Public Spot Instance Dataset Archive Service

KyungHwan Kim*
Computer Science (CS). Kookmin U.
Seoul, South Korea
bryan9801@kookmin.ac.kr

Subin Park*
CS. Kookmin Univ.
Seoul, South Korea
subean@kookmin.ac.kr

Jaeil Hwang
CS. Kookmin Univ.
Seoul, South Korea
jaeil@kookmin.ac.kr

HyeonYoung Lee
CS. Kookmin Univ.
Seoul, South Korea
gusdud0122@kookmin.ac.kr

SeokHyeon Kang
CS. Kookmin Univ.
Seoul, South Korea
kh3654p@kookmin.ac.kr

Kyungyong Lee
CS. Kookmin Univ.
Seoul, South Korea
leeky@kookmin.ac.kr

ABSTRACT

Spot instances offered by major cloud vendors allow users to use cloud instances cost-effectively but with the risk of sudden instance interruption. To enable efficient use of spot instances by users, cloud vendors provide various datasets that reflect the current status of spot instance services, such as savings ratio, interrupt ratio, and instant availability. However, this information is scattered, and they require distinct access mechanisms and pose query constraints. Hence, ordinary users find it difficult to use the dataset to optimize spot instance usage. To resolve this issue, we propose a multi-cloud spot instance dataset service that is publicly available. This will help cloud users and system researchers to use spot instances from multiple cloud vendors to build a cost-efficient and reliable environment expediting cloud system research.

CCS CONCEPTS

• **Computer systems organization → Cloud computing**.

KEYWORDS

spot instance dataset, multi-cloud

ACM Reference Format:
KyungHwan Kim, Subin Park, Jaeil Hwang, HyeonYoung Lee, SeokHyeon Kang, and Kyungyong Lee. 2023. Public Spot Instance Dataset Archive Service. In *Companion Proceedings of the ACM Web Conference 2023 (WWW '23 Companion), April 30–May 04, 2023, Austin, TX, USA*. ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3543873.3587314

1 INTRODUCTION

Cloud computing allows for the flexible use of compute resources through an elastic on-demand billing model. Among the various cloud instance price models, spot instance provides significant cost savings of up to 90% from the on-demand price but carries a risk of unexpected server termination. A public cloud service vendor internally decides the cost savings from spot instances and the server interruption events, while reflecting the surplus computing resources needed to run cloud instances.

Public cloud vendors provide various datasets that present cost savings when using spot instances, such as the spot instance interruption event ratio over the prior period and instant information on spot instance availability, to enable cost-efficient and reliable spot instance usages. This information is crucial for optimizing spot instance usage for various workloads [7]. However, accessing various spot datasets is challenging because different cloud vendors provide their own information in distinct locations using different access mediums, such as programmatic access or a web interface. In addition, some datasets impose query restrictions, making it further difficult to access the entire dataset [8]. Moreover, most spot datasets offer only the latest information, and a lack of historical information prohibits thorough analysis and dataset modeling [2].

A prior work [8] tried to solve this problem by proposing a spot instance dataset archive servic. However, it provides the information only for Amazon Web Service (AWS), and it lacks support for multiple cloud vendors which limits the resource optimization opportunity. To deal with this issue, we develop a web-based spot instance dataset service which covers major cloud vendors, AWS, Microsoft Azure, and Google Cloud Platform (GCP). This allows users to compare easily the spot instances to build an optimal multi-cloud environment. Our proposed dataset service also provides access to historical spot instance datasets, allowing users to model the statistical characteristics of the datasets to enhance spot instance usage. In this study, we compare the characteristics of spot instances from AWS, Azure, and GCP using various spot datasets. To the best of our knowledge, this is the first work to compare spot instances from major cloud vendors. We believe that our publicly available dataset service will allow researchers to conduct further analysis and obtain research outcomes related to building cost-efficient cloud environments.

2 SPOT INSTANCES AND DATASET

The elasticity of using computing resources is an important factor contributing to the success of cloud computing. Cloud vendors should always have significant amounts of computing resources available to ensure that users can easily access these resources at any time. However, this can result in low server utilization as the required computing power changes, especially with a diurnal or weekday usage patterns [6]. To reduce resource wastage and improve resource utilization, public cloud vendors, such as AWS, Azure, and GCP, offer excess computing resources at a lower

*Both authors contributed equally to this research.

WWW '23 Companion, April 30–May 04, 2023, Austin, TX, USA

ACM ISBN 978-1-4503-9419-2/23/04.
https://doi.org/10.1145/3543873.3587314

cost, which are known as *spot instances*. One disadvantage of using spot instances is unexpected instance termination, which occurs when surplus computing resources decrease as regular on-demand instance usage increases. Therefore, cloud vendors publicly offer diverse information to help users estimate the cost savings and the possibility of instance interruption when using spot instances.

2.1 Spot Price and Savings Ratio

The spot price dataset provides timely price information that is dynamic in nature, indicating surplus instances. Users can easily calculate the cost savings ratio by dividing the spot price by the on-demand instance price. When AWS first introduced spot instances, it adopted a uniform price and sealed bidding mechanism [1] where users pay the same spot price regardless of bidding price without knowing other users' bidding price. In the bidding mechanism, instance interruption generally occurs when the spot price, determined by the service provider, becomes higher than the bidding price set by a user. To help users expect the interruption events due to the low bidding price, AWS provided timely spot price information which was updated frequently. By using the spot price dataset, many research, including statistical analysis of price change [1], and proposal for an optimal bidding heuristic with the price suggestion [2, 9], has been conducted.

The spot instances provided by the Azure adopts a similar bidding mechanism to the AWS with spot price dataset. The Azure spot price dataset is not updated as frequently as that of the AWS, and users cannot estimate the out-bidding interruption events from the spot price dataset. GCP adopted a slightly different operation mechanism, and a user does not have to specify the bidding price, and instance interruption events might occur depending on the resource availability. GCP provides a spot price dataset, but it does not provide any indication of instance interruption events.

The spot price datasets provided by the AWS were updated frequently while reflecting the out-of-bid interruption events compared to the Azure and the GCP. However, in 2017, the spot price policy of the AWS changed in a direction similar to that of the Azure and GCP [4]. This change led to less frequent updates of the spot price, resulting in more stability [5]. The new-spot-price dataset does not uncover sudden spot price surging and does not allow users to estimate the instance interruption events from the spot price dataset.

In summary, all three major public cloud vendors provide spot price datasets to help users understand cost savings. However, the spot price information provided by these vendors does not allow users to estimate the possibility of spot instance interruption events.

2.2 Historical Interruption Ratio

The spot price datasets no longer represent instance interruption events. Thus, vendors provide interruption ratio datasets to help users estimate the possibility of interruption. AWS provides the spot instance interruption rate in the past month through its Spot Instance Advisor service. Azure provides the eviction rates for the last 28 days through its Azure Resource Graph service. For both these services, the ratio is categorized into five classes: less than 5%, between 5% and 10%, between 10% and 15%, between 15% and 20%, and more than 20%. In contrast, GCP does not provide the spot instance interruption frequency datasets.

2.3 Instant Availability Information

It should be noted that neither the spot price nor the interruption ratio datasets reflect the timely spot instance availability at the time of the spot instance request. Thus, AWS provides the Spot Placement Score (SPS) to help users calculate the likelihood of a spot request's success before launching instances. The score is provided as a numeric value ranging from 1 to 10, where a higher score implies a higher probability for spot request success. Nevertheless, the internal operation mechanism has not been publicly announced. Lee et. al. [8] thoroughly analyzed the behavior of the SPS operations and concluded that SPS is better at representing timely spot instance availability than the spot instance price and interruption ratio datasets. However, Azure and GCP do not provide instantaneous spot instance availability information.

2.4 Challenges of Accessing Various Datasets

Users can use the aforementioned various spot instance datasets to build a cost-efficient and reliable cloud environment. However, challenges in accessing the dataset include difficulty in accessing various datasets and the fact that they are offered at distinct places. The interruption frequency and spot availability datasets were recently announced, but little is known about them compared to the spot price datasets, which received remarkable interest from spot instance users and research communities.

The dataset access mechanism is not identical across various datasets, and some datasets do not support programmatic access and also imposes a few restrictions that hinder data accessibility. The spot price of the GCP and the interruption frequency of the AWS are officially offered through a webpage without any programmatic access support. The availability of datasets provided by the AWS imposes a few query restrictions that prohibit extensive dataset investigations [8].

Access to historical spot instance prices of the AWS allow the users to enhance their spot instance usages [7]. However, no historical datasets of spot instances are offered by cloud vendors, other than the spot price dataset provided by the AWS. To overcome this issue, Lee et. al. [8] built a system to provide historical datasets of interruption ratio and timely availability information for the AWS, but it still lacks the historical datasets by the Azure and GCP.

3 DATA SERVICE ARCHITECTURE AND IMPLEMENTATION

We implemented the proposed spot instance dataset web service[1] adopting a serverless architecture as shown in Figure 1. The implemented service collects various spot instance datasets from AWS, Azure, and GCP. The data change frequency is different for vendors (Figure 3), and we use either AWS Lambda or an EC2 server to minimize data collection cost. In the case of frequently updated datasets, we collect the dataset every 10 minutes using *cron* a job in a server. In the case of rarely updated datasets, we use the AWS Lambda, which is invoked every hour. The dataset collection is conducted

[1] https://spotlake.ddps.cloud

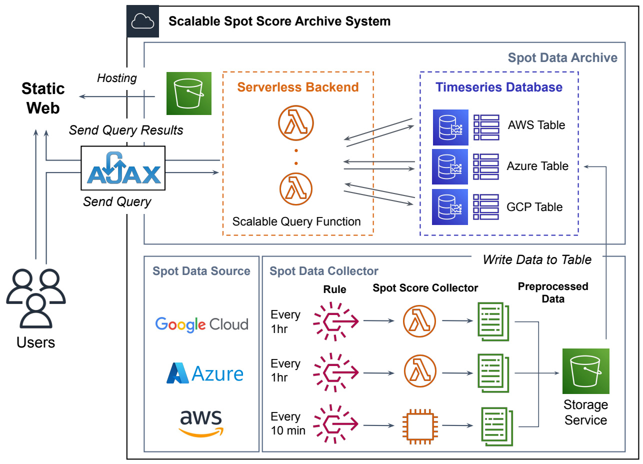


Figure 1: The architecture of data collector implementation using serverless computing

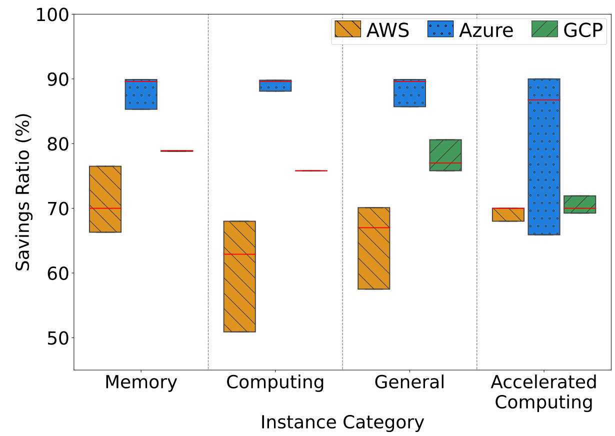
using an AWS *boto3* library, a REST API, or a custom implemented crawler. The collected dataset is stored in a time-series database. The landing HTML web-page is served through an object storage service, Amazon S3. The spot instance datasets and query results are updated dynamically after the initial page loading, and they are served asynchronously using AJAX library.

4 EMPIRICAL ANALYSIS OF DATASETS

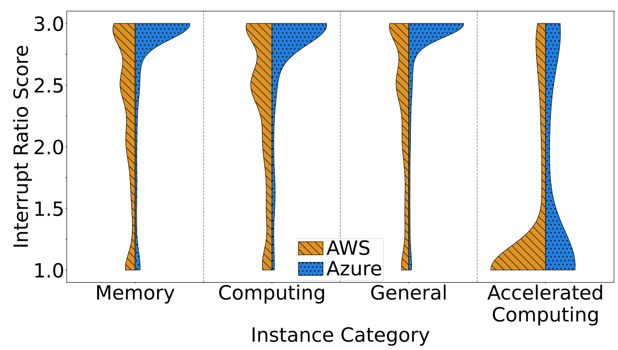
The various datasets from multiple cloud vendors enable comprehensive spot instance analysis and comparisons. In this section, we compare the spot instance datasets provided by multiple vendors. To the best of our knowledge, this is the first work to compare thoroughly public spot instance datasets offered by multiple major cloud vendors, including AWS, Azure, and GCP.

First, we compare the characteristics of spot instance datasets in Figure 2. All of AWS, Azure, and GCP provide spot instance price and on-demand instance price. AWS, Azure, and GCP provide spot instance prices and on-demand instance prices. We calculate the spot instance cost savings using the dataset and present the result in Figure 2a using a dataset ranging from September 15, 2022 to January 5, 2023 per an instance category which is shown in the horizontal axis. Multiple instance types exist in distinct regions in each category, and many savings ratio values are presented using a box plot. The median value is marked with a red line in the middle of the box. We show the values of AWS (upper left diagonal pattern), Azure (dotted pattern), and GCP (upper right diagonal pattern) in each category, from left to right. As shown in the figure, irrespective of the category, Azure showed the most savings, while AWS showed the least savings. We expected instance types with a lower cost savings in the accelerated computing due to the popularity of deep learning and specialized hardware usage [7], but the actual cost savings produced using accelerated computing devices do not show any noticeable difference with other instances.

Figure 2b shows the score distribution for interruption ratio datasets provided by the AWS and Azure using a violin plot format. The vertical axis indicates the interruption ratio as a score ranging



(a) Spot Instance Savings Ratio (Higher is better)

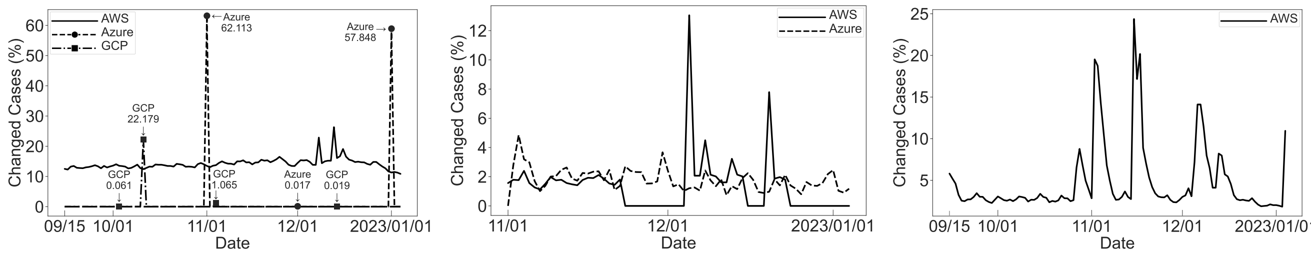


(b) Interruption-Free Score (Higher is better with less interruption)

Figure 2: Value distribution of spot instance datasets from multiple vendors

from 1.0 to 3.0. Both vendors provide the interruption ratio of earlier periods in the range of *less than 5%, 5%-10%, 10%-15%, 15%-20%*, and *more than 20%*, and we match the interruption ratio category to the score of 3.0 to 1.0 with a step of 0.5. Therefore, a higher score implies a more stable spot instance behavior. For each category type presented in the horizontal axis, the upper left diagonal pattern plot on the left presents the interruption ratio score of the AWS, and the dotted pattern on the right presents that of the Azure. As shown in Figure 2b, Azure exhibited a higher interruption ratio score than the AWS across all instance categories. On the average, the interrupt ratio score of AWS is 2.2, and that of Azure is 2.7. The interruption ratio scores of *Accelerated Computing* show a low scores and display different pattern from other categories, implying a higher degree of instance interruption in the category. Based on our analysis of the spot price savings ratios in Figure 2a, the accelerated computing instances were similar to those of other categories. This observation concurs with the recent findings that the spot price cannot be considered as a valid source of information to expect spot instance interruption events [3].

Figure 3 presents the change frequency of spot instance datasets. This change frequency can indicate the timeliness of the provided



(a) Spot instance savings ratio

(b) Interrupt ratio

(c) Instant availability data

Figure 3: The ratio of datasets that have been changed during the observation period

information. Figure 3a presents the spot instance savings value updated ratio of AWS (solid line), Azure (dashed line with round markers), and GCP (dash-dotted line with square markers). The data collection dates are presented along the horizontal axis, while the percentages of instance types that have value change events in each timestamp are presented along the vertical axis. In each observation point in the horizontal axis, we compare an integer saving ratio after rounding with the value in the previous observation point and counting the number of changed cases to calculate the ratio on the vertical axis. As shown in Figure 3, the spot savings of the AWS changes quite regularly across the observation point. The savings of Azure and GCP change rather abruptly compared to those of the AWS. For instance, Azure shows two major spot price changes on Nov. 1, 2022 and Jan. 1, 2023, where over 50% of instances have changed their spot price at once. This observation indicates that users should be more alert about the instant spot price of the AWS as it changes frequently compared to others.

Figure 3b shows the interruption ratio dataset change percentage of the AWS and Azure. We started to collect the Azure interruption ratio dataset from Nov. 1st. 2022 which is the starting date in the horizontal axis. The vertical axis shows the percentage of instance types that have different interruption ratios compared to the previous timestamp. In the observation period, about 2% of the Azure instances showed regular interruption ratio changes. AWS instances show a rather irregular interruption ratio for dataset change events. On Dec. 1, 2022, it shows no dataset changes over a week and thereafter shows a high change ratio. Figure 3c shows the percentage of changes of instant availability dataset provided by the AWS only. We observe a more regular update pattern than the interruption ratio datasets of the AWS, with an occasionally high number of availability score updates.

5 CONCLUSION AND FUTURE WORK

We presented a publicly available spot instance dataset service that provides diverse information from AWS, Azure, and GCP in a single place. The developed system solved many technical challenges encountered during dataset collection, storage, and serving. We believe this service can be used for further analysis and researches in the domain of optimizing the cost efficiency and reliability of using cloud spot instances.

The proposed initial work has scope for improvement. We are working on building a model to predict instance interruptions using historical spot instance datasets. It is challenging to choose proper instance types for a cost and performance optimal environment in a multi-cloud due to the large search space. Thus, we are currently working on an algorithm for spot instance recommendation across multiple cloud vendors using the proposed dataset.

ACKNOWLEDGMENTS

We would like to thank Chaelim Heo, Jeongmyeong Park, and Soohyeok Lee who helped implementing web-crawlers. This work is supported by the National Research Foundation (NRF) Grant funded by the Korean Government (MSIP) (No. NRF-2022R1A5A7000765 and NRF-2020R1A2C1102544), AWS Cloud Credits for Research program, and the SW Star Lab (RS-2022-00144309) of IITP.

REFERENCES

[1] Orna Agmon Ben-Yehuda, Muli Ben-Yehuda, Assaf Schuster, and Dan Tsafrir. 2013. Deconstructing Amazon EC2 Spot Instance Pricing. *ACM Trans. Econ. Comput.* 1, 3, Article 16 (sep 2013), 20 pages. https://doi.org/10.1145/2509413.2509416

[2] Sarah Alkharif, Kyungyong Lee, and Hyeokman Kim. 2018. Time-Series Analysis for Price Prediction of Opportunistic Cloud Computing Resources. In *Proceedings of the 7th International Conference on Emerging Databases*, Wookey Lee, Wonik Choi, Sungwon Jung, and Min Song (Eds.). Springer Singapore, Singapore, 221–229.

[3] Matt Baughman, Simon Caton, Christian Haas, Ryan Chard, Rich Wolski, Ian Foster, and Kyle Chard. 2019. Deconstructing the 2017 Changes to AWS Spot Market Pricing. In *Proceedings of the 10th Workshop on Scientific Cloud Computing* (Phoenix, AZ, USA) *(ScienceCloud '19)*. Association for Computing Machinery, New York, NY, USA, 19–26. https://doi.org/10.1145/3322795.3331465

[4] Amazon Web Services Blogs. 2017. New Amazon EC2 Spot pricing model. https://aws.amazon.com/blogs/compute/new-amazon-ec2-spot-pricing/.

[5] David Irwin, Prashant Shenoy, Pradeep Ambati, Prateek Sharma, Supreeth Shastri, and Ahmed Ali-Eldin. 2019. The Price Is (Not) Right: Reflections on Pricing for Transient Cloud Servers. In *2019 28th International Conference on Computer Communication and Networks (ICCCN)*. 1–9. https://doi.org/10.1109/ICCCN.2019.8846933

[6] Cinar Kilcioglu, Justin M. Rao, Aadharsh Kannan, and R. Preston McAfee. 2017. Usage Patterns and the Economics of the Public Cloud. In *Proceedings of the 26th International Conference on World Wide Web* (Perth, Australia) *(WWW '17)*. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, CHE, 83–91. https://doi.org/10.1145/3038912.3052707

[7] K. Lee and M. Son. 2017. DeepSpotCloud: Leveraging Cross-Region GPU Spot Instances for Deep Learning. In *2017 IEEE 10th International Conference on Cloud Computing (CLOUD)*. 98–105. https://doi.org/10.1109/CLOUD.2017.21

[8] S. Lee, J. Hwang, and K. Lee. 2022. SpotLake: Diverse Spot Instance Dataset Archive Service. In *2022 IEEE International Symposium on Workload Characterization (IISWC)*. IEEE Computer Society, Los Alamitos, CA, USA, 242–255. https://doi.org/10.1109/IISWC55918.2022.00029

[9] Prateek Sharma, David Irwin, and Prashant Shenoy. 2016. How Not to Bid the Cloud. In *8th USENIX Workshop on Hot Topics in Cloud Computing (HotCloud 16)*. USENIX Association, Denver, CO. https://www.usenix.org/conference/hotcloud16/workshop-program/presentation/sharma