Predictive Analytics in Cloud Computing: An ARIMA Model Study on Performance Metrics

Vamsikrishna Bandari
University of south Australia
https://orcid.org/0000-0003-4185-3985

Bandari, V. (2021) Predictive Analytics in Cloud Computing: An ARIMA Model Study on Performance Metrics, Applied Research in Artificial Intelligence and Cloud Computing, 5(1), pp. 55-65

Abstract

Predictive analytics is a key aspect of cloud computing as it helps organizations to anticipate future events and take proactive measures to prevent issues before they occur. In this research, the goal was to perform an ARIMA (AutoRegressive Integrated Moving Average) model to predict cloud performance using various performance metrics. The study utilized ten different performance metrics, such as Response Time, Resource Utilization, Availability, Error Rate, Memory Usage, CPU Utilization, Disk I/O, Network Bandwidth and others to model cloud performance. The aim was to investigate the potential of ARIMA models to predict cloud performance by analyzing the impact of these different performance metrics on the model's accuracy. The study also used four performance criteria, namely LogL (Log Likelihood), AIC (Akaike Information Criterion), BIC (Bayesian Information Criterion), and HQ (Hannan-Quinn Criterion) to evaluate the performance of the ARIMA models. The results of the study showed that the ARIMA model (2,0) and (0,2) had the lowest AIC and BIC values among all the models considered. This indicated that these models were the most suitable for predicting cloud performance, as they had the lowest information loss compared to the other models. The results of the study provided evidence that ARIMA models can effectively predict cloud performance. This research highlights the importance of predictive analytics in cloud computing and the potential for ARIMA models to predict cloud performance. The findings have implications for organizations that rely on cloud computing. However, more research is needed in this area, as the study was limited to only ten performance metrics, and more extensive research is needed to validate the findings and to determine the best approach to predict cloud performance.

Keywords: ARIMA, Cloud computing, Cloud performance, Performance metrices

Introduction

Cloud Performance Monitoring is a process of observing, measuring, and analyzing the performance of cloud-based applications, services, and infrastructure. It involves collecting data on the performance of cloud resources such as compute, memory, storage, and network usage, as well as application performance metrics such as response times, availability, and throughput. This data is then analyzed to identify performance bottlenecks and optimize the cloud environment. Cloud Performance Monitoring also provides visibility into service-level agreements (SLAs) and can help identify any potential issues that may arise due to changes in the underlying infrastructure or application. By monitoring performance, organizations can ensure that their cloud-based services are performing optimally and meeting the needs of their customers.

Cloud performance monitoring is an important part of managing a cloud-based system. It helps to ensure that applications, services, and infrastructure are performing optimally and that any issues are quickly identified and resolved. Cloud performance monitoring can help to reduce downtime, improve reliability, and increase user satisfaction. The first step in cloud performance monitoring is to set up a monitoring system. This system should be able to collect data from all parts of the cloud-based system and provide real-time analysis of performance. This data can then be used to identify any potential issues and to develop strategies to improve performance.

Once the monitoring system is in place, it is important to regularly review the data and take action when necessary. This could involve making changes to the system configuration, adding additional resources, or even scaling back certain services. Regular monitoring can help to ensure that the system is always running optimally and that any issues are quickly identified and addressed.

Cloud performance monitoring can also be used to identify areas of the system that are underperforming. This can help to identify areas where additional resources or changes to the system configuration may be needed. This can help to ensure that the system is running as efficiently as possible and that any potential problems are identified and addressed quickly. Finally, cloud performance monitoring can be used to identify trends in the system. This can help to identify areas where performance is declining and identify potential causes. This can help to ensure that the system is able to meet the demands of its users and that any potential problems are quickly identified and addressed. Overall, cloud performance monitoring is an important part of managing a cloud-based system. It helps to ensure that applications, services, and infrastructure are performing optimally and that any issues are quickly identified and resolved. Regular monitoring and analysis can help to ensure that the system is always running optimally and that any potential problems are identified and addressed quickly.

Predictive analytics in cloud performance is the use of predictive models and algorithms to analyze data sets to predict future events and outcomes. By leveraging the power of the cloud, organizations can gain insights into the performance of their cloud infrastructure, applications, and services. This can help them to better understand their environment, identify potential issues, and plan for the future. Predictive analytics can also provide organizations with insights into the performance of their cloud-based applications and services, enabling them to optimize their operations and improve their performance.

Predictive analytics in cloud performance can be used to identify problems, predict future trends, and optimize performance. For example, an organization can use predictive analytics to identify potential performance issues in their cloud environment, such as latency or resource usage. They can also use predictive analytics to forecast future trends in their cloud-based applications and services, such as user adoption or usage patterns. By leveraging predictive analytics, organizations can gain insights into their cloud-based infrastructure and applications, enabling them to make better decisions and improve their performance.

Cloud performance prediction is an important part of cloud computing. It is used to predict the future performance of cloud systems, such as the amount of resources needed to run applications, the time it will take to complete tasks, and the cost of using the cloud. It is important to accurately predict cloud performance in order to ensure the success of cloud computing initiatives.

There are several methods for predicting cloud performance. One of the most common methods is to use a model-based approach. This approach uses statistical models to predict the performance of a cloud system. The models are based on historical data, such as usage patterns, resource utilization, and cost of services. The models can be used to identify trends in cloud performance and make predictions about future performance.

Another approach for predicting cloud performance is to use machine learning algorithms. Machine learning algorithms can be used to analyze large amounts of data and identify patterns in cloud performance. The algorithms can then be used to make predictions about the future performance of a cloud system. This approach is useful for predicting the performance of complex cloud systems, such as those used in large enterprises.

In addition to model-based and machine learning approaches, there are also methods for predicting cloud performance using simulation. Simulation can be used to create a virtual environment that mimics the real-world environment of a cloud system. This virtual environment can then be used to test different scenarios and predict the performance of the cloud system under various conditions. Simulation is useful for testing new cloud systems or for analyzing the performance of existing cloud systems.

Finally, there are also methods for predicting cloud performance using analytics. Analytics can be used to analyze the usage patterns of cloud systems and identify trends in performance. This can be used to make predictions about the future performance of a cloud system. Analytics can also be used to identify potential problems with a cloud system and suggest ways to improve its performance.

Cloud performance metrics

Cloud performance metrics are important indicators of the health and performance of cloud-based systems and services. These metrics help organizations measure the performance of their cloud-based applications, services, and infrastructure. Common cloud performance metrics include latency, throughput, availability, scalability, cost, and security. Latency measures the time it takes for an application to respond to a request, while throughput measures the rate at which data can be transferred. Availability measures the reliability of an application and scalability measures its ability to scale up or down based on demand. Cost measures the cost of running the application and security measures the security of the application. By monitoring these metrics, organizations can ensure that their cloud-based applications are performing efficiently and cost-effectively.

1. Response Time

Response time is an important factor in the performance of a cloud-based application or service. It is the amount of time it takes for the application or service to respond to a user request. Response time is often measured in milliseconds and can have a significant impact on the user experience. The response time of a cloud-based application or service is influenced by several factors. These include the type of hardware and software being used, the number of users accessing the application or service, the complexity of the user request, and the network latency between the user and the application or service. Response time can also be affected by the amount of traffic on the network, the number of requests that are being processed, and the amount of data that needs to be transferred.

In order to ensure optimal response time, cloud providers need to take into account all of these factors when designing and deploying their applications and services. This includes ensuring that the hardware and software being used is up-to-date and capable of handling the user requests. It also involves ensuring that the network latency is minimized and that the number of requests that are being processed is kept to a minimum. Additionally, cloud providers need to ensure that the data being transferred is kept to a minimum and that the traffic on the network is kept to a minimum.

It is essential for cloud providers to ensure that their applications and services are designed and deployed in a way that minimizes response time. This includes ensuring that the hardware and software being used is up-to-date, that the network latency is minimized, and that the number of requests that are being processed is kept to a minimum. Additionally, cloud providers need to ensure that the amount of data being transferred is kept to a minimum and that the traffic on the network is kept to a minimum. By taking these steps, cloud providers can ensure that their applications and services are able to provide a fast and reliable response time.

2. Resource Utilization

Resource utilization is a key factor in the successful deployment and operation of cloud-based applications and services. It is essential to understand the resource utilization of a cloud-based application or service in order to ensure that the application or service is running efficiently and that maximum performance is being achieved.

The amount of computing resources used by a cloud-based application or service depends on the type of application or service being deployed. For example, a web application may require more memory and CPU resources than a database application. Additionally, the amount of resources used can vary depending on the number of users accessing the application or service, the complexity of the application or service, and the amount of data being processed.

In order to ensure that the cloud-based application or service is running efficiently, it is important to monitor the resource utilization of the application or service. This can be done by using tools such as performance monitoring, which can provide real-time data on the amount of resources being used. This data can then be used to identify potential bottlenecks and to optimize the application or service in order to improve performance.

Additionally, it is important to consider the scalability of the application or service when determining the amount of resources to allocate. Scalability refers to the ability of the application or service to scale up or down depending on the number of users or the amount of data being processed. By ensuring that the application or service is able to scale up or down as needed, it can be ensured that the resources are being used efficiently and that maximum performance is being achieved. It is also important to consider the cost of the resources being used by the cloud-based application or service. By understanding the cost of the resources being used, it can be ensured that the application or service is running efficiently and that the cost of the resources is being minimized. Additionally, by understanding the cost of the resources being used, it can be determined if the application or service is providing a cost-effective solution.

3. Availability

Availability is an important metric when it comes to cloud-based applications and services. It is the percentage of time that the application or service is available and accessible to users. A high availability rate is essential for businesses that rely on cloud-based services for their operations, as any downtime can be costly in terms of lost productivity and revenue.

The availability of a cloud-based service is typically measured by the uptime percentage. This is the amount of time that the service is running and accessible to users, and is typically expressed as a percentage of total time. For example, a service with 99.9% uptime would be available 99.9% of the time, with only 0.1% of the time being unavailable.

To ensure a high level of availability, cloud-based services must use redundant systems and data centers. This means that if one system or data center fails, another can take over and keep the service running. This ensures that the service is always available, even in the event of a system failure.

Availability is also affected by the type of cloud service being used. For example, Infrastructure as a Service (IaaS) typically has a higher availability rate than Platform as a Service (PaaS). This is because IaaS offers more control over the underlying infrastructure, allowing for more redundancy and failover options. Finally, availability is also affected by the provider of the cloud-based service. Different providers have different levels of uptime and reliability, so it is important to choose a provider that can offer a high level of availability. This can be done by researching the provider's service level agreements and performance history.

Availability is an important metric when it comes to cloud-based applications and services. It is important to ensure a high level of availability by using redundant systems, data centers, and choosing a reliable provider. This will ensure that the service is always available and accessible to users, minimizing downtime and ensuring the success of the business.

4. Throughput

Throughput is usually measured in bits per second (bps) or megabits per second (Mbps). Throughput is often used to measure the performance of an application or a network connection. Throughput is an important metric when considering cloud computing performance. It is the rate at which data is transferred over a network in a cloud environment. It is measured in bits per second (bps) and is a measure of how much data can be sent over a given period of time. A higher throughput rate means that more data can be sent more quickly, while a lower throughput rate means that less data can be sent more slowly.

In a cloud computing environment, the throughput rate is largely dependent on the bandwidth of the network and the number of users accessing the network. A higher bandwidth will result in a higher throughput rate, while a lower bandwidth will result in a lower throughput rate. Additionally, the number of users accessing the network will also affect the throughput rate, as more users will require more bandwidth, which will reduce the available bandwidth for each user.

Throughput should be considered when designing a cloud computing system. It is important to ensure that the system is able to handle the expected load, or else the system may become overwhelmed and unable to process requests. Additionally, it is important to ensure that the system is able to handle unexpected spikes in traffic, as this could cause the system to become unresponsive.

Paragraph 4: Additionally, throughput is also important when considering cost. A higher throughput rate will require more bandwidth, which will result in higher costs. Therefore, it is important to consider both the expected and unexpected traffic when designing a cloud computing system, as this will help to ensure that the system is able to handle the expected load and that the costs are kept to a minimum.

It is important to ensure that the system is able to handle the expected load, as well as unexpected spikes in traffic. Additionally, it is important to consider the cost of the system, as a higher throughput rate will require more bandwidth, which will result in higher costs. By carefully considering throughput when designing a cloud computing system, it is possible to ensure that the system is able to handle the expected load and that the costs are kept to a minimum.

5. Latency

Latency determines how quickly data can be transmitted over a network. The amount of latency experienced in a cloud environment is determined by the speed of the connection, the distance between the user and the cloud, and the amount of traffic on the network. In a cloud environment, latency is measured in milliseconds, which is the time it takes for data to travel between the user and the cloud. Latency can be affected by a variety of factors, including the speed of the connection, the amount of traffic on the network, and the distance between the user and the cloud.

The latency of a cloud environment can have a significant impact on the user experience. For instance, if the latency is too high, it can cause delays in the loading of webpages, or slow down the performance of applications. In addition, high latency can also affect the quality of video or audio streaming, as well as the performance of online gaming.

In order to reduce latency, cloud providers often use a variety of techniques, such as caching and content delivery networks. Caching allows data to be stored closer to the user, reducing the amount of time it takes for the data to travel over the network. Content delivery networks help to distribute the data more efficiently, so that it can be accessed more quickly. Finally, cloud providers may also use network optimization techniques, such as route optimization and bandwidth management, to improve the performance of their networks and reduce latency. By optimizing their networks, cloud providers can ensure that their users experience the highest quality of service.

6. Error Rate

Error rate is used to measure the performance of a cloud-based application or service. It is the percentage of requests to the cloud-based application or service that result in an error. This metric is important because it can help identify areas of the application or service that may need improvement, as well as indicate potential issues with the underlying infrastructure.

Error rate can be measured in different ways, depending on the type of application or service. For example, for web applications, the error rate can be calculated by the number of failed requests divided by the total number of requests. This metric can also be used to measure the performance of a database or storage service, by looking at the number of failed queries or failed storage requests.

Error rate can be affected by a number of factors, including the type of application or service being used, the underlying infrastructure, and the amount of traffic being sent to the application or service. It is important to monitor the error rate to ensure that the application or service is performing optimally and that any potential issues are identified and addressed.

There are a number of techniques that can be used to reduce the error rate of a cloud-based application or service. Techniques such as caching, load balancing, and optimization can help reduce the strain on the underlying infrastructure, resulting in fewer errors. Additionally, monitoring the application or service regularly and responding quickly to any errors can help reduce the overall error rate.

Error rate is for measuring the performance of a cloud-based application or service. Monitoring the error rate closely can help identify potential issues and ensure that the application or service is performing optimally. By using techniques such as caching, load balancing, and optimization, as well as responding quickly to any errors, the error rate can be reduced and the overall performance of the application or service improved.

7. Memory Usage

The amount of memory used by a cloud-based application or service can have a significant impact on performance, scalability, and cost. Memory usage is determined by the number of users accessing the application or service, the complexity of the tasks being performed, and the amount of data being stored and processed.

Memory usage is typically measured in terms of RAM (Random Access Memory) or virtual memory. RAM is a type of computer memory that stores data and instructions for quick access by the processor. Virtual memory is a type of memory that is created by the operating system and stored on the hard drive. It is used to extend the amount of RAM available to the processor.

When considering a cloud-based application or service, it is important to understand the amount of memory that will be used. This will help to determine the cost of the service, as well as the performance of the application or service. It is also important to consider the scalability of the application or service. If the memory usage is too high, the application or service may not be able to scale to accommodate more users or data.

Memory usage can be monitored and managed in order to ensure that the application or service is performing optimally. Memory usage can be monitored in real-time or periodically, depending on the application or service. By monitoring memory usage, it is possible to identify areas where memory usage can be reduced or optimized. This can help to reduce the cost of the service and improve the performance of the application or service.

Understanding the amount of memory used by a cloud-based application or service can help to ensure that the application or service is running optimally and cost-effectively. By monitoring and managing memory usage, it is possible to improve the performance, scalability, and cost of the application or service.

8. CPU Utilization

It is the amount of CPU resources being used by the application or service and can have a significant impact on the performance of the application or service. CPU utilization is typically measured in terms of the percentage of CPU resources being used by the application or service.

High CPU utilization can be indicative of a resource-intensive application or service and can lead to increased latency and decreased performance of the application or service. It is important to monitor CPU utilization to ensure that the application or service is running efficiently and is not being over-utilized.

There are several ways to reduce CPU utilization. One way is to optimize the application or service for better performance and resource utilization. This can include reducing the number of requests being made to the application or service and optimizing the code for better efficiency. Additionally, it is important to ensure that the application or service is not running unnecessary tasks or processes that can lead to increased CPU utilization.

Another way to reduce CPU utilization is to scale the application or service. Scaling the application or service can allow for more efficient resource utilization and can help to reduce CPU utilization. Additionally, it is important to ensure that the application or service is running on the most appropriate hardware and is not overburdened with unnecessary resources or tasks.

In addition to reducing CPU utilization, it is important to ensure that the application or service is running on the most appropriate hardware and is not overburdened with unnecessary resources or tasks. Additionally, it is important to ensure that the application or service is running efficiently and is not being over-utilized. Monitoring CPU utilization is an important part of ensuring that the application or service is running efficiently and is not being over-utilized.

9. Disk I/O

Disk I/O refers to the rate at which data is read from or written to disk storage. Disk I/O is especially important for applications that require frequent reads and writes of large data sets. In order to ensure optimal performance, disk I/O should be monitored and managed carefully.

There are a few different ways to measure disk I/O in a cloud environment. The most common metrics used to measure disk I/O include latency, throughput, and IOPS (input/output operations per second). Latency is the amount of time it takes for a disk to respond to a request. Throughput is the amount of data that can be read or written in a given amount of time. IOPS is the number of I/O operations that can be completed in a given amount of time.

Improving disk I/O performance in a cloud environment can be accomplished by selecting the right type of storage for the application. Solid-state drives (SSDs) offer the best performance due to their low latency and high throughput. However, they can be more expensive than traditional hard disk drives (HDDs). It's important to consider the cost of storage when selecting the right type of disk for the application.

Another way to improve disk I/O performance is to use caching. Caching stores frequently used data in memory, which can reduce the amount of time it takes to read and write data to the disk. Caching can also reduce the amount of I/O operations that need to be performed.

It is important to monitor disk I/O performance in order to ensure that the system is running optimally. Monitoring disk I/O can help identify potential bottlenecks and enable administrators to take steps to improve performance. There are a variety of tools available to help monitor disk I/O performance in a cloud environment.

10. Network Bandwidth

It is the measure of how much data can be transferred over a network in a given period of time. Network bandwidth affects the speed and reliability of an application or service, as well as its ability to handle large amounts of data.

When a cloud-based application or service is using more network bandwidth than is available, it can cause slow response times, high latency, and other performance issues. The amount of network bandwidth being used by a cloud-based application or service can vary depending on the type of application or service, the number of users, and the amount of data being transferred.

To ensure the best performance of a cloud-based application or service, it is important to monitor the amount of network bandwidth being used. This can be done by measuring the amount of data transferred over the network in a given period of time. If the amount of network bandwidth being used is too high, it can be reduced by optimizing the application or service, or by increasing the network bandwidth available.

Network bandwidth can also be increased by implementing technologies such as caching and compression. Caching is a technique that stores frequently used data in memory so that it can be accessed quickly, while compression reduces the amount of data that needs to be transferred over the network. These techniques can help to reduce the amount of network bandwidth being used by a cloud-based application or service. It is important to monitor the amount of network bandwidth being used to ensure the best performance of the application or service. Technologies such as caching and compression can also be used to reduce the amount of network bandwidth being used, and to increase the performance of the application or service.

Method

ARIMA (Auto Regressive Integrated Moving Average) is a statistical model used to predict future values of a time series based on past values. It can be used to predict various cloud performance metrics such as CPU utilization, memory utilization, and network traffic. ARIMA models use past values of the time series to predict future values. This is done by fitting the data to an ARIMA model and then using the model to make predictions. The model takes into account the correlation between past values and the current value, as well as any trends or seasonality in the data. By analyzing the data in this way, ARIMA can accurately predict future values of the time series.

ARIMA models are based on the assumption that the current value of a time series can be predicted using a combination of its past values and current values of other time series. The model consists of three components: an autoregressive (AR) component, an integrated (I) component, and a moving average (MA) component. The AR component models the effects of past values of the time series on the current value, while the I component captures any non-stationarity in the data. The MA component captures any short-term fluctuations in the data.

The parameters of an ARIMA model are estimated using a variety of techniques, such as maximum likelihood estimation or the Box-Jenkins method. Once the parameters have been estimated, the model can be used to make predictions about future values of the time series. ARIMA models can also be used to detect outliers and other anomalies in the data. ARIMA models are widely used in many areas, such as economics, finance, and marketing, and are an important tool for forecasting and analyzing time series data.

ARIMA can be used to predict cloud performance metrics such as CPU utilization, memory utilization, and network traffic. By analyzing the data and fitting it to an ARIMA model, the model can be used to make accurate predictions of future values. This can be used to improve the performance of cloud services by predicting when usage is likely to be high or low, and when resources need to be allocated accordingly. ARIMA can also be used to forecast future trends in the data, allowing cloud providers to plan for future capacity needs.

Results

Table 1 provides a comparison of the actual and forested performances of cloud computing for 100 samples. Figure 2 and 3 show the comparison of ARIMA models to predict the cloud performance for top models. The ARIMA model selection table presents various ARIMA models and their performance criteria, including LogL, AIC, BIC, and HQ. These criteria are used to evaluate the goodness of fit of the models and to determine which model is the best fit for the data. The LogL value in the table represents the log likelihood of the model.

The log likelihood measures the likelihood that the model's parameters would have produced the observed data. A higher LogL value indicates a better fit of the model to the data. In this table, the ARIMA model (2,3) has the highest LogL value of 1237.448, meaning that this model provides the best fit according to the log likelihood criterion.

The AIC and BIC values in the table represent the Akaike Information Criterion and Bayesian Information Criterion, respectively. These criteria balance the fit of the model to the data and the complexity of the model. A lower AIC and BIC value indicates a better fit of the model to the data. From the table, it can be seen that the ARIMA models (2,0) and (0,2) have the lowest AIC and BIC values, meaning that these models are preferred over others as they have the best fit to the data according to these information criteria

The HQ value in the table represents the Hannan-Quinn Criterion. This criterion is similar to AIC and BIC, but it places more weight on the complexity of the model. A lower HQ value indicates a better fit of the model to the data. From the table, it can be seen that the ARIMA models (2,0), (0,2), and (1,0) have the lowest HQ values. All four criteria should be considered to make the final decision on the best ARIMA model. While the ARIMA models (2,0) and (0,2) have the lowest AIC and BIC values, and the lowest HQ values, it is recommended to also consider the LogL value to ensure that the best model is selected. A balance between the model's fit to the data and its complexity must be achieved in order to select the most appropriate ARIMA model for the data.

Figure 1.

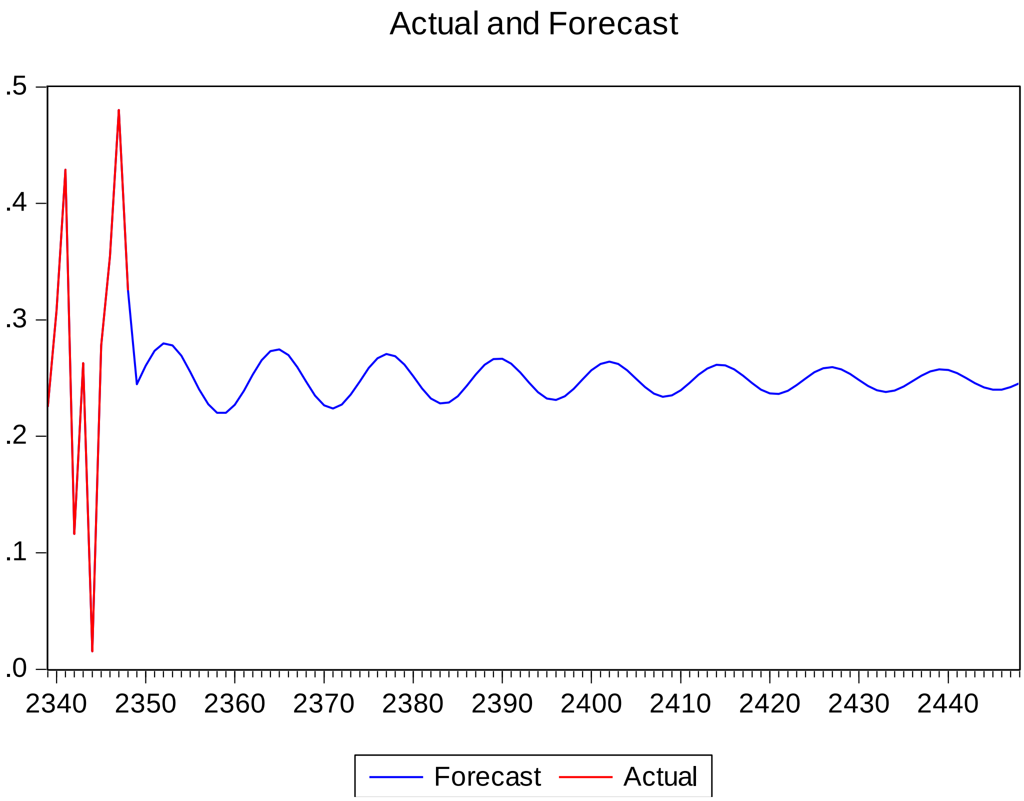

Actual and Forecast
.5
.4
.3
.2
.1
.0
2340
2350
2360
2370
2380
2390
2400
2410
2420
2430
2440
Forecast
Actual

Table 1. ARIM model criteria

Model	LogL	AIC*	BIC	HQ
(2,3)	1237.448	-1.04808	-1.03091	-1.04183
(0,0)	1229.99	-1.04599	-1.04108	-1.0442
(1,0)	1229.99	-1.04514	-1.03778	-1.04246
(0,1)	1229.99	-1.04514	-1.03778	-1.04246
(3,3)	1234.492	-1.04471	-1.02508	-1.03756
(2,0)	1229.99	-1.04429	-1.03447	-1.04071
(0,2)	1229.99	-1.04429	-1.03447	-1.04071
(1,1)	1229.99	-1.04428	-1.03447	-1.04071
(4,0)	1231.768	-1.0441	-1.02937	-1.03873
(0,4)	1231.766	-1.04409	-1.02937	-1.03873
(2,4)	1233.174	-1.04359	-1.02396	-1.03644
(4,2)	1233.167	-1.04358	-1.02395	-1.03644
(1,4)	1232.104	-1.04353	-1.02635	-1.03727
(4,1)	1232.103	-1.04353	-1.02635	-1.03727
(0,3)	1230.025	-1.04346	-1.03119	-1.03899
(3,0)	1230.023	-1.04346	-1.03119	-1.03899
(1,2)	1229.99	-1.04343	-1.03116	-1.03897
(2,1)	1229.99	-1.04343	-1.03116	-1.03897
(3,2)	1231.827	-1.04329	-1.02612	-1.03704
(1,3)	1230.415	-1.04294	-1.02822	-1.03758
(3,1)	1230.398	-1.04293	-1.02821	-1.03757
(4,3)	1233.183	-1.04275	-1.02066	-1.0347
(3,4)	1233.182	-1.04275	-1.02066	-1.0347
(2,2)	1229.99	-1.04258	-1.02786	-1.03722
(4,4)	1233.183	-1.04189	-1.01736	-1.03296

Figure 2.

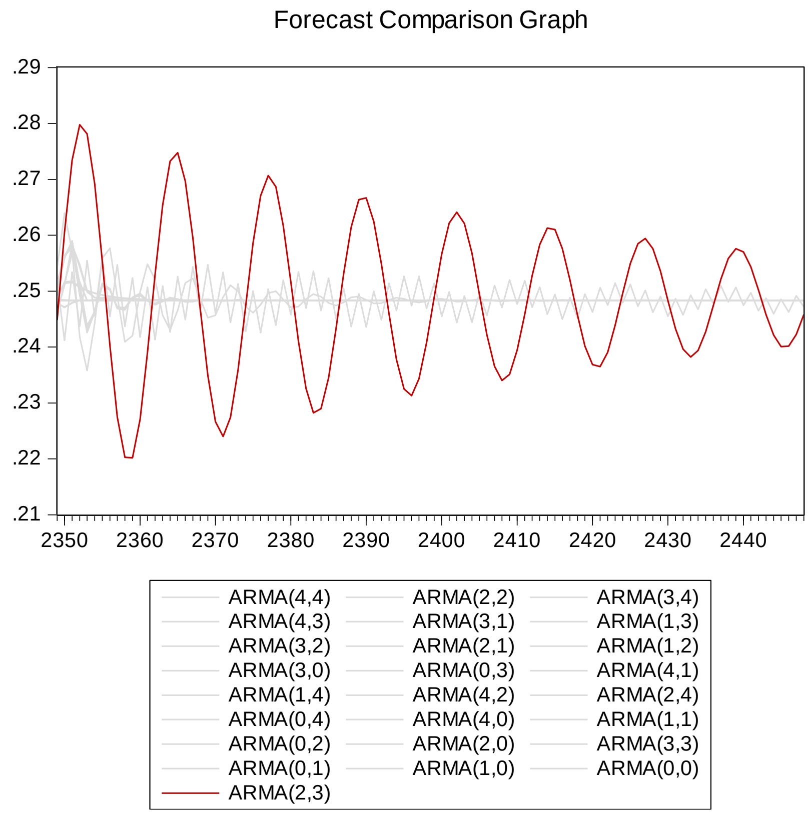
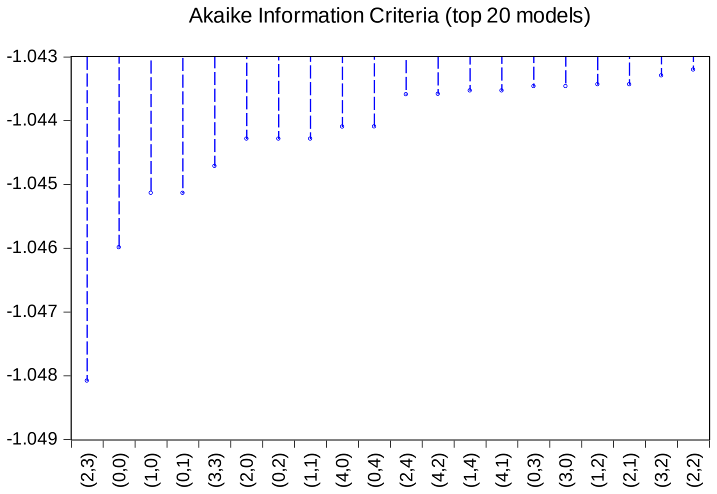


Figure 3.



Conclusion

Predicting cloud performance can be a challenging task for many organizations. Cloud computing is one of the most important technologies of the modern era, with its ability to provide on-demand computing resources and services. But predicting the performance of cloud-based applications and services can be difficult due to the number of variables and the dynamic nature of cloud environments.

One of the biggest challenges in predicting cloud performance is the lack of visibility into the underlying infrastructure. Many cloud providers operate on a "black box" model, meaning that the customer has no visibility into the underlying hardware or software that powers the cloud. This lack of visibility makes it difficult to accurately predict the performance of the cloud environment.

Another challenge in predicting cloud performance is the dynamic nature of cloud environments. Cloud providers are constantly changing and updating their infrastructure, which can affect the performance of applications and services running on the cloud. This makes it difficult to accurately predict the performance of applications and services in the cloud.

A third challenge in predicting cloud performance is the wide range of cloud services and applications available. Each cloud provider offers different services and applications, and each of these services and applications have different performance characteristics. This makes it difficult to accurately predict the performance of applications and services across different cloud providers.

Finally, predicting cloud performance can be complicated by the fact that cloud environments are often shared by multiple customers. This means that the performance of applications and services can be affected by the actions of other customers on the same cloud environment. This can make it difficult to accurately predict the performance of applications and services running on the cloud.

Predicting cloud performance can be a challenging task for many organizations. The lack of visibility into the underlying infrastructure, the dynamic nature of cloud environments, the wide range of cloud services and applications available, and the fact that cloud environments are often shared by multiple customers can all make it difficult to accurately predict the performance of applications and services running on the cloud.

The future of predicting cloud performance will be heavily reliant on the use of artificial intelligence (AI) and machine learning (ML). AI and ML are rapidly becoming the go-to methods for predicting cloud performance, as they are able to analyze vast amounts of data and generate accurate predictions. AI and ML can be used to identify patterns in cloud performance, and then use those patterns to accurately predict future performance. This type of predictive analysis can be used to identify potential issues and resolve them before they become major problems.

In the future, predictive analytics will be used to identify potential performance bottlenecks and other issues that could lead to degraded performance. AI and ML will be used to analyze historical data and identify trends that can be used to anticipate future performance issues. This type of predictive analysis will be invaluable in helping organizations proactively address potential issues before they become serious problems.

In addition, predictive analytics will be used to identify potential cost savings. AI and ML can be used to analyze usage patterns and identify areas where costs can be reduced. This type of analysis can help organizations optimize their cloud usage and reduce their overall cloud costs.

In the future, predictive analytics will also be used to identify potential security risks. AI and ML can be used to analyze cloud usage patterns and identify areas where security risks might be present. This type of analysis can help organizations stay ahead of potential security threats and take the necessary steps to mitigate them before they become serious issues. The future of predicting cloud performance will be heavily reliant on the use of AI and ML. This type of predictive analysis will be invaluable in helping organizations identify potential performance issues, cost savings, and security risks. By using predictive analytics, organizations will be able to stay ahead of potential problems and ensure their cloud performance remains optimal.

More research is needed in predicting cloud performance because cloud computing is a rapidly evolving technology with new innovations and services being introduced at a rapid pace. As such, it is difficult to accurately predict the performance of cloud-based applications and services. Additionally, the cloud environment is highly dynamic and can be affected by many external factors, such as network latency, resource availability, and other environmental variables. Thus, there is a need for more research to develop better models and algorithms to accurately predict the performance of cloud-based applications and services.

ARAIC-2021

References

[1] M. Yousif and L. Schubert, Eds., *Cloud computing*, 2013th ed. Cham, Switzerland: Springer International Publishing, 2013.

[2] V. C. M. Leung, R. X. Lai, M. Chen, and J. Wan, Eds., *Cloud computing*, 2015th ed. Basel, Switzerland: Springer International Publishing, 2015.

[3] N. K. Sehgal and P. C. P. Bhatt, *Cloud computing*, 1st ed. Basel, Switzerland: Springer International Publishing, 2018.

[4] “Investigating the Impacts of Cloud Computing on Firm Profitability,”

Reviews of Contemporary Business Analytics, vol. 2, no. 1, pp. 20–32, 2019.

[5] J. Ruiter and M. Warnier, “Privacy Regulations for Cloud Computing: Compliance and Implementation in Theory and Practice,” in *Computers, Privacy and Data Protection: an Element of Choice*, S. Gutwirth, Y. Poullet, P. De Hert, and R. Leenes, Eds. Dordrecht: Springer Netherlands, 2011, pp. 361–376.

[6] P. V. Garach and R. Thakkar, “A survey on FOG computing for smart waste management system,” in *2017 International Conference on Intelligent Communication and Computational Techniques (ICCT)*, 2017, pp. 272–278.

[7] M. Giacobbe, C. Puliafito, and M. Scarpa, “The Big Bucket: An IoT Cloud Solution for Smart Waste Management in Smart Cities,” in *Advances in Service-Oriented and Cloud Computing*, 2018, pp. 43–58.

[8] V. Bandari, “Cloud Workload Forecasting with Holt-Winters, State Space Model, and GRU,” *Journal of Artificial Intelligence and Machine Learning in Management*, vol. 4, no. 1, pp. 27–41, 2020.

[9] V. Bandari, “The Impact of Artificial Intelligence on the Revenue Growth of Small Businesses in Developing Countries: An Empirical Study,” *Reviews of Contemporary Business Analytics*, vol. 2, no. 1, pp. 33–44, 2019.

[10] V. Bandari, “Proactive Fault Tolerance Through Cloud Failure Prediction Using Machine Learning,” *ResearchBerg Review of Science and Technology*, vol. 3, no. 1, pp. 51–65, 2020.

[11] S. Sahana, T. Mukherjee, and D. Sarddar, “A conceptual framework towards implementing a cloud-based dynamic load balancer using a weighted Round-Robin algorithm,” *Int. J. Cloud Appl. Comput.*, vol. 10, no. 2, pp. 22–35, Apr. 2020.

[12] S. P. Ahuja, E. Czarnecki, and S. Willison, “Multi-factor performance comparison of Amazon Web Services Elastic Compute Cluster and Google Cloud Platform Compute Engine,” *Int. J. Cloud Appl. Comput.*, vol. 10, no. 3, pp. 1–16, Jul. 2020.

[13] V. Bandari, “The Adoption Of Next Generation Computing Architectures: A Meta Learning On The Adoption Of Fog, Mobile Edge, Serverless, And SoftwareDefined Computing,” *ssraml*, vol. 2, no. 2, pp. 1–15, 2019.

[14] G. I. Shidaganti, A. S. Inamdar, S. V. Rai, and A. M. Rajeev, “SCEF: A model for prevention of DDoS attacks from the cloud,” *Int. J. Cloud Appl. Comput.*, vol. 10, no. 3, pp. 67–80, Jul. 2020.

[15] H. Le Ngoc, T. N. Thi Huyen, X. Phi Nguyen, and C. Hung Tran, “MCCVA: A new approach using SVM and kmeans for load balancing on cloud,” *Int. J. Cloud Comput. Serv. Archit.*, vol. 10, no. 3, pp. 1–14, Jun. 2020.

[16] R. Beraldi, H. Alnuweiri, and A. Mtibaa, “A power-of-two choices based algorithm for fog computing,” *IEEE Trans. Cloud Comput.*, vol. 8, no. 3, pp. 698–709, Jul. 2020.

[17] H. Suleiman and O. Basir, “SLA-driven load scheduling in multi-tier cloud computing: Financial impact considerations,” *Int. J. Cloud Comput. Serv. Archit.*, vol. 10, no. 2, pp. 1–24, Apr. 2020.

[18] M. Yuvaraj, *Cloud computing in libraries*, 1st ed. K.G. Saur Verlag, 2020.

[19] J. Hoffman, *Cloud Computing Certifications*. Book Collection, 2020.

[20] P. Sharma, M. Sharma, and M. Elhoseny, *Applications of cloud computing*. London, England: CRC Press, 2020.

[21] D. Muller, *Cloud Computing*. Duncker & Humblot, 2020.

[22] Q. Zhang, Y. Wang, and L.-J. Zhang, Eds., *Cloud computing - CLOUD 2020*, 1st ed. Cham, Switzerland: Springer Nature, 2020.