 자원을 요청하는 사용자에게 동일한 성능 명세를 가지는 예약형 인스턴스를 대신 제공하게 되면 사용자는 절감된 비용으로 동일한 서비스를 제공받을 수 있다. 또한 다중 사용자가 브로커에 연동되므로 상대적으로 할당 시간이 긴 예약형 인스턴스의 낭비 문제도 해결할 수 있다. 즉 예를 들어, 클라우드 서비스 사용자가 10 명이고 각각의 응용 처리 시간이 전부 동일하게 12 시간이며, 오버랩 되는 시간이 1 시간이라고 가정하자. 또한 각 응용에 대해 요구되는 인스턴스의 성능 명세는 전부 동일하게  $I$  라고 하자. 그림 2 에서 이와 같은 가정에서의 시간 흐름도를 나타낸다.

이 경우 기존의 클라우드 브로커링 시스템이 없을 때 각 클라우드 서비스 사용자는 응용을 처리하기 위하여 개별적으로 온디맨드 가상 자원 인스턴스 (예약형 가상 자원 인스턴스는 최소 할당 시간이 길기 때문에 각 서비스 사용자가 개별적으로 할당받으면 자원 사용 비용이 오히려 크게 증가한다) 를 하나씩 할당 받아야 하므로 필요한 총 비용은  $tnC_{on}^d = 12 \times 100 \times C_{on}^d = 1200C_{on}^d$  이 된다. 그러나 클라우드 브로커링 시스템이 예약형 인스턴스를 미리 확보하여 클라우드 서비스 사용자에게 제공하는 경우 예약형 인스턴스의 기본 최소 할당 시간을 1달 (=30일) 로 가정할 때[2], 오버랩되는 시간을 고려하면 2 개의 예약형 가상 자원 인스턴스가 요구되므로 필요한 총 비용은  $tnC_{resv}^d = 24 \times 30 \times 2 \times C_{resv}^d = 1440C_{resv}^d$  가 된다. 일반적으로

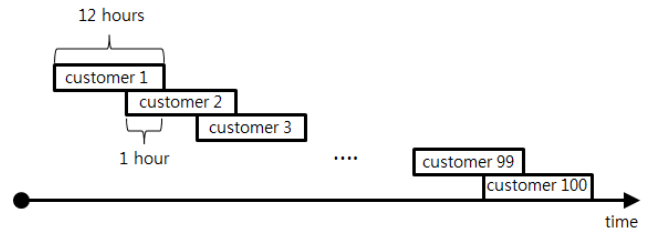


그림 2. 다중 서비스 사용자의 응용 처리 스케줄 시나리오

자원 할당 시간이 한 달일 때,  $C_{resv}^d$  는  $C_{on}^d$  의 약 60% 비용을 가지므로,  $C_{resv}^d = 0.6C_{on}^d$  이 되며, 클라우드 브로커링 시스템에서 요구되는 자원 사용 비용은 최종적으로  $864C_{on}^d$  가 되므로 동일한 서비스로 응용을 처리하면서도 동시에 기존 브로커링 시스템이 없는 경우보다 비용이 절감될 수 있음을 확인할 수 있다.

### 3. 결론

본 논문에서는 예약형 가상 자원 인스턴스를 기반으로 자원 관리 기법을 수행하는 클라우드 브로커링 시스템의 구조를 보이고 본 구조에서의 자원 사용 비용 절감 가능성을 확인하였다. 즉 다중 클라우드 서비스 사용자와 연동하는 클라우드 브로커링 시스템에서는 기존의 자원 관리 연구와 같이 단일 클라우드 서비스 사용자 관점에서의 최적화 기법의 한계점을 해결하고 예약형 가상 자원을 통한 서비스 비용 절감을 달성할 수 있다. 추후 연구에서는 본 연구에서 제시된 구조를 토대로 보다 일반적인 서비스 요청 시나리오를 다양하게 고려하여 이론 모델을 세우고, 이를 기반으로 보다 정교한 자원 스케줄링 기법을 제안하고자 한다.

### Acknowledgment

이 논문은 2012 년도 정부(교육과학기술부)의 재원으로 한국연구재단-클라우드 Collaboration 기술 사업과 BK21 사업의 지원을 받아 수행된 연구임(No. 2012-0006425)

### 참고문헌

- [1] Amazon EC2 (2013), <http://aws.amazon.com/ec2/>
- [2] GoGrid (2013), <http://www.gogrid.com/>
- [3] D. Kang, S. Kim, Y. Ren, B. Kim, W. Kim, Y. Kim, C. Youn, and C. Jeong, "Enhancing a Strategy of Virtualized Resource Assignment in Adaptive Resource Cloud Framework," Proc. ACM Int'l Conf on Ubiquitous Information Management and Communication, (ICUIMC), 2013.
- [4] S. Chaisiri, B. Lee, and D. Niyato, "Optimal Virtual Machine Placement across Multiple Cloud Providers," Proc. IEEE Asia-Pacific Services Computing Conf. (APSCC), 2009