Journal of Green Engineering (JGE)

Volume-10, Issue-3, March 2020

Docker Container in Green Cloud Platform for Workload Prediction

[1]P. Akilandeswari , [2]Medha Manoj Panikkasseri, [3]H. Srimathi

[1]Assistant professor,Department of Computer Science and Engineering, SRM IST, Chennai, TamilNadu, India. E-mail:akilasharanju@gmail.com
[2]Student,Department of Computer Science and Engineering, SRM IST, Chennai, TamilNadu, India. E-mail:medhamanojpanik@gmail.com
[3]Professor,Department of Computer Science and Engineering, SRMIST, Chennai, TamilNadu, India. E-mail: srimath@srmist.edu.in

Abstract

As computing technology grows and deployed application varies over time the services offered by the cloud providers become tough competition to give better quality of service. Adoption of instances based on resource scaling and variation in workload handling becomes quite challenging task. The literature survey reveals different workload run on different cloud being managed by hypervisor called virtual machine monitor. These are system containers called virtual machines, to run multiple processors in parallel. To handle big data and AI workloads these virtual machines are quite challenging in scalability to provide better quality of services. In this paper we addressed the problem using application containers called Docker containers. Prediction of workloads using ARIMA and exponential smoothing forecast methods are implemented and better memory and CPU usage is achieved when docker runs image instances and compared with benchmark workloads. Simulation result shows that using docker container provides better quality of service than running the same workload with virtual machine. In the future streaming data can be analysed with cloud container by focuses automatically deploying workloads inside application containers.

Keywords: Cloud platform, Docker container, Workload prediction, ARIMA, Exponential smoothing forecast.

Journal of Green Engineering, Vol. 10_3,575-592.

Alpha Publishers

1 Introduction

The emerging computing requirements of scalable resources and cost effectiveness are provided by cloud infrastructure. Making capacity supporting environment closely matching needs of new workloads need knowledge and insights to handle large volume and variety of workloads. They aid in creating new business plans and strategies by analysing the workload pattern observed from the historical data and predicting the best method to follow for better outcome. The analytics learned from data can be much helpful when used in predicting the activities that involve environmental conditions such as climate and for scientific research. Transform every industry from financial services, retail, and healthcare to government very rapidly assigned in a particular sequence on virtual machines due to the amount of data generated at an exponential rate. The different computing workloads are categorized as static workload does not changes with time epoch such as private websites or websites by small and medium companies, Dynamic workload [29] has variation as time changes, periodic workload as repeating behaviour of workload pattern, unpredictable workload like online traffic during rush hours and web searching patterns. In this paper the workload considered are transactional data.

1.1 Need of Docker Container

Virtual machine can be defined as an emulation of Computer system. They can be implemented as hardware or software or both and run on top of a hypervisor which duplicates the underlying physical hardware resources. The VM contains all the components required to run apps, for computing, storage etc. The hardware resources that are virtualized are pooled together and made available to apps running on VM as discussed in paper [37]. There arises problem when the workload needs to be migrated between different machines. The entire OS have to be migrated. Therefore, effective utilization of resources is not possible always which result in wastage. Virtual machines also take up time for starting up. Since developers and consumers require faster access and computing, and the apps developed today are modular for increasing flexibility, easier release and change. These gave way to the popularization of Containers which virtualizes at the OS level with multiple containers running on top of OS kernel.

The various reasons for using Containers over virtual machines are consistent environments that are isolated from applications, have ability to run anywhere, on different operating systems. Docker is a most popular open source container format to run and create cloud environment.

Container basically make use of images have details to execute on container engine operating system at runtime .Containerized applications can be composed of several container images. A container can have multiple images running simultaneously, and failed ones can be replaced by new ones without disrupting the operation. The various reasons for using Containers over virtual machines are consistent environments that are isolated from applications, ability to run anywhere, ie , on different operating systems or on premise etc.

2 Related Works

This literature discusses techniques for characterizing and predicting the workload as stated in these papers [1][11]. In a specified time in order to meet cloud consumer computing services anticipations in the existence of variable workloads assigned on cloud, there is a need to predict cloud resources.

The workload prediction analysed from various papers like [3-6][23] [35], proposed budget constraints workload prediction, the methodology ARMAX available resources at each VM [34] based on control theory. Drawback in this paper continuous data set is not considered for moving average. Chu-Fu Wang et al [6], discussed, predicting dense and sparse jobs from the workload information available at the time of job submission on the cloud. They developed prediction based energy conserving resource allocation method (ECRASP) near future to make adequate decision making. Drawback in this paper is resource allocation is implemented for energy conservation compared with conventional algorithm.

Baldan et al [8], discussed workload analysed series are non-linear data applied to several realistic workload datasets from different data centres. Because of forecasting the cloud resources are adopted as the few elements, like a specific cost function, statistical tests. But seasonal data is not considered penalty cost included in service level agreements. Calheiros et al [26], Workload Prediction Using ARIMA Model proactive resource provisioning is identified for cloud QOS with 91 percentages and this work is implemented using virtualization of processor, in the future work application assigned on virtualized operating system will lead better results.

Yazhou Hu et al.[38]have proposed three models to predict workload by examining time series data. First, Time series model, which analyse time stamped data using different mathematical models such as AR, MA, ARMA, ARIMA [24] ,DM, MM . Secondly, Kalman Filter model that forecast true data from historical data by using two steps, prediction and update, which works in real time. Thirdly, Pattern matching model, that matches.The review analased in [19] specifies list of cloud providers with various task like periodic task, repetitive task, APEX jobs addressed in cloud environment.

sequences with historical pattern by pre-processing and match. Fourthly, they have also put forward a trigger strategy from the results of predicted workload data. This decides when to activate the elasticity [35],[36] mechanism This depend on factors such as rising tendency and CPU workload as discussed [16-18] and elobarate about workload charcterization.They have evaluated the models and trigger strategy for accuracy and reduced error.

Box et al. [9] have tried to analyse the forecast problem and it's no symmetric nature and to find the best one in time series forecasting. They have proposed a combination of tools to tackle the problem of forecasting using cost and statistical data. Initially, the visualization of time series is done and ACF and PACF is analysed. Next non-seasonal study is done using ARIMA and ETS models as a first-time study and other regression models are built from results. Thirdly, a similar study is done with seasonality [30] . Different models are evaluated using a same dataset to find the best model. The result is evaluated by applying to datacentres. This study has also achieved cost reduction in over provisioning and under provisioning, which is a major factor in achieving elasticity. Crone et al [10] designed for time series data forecasting is implemented using neural network 3 called NN3 provides results that outshine theta method equivalent to Exponential smoothing with variation relatively simple in machine learning.

Arijith Khan et al.[4] have developed a mechanism that characterize and predict the workload continuously. The authors have applied multiple time series approach, which examine workload among a group of virtual machines rather than single VM. The authors have proposed a new method for characterizing correlated patterns of workload due to the dependencies of applications running on different VMs. A co-clustering algorithm is used to group these patterns among different VM groups and also the time period when these patterns arise. Then, Hidden Markov model approach is used to find temporal correlations which can help predict individual VM workload from the previous step. They have evaluated the approach using 21 days of CPU utilization data from real time enterprise. This approach has shown to have accuracy of 73 percent compared to 55 percent of single time series approach.

Rodrigo et al [26] predicts the workload using Autoregressive integrated moving average (ARIMA) model, which helps in proactive provisioning of resources and increase the Quality of Service. In [6] [7], the authors have tried to optimally allocate resources for different application in cloudlets. They tried to decrease the response time of applications based on IoT by previously deciding the cloudlets before deploying as different applications have different QoS [20]. The main components of the system that predicts and updates the model on run are application provisioner, Load Predictor and Performance Modeller and workload analyser. They were able to achieve 91 percent accuracy as a result.

John Panneerselvam et al. [14] have tried to reduce the energy consumption by excessive use of resources by implementing auto-scaling. Firstly, they have categorised the workload into static workloads, periodic workloads, unpredictable workloads and continuously changing workload according to pattern of arrival. Then the two different modelling techniques, Bayesian modelling and Markov modelling are applied to google cluster data [3]to evaluate the efficiencies of these techniques.

Workloads to the cloud environment to get access to the elastic resources that can be used according to demand concept is stated by [34-36]. They have characterised the workload based on the job trace which are collected in the production system which contains the jobs that are completed before a mentioned date.Resource demand handled in [25] states that quality metrics are important for providing cloud services.

Hui Zhang et al. [12] have proposed a cloud computing model, which has an automatic and intelligent workload for managing and characterizing the workload. It separates the workload into base and flash crowd workload and utilizes a fast-frequent data detection algorithm that segregates the workload based on volume and data content The authors have targeted applications which are internet-based and have scaling-out architecture such as YouTube. The aim of this paper is to attain QoS and resource efficiency during computing of highly dynamic workloads. They have evaluated the model using hybrid test bed with local server and Amazon AWS.

The authors in [2] [5][28] have proposed hybrid prediction strategy which checks the type of workload and uses the appropriate prediction algorithm. They first check whether the workload belongs to period or trend using autocorrelation coefficients and Hurst components. Then linear regression is used to replace missing data. Galante et al, [27] discusses analytical technique to forecast the application load and computing system load before making scaling decisionsPrediction of self adaptive resource provisioning is addressed in these publications [15][21] addresses constraints on cost to process in the cloud environment. Advantages of container than VM have been identified and developed in these papers [31-33][39-41] contribute basic idea about docker container.

Keung et al. [13] have developed a resource allocation scheme focused on conserving energy in the data centres which involves two methods, prediction mechanism and job allocation mechanism, which forecast the arrival trend of jobs in the future. Exponential smoothing method is used for forecasting the status of forthcoming jobs. This helps in faster allocation and much power conservation compared to other method. In this survey paper author [22] identifies different cloud platform infrastructure, platform and services delivered by cloud providers and task scheduling methods are discussed. Related to environment, virtual machines and task as workloads are studied. In all above papers metric addressed are performance, CPU utilization, response time, arrival pattern, workflow, Job parameters, Service time and volume of data for various workload pattern. In this work transactional data set workload that change dynamically.

The comparision of VM based cloud and docker container gives how these extensive approaches to gain depth knolege on same cloud environment

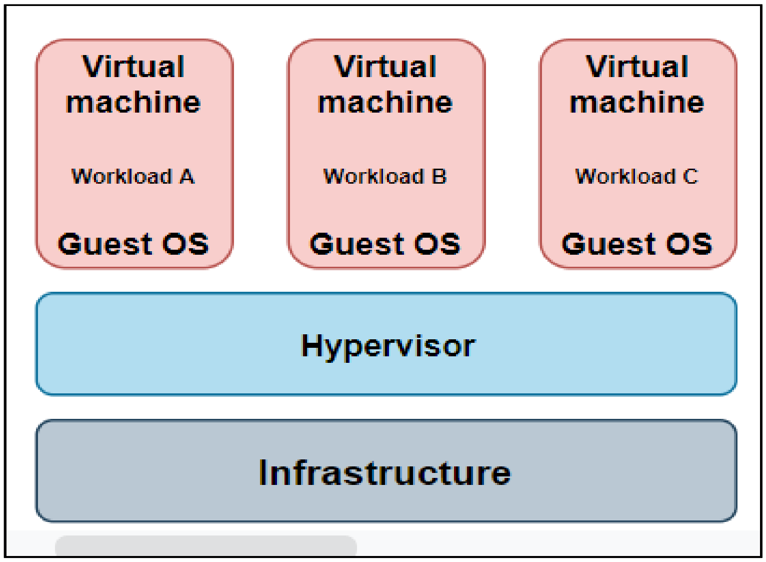

Fig 1. Cloud virtual machines

The above figure 1 shows cloud virtual machines that are managed by hypervisor called virtual machine monitor. Multiple processes can be executed on system containers.

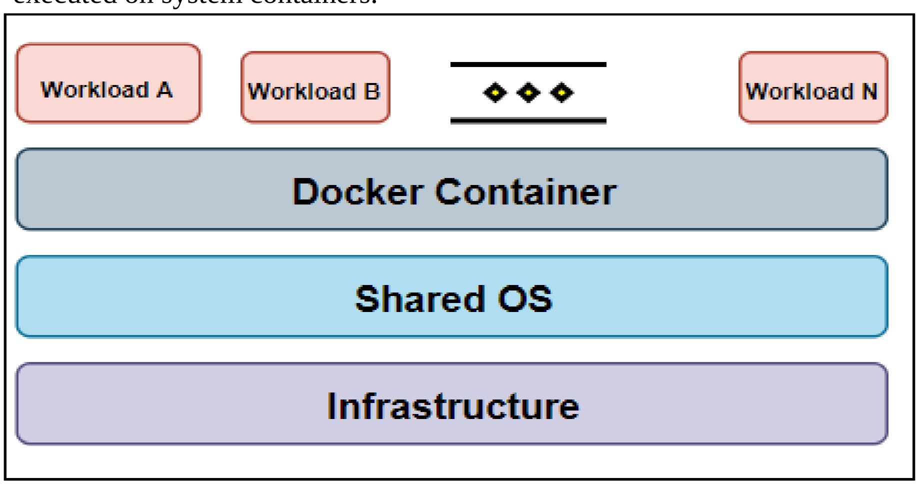

Fig 2 Docker Container

The figure 2 depicts docker container have application packages in a single service to deploy the workload inside the application container.

2.1 Drawbacks in the Existing System

In the prediction of data used in the above studies, the workload are deployed in the virtual machines which are comparatively slower and occupies space since it is equipped with all packages although unnecessary. The uses of containers which are better in terms of start-up time, easy usage, and portability are not made use of for analyzing data. The analysis is comparatively slow and the exact required CPU utilization, memory space required or threads created are difficult to analyze using Virtual machines since the entire operating system is build which may contain unnecessary packages and memory which is not really used by the application. if Containers are replaced for virtual machines, the exact metrics can be calculated. This helps in the future use of containers for similar type of analysis.

3 Proposed Methodology

3.1 *Data Set*

The dataset chosen here is a transactional dataset of a Company with clients in different cities in United States of America. The company does transaction of various supplies such as Furniture, Office supplies, technology with subcategories as chairs, tables, telephones, Binders etc. The dataset contains attributes such as Row ID, Order ID, Order Date, Ship Date, Ship Mode, Customer ID, Customer Name, Segment, Country, product ID, Category, Sub-Category, Product Name ,Sales Quantity, Discount, Profit. The data is collected from 2014 to 2017 on a weekly basis.

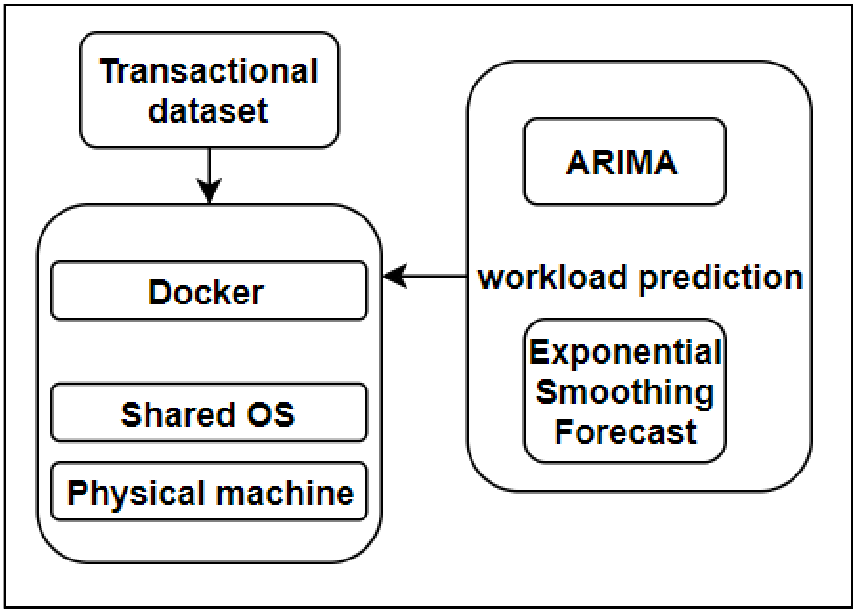

Fig 3 Prediction of Workload using docker container for transactional data

To predict workload pattern for variation in the transactiona data set with forecasting methods are shown in the figure 3.

3.1 Deployment on DOCKER

The workload pattern observed is assigned on the Docker client (docker) is the primary way that many Docker users interact with Docker, commands such as docker run, the client sends these commands to dockerd, that carries command uses the Docker API. The Docker client can communicate with more than one daemon. The dataset used for this project contains transactional data of a store situated in different cities. The features are ship mode, customer ID, segment, region, category, sales etc. The chosen variable is furniture and this project tries to build and fit an ARIMA model for the same data and forecast the future sales for the store in all cities for a time period of 9-10 years. The time series prediction is majorly used in sales for prediction that helps stores customize their inventory or change the business models and strategies according to the results.

Time series analysis is data points that are recorded at different points in time. They are useful for the fact that they help uncover structures that help produce the observed data and also fit a model and produce a forecast for future. Time series is commonly used in studies regarding prediction of workloads. The different methods of time series analysis are given below.

3.1.1 ARIMA Model

It predicts the value at the next step using observed patterns from the previous step. It is a combination of Autoregressive and moving average model, applied to stationery data or the available data is made stationery by continuous process,

$$X_t = c + \sum_{t=1}^{p} y\, \phi i\, ^p X_{t-i} + \varepsilon_t \qquad (1)$$

Exponential Smoothing: Produces smoothed out data by removing much of the noise. They assign relatively more weight in recent observations than in older ones.

$$F_{t+1} = a \sum_{k=0}^{t-1} y\, (1-\propto)^k y_{t-k} + (1-\propto)^t y_1 \qquad (2)$$

The same dataset is used for forecasting using ARIMA model and Exponential smoothing model. The jupyter notebooks environment, in which the data analysis is carried out deployed to the Docker container from the local host. Both the notebooks are hosted via two images with inbuilt packages such as numpy, pandas and seaborn to run a data analytic project. The performance of the containers, analysing the dataset using two different models is checked in runtime.

After forecasting the result in Jupyter environment, the Jupyter notebook is connected with the Docker Container on local computer. The image Jupyter/ data science notebook is pulled from the docker hub to run the jupyter notebook which contains packages necessary to run the ARIMA model and exponential smoothing model. The notebook is deployed using container image which is run on docker. The real time metrics are calculated such as CPU % utilization, Memory usage, Net I/O, Block I/O, PIDS

The performance of time series models, namely, ARIMA and Exponential smoothing when it is deployed in a Docker container and uses the same data set with following criteria,

1. Dataset collected should be in CSV format.
2. Dataset collected should be in Date Time Index format.
3. Virtualization must be enabled in the BIOS.

4 Experimental Setup

Anaconda environment used Physical server or virtual machine. CPU: 2 x 64-bit, 2.8 GHz, 8.00 GT/s CPUs or better. Verify machine architecture. Memory: minimum RAM size of 32 GB, or 16 GB RAM with 1600 MHz DDR3 installed, for a typical installation with 50 regular users. Verify memory requirements. Storage: Recommended minimum of 100 GB, or 300 GB if planned to mirror both Anaconda Repository, which is approximately

90 GB, and the PyPI repository, which is approximately 100 GB, or at least 1 TB for an air gapped environment. Additional space is recommended if Repository is used to store packages built. This study implements the ARIMA model and Exponential smoothing method on same dataset and analyse the performance it gives after deploying it in Docker Container.

4.1 Software Used

Operating System: Windows 7 or higher,MacOS
Cloud: Amazon web services,Microsoft Azure
Programming: Python 3.6 and related libraries
Environment: Anaconda
Container Software: Docker Enterprise/ Docker Community Edition/ Docker Toolbox
Servers for Running Docker:Ubuntu/SUSE Linux Enterprise Server/Red Hat Enterprise
Linux/Oracle Linux/Fedora/ CentOS/ Debian

The ARIMA model and exponential smoothing model are analysed in jupyter notebook. The above two notebooks are deployed as two container using Docker Toolbox(Windows Home edition compatible). Docker run command pulls the jupyter/scipy-notebook image from Docker Hub if it is not already present on the local host. It then starts a container running a Jupyter Notebook server and exposes the server on host port 8888. The server logs appear in the terminal

The token id in a browser loads the Jupyter Notebook dashboard page, where hostname is the name of the computer running docker and token is the secret token printed in the console. The container remains intact for restart after the notebook server exits. The image can be build according to the requirements of the application and software it is running. Fig 4 shows Docker imaga pull.

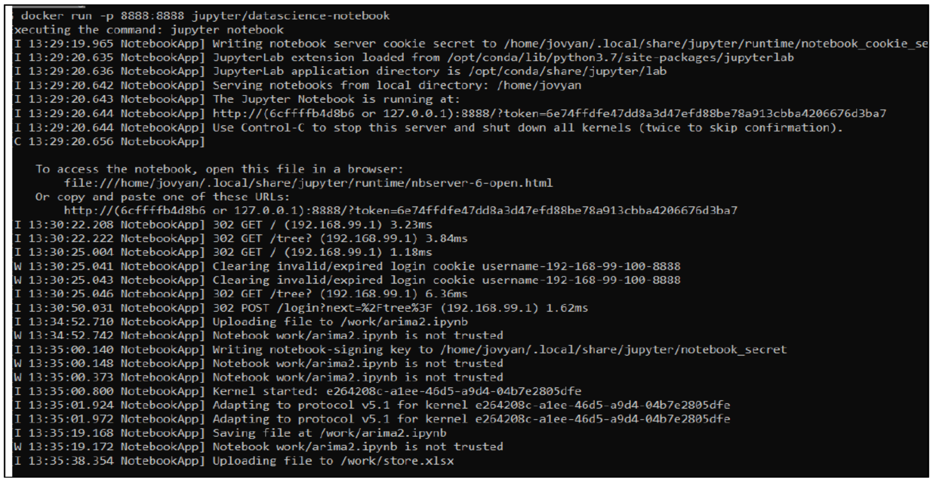

Fig 4 Docker imaga pull

From pylab configuration executable files to decompose time critical task with decomposion.tsa.seasonal_decompose starting from bottom to top in y-axis ,variations in workload residual,seasonal,treand and observed pattern and in X-axis,timeline for four years are depicted.

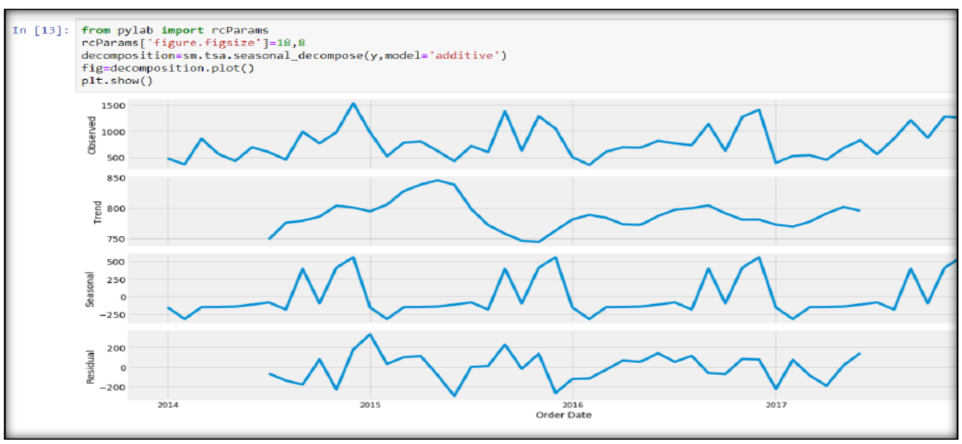

Fig 5 Workload variationfor transactional data

The above figure 5 shows decomposition of Residual, Seasonal, Trend and Observed pattern applied transactional data set are shown for four years {2014-2017}.

4.2 Experimental Results

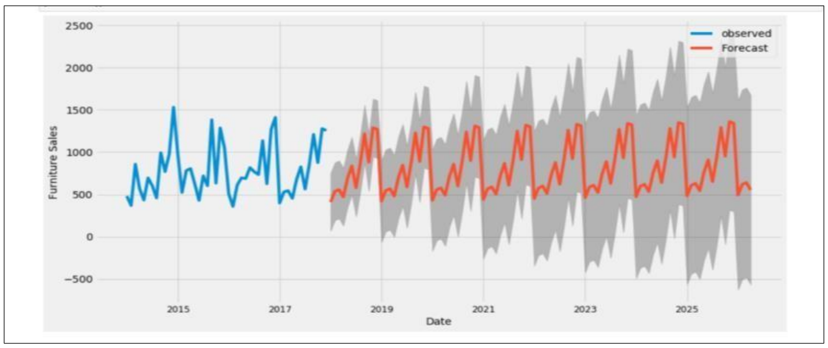

Fig 6 ARIMA model forecast

Figure 6 shows ARIMA Model prediction of transactional data observed and predicted data for next few years.

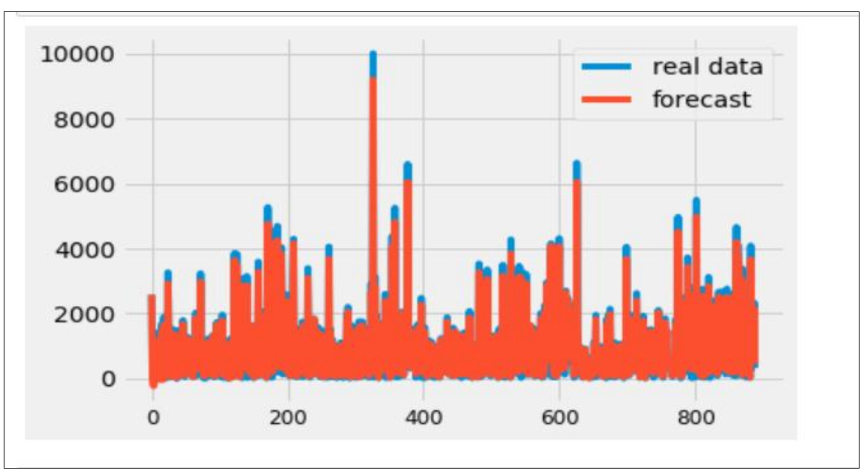

Fig 7 Exponential Smoothing forecast

Figure 7 shows Exponential smoothing forecast of transactional data observed and predicted data for next few years.ARIMA and Exponential

Smoothing method is proposed to be used in this project for the data analysis. Different types of workload can be compared and the most suitable one for time series approach is used for the part. A suitable interval of time can be set as a time stamp for monitoring the data.

Table 1 Docker runtime statistics for memory and processor utilization

Container ID/ Forecast Method	CPU %	Mem.Usage/Limit	Mem %	NET I/O (BW limitation)	Block I/O	PIDS (Process ID)
B7d3257e Exponential smoothing	0.06	181.6MIB/989.4MIB	18.36	6.52/14.8MB	72 MB	13
	0.07	180.6MIB/989.4MIB	20.36	6.52/14.8MB	72MB	12
	0.04	183.6MIB/989.4MIB	24.36	6.52/14.8MB	72MB	10
	0.05	184.6MIB/989.4MIB	20.61	6.52/14.8MB	72 MB	11
E6fb30dc2 ARIMA model	0.06	172.8 MIB/989.4MIB	17.46	3.83/8.34MB	35MB	13
	0.07	172.8 MIB/989.4MIB	17.31	3.83/8.34MB	35 MB	12
	0.04	172.8 MIB/989.4MIB	16.26	3.83/8.34MB	35 MB	10
	0.05	172.8 MIB/989.4MIB	15.26	3.83/8.34MB	35 MB	11

From the Docker runtime stats it is shown that Exponential smoothing uses more memory than ARIMA model and have more CPU utilization during runtime for the given time series dataset. The future resource utilization can also be predicted from the metrics that are related to the CPU usage and utilization of container. ARIMA model uses better memory and CPU usage than exponential smoothing method. Table 1 show that Docker runtime statistics for memory and processor utilization

5 Conclusions

To run the data analytics applications on independent infrastructure platform that allows optimized time series models best suited for mitigating applications and environmental conflicts. This paper analyses and examines various studies that have been conducted on characterization of workloads and their prediction in the cloud environment using different metrics such as performance, CPU utilization ,volume which are obtained from real data centres. Also, the paper tries to implement the dataset and analyse them using prediction models and deploy it in real time Docker container to get performance metrics. As a future work, the paper plans to deploy the workloads, streaming data in two different containers and compare their performance metrics and analyse which container is suitable for time series.

References

[1] J. Guo, J. Wu, J. Na and B. Zhang, "A Type-Aware Workload Prediction Strategy for Non-stationary Cloud Service”, IEEE 10th Conference on Service-Oriented Computing and Applications (SOCA), pp. 98-103, 2017.

[2] A. Adegboyega, "Time-series models for cloud workload prediction: A comparison," IFIP/IEEE Symposium on Integrated Network and Service Management (IM), pp. 298-307, 2017.

[3] A. K. Mishra, J. L. Hellerstein, W. Cirne, and C. R. Das, “Towards characterizing cloud backend workloads: Insights from google computer clusters,” ACM SIGMETRICS Performance Evaluation Review, vol. 37, no. 4, pp. 34–41, 2010.

[4] A. Khan, X. Yan, S. Tao and N. Anerousis, "Workload characterization and prediction in the cloud: A multiple time series approach”, IEEE Network Operations and Management Symposium, Maui, HI, pp. 1287-1294, 2012.

[5] Abdullah Mohammed Al-Faifi , Biao Song , Mohammad Mehedi Hassan , Atif Alamri , Abdu Gumaei “Performance prediction model for cloud service selection from smart data “,Future Generation Computer Systems,vol.85,pp.97-106, 2018.

[6] C. Wang, W. Hung and C. Yang, "A prediction based energy conserving resources allocation scheme for cloud computing”, IEEE International Conference on Granular Computing (GrC), pp. 320-324, 2014.

[7] E. Caron, F. Desprez and A. Muresan, "Forecasting for Grid and Cloud Computing On-Demand Resources Based on Pattern Matching”, IEEE Second International Conference on Cloud Computing Technology and Science, pp. 456-463, 2010.

[8] F. J. Baldan, S. Ramirez-Gallego, C. Bergmeir, F. Herrera and J. M. BenitezSanchez, "A Forecasting Methodology for Workload Forecasting in Cloud Systems”, IEEE Transactions on Cloud Computing , Vol. 6,no.4,pp.929-941, 2018.
[9] G. E. P. Box, G. M. Jenkins, and G. C. Reinsel, “Time Series Analysis: Forecasting and Control”, 4th ed. Hoboken, NJ, USA,Wiley, 2008.
[10] S. Crone, M. Hibon, and K. Nikolopoulos, “Advances in forecasting with neural networks? empirical evidence from the NN3 competition on time series prediction,” International Journal of Forecasting, vol. 27, no. 3, pp. 635–660, 2011.
[11]Hoffinann G, Trivedi K S, Malek M. “A best practice guide to resource predicting for computing systems”,IEEE Transactions on Reliability, vol.56,no.4,pp. 615-628, 2007.
[12]Hui Zhang, Guofei Jiang, Kenji Yoshihira, and Haifeng Chen”Proactive Workload Management in Hybrid Cloud Computing”, IEEE Transactions on Network And Service Management, vol.11,no.1,pp.90-100, 2014.
[13]Islam S, Keung J, Lee K, et al. “Empirical prediction models for adaptive resource provisioning in the cloud”, Future Generation Computer Systems,vol.28,no.1,pp.155-162,2012.
[14]J. Panneerselvam, L. Liu, N. Antonopoulos and Y. Bo, "Workload Analysis for the Scope of User Demand Prediction Model Evaluations in Cloud Environments”, IEEE/ACM Conference on Utility and Cloud Computing, pp. 883-889, 2014.
[15] Jiang Y, Perng C, Li T, et al. “ASAP: A self-adaptive prediction system for instant cloud resource demand provisioning”, IEEE 11th International Conference on Data Mining (ICDM), pp. 1104-1109,2011.
[16]M. Calzarossa and G. Serazzi, “Workload characterization: A survey”, IEEE Procedings, vol. 81, no. 8, pp. 1136–1150, 1993.
[17]B. Downey and D. G. Feitelson, “The elusive goal of workload characterization”, ACM SIGMETRICS Performance Evaluation Review, vol. 26, no. 4, pp. 14–29, 1999.
[18]N. Roy, A. Dubey, and A. Gokhale, “Efficient auto scaling in the cloud using predictive models for workload forecasting”, in Proc. 4th Int. Conf. Cloud Computing,pp. 500–507 ,2011.
[19]P. Akilandeswari, H.Srimathi “Survey and Analysis on Task Scheduling in Cloud Environment”, Indian Journal of Science and Technology, vol.9,no.37, 2016.
[20]Q. Fan and N. Ansari, "Application Aware Workload Allocation for Edge Computing-Based IoT", in IEEE Internet of Things Journal, vol. 5, no. 3, pp. 2146-2153, 2018.
[21]Q. Zhu and G. Agrawal, “Resource provisioning with budget constraints for adaptive applications in cloud environments”, IEEE transactions on service computing, vol.5,pp 497-511,2012.

[22]Feng,Guofu and Rajkumar Buyya."Maximum revenue-oriented resource allocation in cloud", International Journal of Grid and Utility Computing, vol.7, no.1, pp. 12-21, 2016.

[23]R. Hyndman and G. Athanasopoulos, "Forecasting: Principles and practice", OTexts,pp.1- 380,2018

[24]R. N. Calheiros, E. Masoumi, R. Ranjan and R. Buyya, "Workload Prediction Using ARIMA Model and Its Impact on Cloud Applications' QoS", IEEE Transactions on Cloud Computing, vol. 3, no. 4, pp. 449-458, 2015.

[25]Reig G, Guitart J," On the anticipation of Resource Demands to fulfill the QOS of SAAS Web Applications", Proceedings of the ACM/IEEE 13th International Conference on Grid Computing, IEEE Computer Society,pp. 147-154,2012

[26]Roy N, Dubey A, Gokhale "A. Efficient auto scaling in the cloud using predictive models for workload predicting", IEEE 4th International Conference on Cloud Computing ,2011..

[27] G. Galante and L. C. E. d. Bona, "A survey on cloud computing elasticity," in Proceedings of the 2012 IEEE/ACM Fifth International Conference on Utility and Cloud Computing, ser. UCC '12, pp. 263–270. 2012

[28]S. K. Sahi and V. S. Dhaka, "A survey paper on workload prediction requirements of cloud computing", 3rd International Conference on Computing for Sustainable Global Development (INDIACom), New Delhi, pp. 254-258, 2016.

[29]S. Vakilinia, M. Cheriet and J. Rajkumar, "Dynamic resource allocation of smart home workloads in the cloud",12th International Conference on Network and Service Management (CNSM), pp. 367-370, 2016.

[30]V. G. Tran, V. Debusschere, and S. Bacha, "Hourly server workload forecasting up to 168 hours ahead using seasonal ARIMA model", in Proc. 13th Int. Conf. Ind. Technology, pp. 1127–1131, 2012.

[31]W. Tang ,"Workload characterization for MG-RAST meta genomic data analytics service in the cloud", IEEE International Conference on Big Data, 2014.

[32]Xu W, Zhu X, Singhal S, "Predictive control for dynamic resource allocation in enterprise data centers", Network Operations and Management Symposium, 10th IEEE/IFIP,2006.

[33]Y. Chen, A. Ganapathi, R. Griffith, and R. Katz, "Towards understanding cloud performance tradeoffs using statistical workload analysis and replay", EECS, 2010.

[34] Y. Wang, R. Yang, T. Wo, W. Jiang, and C. Hu, "Improving utilization through dynamic VM resource allocation in hybrid cloud environment",

20th IEEE International Conference on Parallel and Distributed Systems (ICPADS), pp. 241–248, 2014.
[35]Yazhou Hu, Bo Deng, Fuyang Peng and Dongxia Wang, "Workload prediction for cloud computing elasticity mechanism", IEEE International Conference on Cloud Computing and Big Data Analysis (ICCCBDA), pp. 244-249, 2016.
[36] Z. Shen, S. Subbiah, X. Gu, and J. Wilkes, "CloudScale: elastic resource scaling for multi-tenant cloud systems", Proceedings of the 2nd ACM Symposium on Cloud Computing,pp. 1–14, 2011.
[37]Kyoung-Taek Seo, Hyun-Seo Hwang, Il-Young Moon, Oh-Young Kwon, Byeong-Jun Kim" Performance Comparison Analysis of Linux Container and Virtual Machine for Building Cloud", Advanced Science and TechnologyLetters,Vol.66, pp.105-111, 2014.
[38]Paolo Di Tommaso, Emilio Palumbo, Maria Chatzou,Pablo Prieto, Michael L. Heuer and Cedric Notredame "The impact of Docker containers on the performance of genomic pipelines",PeerJ — the Journal of Life and Environmental Sciences ,2015.
[39]Hung L-H, Kristiyanto D, Lee SB, Yeung KY " GUIdock: Using Docker Containers with a Common Graphics User Interface to Address the Reproducibility of Research", PLoS ONE,vol.11,no.4, 2016.
[40]Babak Bashari Rad, Harrison John Bhatti, Mohammad Ahmadi,"An Introduction to Docker and Analysis of its Performance", International Journal of Computer Science and Network Security, vol.17, No.3, 2017.
[41]Qi Zhang and Ling Liu and Calton Pu and Qiwei Dou and Liren Wu and Wei Zhou "A Comparative Study of Containers and Virtual Machines in Big Data Environment", IEEE 11th International Conference on Cloud Computing,pp. 178-185, 2018.

Biographies

Mrs.Akilandeswari P received the Engineering degree in Electronics and Communication engineering from MK University, Tamil Nadu and Master Degree in Computer Science and Engineering from Anna University, Tamil Nadu. She is doing Ph.D in Computer Science and Engineering in the research area of Cloud computing for stochastic task scheduling. She is currently an Assistant Professor in Department of Computer Science and

Engineering, SRM University. Her research includes Scalable Computing, Big data and IOT.

Medha Manoj Panikkasseri is a bachelor degree student from SRM Institute of Science and Technology (formerly known as SRM University). She is expertise in Databases, Data Mining .Cloud computing.

H.Srimathi, Professor, SRM Institute of Science and Technology has two decades of experience in higher education and services. since 1999 and served in various domains such as academics and administration. She is passionate about the studies on higher education systems, qualification framework, and academic mobility.