International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

AWS Spot Instances: A Cloud Computing Cost Investigation Across AWS Regions

Parul Dubey*[1], Dr. Arvind Kumar Tiwari [2]

Submitted: 15/08/2023 **Revised:** 05/10/2023 **Accepted:** 19/10/2023

Abstract: Renting virtual machines (VMs) on AWS offers cost-effective and scalable computing resources with global reach, allowing businesses to easily scale up or down based on their needs. With a wide range of VM types, robust security features, high reliability, and easy management tools, businesses can deploy VMs in multiple regions worldwide, ensuring optimal performance, security, and operational efficiency. Making a choice about these VM's is an important task. This research paper analysed the costs of 17 different instances across 17 different regions in the AWS cloud platform. The analysis considered instances from three different groups, and in all three groups, us-east-2 had lower costs compared to other regions. Specifically, in the general-purpose instances category, us-east-2a consistently offered lower prices than other regions throughout the considered time period, with some fluctuations for certain instances. The m7g.2xlarge instance, available in limited regions, was found to be offered at a minimum cost in us-east-2. Additionally, randomly selected instances from different categories, also showed that us-east-2 had lower costs compared to other regions for most instances. Although t2.large instances were not available in us-east-2a, they were found to be cheaper in us-west-2. This research suggests that organizations seeking cost-effective cloud computing solutions may benefit from selecting us-east-2 as their preferred AWS cloud platform region.

Keywords: *Amazon Web Services, Availability Zone (AZ), Elastic Compute Cloud, Regions, Spot Instances, Virtual Machines.*

1. Introduction

The Amazon Elastic Compute Cloud (EC2) is a web service that is hosted in the cloud and provides scalable computing capabilities. Its goal is to make it easier for programmers to do web-scale computing. Amazon EC2 permits "computing" in the cloud, while Amazon S3 enables "storage" in the cloud. Both of these capabilities are offered by Amazon. Because of the intuitive nature of Amazon EC2's web service interface, obtaining and configuring capacity on the platform requires just a minimal amount of effort. Due to this feature, users are able to take advantage of Amazon's tried-and-true computing environment while still retaining complete authority over their own computer assets. Because extra server instances can be procured and started up in a matter of minutes using Amazon EC2, customers are able to promptly scale up or decrease their current server capacity in response to changes in their computing requirements. The traditional approach to computing's cost-benefit analysis is revolutionized by the use of Amazon EC2.

We'll look at the work done by others to analyze EC2 instance pricing in the next portions of this research. In this article, we'll go through the many approaches to EC2 instance pricing that may be implemented inside an AWS account. After this, we'll examine spot instance price data from the last 60 days. The cost of renting spot instances will be analyzed with a focus on the role of different geographical locations.

2. Literature Review

Spot instances provide users the chance to bid on EC2's unused computing power at significantly reduced rates. The user's instance will be shut down if alternative bids for the available capacity are higher. One of the articles [1] presented one of the first comprehensive studies of the effect of geography on the total expenses of deploying Spot Instances. In the study, they examined Spot Instance price information from all AWS regions for the last 2 months. They analyse weekly price trends and compare them to the larger AWS region. Spot Instance prices vary widely among regions, suggesting the importance of geography in setting these rates.

In one of the articles [2] provided a broad overview of the research into spot instances by first analysing the evolution of pricing models for these instances, then outlining approaches that might increase their availability, and then explaining how to make the most of them. Spot instances may be purchased on a pay-as-you-go basis and are much less expensive than their on-demand counterparts of the

[1] *Dr C V Raman University, Bilaspur, Chattishgarh, India*
ORCID ID : 0000-0001-8903-6664
[2] *Dr C V Raman University, Bilaspur, Chattishgarh, India*
ORCID ID : 0009-0007-7916-2369
** Corresponding Author Email: dubeyparul29@gmail.com*

same specifications. Users that need a lot of computational power have taken notice of these low prices. The problem is that they can't promise that spot instances will always be available. This has led to much discussion on how to increase the number of available spot instances.

Researcher methodology was proposed for the investigation of the EC2 spot instance infrastructure that utilizes the pricing history of such resources to identify the availability zones for spot instances and then constructs price prediction models tailored to each class [3]. Extensive experimental procedures are used to verify the presented models. As a consequence of using these models, researchers can create resource provisioning strategies that make the most economical use of the Spot instance infrastructure. Finally, they addressed the recent adjustments Amazon has made to the Spot instance model, and how their approach may accommodate them.

Users may improve the method's dependability by pre-preparing bid prices and execution time based on an accurate estimate of spot pricing. For this, researchers [4] history to build a model for predicting the future price using a suggested modified gated recurrent unit and dropout technique. The results of the tests demonstrated that the suggested approach performs better and more accurately than more advanced methods.

Researchers [5] suggested using a variety of machine learning methods to predict the cost of EC2 spot instances in the future. Among them are the more common - multilayer perceptrons, K-nearest neighbours, additional trees, and random forests, as well as the more complex linear, ridge, and lasso regressions. Root-mean-squared errors vary from near zero in the cheapest examples to well over 60 in the costliest ones, indicating a wide variation in the acquired performance.

3. Problem Statement

AWS is expanding its infrastructure. With the increase in number of regions and AZ's, will there be increase or decrease in the pricing of spot instances. Is there any specific region where the instances are going to be cheaper than anywhere else. Spot instances are available in different regions in AWS. There are several instance types available under this offering in bidding strategy. The user is having choice in region as well, then which region should the user pick for a particular instance to get the best price on the spot instance.

4. Pricing Options with EC2 Instances

Billing per second guarantees that consumers only pay for the time they really use. Users that utilize EC2's per-second pricing model will not be charged for any time that they do not use. Instead of optimizing for hourly usage, concentrate on improving the programs, which is especially critical for instances that run for various lengths of time.

As per Amazon EC2 - Secure and Resizable Compute Capacity – Amazon Web Services, 2023 - EC2 billing is done in one-second increments, with a minimum fee of one minute. Amazon Elastic Block Store (EBS) volumes are likewise billed in 60-second increments per second. For on-demand, savings plan, reserved, and spot instances that users install, customers may opt to be paid by the second. All AZs and regions provide Amazon Linux, Windows, and Ubuntu. Figure 1 shows various pricing options available for EC2 instances in AWS.

4.1. On-Demand

The expenses of using on-demand computing are often invoiced either by the hour or by the second; however, this might vary depending on the kind of instance that is used. There are no short-term commitments or one-time fees associated with this transaction. Users have the ability to scale up or down the computing power according to the requirements of the application, and customers will only be charged the hourly rates for the instances that they actually utilize. It is suggested that we use an on-demand instance when:

- Customers that value Amazon EC2's cheap prices and adaptability but who don't want to commit to a long-term contract or pay anything up front
- Interruption-sensitive applications with transient, peak, or unexpected loads
- Newly created or tested applications on Amazon EC2

4.2. Spot Instances

With Spot Instances, users may save up to 90 percent off the price of On-Demand pricing by requesting unused computing capacity from Amazon EC2.It is recommended to use spot instances when:

- Applications that allow for the setting of a time frame
- Some applications can only be made financially feasible with very low computing costs.
- Users that have an urgent need for a significant amount of additional computing capacity

4.3. Savings Plans

Savings Plans provide discounted prices for Amazon EC2 and Fargate in exchange for a commitment to a consistent quantity of consumption (measured in $/hour) over a period of one or three years.

4.4. Dedicated Host

Users have the option to hire out a "Dedicated Host," which is a genuine EC2 server, for use on an exclusive basis. Any current server-bound software licences (depending on the

conditions of the licence), may be utilised on a dedicated host, which will save the company money and make it simpler to comply with laws. Examples of such software include:

- Available for purchase on an as-needed basis (hourly).
- Users are able to save up to 70 percent off the standard, On-Demand price by making reservations.

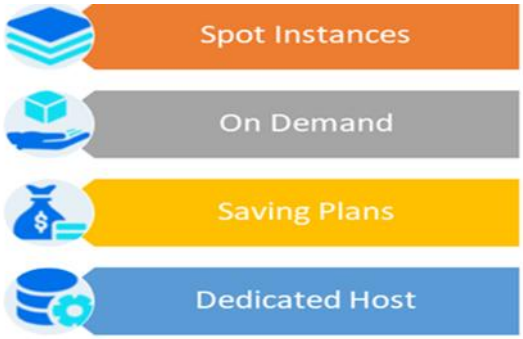

Fig. 1. EC2 Pricing options.

5. AWS Expansion

Amazon has data centres in a great number of places all around the world. These regions are comprised of Amazon Web Services (AWS) Regions, Availability Zones, and Local Zones. A number of Amazon Web Services (AWS) Regions are located in various regions of the globe. Within each AWS Region, there are several distinct physical locations known as Availability Zones.

As per AWS (Global Infrastructure, 2023), AWS Cloud now covers 96 Availability Zones across 30 geographic areas across the globe, and the company has announced plans for 15 more Availability Zones and 5 additional AWS Regions in the countries of Australia, Canada, Israel, New Zealand, and Thailand.

5.1. Regions

AWS offers the concept of a region, which can be thought of as a geographical location anywhere on the planet where a cluster of data centers is located. Each collection of logical data centers has been given the name of an Availability Zone by the company. Each Amazon Web Services (AWS) Region is made up of a minimum of three AZs that are geographically close to one another but remain physically isolated from one another. The numerous availability zones (AZ) architecture that is present in each AWS Region delivers advantages to customers, in contrast to the practices of many other cloud service providers that treat a region as a single data centre. Each AZ is connected to the others over low-latency, redundant networks, and each AZ also has its own source of power, cooling, and security staff. Customers of Amazon Web Services that are most concerned with high availability may choose to design their applications to run across a number of AZs in order to achieve an even higher level of fault-tolerance. AWS takes measures to guarantee that each of our regions adheres to all applicable regulations and offers the greatest level of data protection possible.

AWS has a more extensive global presence than any other cloud provider, and in order to maintain this global presence and ensure that all of its customers are served no matter where they are located, AWS is constantly expanding into new regions. Amazon Web Services (AWS) is comprised of a large number of regions located all over the globe. These locations may be found on a number of different continents as well as maritime regions.

Our facilities for this investigation are in India. As a result, we will only be able to analyze certain regions. Only 17 of AWS's currently accessible 30 regions are really operational in India. These regions can be seen in the plotted graphs.

5.2. Availability Zone

An Amazon Web Services (AWS) Availability Zone (AZ) is made up of one or more data centres located within a Region that are geographically independent from one another and share redundancy in their architecture. Users that make use of AZs are able to run production applications and databases that are more resistant to the occurrence of failure, more scalable, and more available than those programmes and databases that are operated out of a single data centre. A particular AWS Region's availability zones (AZs) are connected by fully redundant, dedicated metro fibre that not only provides high throughput but also minimal latency for users. Every data transmission that takes place between AZs is encrypted. Because of the network's appropriate performance, synchronous replication may take place across availability zones. AZs make it easier to split high-availability applications into many sections. Businesses are able to better withstand the effects of natural catastrophes and other types of interruptions by spreading their applications over a large number of AZs. Despite the fact that no two AZs are more than 100 kilometres (60 miles) apart from one another, the gap that separates them is rather large and consists of several kilometres.

6. EC2 Spot Best Practices

Spot Instances on Amazon EC2 are unused EC2 computing resources in the AWS Cloud that may be used for up to 90 percent less than the standard on-demand pricing (Best Practices for EC2 Spot - Amazon Elastic Compute Cloud, 2023). Spot instances are similar to on-demand instances, with the exception that Amazon EC2 may terminate them with a two-minute notice if it wants to reclaim the capacity.

Spot instances are suitable for applications that must be stateless, fault-tolerant, and extremely flexible. Spot Instances, for example, are optimized for certain sorts of workloads, such as huge data, containerized workloads, continuous integration and continuous deployment, stateless

web servers, HPC, and rendering.

Spot instances function in the same way as on-demand instances when they are enabled. Spot does not promise that active instances will be available for the full duration of the time it takes to complete the workloads. In addition, we cannot assume that Spot will always provide us with the aggregate capacity or instances we need instantly. Depending on supply and demand, spot instances may be unavailable at times or have varying capacities. Moreover, there is no assurance that I will continue to do well.

Workloads that are not flexible, stateful, fault-tolerant, or loosely connected between instance nodes are not a good fit for spot instances. They also shouldn't be used for workloads that can't survive for any length of time without the designated capacity. AWS does not advise utilizing spot instances for this kind of work or trying to fail over to on-demand instances in the event of an outage since they were not designed to handle such workloads. The EC2 best practices can be seen in Figure 2.

6.1. Make sure specific instances are ready for disruptions:

Building software that is fault-tolerant is the best way to deal with Spot Instance interruptions gracefully. This may be accomplished by using recommendations for rebalancing EC2 instances and spot instance interruption warnings.

When a spot instance is about to be terminated, customers will get a warning through a newly introduced signal called an EC2 instance rebalancing suggestion. Users may take preventative measures before the spot instance's two-minute interruption notice even begins. New or existing spot instances that are not at increased risk of interruption may be utilized to rebalance the workload. Capacity rebalancing in auto-scaling groups and Spot Fleet have been used to make this new signal accessible to everyone.

Amazon EC2 will provide a two-minute notice before shutting down a spot instance. If the workload is "time-flexible," users may set up Spot Instances to be halted or hibernated rather than terminated during an outage. In the event of an outage, Spot Instances on Amazon EC2 will be put on hold and resume running once capacity is restored.

Amazon EventBridge rules that record rebalancing suggestions and interruption warnings, then either save the current state of the workload at a checkpoint or gracefully deal with the interruption, come highly recommended by industry experts.

6.2. Be adaptable in terms of instances and AZ's:

Idle Amazon EC2 instances of the same Availability Zone and instance type make up a Spot capacity. The Availability Zones and instance types in which the workload is deployed should be adaptive. This increases Spot's chances of identifying and assigning the required amount of computer resources. Do not restrict oneself to asking c5.large when larges from other families, may suffice.

By analysing compute needs, users may evaluate how versatile a firm can be in terms of instance type. If the workload may be vertically extended, requests should include high configuration instance types with more vCPUs and RAM. Although older generation instance types are not as well-liked by On-Demand clients, they should be taken into account if horizontal expansion is required.

It is recommended that we have access to at least ten distinct instance types for every activity. It is also necessary to verify that the suitable Availability Zones for the workload have been selected and configured for use in the Virtual Private Cloud.

6.3. Utilize Auto Scaling groups or Spot Fleet:

Instead of addressing resources on an instance-by-instance basis, Spot allows users to do so by concentrating on the total capacity of the environment. This might be expressed as virtual CPUs, memory, space, or available bandwidth. Spot Fleet and Auto Scaling Groups enable users to deploy and maintain a target capacity. This is done automatically, requesting replacement resources in the event of an interruption or the manual cancellation of an earlier request. All users need to do to set up an auto-scaling group or a spot fleet is specifying what instance types and target capacity the application needs.

6.4. Utilize price and capacity optimum allocation technique:

By using allocation strategies in auto-scaling groups, one may supply the desired capacity without having to actively seek out spot capacity pools with available resources. The price-capacity-optimized technique is highly recommended, as it will automatically supply instances from the most cost-effective and

readily available spot capacity pools. Spot Fleet also provides an optimal allocation approach that takes into account both costs and available resources. The likelihood of having the spot instances repossessed is reduced if that capacity is obtained from pools that are operating at full capacity.

6.5. Rebalance the capacity proactively

Capacity Rebalancing lets users to keep their workloads online even during peak demand by adding a new Spot Instance to the fleet before a current Spot Instance receives the two-minute Spot Instance interruption notice. When Capacity Rebalancing and Auto Scaling are activated, or when Spot Fleet replaces Spot Instances that have obtained a rebalancing recommendation, the workload may be rebalanced to new Spot Instances that are not at greater risk of interruption.

It is an add-on to both the mixed-instances policy and the price-capacity-optimized allocation approach, both of which aim to locate the most efficient reserves.

6.6. Manage spot instances with ease using built-in AWS tools.

Spot is integrated with other AWS services to reduce computing expenses without the need to maintain individual instances or fleets. AWS recommends the following services for related workloads: Elastic MapReduce, Elastic Container Service, Elastic Kubernetes Service, Amazon SageMaker, Amazon Elastic Beanstalk, Amazon GameLift, and the other AWS services.

Fig. 2. EC2 Spot best practices

7. Dataset Description

For this research we have taken data for 17 different types of EC2 spot instances. They are listed in table 1. The configuration for the instances can be seen in table 1. Details about the count of the instance in all regions and the OS based count can be seen in table 2. The dataset includes total 664584 entries which includes the price for spot instances shown in table 2. The time period for which these data are collected are from Dec22 ,2022 to Feb 21 ,2023. They are grouped into three categories group 1,2 and 3 respectively. The counts of them can be found in table 2.

Table 1. Details of instances used in the analysis

SI.No	Instance Name	Instance No.	Type	vCPU	Mem (GiB)	Count of Instance
1	c5.4xlarge	I1	Compute Optimized	4	8	39003
2	m7g.2xlarge	I2	General Purpose	8	32	980
3	r7g.xlarge	I3	Memory Optimized	8	64	758
4	i2.2xlarge	I4	Storage Optimized	8	61	18864
5	i3.xlarge	I5	Storage Optimized	4	30.5	45270
6	c4.large	I6	Compute Optimized	2	3.75	35549
7	c4.xlarge	I7	Compute Optimized	4	7.5	39206
8	c4.2xlarge	I8	Compute Optimized	8	15	48299
9	c4.4xlarge	I9	Compute Optimized	16	30	45368
10	m4.large	I10	General Purpose	2	8	35078
11	m4.xlarge	I11	General Purpose	4	16	74876
12	m4.2xlarge	I12	General Purpose	8	32	53234
13	m5.large	I13	General Purpose	2	8	41938

14	m5.xlarge	I14	General Purpose	4	16	37025
15	m5.2xlarge	I15	General Purpose	8	32	56203
16	m5.4xlarge	I16	General Purpose	16	64	75704
17	t2.large	I17	General Purpose	2	36	17229

Table 2. Count of instances of different OS in different groups

Si. No	Group Number	Instance No.	UNIX/Linux Count	Windows Count
1	Group 1	I6	35549	5375
		I9	13182	5801
		I8	14207	5654
		I7	11177	5651
		I1	10759	6691
		Total	84874	29172
2	Group 2	I17	4774	2896
		I2	325	0
		I14	9955	7163
		I13	11225	6601
		I16	22765	7383
		I15	16295	7291
		Total	65339	31334
2	Group 3	I11	21257	11053
		I10	9881	5411
		I12	15835	5705
		I4	4825	4371
		I5	12705	7127
		I3	251	0
		Total	64754	33667

8. Result Analysis and Discussion

AWS has been rapidly expanding its infrastructure across the globe in recent years to better serve its customers and meet the increasing demand for cloud services. The company has added several new regions and availability zones to its global network, which has helped AWS to improve performance, reduce latency, and enhance reliability for its customers.

Over the past few years, AWS has launched new regions in key markets such as Asia Pacific, the Middle East, Africa, Europe, and South America. These regions include Hong Kong, Bahrain, Cape Town, Milan, Osaka, and São Paulo State. AWS has also announced plans to launch additional regions in Jakarta, Zurich, and Madrid.

The expansion of AWS regions has provided significant benefits to customers by bringing AWS services closer to their end-users, enabling them to store their data locally, and ensuring compliance with local data protection laws. In the studies conducted earlier, these regions do not exist. Our analysis will expand over a greater number of regions as compare to previous studies and will also include a greater number of instances for a better and generalized conclusion.

We have analysed the price for the 17 different instances shown in table 1. Two different categories are made on the basis of the operating system of the instance, namely Linux/UNIX and Windows. Out of the total data extracted for the past 60 days, we are using 214967 Linux/UNIX instances and 94173 Windows instances.

The instance will then be analysed for the prices on all 60 days. The prices for the instance vary over time as well as over region of selection.

8.1. Linux/UNIX

Linux/UNIX spot instances on Amazon Web Services (AWS) are a cost-effective way of running workloads that can tolerate interruptions or have flexible start and end times. Spot instances are spare computing capacity offered by AWS at a significantly less price when compared to on-demand instances. These instances are ideal for workloads that are fault-tolerant, such as batch jobs, data analysis, and machine learning workloads, as they can be interrupted when the spot price exceeds the bid price, also when AWS needs the capacity back.

Linux/UNIX spot instances are instances that operate on the Linux or UNIX operating systems. The difference in cost between these and on-demand instances is reflected in their names. A spot instance's cost may be adjusted in response to fluctuations in the market value of available computing resources hosted in Amazon Web Services' cloud. If a user places a bid that is greater than the current spot price for the instance capacity they need, they will be able to utilize the capacity until their offer is outbid.

Spot instances may save our money, which is one of its many advantages. The price difference between spot and on-demand instances might be as high as 90%. As a result, they

are best suited

for non-urgent tasks that can withstand brief interruptions. Users should be aware that spot instances are not always available and may be temporarily unavailable.

Spot instances for Linux and UNIX may be requested through the AWS SDK or the AWS Management Console. Users must provide a bid price, AMI ID, and instance type. After the request is made, AWS compares the bid price to the current spot price and starts the instance if the bid is greater.

The spot instance prices are obtained in USD. We have further multiplied the value by 1000 to plot the graphs for difference. In the plot we have taken the dates in X axis and the price of the instance in Y axis. We found the individual graphs for different instances. In figure 3, we can see compute optimized instances US-east-2 shows less cost as compare to all other regions we have considered.

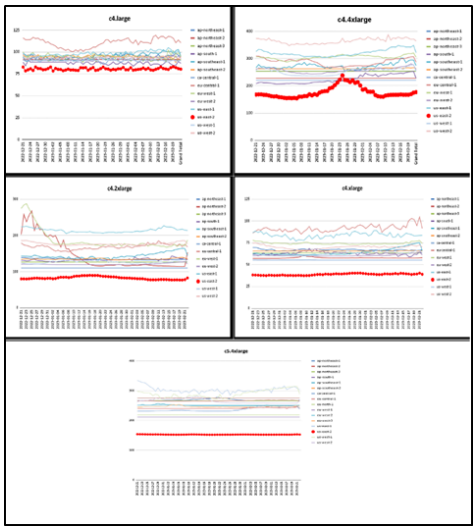

Fig. 3. UNIX/Linux Spot instance price plot for group-1

In another Category we have considered general purpose instances which includes six different instances. The details of the price analysis can be seen figure 4. Here we notice that us-east-2a shows less price than all other regions throughout the time period taken into consideration. few fluctuations with m5.4xlarge and m5.2xlarge for a short period of time. We can also see that m7g.2xlarge is found less and in few regions only, but again we can observe that us-east-2 is the one which offers this at a minimum cost.

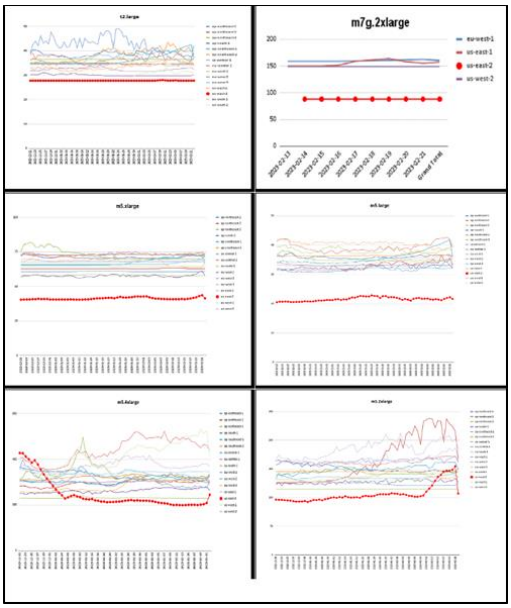

Fig. 4. UNIX/Linux Spot instance -price plot for group-2

We have further taken few instances randomly from different categories and found that us-east-2 is found to be cheaper in this as well. In this category we have included general purpose, memory optimized and storage optimized instances. The details of the analysis can be seen in figure 5. Other than i2.2xlarge all the instances shows that us-east-2 is cheaper throughout.

8.2. Windows Instances

Windows spot instances on Amazon Web Services (AWS) are a cost-effective way to run Windows-based workloads that can tolerate interruptions or have flexible start and end times. These instances are available at significantly reduced prices compared to on-demand instances, making them ideal for workloads that are fault-tolerant and can handle interruptions.

Similar to on-demand instances, Windows spot instances run on the Windows operating system and provide the same computing capacity as regular instances. However, the pricing model for spot instances is different. The price of a Windows spot instance is determined by the supply and demand of spare capacity in the AWS cloud, and users can bid for the instance capacity they require. If their bid is higher than the current spot price, they can use the capacity until their bid is exceeded.

Fig 5. UNIX/Linux Spot instance- price plot for group-3

One of the main advantages of using Windows spot instances is their cost-effectiveness, with potential savings of up to 90% compared to on-demand instances. This makes them ideal for workloads that are not time-critical and can tolerate interruptions. However, users should be aware that the availability of spot instances is not guaranteed, and they may be interrupted at any time if the spot price exceeds their bid.

To use Windows spot instances, users need to create a spot instance request in the AWS Management Console or through the AWS SDK. They need to specify the instance type, the Windows Amazon Machine Image (AMI) ID, and the bid price. Once the request is submitted, AWS matches the bid price with the current spot price, and if the bid is higher, the instance is launched, and users can start using it for their Windows-based workload.

Similar to Linux instances we have done the analysis for same groups of instances here as well. For the first group, taking compute instances into consideration we have obtained figure 6. In the figure we can clearly see that the observation which was made earlier for Linux instances same will hold for windows instances as well. Windows instances are found less in number and the cost for windows OS are higher as compare to Linux/UNIX. The cost remains same throughout the considered time line.

For the second group of same general purpose instance, we found that t2.large instances were not found in us-east-2a but they were found with cheapest price in us-west-2 . This can be seen in figure 7.

For the third group of instances we observed that i2.2xlarge instances shows large fluctuations in the prices in us-east-2 and they were found cheapest in eu-west-1 followed by us-west-1. Similarly, for i3.xlarge we can say that us-west-2 was cheapest and it was also observed that the instances were not found in us-east-2. We can see all the details in figure 8.

These findings highlight that us-east-2 consistently offers the most cost-effective options for a wide range of instances, including both

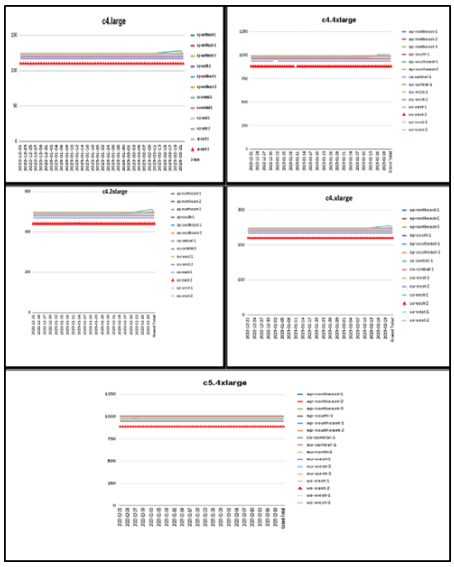

Fig 6. Windows Spot instance- price plot for group-1

Linux/UNIX and Windows instances. Organizations looking for affordable options for their cloud computing needs may benefit from considering us-east-2 as their preferred region in the AWS cloud platform.

9. Future Scope

The future scope of the analysis mentioned could be the following mentioned below.

1. Continual monitoring of the AWS infrastructure changes: As AWS continues to expand its services, regions, and availability zones, the analysis may need to be updated to account for these changes. Therefore, it is essential to continually monitor and analyze AWS's infrastructure to provide businesses with accurate cost-saving recommendations.

2. Optimization of cost savings: The analysis could be extended to

include other factors such as the size of the instance, the type of instance, and the duration of usage to optimize cost savings. As these factors change, it will be necessary to update the analysis to ensure that businesses are still achieving the best possible cost savings.

3. Comparison with other cloud providers: While the analysis focuses on AWS, it may be beneficial to compare

AWS with other cloud providers to determine which provider offers the most cost-effective solutions for businesses. This analysis could be extended to compare other cloud providers' infrastructure, regions, and pricing models to AWS.

4. Impact of region on performance: The analysis could be expanded to include an assessment of how the region's selection affects the performance of the instances. Businesses can use this analysis to choose the region that provides the best balance of cost

5. Analysis for different time periods: The analysis may need to be repeated periodically to reflect changes in the AWS infrastructure

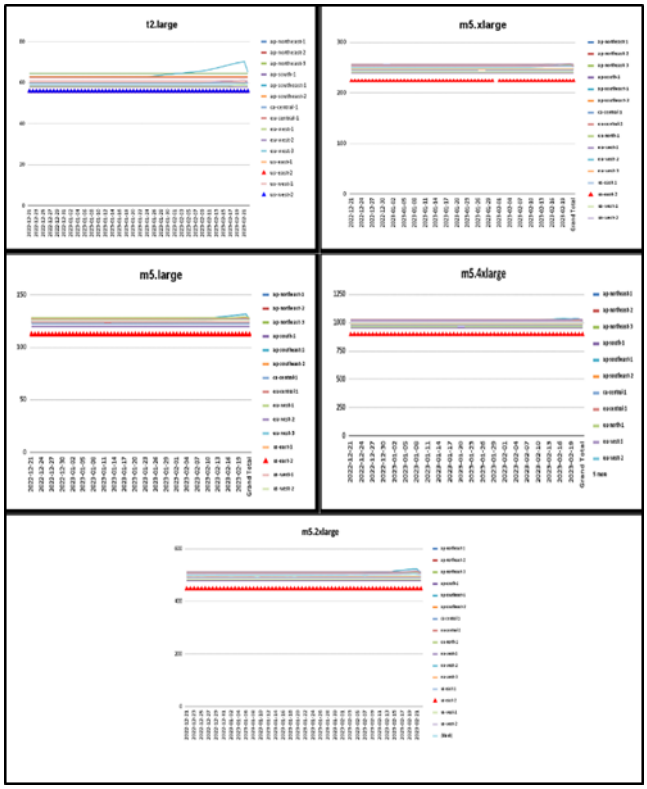

Fig 7. Windows Spot instance- price plot for group-2

and pricing models. By analyzing data from different time periods, businesses can gain a better understanding of how AWS pricing and infrastructure evolve over time and make informed decisions.

Overall, this analysis can be an excellent resource for businesses looking to expand their operations with AWS, and it's vital to keep monitoring changes in the infrastructure and pricing models to provide up-to-date and relevant recommendations about their cloud strategy.

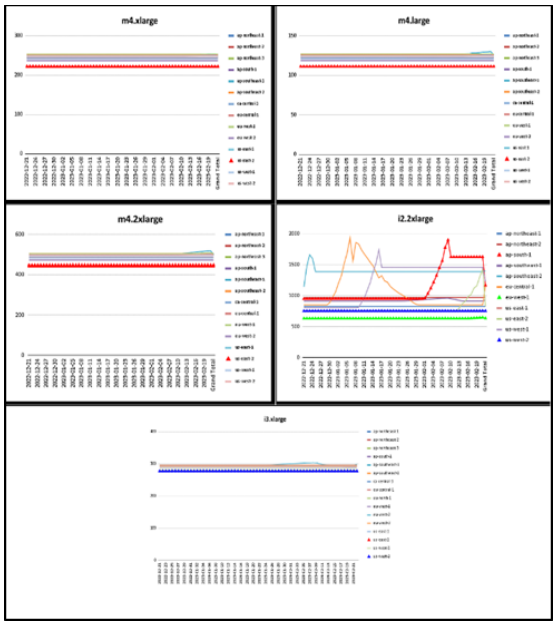

Fig 8. Windows Spot instance- price plot for group-3

10. Conclusion

The research analysed the costs of 17 different instances across 17 different regions in the AWS cloud platform. The findings consistently showed that us-east-2 had the cheapest instances for both Linux/UNIX and Windows instances. For instances of all three groups, us-east-2 was found to have lower costs compared to other regions. In the general-purpose instances category, us-east-2a consistently offered lower prices than other regions throughout the considered time period, with some fluctuations for certain instances. The m7g.2xlarge instance was found to be available in limited regions but was offered at a minimum cost in us-east-2. Randomly selected instances from different categories, including general purpose, memory optimized, and storage optimized, also showed that us-east-2 had lower costs compared to other regions for most instances. However, instances which were not available in us-east-2a but were cheapest in us-west-2. Organizations searching for cost-effective cloud computing solutions may benefit from selecting us-east-2 as their chosen AWS cloud platform region.

Author contributions

Parul Dubey: Data curation, Writing-Original draft preparation, Software, Validation, Field study

Dr Arvind Kumar Tiwari: Visualization, Investigation, Writing-Reviewing and Editing.

Conflicts of interest

There are no conflicts of interest among the authors.

References

[1] N. Ekwe-Ekwe and A. Barker, “Location, Location,

Location: Exploring Amazon EC2 Spot Instance Pricing Across Geographical Regions," 2018 18th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGRID), May 2018, Published, doi: 10.1109/ccgrid.2018.00059.

[2] L. Lin, L. Pan, and S. Liu, "Methods for improving the availability of spot instances: A survey," Computers in Industry, vol. 141, p. 103718, Oct. 2022, doi: 10.1016/j.compind.2022.103718.

[3] J. Fabra, J. Ezpeleta, and P. Álvarez, "Reducing the price of resource provisioning using EC2 spot instances with prediction models," Future Generation Computer Systems, vol. 96, pp. 348–367, Jul. 2019, doi: 10.1016/j.future.2019.01.025.

[4] S. S. Nezamdoust, M. A. Pourmina, and F. Razzazi, "Optimal prediction of cloud spot instance price utilizing deep learning," The Journal of Supercomputing, vol. 79, no. 7, pp. 7626–7647, Dec. 2022, doi: 10.1007/s11227-022-04970-x.

[5] A. Baldominos Gómez, Y. Saez, D. Quintana, and P. Isasi, "AWS PredSpot: Machine Learning for Predicting the Price of Spot Instances in AWS Cloud," International Journal of Interactive Multimedia and Artificial Intelligence, vol. 7, no. 3, p. 65, 2022, doi: 10.9781/ijimai.2022.02.003.

[6] "Amazon EC2 - Secure and resizable compute capacity – Amazon Web Services," Amazon Web Services, Inc. https://aws.amazon.com/ec2/pricing/

[7] "Global Infrastructure," Amazon Web Services, Inc. https://aws.amazon.com/about-aws/global-infrastructure/

[8] "Best practices for EC2 Spot - Amazon Elastic Compute Cloud," Best practices for EC2 Spot - Amazon Elastic Compute Cloud. https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/spot-best-practices.html