

다양한 클라우드 업체의 스팟 인스턴스 가격 및 특성 분석*

김정환, 이현영, 황재일, 이경용
국민대학교 소프트웨어학부
{bryan9801, gusdud0122, jaeil, leeky}@kookmin.ac.kr

Analyzing Spot Instance Price and Characteristics of Various Cloud Vendors

Kyunghwan Kim, Hyeonyoung Lee, Jaehil Hwang, Kyungyong Lee
Department of Computer Science, Kookmin University

요 약

클라우드 컴퓨팅에서 스팟 인스턴스를 사용하는 것은 비용 효율적이다. 따라서 많은 이용자들은 스팟 인스턴스를 활용하여 워크로드를 실행하려고 한다. 클라우드 서비스 제공 업체별로 스팟 인스턴스는 종류는 다양하고 그 특성도 다르다. 하지만 업체들은 스팟 인스턴스 가격의 최신 데이터만을 제한된 방식으로 제공하기에, 사용자는 스팟 인스턴스의 가격을 예측하고 적절한 인스턴스를 선택하는데 어려움이 따른다. 본 논문에서는 대표적인 클라우드 서비스 제공 업체 Amazon Web Services(AWS), Google Cloud Platforms(GCP), Microsoft Azure 3개 업체의 스팟 가격 데이터를 수집하고 시각화하여 각 업체의 스팟 인스턴스 가격 특성을 분석하였고, 업체별 스팟 인스턴스의 가격 및 특성에서 유의미한 차이를 확인했다.

1. 서 론

최근 소셜 네트워크, 온라인 스트리밍 플랫폼 등 다양한 응용 서비스 사례들이 클라우드 환경에서 구축 및 서비스되고 있다. 또한 인공지능을 응용한 서비스 구축 시에도 클라우드 컴퓨팅에서 제공되는 자원이 다양하게 활용되고 있다. 퍼블릭 클라우드 리소스에 대한 활용 수요가 급증하면서, 클라우드 서비스 제공 업체들은 클라우드 자원의 가격, 성능 및 안정성을 조정하며 다양한 과금 체계를 제공하고 있다. 대표적인 클라우드 제공 업체인 AWS는 Elastic Compute Cloud (EC2) 인스턴스 과금 체계로 온디맨드 인스턴스, 예약 인스턴스, 전용 호스트 및 스팟 인스턴스 등을 제공한다. 온디맨드 인스턴스는 사용 단위에 따라 고정된 가격을 지불하는 종량제 요금으로, 많은 퍼블릭 클라우드 서비스 제공 업체에서 채택하고 있다. 하지만 온디맨드 인스턴스의 가격은 잘 변동되지 않기 때문에 사용자가 저렴한 가격에 인스턴스를 사용하기 어렵다. 스팟 인스턴스는 퍼블릭 클라우드 컴퓨팅 자원의 미사용 용량을 온디맨드에 비해 저렴한 가격으로 제공한다는 장점이 있다. 일반적으로 시작 및 종료 시각이 중요하지 않은 중단 가능한 워크로드나 예측할 수 없는 대용량 컴퓨팅 용량이 필요한 워크로드에서 주로 사용된다.

스팟 인스턴스는 온디맨드에 비해 낮은 가격에 서비스를 제공하지만, 컴퓨팅 자원에 대한 안정성과 신뢰도는 낮다. 또한, 주요 업체마다 스팟 인스턴스의 할인율과 제공하는 정보에 차이가 있다. 따라서 사용자는 각 제공 업체별로 제공하는 스팟 인스턴스의 특성을 파악하여 워크로드별로 적합한 업체의 스팟 인스턴스를 사용할지 결정할 수 있다. 본 논문에서는 다양한 퍼블릭 클라우드 서비스 제공 업체의 스팟 인스턴스의 특성을 비교하고자 한다. AWS, GCP, Azure 3개 업체의 스팟 인스턴스 자원 정보를 API를 통해 수집하여 업체별로 제공하는 정보의 차이를 비교한다. 또한, 일정 기간의 업체별 스팟 인스턴스의 가격 변화 추이를 분석하고 가격 변화량의 차이를 비교함으로써 업체 간 스팟 인스턴스의 가격 특성을 분석한다. 그 결과, AWS는 잦은 스팟 인스턴스 가격 변동을 보여줬으며, GCP는 그에 비해 스팟 인스턴스 변화율이 적었다. Azure는 분석 기간 동안 스팟 인스턴스 가격 변화를 보여주지 않았다.

2. 스팟 인스턴스

2-1. 스팟 인스턴스의 개념

스팟 인스턴스는 클라우드의 유휴 자원을 온디맨드에 비해 상당히 낮은 가격으로 가상 머신을 이용할 수 있는 자원이다. 클라우드 서비스 제공 업체들이 일반적으로 제공하는 종량제 요금인 온디맨드 인스턴스는 정적인 가격 정책을 채택한다. 고객이 온디맨드 인스턴스를 사용하면 지불해야 할 금액을 정확히 예측할 수 있지만, 업체에서는 시장의 수요와 공급을 자원 가격에 제대로 반영할 수 없다. 반면에 스팟 인스턴스는 동적인 가격정책을 사용하기 때문에 수요와 공급 반영에 있어서 유연하다. 즉, 자원의 수요가 높을 때 업체가 가격을 인상할 수 있고, 수요가 낮을 때 고객은 자원을 저렴하게 사용할 수 있다. 사용자는 스팟 인스턴스를 시작할 가격을 지정할 수 있다.[1] 스팟의 시장 가격이 사용자가 지정한 가격보다 낮을 때 인스턴스가 시작하고, 그 반대의 경우에는 인스턴스가 종료된다. 이러한 특징 때문에 스팟 인스턴스는 내결함성이 보장되거나, 긴급하게 대량의 컴퓨팅 용량이 필요한 워크로드에 적합하다.

2-2. 클라우드 서비스 제공 업체별 스팟 인스턴스 특성

AWS에서는 'Amazon EC2 Spot Instances' 라는 명칭으로 스팟 인스턴스를 제공한다. AWS의 스팟 인스턴스는 온디맨드 가격 대비 최대 90%의 할인율을 제공한다. EC2 Auto Scaling을 사용하여 스팟 인스턴스 전반에 용량을 프로비저닝 할 수 있으며, Amazon ECS, AWS CloudFormation 등 AWS가 제공하는 서비스와 통합되어 유지관리가 용이하다. AWS에서는 스팟 인스턴스가 종료되기 2분 전에 사용자에게 알림을 보낸다. 사용자는 'Spot Instance Advisor' 서비스를 사용해 온디맨드와 스팟 인스턴스 가격을 비교할 수 있다.

GCP에서는 'Spot VM' 라는 명칭으로 스팟 인스턴스를 제공한다. 이전에는 'Preemptible VM' 이라는 명칭으로 유사한 가격 정책의 인스턴스를 제공했으나, 최근에는 가상 머신의 24시간으로 제한되었던 최대 실행 시간 규제를 제거한 'Spot VM'으로 변경하고 있다. GCP의 스팟 인스턴스는 온디맨드 가격 대비 최대 91%의 할인율을 제공한다. GCP에서는 종료 스크립트를 활용하여 가상 머신의 상태를 보존하는 기능을 사용할 수 있다.

Azure에서는 'Spot Virtual Machines(VMs)' 라는 명칭으로 스팟 인스턴스를 제공한다. Azure는 온디맨드 가격 대비 최대

*본 연구는 과학기술정보통신부 및 정보통신기획평가원의 SW스타랩(RS-2022-00144309) 및 한국연구재단 과학기술분야 기초연구사업의(NRF-2020R1A2C1102544) 지원을 받아 수행되었음.

90%의 할인율을 제공한다. 사용자는 정책을 설정하여 지정 가격보다 스팟 인스턴스 가격이 높아졌을 때, 가상 머신 중지 혹은 제거를 선택할 수 있다. 가상 머신이 제거되기 30초 전에 알림을 전송한다. 사용자는 Portal 서비스[5]에서 인접 리전의 최근 3개월간 스팟 가격 변화 추이를 살펴볼 수 있다.

본 논문에서는 3개 업체의 스팟 자원을 '스팟 인스턴스'로 공통되게 칭한다.

3. 데이터 수집 모듈 및 업체별 데이터 특성

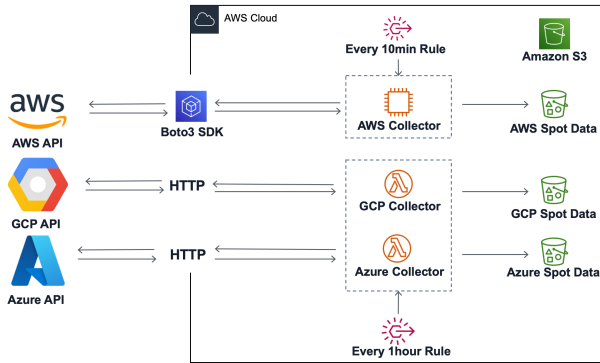


그림 1. 스팟 데이터 수집 시스템 구조도

세 클라우드 서비스 제공 업체 모두 스팟 인스턴스 정보를 제공하지만, 제공 방법과 데이터셋에 차이가 있다. AWS의 경우 SDK로, GCP와 Azure는 HTTP 요청을 통해 데이터를 수집할 수 있다. 따라서 업체별로 다른 자원을 사용하여 수집 모듈을 구현하였다.

3-1. 데이터 수집 모듈 구조

본 논문에서는 AWS Lambda와 EC2를 기반으로 데이터를 수집한다. 그림 1은 업체별 스팟 데이터 수집 시스템의 구조도이다. 수집 모듈의 내부는 데이터를 수집, 전처리, 저장하는 3 단계로 구분된다. AWS collector는 매 10분마다 crontab을 이용하며, GCP와 Azure는 매 1시간마다 AWS EventBridge 트리거를 작동시켜 수집 모듈이 실행된다. 각 업체별로 스팟 데이터를 수집하기 위해 AWS 데이터 수집은 Boto3 라이브러리를 사용하고 GCP와 Azure의 데이터 수집은 업체에서 제공하는 REST API를 활용한다. GCP에서는 추가적으로 VM 인스턴스 가격을 제공하는 페이지[4]를 크롤링하여 두 개의 데이터 리소스를 가진다. 수집한 데이터는 csv형식을 gzip으로 압축 후, 수집 시점을 파일 경로로 하여 S3에 저장한다.

AWS 수집 모듈의 수집 주기와 환경이 다른 이유는 스팟 인스턴스의 가격이 매우 빈번하게 변화했기 때문이다. 다른 업체들의 수집 모듈처럼 한 시간의 주기로 동작하면 포착하지 못하는 데이터 변화가 많을 수밖에 없다. 또한, 서버리스의 단점 중 하나인 콜드스타트로 인해 10분 간격의 데이터 수집에 영향을 미칠 수 있어 AWS 수집 모듈은 콜드스타트에서 자유로운 EC2에서 수집 모듈을 실행한다.

3-2. 업체별 수집 데이터의 특성

표 1은 수집된 3개 클라우드 서비스 제공 업체에서 수집하는 데이터의 유형이다. 세 업체 모두 인스턴스 타입, 리전, 온디맨드 가격, 스팟 가격과 수집 시점 정보를 저장한다. AWS의 경우 추가적으로 SPS와 IF정보를 수집한다. SPS란 'Spot Placement Score'를 의미하며, 이는 스팟 요청이 리전 또는 가용 영역에서 성공할 가능성을 나타낸다. 스팟 자원은 잉여 컴퓨팅 자원으로 부터 제공되기 때문에 자원의 크기는 가변적이다. 따라서 사용

Vendor	Dataset
AWS	Date, Instance Type, Region, Ondemand Price(\$), Spot Price(\$), SPS, IF
GCP	Date, Instance Type, Region, Ondemand Price(\$), Spot Price(\$)
Azure	Date, Instance Type, Region, Ondemand Price(\$), Spot Price(\$)

표 1. 제공업체별 수집 데이터 유형

자가 항상 필요한 스팟 자원을 확보할 수 있는지는 장담할 수 없다. 사용자는 SPS를 이용하여 스팟 요청 성공 가능성이 높은 리전이나 가용 영역을 알아낼 수 있다. IF는 'Interrupt Frequency'로, 스팟 인스턴스가 중단될 확률을 의미한다. AWS의 Spot Instance Advisor[2]를 통해 알 수 있으며, 현재 프로그래밍 방식 액세스는 지원하지 않기 때문에 웹 사이트를 통해서만 접근할 수 있다. 사용자는 SPS와 IF 정보를 통해 스팟 인스턴스에 대한 새로운 통찰을 얻을 수 있다. 따라서, AWS의 스팟 인스턴스 자원 사용을 고려할 때 단순히 과거의 가격에만 의존하는 것이 아니라 보다 객관적인 지표로 인스턴스 가용성에 확신을 얻을 수 있다.[3]

4. 분석 및 평가

4-1. 데이터 수집 환경

3가지 클라우드 서비스 제공 업체의 스팟 데이터를 수집하기 위해서 그림 1과 같은 구조의 AWS 환경의 EC2와 서버리스 기반 수집 시스템을 구현하였다. AWS 스팟 데이터 수집 모듈은 EC2 t3.micro인스턴스 OS는 ubuntu 22.04, x86 아키텍처 vCPU2개, 1GiB 메모리, 8GB EBS의 인스턴스에서 python 3.9 런타임, boto3 1.24.35 버전으로 실행된다.

AWS Lambda에서 작동되는 Azure, GCP의 스팟 데이터 수집 모듈은 python 3.9 런타임의 pandas 1.5.0 패키지로 작동한다. AWS Lambda는 amazon/aws-lambda-python:3.9 이미지 기반의 도커 컨테이너에서 1024MB 메모리 설정으로 실행된다. 수집된 모든 데이터는 AWS S3에 저장되며 위 모든 자원은 us-west-2 (oregon) 지역에 구성하였다.

4-2. AWS, GCP, Azure의 스팟 가격 변화 빈도 비교

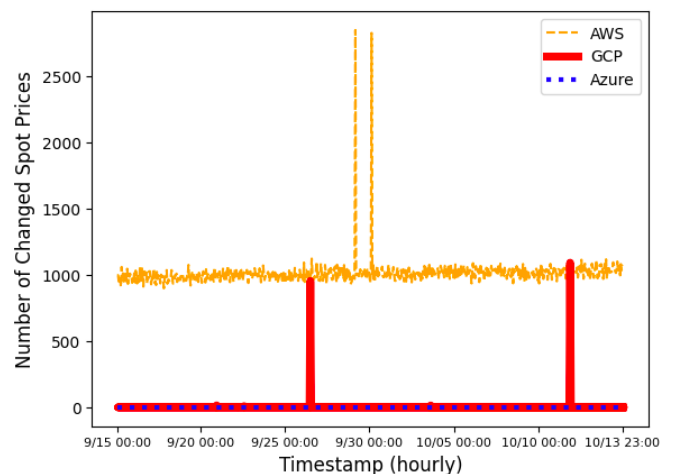


그림 2. 각 제공업체별 시간당 스팟 가격 변화 횟수

그림 2는 제공 업체별 스팟 인스턴스 가격 데이터의 변화 추이를 비교한다. 수집 기간은 2022년 9월 15일부터 2022년 10월 14일까지로, 총 30일간의 데이터를 비교했다. 세 제공 업체의 스팟 인스턴스의 가격이 변화하는 양상이 다른 것을 알 수 있는데, AWS는 지속적으로 변화하며, GCP는 특정 시점에서 데이

터가 크게 변화했다. Azure의 경우 수집 기간 동안 한 번도 데이터가 변화하지 않았다.

AWS의 경우 평균적으로 시간 당 1011개의 데이터가 변화했으며, 전체 수집 데이터 대비 평균적으로 5%의 데이터가 변화했다. 변화된 데이터가 없는 수집 시점은 존재하지 않았다. 눈에 띄게 변화가 발생한 두 시점에는 각각 UTC 기준 9월 29일 04:00에 2846개, 9월 30일 03:00에 2824 개가 변화했다. 두 변화 데이터 중 742개의 AZ정보가 NaN 처리가 되어 발생한 변화였다. 이상치가 존재한 데이터셋은 모두 일치했기에, API 수집이 불안정 했던 것으로 예상된다. 그 외의 데이터는 약 35%의 동일한 인스턴스 타입의 변화가 발생하였고, 65%는 서로 다른 인스턴스들의 변화가 발생하였다.

GCP의 스팟 인스턴스의 가격은 AWS만큼 빈번하게 변화하지는 않았다. 하지만, 해당 기간 동안 눈에 띄는 변화를 보인 시점이 두 번 존재했다. 9월 26일 오후 12시에 953개의 스팟 인스턴스 가격이 변화하였으며, 전체 수집 데이터 개수 대비 약 19.5%의 변화율을 보였다. 이후 10월 11일 21시에 1091개의 스팟 인스턴스 가격이 변화하였고, 전체 수집 데이터 개수 대비 약 21.5%가 변화했다. 흥미로운 점은, 첫 번째 시점에서 변화한 스팟 인스턴스 데이터가 모두 두 번째 시점에서도 변화했다는 것이다. 이 중 첫 번째 시점의 스팟 가격보다 두 번째 시점의 스팟 가격이 증가한 데이터가 약 39%를 차지했으며, 평균적으로 기존 스팟 가격 대비 140%로 증가하였다. 나머지 61% 데이터는 두 번째 시점에서 스팟 가격이 감소했고, 기존 스팟 가격 대비 80%로 감소하는 모습을 보였다. 두 번의 변화가 발생한 인스턴스 타입은 범용 제품군인 N1, N2, N2D, E2, T2D 계열과 컴퓨팅 최적화 제품군인 C2와 C2D 계열로 분류할 수 있다. 첫 번째 데이터 변화와 두 번째 데이터 변화를 비교하였을 때, N2D와 E2 제품군의 모든 스팟 인스턴스 가격은 증가하였다. T2D 제품군의 경우에는 20%는 증가하였고, 80%는 감소하였다. 나머지 제품군은 모두 가격이 감소하였다.

4-3. 업체별 스팟 가격의 할인율 비교

Vendor	Date	Instance Type	Region	Ondemand Price(\$)	Spot Price(\$)	Savings	SPS	IF
AWS	2022/09/30 00:00	m5.xlarge	ap Northeast 2	0.236	0.0629	73	3	2.0
GCP	2022/09/30 00:00	n1 Standard 4	asia northeast 3	0.244	0.053	78	X	X
Azure	2022/09/30 00:00	D4s_v3	koreacentral	0.246	0.0247	89	X	X

표 2. 제공업체별 스팟 인스턴스 수집 데이터

표 2는 AWS에서 균형있는 컴퓨팅, 메모리 및 네트워킹 리소스를 제공하는 범용 인스턴스 타입인 m5.xlarge와 그와 동일한 vCPU, memory 용량을 가진 GCP의 n1-standard-4, Azure의 Standard_D4s_v3 인스턴스를 선택하여 온디맨드와 스팟 가격을 비교하였다.

표에는 수집한 데이터의 온디맨드 가격 대비 스팟 가격의 할인율을 나타내는 Savings를 나타내었다. 온디맨드 가격의 경우 GCP, Azure는 비슷하며, AWS가 가장 저렴했다. 하지만 스팟 가격을 비교해 볼 때 Azure가 가장 저렴했고 GCP, AWS 순서로 스팟 가격이 높았다. 표에서 제시하지 않은 다른 인스턴스들도 비슷한 vCPU 개수, 메모리 용량을 가지고 있다면 위 표와 비슷한 결과를 보였다.

위 결과로 미루어 보았을 때 동일한 하드웨어 자원일 경우 Azure에서 가장 저렴하게 스팟 인스턴스를 이용할 수 있다. 다만 스팟 가격의 변화가 없었기 때문에 실제로 유휴 자원의 수요 공급을 잘 반영하지 못해 예상하지 못한 중단이 발생할 확률이 클 가능성도 고려해야한다. AWS는 다른 제공업체에 비해

상대적으로 Savings는 낮지만 SPS, IF 등 스팟 인스턴스 가용성에 대해 더 자세한 정보를 제공해주고, 이러한 정보들을 바탕으로 자주 변하는 AWS 스팟 인스턴스의 안정성을 예측하는 데 도움을 준다. GCP는 Azure와 AWS의 중간 정도의 스팟 가격을 보여주었다. 그러나 GCP는 스팟 인스턴스에 연결된 GPU, 로컬 SSD, 고정 외부 IP 주소에도 할인이 적용된다.[6] 이점을 이용해 다른 업체에 비해 저렴한 가격으로 GPU연산을 하거나 추가적인 스토리지가 필요한 워크로드의 경우 GCP의 스팟 인스턴스의 이용을 고려할 수 있다.

5. 결론 및 향후 계획

본 연구에서는 스팟 가격 데이터 수집 모듈을 통해 데이터를 수집, 전처리, 저장하고 시각화하여 각 클라우드 서비스 제공업체별 스팟 데이터 변화와 특성을 고찰하였다. 이후 서술하는 항목들에서 추가적인 연구와 발전이 가능하다.

첫째, 스팟 인스턴스를 사용하면 유휴 클라우드 자원을 온디맨드 인스턴스에 비해 더 저렴하게 사용할 수 있으나, 스팟 가격의 변동에 따라 언제든지 스팟 인스턴스가 업체에 의해 중단될 수 있는 가용성의 문제를 가지고 있다. 이 문제를 해결하기 위해 업체별 스팟 가격 및 안정성 데이터를 수집하고 전처리한 후 저장, 시각화함으로써 업체 별 스팟 인스턴스의 가용성 및 가격 변화를 예측할 수 있다. 사용자는 이 정보를 참조하여 앱 및 워크로드를 실행할 때 비용 효율적이고 가용성을 예측할 수 있는 인스턴스 및 업체를 선택하는데 도움이 될 수 있다.

둘째, 대부분의 업체가 스팟 가격 데이터를 제한된 방식으로 최신 정보만 제공한다는 점을 고려했을 때, 제안된 데이터 수집 모듈을 통해 스팟 데이터 세트를 분석한다면 수많은 연구 기회를 가진 다양한 업체의 클라우드 사용이 크게 향상될 것이다. 추후 AWS GCP, Azure 뿐만 아닌 알리바바, 화웨이, 네이버, IBM 등 다양한 업체의 스팟 인스턴스 가격 및 특성을 분석하여 클라우드 컴퓨팅의 발전 방향이라고 여겨지는 멀티 클라우드 서비스를 연구하는데 도움을 줄 것이다.

셋째, 현재 클라우드 이용자가 앱 및 워크로드에 적합한 업체, 인스턴스를 선택하는 것에는 많은 고민이 따른다. 현재 클라우드 시장에는 수많은 업체와 그만큼 다양한 인스턴스 타입이 있다. 다양한 인스턴스 크기와 옵션, 유형에 따라 그 가격과 성능은 천차만별이다. 이 중 어떤 인스턴스 타입이 실행시킬 앱 및 워크로드에 적합한지 선택하는 것은 매우 어려운 일이다. 이를 위해 수집된 가격 데이터를 기반으로 하드웨어 특성, 가용성 정보 등을 추가해 앱 및 워크로드별 인스턴스를 추천하는 서비스를 구축 할 수 있다.

참고문헌

- [1] Javadi B, Thulasiram RK, Buyya R.Characterizing spot price dynamics in public cloud environments [Internet]. Future Generation Computer Systems. 29(4), 988-999 (2013).
- [2] Retrieved October 26, 2022, from <https://aws.amazon.com/ko/ec2/spot/instance-advisor>
- [3] Sungjae Lee, Jaeil Hwang, and Kyungyong Lee, IISWC 2022 Artifacts for “SpotLake:Diverse Spot Instance Dataset Archive.”
- [4] Retrieved October 26, 2022, from <https://cloud.google.com/compute/pricing#vm-instance-pricing>
- [5] Retrieved October 26, 2022, from <https://portal.azure.com/#create/Microsoft.VirtualMachine>
- [6] Retrieved October 26, 2022, from <https://Cloud.Google.Com/Compute/Docs/Instances/Spot>.