ANALYSIS AND PREDICTION OF WORKLOADS IN CLOUD ENVIRONMENT

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

***ABSTRACT -- ****Cloud computing technologies over the years have helped meet the changing needs for resources and computing capabilities of different organizations globally. By migration of workloads to cloud platform, (public, private or hybrid) there have been significant increase in Data Analytics applications to gain valuable insights from data. The rising amounts of workload have led to the necessity of predicting the optimal choice of platform for performing analytics in cloud to achieve better results. The increases in popularity of containers have made them an efficient choice for computing and prediction over traditional Virtual machines. This paper tries to analyse the different workloads and the types of model to predict them with the prospect of running workloads in a containerised environment.*

Keywords – *Workload, prediction, cloud, performance.*

I. INTRODUCTION

Big Data Analysis is transforming every industry from financial services, retail, healthcare to government very rapidly due to the amount of data generated at an exponential rate. The knowledge and insights gained from these large volumes and varieties of data are redefining and innovating the industries to an efficient environment which is highly profitable and futuristic. They aid in creating new business plans and strategies by analysing the 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.

Cloud computing have become indispensable part of data analytics today due to the need of high capacity servers on premise to serve the high demands. Everyone opts to migrate to cloud or start the business on cloud due to its many benefits which are paying for the amount of computing used (pay-as-you-go model), more computing capabilities compared to onpremise cloud and less cost and energy compared to normal servers .Due to the demand for more resources, there needs to be a proper provisioning of resources to the tasks that are to be computed in the cloud. Different kinds of workload patterns can be analysed in the cloud with different deployment models such as public, private and hybrid which are interrelated with each other. Predicting the workload beforehand is necessary for allocating required resources and in choosing most probable deployment model according to the workload and other characteristics.

[1] *Computer Science and Engineering, SRM University, Chennai - 603203, Tamil Nadu, India, akilandeswari.p@ktr.srmuniv.ac.in*

[2] *Computer Science and Engineering, SRM University, Chennai - 603203, Tamil Nadu, India, medhamanojpanik@gmail.com*

[3] *Computer Science and Engineering, SRM University, Chennai - 603203, Tamil Nadu, India, srimathi.h@ktr.srmuniv.ac.in*

Received: 21 Dec 2019 | Revised: 18 Jan 2020 | Accepted: 05 Feb 2020

II. TYPES OF WORKLOADS ANALYSED IN THE CLOUD

1) Static workloads: They have constant behaviour in the cloud pattern for over long periods of time with flat utilization profiles. It is experienced by many applications that do not fully employ a single server. Examples are private websites or websites by small and medium companies. The number of resources required for this type of workload remains constant and hence provisioning is easy.

2) Periodic workload: They have repeating behaviour of workload pattern. The volume of workload keeps increasing in periodic intervals of time. An example is financial data uploaded once in a month or quarterly. There arises a need to predict the workload during peak and non-peak hours for proper resource provisioning

3) Once in a lifetime workload: The utilization of resources occurs very rarely, ie once in long time . Usually resource is required only once.

The applications that are migrated to the cloud for the first time have this pattern. The peak rise in workload can be predicted beforehand and can be handled properly. Example is one time digitalization of paper magazines.

4) Unpredictable workload: They are similar to periodic workload but with random and unknown rate of change of workloads. The pattern of future workload is almost impossible to predict. It is therefore difficult to allocate resources for the same. Examples are unpredicted traffic during rush hours and web searching patterns.

5) Continuously Changing workloads: They have either increasing or decreasing volume of workload arrival in one direction. Newly launched business has increasing pattern and legacy applications of old products show decreasing pattern. The rate of growth or shrink needs to be predicted. Example is the marketing activities and sales after the launch of a new product until the next one comes out.

III. CLOUD DEPLOYMENT MODELS

The cloud deployment models are depending on the users who access them and how the resources are shared between customers. The cross-cloud environment in the project refers to the common operating environment across private and public cloud.

1. Private: Private clouds are virtualized resources that are deployed within a private network. They are managed by the organization within its data centre or by a third party as an outsourced private cloud. They are only accessible to the internal members of the organization and provide high level of security. Examples are VMware product suit comprised of vCloud, vSphere and ESX. Open Source software such as Eucalyptus, Open Nebula or Open stack.

2. Public: Public clouds are virtualized resources that are deployed outside the organization’s private network and are managed by third parties and are accessible to everyone. Due to sharing of resources between large diverse groups of people, peak workloads can be handled. Examples are Amazon Elastic Compute Cloud (EC2), Google cloud, Microsoft Windows Azure.

3. Hybrid: Hybrid clouds are combination of private and public model, i.e. virtual machines are hosted on both private and public cloud(on- premise or off-premise), with orchestration between the two platforms.

IV. CLOUD SERVICE MODELS:

1. Infrastructure as a Service: This model delivers the computing infrastructure for the customers to build their enterprises. Virtualization, servers, storage , networking are managed by the service provider based on a pay per go model. Examples are AWS EC2, Google compute engine.
2. Platform as a Service: In this model the infrastructure is owned by a third party apart from the service provider and they provide and hardware and software required for application development in the cloud. Examples are AWS lambda, Windows azure.
3. Software as a service: The users use the already developed software that is deployed on the cloud for their own personal or professional use. All the infrastructure, software are managed by service provider or the software is hosted by third party. Examples are DropBox, Google apps etc.

V. VIRTUAL MACHINES AND CONTAINERS

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. 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 since the apps developed today are modular for increasing flexibility and 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, ability to run anywhere, ie , on different operating systems or on premise etc. Docker is a most popular open source container

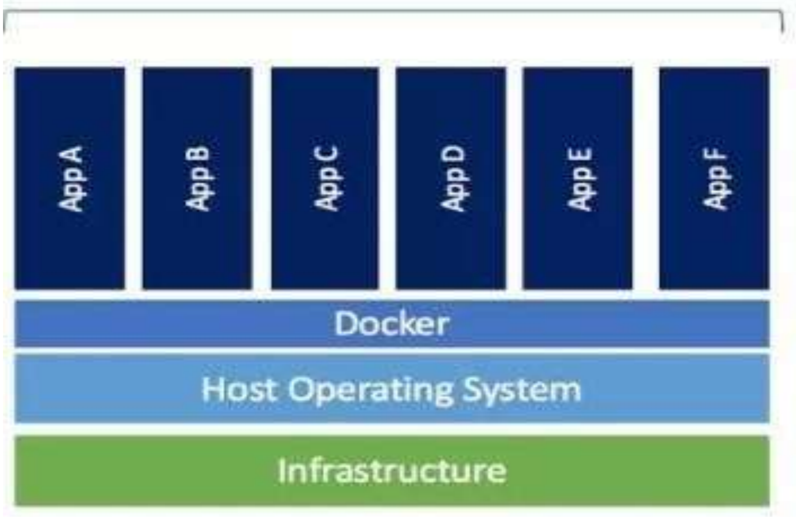


Fig. Container

Forecasting methodologies (Time Series)

Time series analysis are 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.

1. Autoregressive model: It predicts the value at the next step using observed patterns from the previous step.

$$X_t = c + \sum_{i=1}^{p} \varphi_i X_{t-i} + \varepsilon_t$$

2. Moving Average model: It predicts the next step in the sequence by considering the current and the past value of a variable. Can identify whether the pattern is uptrend or downtrend easily.

$$X_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}$$

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

$$s_t = \alpha \cdot x_t + (1 - \alpha) \cdot s_{t-1} = s_{t-1} + \alpha \cdot (x_t - s_{t-1})$$

4. ARIMA model: It is combination of Autoregressive and moving average model. This model is applied to stationery data or the available data is made stationery by continuous process.
5. Neural Networks: Predicting the future values of the data set using training neural network layer according to the number of parameters.

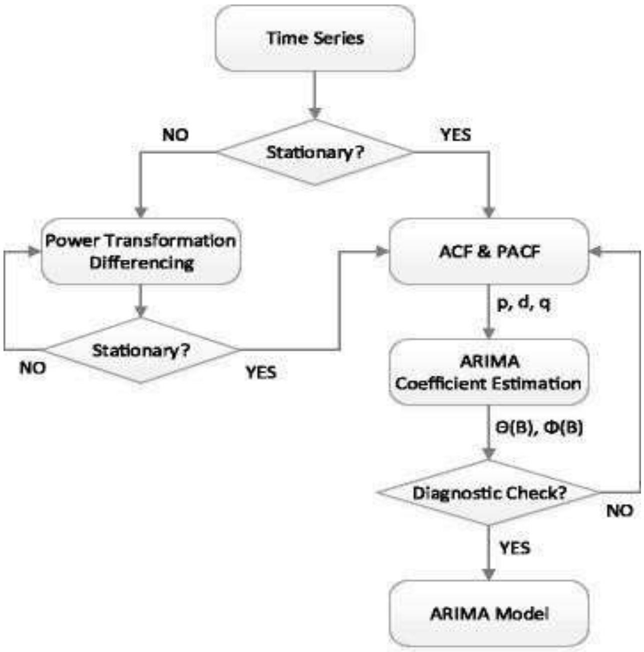


Fig 1 ARIMA model for time series workload

VI. LITERATURE SURVEY

This literature discusses techniques for characterizing and predicting the workload

A. Characterisation and prediction of workload characteristics

Abdullah Mohammed Al-Faifi et al. [1] have incorporated the performance of the organization for selecting cloud vendor. They have attempted to take into consideration the user's requirement when deploying workload in cloud for the first time or migrating between clouds. Therefore , the authors have tried to automate the selection process for the cloud environment based on workload pattern in a smart home data and resource utilization from node data. The model is based on Naïve Bayes for performance prediction of the workload. Kernel Density estimation is also done to increase the accuracy of the results. The class label used for Naïve Bayes is performance (CPU utilization, Response time and Memory utilization)and the features are workload parameters(memory capacity, average no of jobs etc)

Yazhau Hu et al [2] 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 ,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 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 mechanism This depend on factors such as rising tendency and CPU workload.They have evaluated the models and trigger strategy for accuracy and reduced error.

Francisco Javier Bald ‘an et al.[3] have tried to analyse the forecast problem and its non- 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 function, statistical tests, visual analysis etc. There are different steps taking part in this methodology. Firstly, the visualization of time series is done and ACF and PACF is analysed. Secondly, a nonseasonal 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. 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.

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% compared to 55% of single time series approach.

Rodrigo N. Calherios et al. [5] predicts the workload using ARIMA(Auto Regressive Integrated Moving Average) model, which helps in proactive provisioning of resources and increase the Quality of Service.

In [6], 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. The main components of the system that predicts and updates the model on run are, Application provisioner, Load Predictor and Performance Modeler and workload analyser. They were able to achieve 91 percent accuracy as a result.

John Pnneerselvam et al. [7] have tried to reduce the energy consumption by excessive use of resources by implementing auto-scaling. Firstly, they have categorised the workloadinto 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 (CPU intensive, memory intensive and both)

Naïve Bayes classifier and Hidden Markov model are modelled in MATLAB to evaluate the efficiencies of Markov and Bayesian techniques.

Wei Tang et al. [8] have ported the MG-RAST workloads to the cloud environment to get access to the elastic resources that can be used according to demand. 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.

Hui Zhang et al. [9] have proposed a hybrid cloud computing model, that has an automatic and intelligent workload factoring service 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.

Table I. Workload Prediction metrics and algorithms

S.NO.	STUDY BY	METRICS	DATA SET	METHODOLOGY USED
1	Abdullah Mohammed Al-Faifi et al[1]	Performance	CPU utilization data	Naïve Bayes classifier along with kernel density estimations.
2	Yazhau Hu et al [2]	Performance, CPU utilization	Virtual machine performance data, CPU utilization	Kalman filter model

3	Francisco Javier Bald’an et al.[3]	CPU utilization	CPU usage data from Google cluster, LANL cluster, University Gaia cluster , Sharnet whale	Forecasting methodology enhanced with elements such as specific cost function, statistical tests, visual analysis etc.
4	Arijith Khan et al. [4]	Performance	CPU utilization time series	Hidden Markov Model based predictors.
5	Rodrigo N. Calherios et al.	Performance	Request to web servers from Wikipedi a	ARIMA model

Received: 21 Dec 2019 | Revised: 18 Jan 2020 | Accepted: 05 Feb 2020

6	Qiang Fan et al. [6]	Response time	Sensor data Iot	AREA algorithm for allocation..
7	John Panneerselvam et al. [7]	Pattern of arrival	7 hours of Google cluster data	Markov Modelling and
8	Wei Tang et al. [8]	workflow	DNA sequence data	Scalable platform where resources can be used on demand.
9	Hui Zhang et al. [9]	volume	Video streaming data	Fast frequent data item
10	JunGho et al. [10]	volume	Recommendation service and recognition service data of audio navigator	Linear regression combined with ARMA and SVR .
11	Chu-Fu et al. [11]	Job info	Cluster data	Prediction based energy conserving resource allocation method (ECRASP)
12	Shahin Vakilin et al[12]	Service time	Smart device data	Markovian Poisson process

The above table shows different metrics used for workload prediction and algorithms used in cloud platform.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 testbed with local server and Amazon AWS.

The authors in [10] 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. It adopts the method of linear regression with ARMA to forecast the trend and SVR method for predicting the period workload.Chu-Fu et al. [11] have developed a resource allocation scheme focused on conserving energy in the data centres which involves two methods, prediction mechanism and job

Received: 21 Dec 2019 | Revised: 18 Jan 2020 | Accepted: 05 Feb 2020

allocation mechanism, which forecast the arrival trend of jobs in the future. Exponential smoothing method is used for forecasting the status of forthcoming jobs.

Shahin Vakilinia et al[12] have developed a mechanism for finding the optimum number of VM's to satisfy the time constraints(execution time) of smart home applications running in a cloud platform. Firstly, M/M/c queue model is used to find smart home task's response time with slight change over rate of arrival of tasks. Then Markovian Modulated Poisson Process (MMPP) is applied to extend and use it for other type of advanced workloads to be processed. They provide the number of virtual machines that could optimally satisfy the execution time and calculate the total service time of the application. Simulation is done to evaluate the approach.

VII. IMPLEMENTATION

Proposed Architecture: The proposed architecture predicts the workload pattern and performance comparison suited for different workload prediction models.

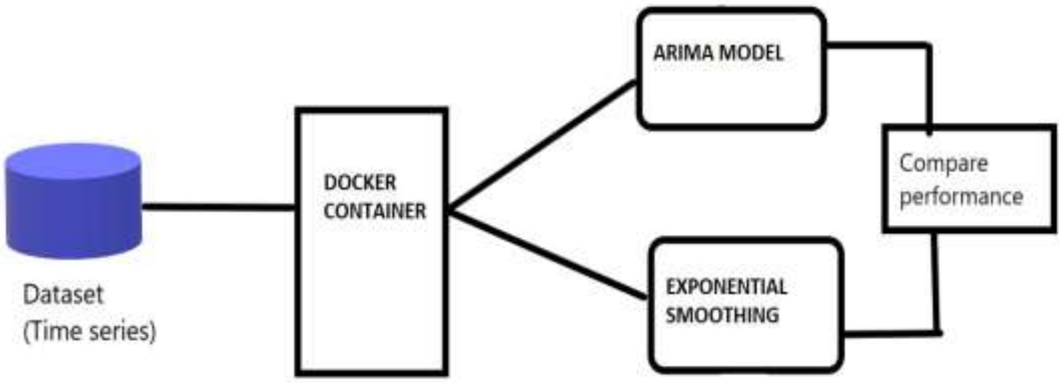
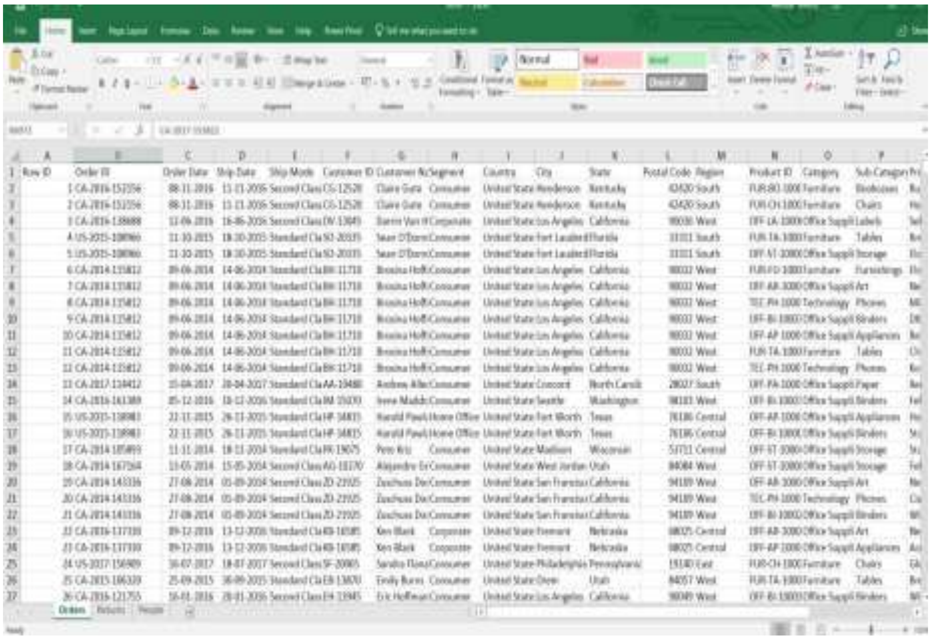


Fig 2. Proposed architecture.

This paper tries to implement the ARIMA model and Exponential smoothing method on same dataset and analyse the performance it gives after deploying it in Docker Container

Module 1: The dataset used for this project contains data of a store situated in different cities of United States of America. The different features are ship mode, customer ID, segment, region ,category, sales etc. The chosen variable is furniture and this project tries to build and fit a Arima model for the same data and forecast the future sales for the store in all cities for a time period of atleast 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.



	Row ID	Order ID	Order Date	Ship Date	Ship Mode	Customer ID	Customer N	Segment	Country	City	State	Postal Code	Region	Product ID	Category	Sub-Categor
2	1	CA-2016-152156	08-11-2016	11-11-2016	Second Class	CG-12520	Claire Gute	Consumer	United State	Henderson	Kentucky	42420	South	FUR-BO-1000	Furniture	Bookcases
3	2	CA-2016-152156	08-11-2016	11-11-2016	Second Class	CG-12520	Claire Gute	Consumer	United State	Henderson	Kentucky	42420	South	FUR-CH-1000	Furniture	Chairs
4	3	CA-2016-138688	12-06-2016	16-06-2016	Second Class	DV-13045	Darrin Van H	Corporate	United State	Los Angeles	California	90036	West	OFF-LA-1000	Office Suppli	Labels
5	4	US-2015-108966	11-10-2015	18-10-2015	Standard Cla	SO-20335	Sean O'Donn	Consumer	United State	Fort Lauderd	Florida	33311	South	FUR-TA-1000	Furniture	Tables
6	5	US-2015-108966	11-10-2015	18-10-2015	Standard Cla	SO-20335	Sean O'Donn	Consumer	United State	Fort Lauderd	Florida	33311	South	OFF-ST-1000	Office Suppli	Storage
7	6	CA-2014-115812	09-06-2014	14-06-2014	Standard Cla	BH-11710	Brosina Hoff	Consumer	United State	Los Angeles	California	90032	West	FUR-FU-1000	Furniture	Furnishings
8	7	CA-2014-115812	09-06-2014	14-06-2014	Standard Cla	BH-11710	Brosina Hoff	Consumer	United State	Los Angeles	California	90032	West	OFF-AR-1000	Office Suppli	Art
9	8	CA-2014-115812	09-06-2014	14-06-2014	Standard Cla	BH-11710	Brosina Hoff	Consumer	United State	Los Angeles	California	90032	West	TEC-PH-1000	Technology	Phones
10	9	CA-2014-115812	09-06-2014	14-06-2014	Standard Cla	BH-11710	Brosina Hoff	Consumer	United State	Los Angeles	California	90032	West	OFF-BI-1000	Office Suppli	Binders
11	10	CA-2014-115812	09-06-2014	14-06-2014	Standard Cla	BH-11710	Brosina Hoff	Consumer	United State	Los Angeles	California	90032	West	OFF-AP-1000	Office Suppli	Appliances
12	11	CA-2014-115812	09-06-2014	14-06-2014	Standard Cla	BH-11710	Brosina Hoff	Consumer	United State	Los Angeles	California	90032	West	FUR-TA-1000	Furniture	Tables
13	12	CA-2014-115812	09-06-2014	14-06-2014	Standard Cla	BH-11710	Brosina Hoff	Consumer	United State	Los Angeles	California	90032	West	TEC-PH-1000	Technology	Phones
14	13	CA-2017-114412	15-04-2017	20-04-2017	Standard Cla	AA-10480	Andrew Alle	Consumer	United State	Concord	North Caroli	28027	South	OFF-PA-1000	Office Suppli	Paper
15	14	CA-2016-161389	05-12-2016	10-12-2016	Standard Cla	IM-15070	Irene Madd	Consumer	United State	Seattle	Washington	98103	West	OFF-BI-1000	Office Suppli	Binders
16	15	US-2015-118983	22-11-2015	26-11-2015	Standard Cla	HP-14815	Harold Pawl	Home Office	United State	Fort Worth	Texas	76106	Central	OFF-AP-1000	Office Suppli	Appliances
17	16	US-2015-118983	22-11-2015	26-11-2015	Standard Cla	HP-14815	Harold Pawl	Home Office	United State	Fort Worth	Texas	76106	Central	OFF-BI-1000	Office Suppli	Binders
18	17	CA-2014-105893	11-11-2014	18-11-2014	Standard Cla	PK-19075	Pete Kriz	Consumer	United State	Madison	Wisconsin	53711	Central	OFF-ST-1000	Office Suppli	Storage
19	18	CA-2014-167164	13-05-2014	15-05-2014	Second Class	AG-10270	Alejandro Gr	Consumer	United State	West Jordan	Utah	84084	West	OFF-ST-1000	Office Suppli	Storage
20	19	CA-2014-143336	27-08-2014	01-09-2014	Second Class	ZD-21925	Zuschuss Do	Consumer	United State	San Francisc	California	94109	West	OFF-AR-1000	Office Suppli	Art
21	20	CA-2014-143336	27-08-2014	01-09-2014	Second Class	ZD-21925	Zuschuss Do	Consumer	United State	San Francisc	California	94109	West	TEC-PH-1000	Technology	Phones
22	21	CA-2014-143336	27-08-2014	01-09-2014	Second Class	ZD-21925	Zuschuss Do	Consumer	United State	San Francisc	California	94109	West	OFF-BI-1000	Office Suppli	Binders
23	22	CA-2016-137330	09-12-2016	13-12-2016	Standard Cla	KB-16585	Ken Black	Corporate	United State	Fremont	Nebraska	68025	Central	OFF-AR-1000	Office Suppli	Art
24	23	CA-2016-137330	09-12-2016	13-12-2016	Standard Cla	KB-16585	Ken Black	Corporate	United State	Fremont	Nebraska	68025	Central	OFF-AP-1000	Office Suppli	Appliances
25	24	US-2017-156909	16-07-2017	18-07-2017	Second Class	SF-20065	Sandra Flana	Consumer	United State	Philadelphia	Pennsylvania	19140	East	FUR-CH-1000	Furniture	Chairs
26	25	CA-2015-106320	25-09-2015	30-09-2015	Standard Cla	EB-13870	Emily Burns	Consumer	United State	Orem	Utah	84057	West	FUR-TA-1000	Furniture	Tables
27	26	CA-2016-121755	16-01-2016	20-01-2016	Second Class	EH-13945	Eric Hoffman	Consumer	United State	Los Angeles	California	90049	West	OFF-BI-1000	Office Suppli	Binders

Fig 3. Dataset

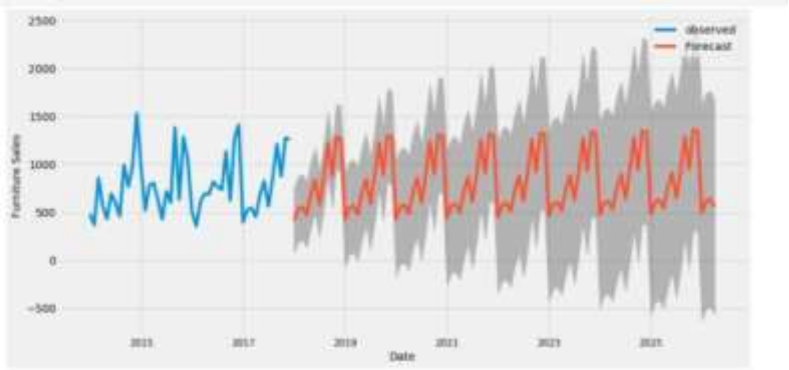


Fig 4. Results

Module 2: After forecasting the result in Jupyter environment, the jupyter notebook is connected with the Docker Container on local computer. The image jupyter/ datascience- notebook is pulled from the docker hub to run the jupyter notebook which contains packages necessary to run the ARIMA model. The notebook is deployed using container image which is run on docker. The real time metrics are calculated using commands.

jupyter/datascience-notebook ☆

By jupyter • Updated a day ago

Jupyter Notebook Data Science Stack from https://github.com/jupyter/docker-stacks

Container

Fig 5. Docker image

Received: 21 Dec 2019 | Revised: 18 Jan 2020 | Accepted: 05 Feb 2020

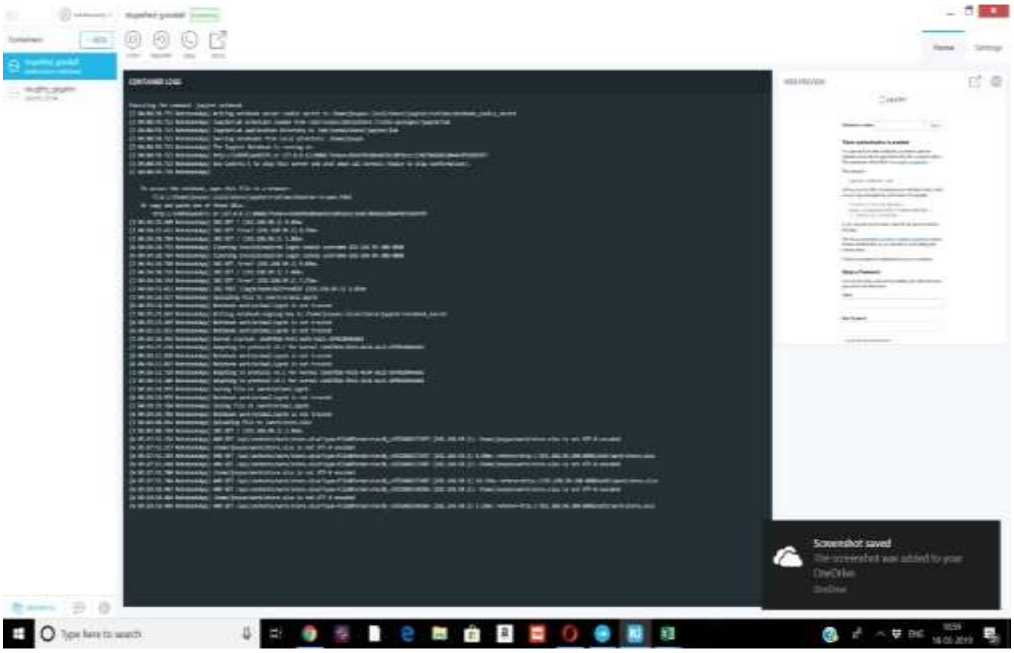


Fig 6. Docker Connection

The result from the metrics can serve as a benchmark while dealing with similar data in the near future. Major time, energy and computing resources can be saved by knowing beforehand the most optimal solution rather than trying different methods.

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 forare related to the CPU usage and utilization of container.

VIII. CONCLUSION AND FUTURE WORK

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 or nodes. Also, the paper tries to implement the dataset and analyse them using different 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 in two different containers and compare their performance metrics and analyse which container is suitable for time series.

REFERENCES

1. AL-Faifi, Abdullah & Song, Biao & Hassan, Mohammad & Alamri, Atif & Gumaei, Abdu. (2018). Performance prediction model for cloud service selection from smart data. Future Generation Computer Systems.10.1016/j.future.2018.03.015.
2. Yazhou Hu, Bo Deng, Fuyang Peng and Dongxia Wang, "Workload prediction for cloud computing elasticity mechanism," 2016 IEEE International Conference on Cloud Computing and Big Data

Analysis (ICCCBDA), Chengdu, 2016, pp. 244249

3. F. J. Baldan, S. Ramirez-Gallego, C. Bergmeir, F. Herrera and J. M.BenitezSanchez, "A Forecasting Methodology for Workload Forecasting in CloudSystems," in IEEE Transactions on Cloud Computing
4. A. Khan, X. Yan, S. Tao and N. Anerousis, "Workload characterization and prediction in the cloud: A multiple time series approach," 2012 IEEE Network Operations and Management Symposium, Maui, HI, 2012, pp. 1287-1294.
5. R. N. Calheiros, E. Masoumi, R. Ranjan and R. Buyya, "Workload Prediction Using ARIMA Model and Its Impact on Cloud Applications' QoS," in IEEE Transactions on Cloud Computing, vol. 3, no. 4, pp. 449-458, 1 Oct.-Dec. 2015.
6. Q. Fan and N. Ansari, "Application Aware Workload Allocation for EdgeComputing-Based IoT," in IEEE Internet of Things Journal, vol. 5, no. 3, pp.2146-2153, June 2018.
7. J. Panneerselvam, L. Liu, N. Antonopoulos and Y. Bo, "Workload Analysis for theScope of User Demand Prediction Model Evaluations in Cloud Environments,"2014 IEEE/ACM 7th International Conference on Utility and Cloud Computing, London, 2014, pp. 883-889.
8. W. Tang et al., "Workload characterization for MG-RAST metagenomic data analytics service in the cloud," 2014 IEEE International Conference on Big Data (Big Data), Washington, DC, 2014, pp. 56-63
9. H. Zhang, G. Jiang, K. Yoshihira and H. Chen, "Proactive Workload Management in Hybrid Cloud Computing," in IEEE Transactions on Network and Service Management, vol. 11, no. 1, pp. 90-100, March 2014Service", 3D Digital Imaging and Modeling, International Conference on, vol. , no., pp. 98103, Nov., 2017.
10. C. Wang, W. Hung and C. Yang, "A prediction based energy conserving resources allocation scheme for cloud computing," 2014 IEEE International Conference on Granular Computing (GrC), Noboribetsu, 2014, pp. 320-324.
11. S. Vakilinia, M. Cheriet and J. Rajkumar, "Dynamic resource allocation of smart home workloads in the cloud," 2016 12th International Conference on Network and Service Management (CNSM), Montreal, QC, 2016, pp. 367-370.
12. A. Adegboyega, "Time-series models for cloud workload prediction: A comparison," 2017 IFIP/IEEE Symposium on Integrated Network and Service Management (IM), Lisbon, 2017, pp. 298-307.
13. 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.
14. E. Caron, F. Desprez, and A. Muresan, "Forecasting for grid and cloud computing on-demand resources based on pattern matching," in Proc. 2nd IEEE Int. Conf. Cloud Comput. Technol. Sci., Dec. 2010, pp. 456–463.
15. G. E. P. Box,, G. M. Jenkins, and G. C. Reinsel, Time Series Analysis: Forecasting and Control, 4th ed. Hoboken, NJ, USA: Wiley, 2008
16. Hoffinann G, Trivedi K S, Malek M. A best practice guide to resource predicting for computing systems[J]. Reliability, IEEE Transactions on, 2007,56(4): 615-628.
17. Islam S, Keung J, Lee K, et al. Empirical prediction models for adaptive resource provisioning

in the cloud. Future Generation Computer Systems, 2012,28(1): 155-162.

18. Jiang Y, Perng C, Li T, et al. Asap: A self-adaptive prediction system for instant cloud resource demand provisioning, Data Mining (ICDM), 2011 IEEE 11th International Conference on. IEEE, 2011: 1104-1109

19. Jiang Y, Perng C, Li T, et al. Asap: A self-adaptive prediction system for instant cloud resource demand provisioning, Data Mining (ICDM), 2011 IEEE 11th International Conference on. IEEE, 2011: 1104-1109.

20. M. Calzarossa and G. Serazzi, “Workload characterization: A survey,” Proc. of the IEEE, vol. 81, no. 8, pp. 1136–1150, 1993.A. 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.

21. M. Calzarossa and G. Serazzi, “Workload characterization: A survey,” Proc. of the IEEE, vol. 81, no. 8, pp. 1136–1150, 1993.

22. N. Roy, A. Dubey, and A. Gokhale, “Efficient autoscaling in the cloud using predictive models for workload forecasting,” in Proc. 4th Int. Conf. Cloud Comput., Jul. 2011, pp. 500–507.

23. Q. Zhu and G. Agrawal, “Resource provisioning with budget constraints for adaptive in cloud.” International Journal of Grid and Utility Computing, vol 7, no.1, pp1221,2016.

24. R. Hyndman and G. Athanasopoulos, “Forecasting: Principles and practice,2013. [Online]. Available: http://otexts.org/fpp/

25. Reig G, Guitart J. On the anticipation of resource demands to fulfill the qos of saas web applications, Proceedings of the 2012 ACM/IEEE 13th International Conference on Grid Computing. IEEE Computer Society, 2012: 147-154.

26. Roy N, Dubey A, Gokhale A. Efficient autoscaling in the cloud using predictive models for workload predicting, Cloud Computing (CLOUD), 2011 IEEE International Conference on. IEEE, 2011: 500507.

27. Roy N, Dubey A, Gokhale A. Efficient autoscaling in the cloud using predictive models for workload predicting, Cloud Computing (CLOUD), 2011 IEEE International Conference on. IEEE, 2011: 500507.

28. S. Islam, J. Keung, K. Lee, and A. Liu, “Empirical prediction models for adaptive resource provisioning in the Cloud,” Future Gener. Comput. Syst., vol. 28, no. 1, pp. 155–162, 2012.

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

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. Technol., Mar. 2012, pp. 1127–1131.

31. Xu W, Zhu X, Singhal S, et al. Predictive control for dynamic resource allocation in enterprise data centers, Network Operations and Management Symposium, 2006. NOMS 2006. 10th IEEE/IFIP. IEEE, 2006: 115-12

32. Y. Chen, A. Ganapathi, R. Griffith, and R. Katz, “Towards understanding cloud performance tradeoffs using statistical workload analysis and replay,” EECS, UC Berkeley, Tech. Rep., 2010.

33. P. Akilandeswari, H.Srimathi Survey and Analysis on Task Scheduling in Cloud Environment, Indian

Journal of Science and Technology, Vol 9(37), 2016.

34. Y. Wang, R. Yang, T. Wo, W. Jiang, and C. Hu, “Improving utilization through dynamic VM resource allocation in hybrid cloud environment,” in Proc. 20th IEEE Int. Conf. Parallel Distrib. Syst. (ICPADS), Hsinchu, Taiwan, 2014, pp. 241–248
35. Z. Shen, S. Subbiah, X. Gu, and J. Wilkes, “CloudScale: elastic resource scaling for multi-tenant cloud systems,” in Proceedings of the 2nd ACM Symposium on Cloud Computing, ser. SOCC ’11, 2011, pp. 5:1–5:14.