

하드웨어 성능 및 비용을 고려한 최적의 클라우드 인스턴스 선택 방안

김경환, 남채원, 이경용
국민대학교 컴퓨터공학과

{bryan9801, rabbit7653, leeky}@kookmin.ac.kr

Towards Optimal Cloud Instance Selection Considering Hardware Performance and Cost

Kyunghwan Kim, Chaewon Nam, Kyungyong Lee
Department of Computer Science, Kookmin University

요 약

여러 클라우드 제공 업체들은 컴퓨팅 리소스를 최적화하여 다양한 인스턴스 타입으로 사용자에게 제공한다. 그러나 클라우드 제공 업체는 인스턴스에 대한 정량적인 성능 지표를 제공하고 있지 않아 사용자는 수많은 인스턴스 중 자신의 응용 프로그램에 맞는 최적의 인스턴스를 선택에 많은 어려움을 겪는다. 본 논문에서는 클라우드 제공 업체별 인스턴스를 정량적 지표로 비교하기 위해 널리 사용되는 범용 인스턴스들의 성능을 CoreMark 벤치마크를 이용하여 CPU 성능을 측정하고 SpotLake 데이터 세트를 이용하여 인스턴스 가격을 분석해 가격과 성능을 고려한 최적의 인스턴스를 선택하는 방법에 대해 논의한다.

1. 서론

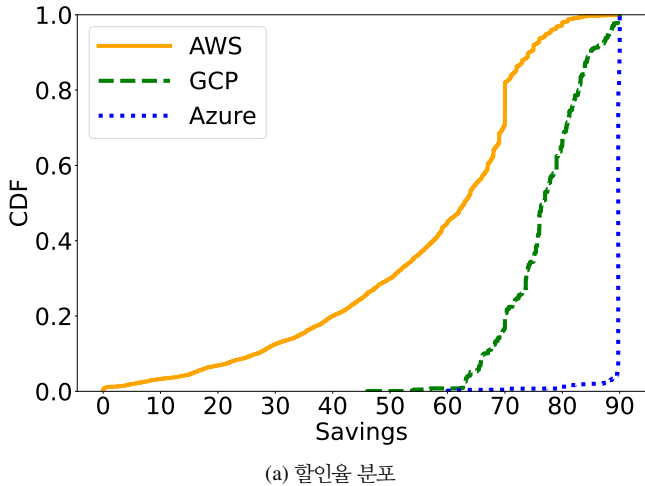
Amazon Web Services(AWS), Google Cloud Platform(GCP) 및 Azure와 같은 클라우드 제공 업체들은 다양한 사용자의 응용 프로그램에 맞춰 최적화된 컴퓨팅 리소스를 제공한다. 또한 다양한 과금 모델을 통해 컴퓨팅 리소스를 유연하게 사용할 수 있도록 한다. 대표적으로 온디맨드 인스턴스와 스팟 인스턴스가 있다. 온디맨드 인스턴스는 리소스 안정성이 보장되고 사용한 만큼 요금을 지불하며 가격은 대체로 안정적이고 매우 드물게 변화한다. 스팟 인스턴스는 언제든지 중단될 수 있지만 온디맨드 인스턴스에 비해 최대 90% 저렴한 비용으로 여분의 컴퓨팅 리소스를 제공한다. 유연한 컴퓨팅 리소스 사용은 클라우드 시장을 발전시켰고 이에 따라 인스턴스 종류가 다양해지고 복잡해졌다. 이에 클라우드 제공 업체는 수많은 인스턴스 중 적절한 인스턴스를 선택할 수 있도록 범용, 컴퓨팅 최적화, 메모리 최적화, 가속화 컴퓨팅 4가지로 인스턴스를 분류하고 출시 시기별로 세대를 나누어 사용자에게 최적의 인스턴스를 선택할 수 있도록 하지만 인스턴스별 상세한 성능에 대한 지표는 제공하고 있지 않다. 성능 지표 부재, 그리고 복잡한 인스턴스 유형과 동적으로 변화하는 가격은 멀티 클라우드 환경에서 사용자가 응용 프로그램에 알맞은 최적의 인스턴스 선택하는 데 많은 어려움을 느끼게 한다.

본 논문에서는 이러한 문제를 해결하기 위해 대표적인 CPU 벤치마크 도구인 CoreMark [1] 이용하여 가장 빈번하게 사용되는 범용 인스턴스들을 각 클라우드 제공 업체별로 선정하여 CPU 성능을 측정한다. 또한 인스턴스 가격, 안정성 정보를 제공하는 SpotLake [2]의 데이터 세트로 클라우드 제공 업체별 가격 데이터를 분석하고 측정된 인스턴스들의 CoreMark 점수를 이용해 통일된 지표로 인스턴스들의 성능을 평가한다. 그 결과 클라우드 제공업체 간, 인스턴스 세대 간의 가격 및 성능에 유의미한 차이가 있었고 이를 고려한 최적의 인스턴스 선택 방법을 논의한다.

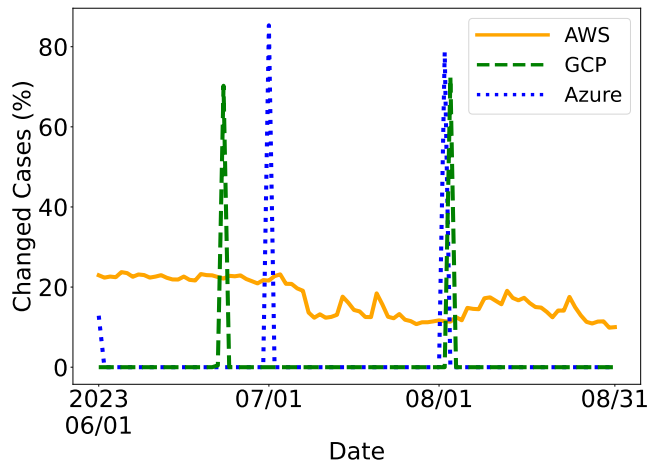
2. SpotLake 데이터 세트

AWS, Azure, GCP와 같은 클라우드 공급 업체에서는 사용자에게 비용 절감 효과와 인스턴스 중단 가능성을 예측할 수 있도록 하는 스팟 인스턴스에 대한 다양한 데이터를 제공한다. 우선 지난 28일간 스팟 인스턴스의 중단 비율을 나타내는 과거 중단 비율 데이터 세트(Frequency of Interruption)를 제공한다. 5% 미만, 5%에서 10%, 10%에서 15%, 15%에서 20%, 그리고 20% 초과까지 총 5가지 범위로 분류하며, AWS와 Azure는 제공하지만 GCP에서는 제공하지 않는다. 그러나 과거 중단 비율 데이터 세트는 현재의 스팟 인스턴스 가용성을 실시간으로 반영하지 못한다는 점을 유의해야 한다. 다음으로는 사용자가 인스턴스 시작 직전 스팟 인스턴스 요청의 성공 가능성을 계산할 수 있도록 AWS에서 제공하는 스팟 배치 점수(Spot Placement Score)가 있다. 점수는 1,2,3 수치로 나타내며 점수가 높을수록 스팟 요청이 성공할 확률이 높다는 것을 의미한다. 이처럼 다양한 스팟 인스턴스 데이터 세트들을 통해 사용자는 효율적이고 저렴한 가격으로 클라우드 환경을 구축할 수 있다. 하지만 이러한 스팟 데이터 세트들은 각 클라우드 제공 업체에 분산되어 있고, 웹 또는 API 쿼리 등 접근 방법이 복잡하며 일부 최신 데이터만 제공하도록 쿼리가 제한되기도 한다. 게다가 스팟 가격과 안정성은 수요와 공급에 따라 항상 변화하기 때문에 사용자가 각각의 데이터들을 비교하여 본인에게 맞는 인스턴스를 선택하기란 상당히 어려운 일이다.

이러한 문제를 해결하기 위해 SpotLake라는 통합 스팟 데이터 아카이브 서비스가 제안되었다. SpotLake는 다양한 클라우드 제공업체로부터 수집한 스팟 배치 점수, 스팟 중단 빈도, 스팟 및 온디맨드 가격 등 다양한 인스턴스 정보들을 웹 페이지로 제공한다. 이 서비스는 공개적으로 이용이 가능하며 과거의 데이터도 포함하고 있다. 이처럼 SpotLake를 통해 사용자는 스팟 인스턴스 가격 및 안정성의 시계열 특징을 분석하여 비용 효율적이고 안정적인 인스턴스를 손쉽게 선택할 수 있다.



(a) 할인율 분포



(b) 할인율 변화 빈도

그림 1: 각 클라우드 제공업체별 할인율의 분포와 변화 빈도

3. CoreMark를 이용한 클라우드 인스턴스 성능 측정

클라우드 인스턴스들의 성능을 정량적으로 비교하기 위해 CoreMark 벤치마크를 사용한다. CoreMark는 EEMBC에서 개발한 오픈소스 CPU 벤치마크 도구다. 행렬 처리, 스테이트 머신, 순환 중복 검사의 워크로드를 실행하여 CPU의 성능을 점수로 표현한다. CoreMark점수를 비교하기 위해 2023년 8월 기준 각 클라우드 벤더별 범용 유형의 최신 인스턴스인 AWS의 M6i, GCP의 N2-standard 그리고 Azure의 Dsv5 인스턴스를 선택하였다. 해당 인스턴스들은 모두 인텔 3세대 제온 스케일러블(Ice Lake) 프로세서를 탑재하고 있다. 그리고 인스턴스 세대 간 성능 변화를 파악하기 위해 선택한 인스턴스의 이전 세대 인스턴스도 포함하여 비교하였다. AWS에서는 M4와 M5, GCP에서는 N1, 그리고 Azure에서는 Dsv4와 Dsv3 인스턴스를 선택하였고, 인스턴스별로 2, 4, 8, 16, 32, 64개 vCPU 인스턴스 크기로 측정하였다. 모든 인스턴스는 동일하게 인스턴스의 vCPU 수와 동일한 수의 쓰레드를 지정하였고 Ubuntu 18.04 운영체제에서 실행되었다.

4. 분석 및 평가

본 절에서는 클라우드 제공업체별 인스턴스 비교를 위해 인스턴스 가격 분포와 변화 빈도를 분석하고, 측정된 CoreMark 점수를 정규화하여 인스턴스 가격과 함께 그 특성을 분석한다. 그리고 인스턴스를 선택할 때 어떤 점을 고려해야 하는지 논의한다.

4.1 가격분석

그림 1은 클라우드 제공업체별 분포와 변화 빈도를 보여준다. SpotLake [2]에서 제공하는 2023년 6월 1일부터 2023년 8월 31일까지의 AWS(주황색 실선), GCP(초록색 파선), Azure(파란색 점선)의 스팟, 온디맨드 인스턴스의 가격 데이터 세트를 사용한다. 먼저 그림 1a는 클라우드 제공업체별 할인율의 누적 분포 함수를 나타낸다. 할인율은 스팟 인스턴스 가격이 온디맨드 인스턴스 가격에 비해 얼마나 할인된 가격으로 제공되는지 나타낸 지표로 $(1.0 - \frac{SpotPrice}{On-demandPrice}) \times 100$ 으로 계산된다. 그림에서 보이는 바와 같이 Azure가 가장 높은 할인율 분포를 보인 반면 AWS가 가장 낮은 할인율 분포를 보였다. GCP의 할인율은 그 중간에 분포하고 있었다. 그림 1b는 클라우드 제공 업체별 할인율의 변화 빈도를 나타낸다. X축은 관측 기간의 시간을 일 단위로 나타내며, Y축은 전체 할인율 데이터 중에서 변화한 비율로 이전 관측 시점에서 현재 관측 시점 할인율이 변화한 비율이다. 그림에서 알 수 있듯이 AWS의 할인율은 매일 변화하며 이는 AWS 스팟 인스턴스 가격의 지속적 변화를 나타낸다. 반면 Azure와 GCP의 할인율은 AWS에 비해 매우 급격히 변화하며 특히 Azure는 매월 1일에 할인율이 변하는 특징을 보였다. 이는 AWS의 스팟 가격은 지속적으로 변화하므로 더욱 주의 깊게 관심을 가져야 함을 의미하고 Azure 및 GCP의 스팟 가격은 드물게 변화하므로 스팟 가격이 안정적이라고 할 수 있다.

4.2 성능분석

그림 2는 3절에서 측정된 CoreMark점수를 코어당 성능 그리고 가격 대비 성능으로 비교하기 위해 해당 인스턴스 유형의 vCPU 수, 스팟 가격, 온디맨드 가격 3가지 정규화 방법으로 $\frac{CoreMarkScore}{NormalizationMethod}$ 와 같이 계산된다. 계산된 값들은 최소-최대 정규화로 다시 정규화된 후, 이를 막대그래프로 시각화하였다. 막대는 정규화한 CoreMark점수의 중앙값이다.

그림 2a는 CoreMark점수를 vCPU개수로 정규화 한 것이다. vCPU 개수당 성능은 AWS의 최신 세대 인스턴스인 M6i가 가장 좋은 성능을 보였다. GCP의 N2에 비해 18%, Azure의 Dsv4보다 16% 더 좋은 성능을 보였다. 대체적으로 최신 세대의 인스턴스일수록 성능이 더 좋았으나 Azure의 경우 최신 세대인 Dsv5보다 이전 세대인 Dsv4가 30% 더 높게 나타났다. 이는 최신 세대의 인스턴스가 무조건적으로 좋은 성능을 보이지 않음을 나타내고 인스턴스 선택에 정량적 지표의 필요성을 나타낸다. 그림 2b

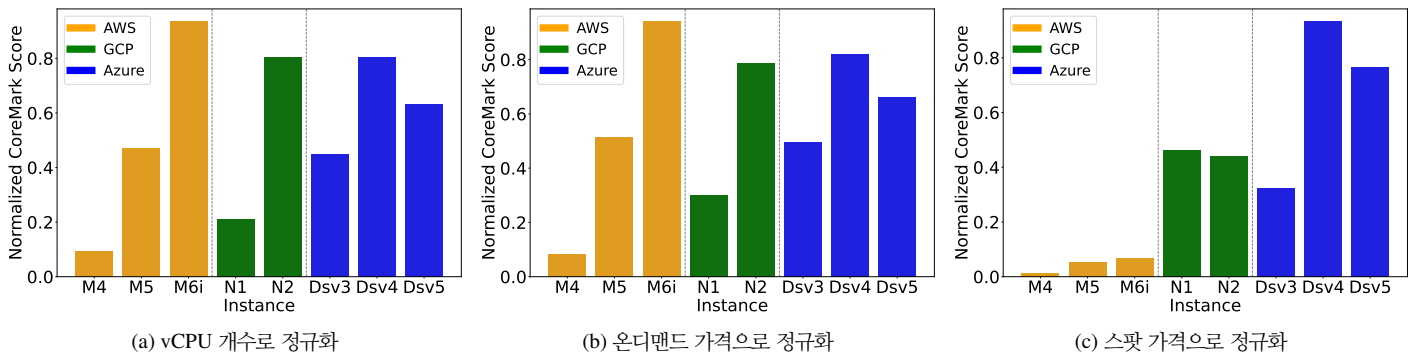


그림 2: 정규화에 따른 CoreMark 점수

은 CoreMark 점수를 온디맨드 가격으로 정규화한 것이다. 온디맨드는 vCPU 개수로 정규화 했을 때와 패턴의 차이는 보이지 않았다. 이는 성능에 따라 온디맨드 가격이 책정되었다고 볼 수 있다. 그림 2c은 CoreMark 점수를 스팟 가격으로 정규화 한 것이다. 기존 vCPU 개수와 온디맨드 가격으로 정규화했을 때와 다른 흥미로운 결과가 나타났다. Azure 인스턴스들이 가격 대비 성능이 가장 좋았고 그다음으로 GCP, AWS 순서를 보였다. 특히 AWS의 스팟 가격 대비 성능이 상대적으로 가장 낮은 것으로 나타났는데 Azure의 Dsv4는 AWS의 M6i보다 14배 더 좋은 가격 대비 성능을 보였고 다른 인스턴스들도 마찬가지였다. 이는 그림 1a에서 분석한 바와 같이 Azure의 할인율이 높고 AWS의 할인율이 상대적으로 낮기 때문에 나타난다.

4.3 인스턴스 선택에 대한 논의

온디맨드 인스턴스로 사용한다면 AWS의 최신 인스턴스인 M6i가 가장 좋은 선택지가 될 수 있다. 그러나 스팟으로 사용할 경우 가장 낮은 가격 대비 성능을 보여 좋은 선택지가 아닐 수 있다. Azure 인스턴스는 스팟으로 사용할 시 가격 대비 성능이 가장 좋았다. 또한 가격의 변화가 매우 드물게 관측되었으므로 가격 최적화가 가장 중요하고 자신의 응용 프로그램이 언제든지 중단되어도 상관없다면 Azure의 스팟 인스턴스를 사용하는 것이 유리하다. GCP의 경우는 온디맨드, 스팟 모두 가격 대비 성능이 다른 클라우드 제공 업체의 인스턴스에 비해 뛰어나지 않았지만, 사용자가 직접 인스턴스의 vCPU 수, 메모리 용량을 지정할 수 있다. [3] 또한 N1 인스턴스의 경우 GPU를 추가할 수 있어 특정 vCPU, 메모리, 혹은 GPU가 필요한 워크로드의 경우 GCP의 인스턴스 선택을 고려할 수 있다.

5. 결론 및 향후 계획

본 연구에서는 CoreMark 벤치마크 도구를 이용하여 AWS, GCP, Azure의 범용 인스턴스들의 성능을 측정하여 클라우드 제공업체별, 세대별 인스턴스를 비교하였고 SpotLake 데이터 세트

를 이용하여 각 인스턴스의 성능, 가격, 안정성의 균형에서 최적의 인스턴스 선택 방법에 대해 고찰하였다. 분석 결과 최신 세대의 인스턴스를 사용한다고 할지라도 무조건 성능이 좋지 않았으며, 온디맨드 인스턴스의 경우 AWS의 M6i가 가장 좋은 성능을 보였고 스팟 인스턴스는 스팟 할인율이 높은 Azure의 인스턴스들이 이용할 시 최대 14배 가격 대비 성능이 좋은 것을 발견하였다. 추가적인 연구로 본 논문에서 다룬 일부의 범용 인스턴스의 CPU 성능 외에도 다양한 유형의 인스턴스에 대한 메모리, 스토리지, GPU 등 다양한 하드웨어 성능을 공정하게 측정해 비교 지표로 제공함으로써 사용자에게 다양한 클라우드 제공 업체의 인스턴스 사용을 장려하여 멀티 클라우드 발전에 도움이 될 것이라 믿는다.

6. 사사

본 연구는 과학기술정보통신부 및 정보통신기획평가원의 SW 중심대학 (2022-0-00964), SW스타랩 (RS-2022-00144309) 및 한국연구재단의 개인연구사업 (NRF-2020R1A2C1102544)의 지원을 받아 수행됨.

참고 문헌

- [1] S. Gal-On and M. Levy, "Exploring coremark a benchmark maximizing simplicity and efficacy," *The Embedded Microprocessor Benchmark Consortium*, 2012.
- [2] S. Lee, J. Hwang, and K. Lee, "Spotlake: Diverse spot instance dataset archive service," in *2022 IEEE International Symposium on Workload Characterization (IISWC)*, (Los Alamitos, CA, USA), pp. 242–255, IEEE Computer Society, nov 2022.
- [3] G. C. P. Documents, "Create a vm with a custom machine type." <https://cloud.google.com/compute/docs/instances/creating-instance-with-custom-machine-type>, 2023.