International Journal of Future Generation Communication and Networking
Vol. 13, No. 1, (2020), pp. 1072-1080

ONTOLOGY BASED CONTEXT-AWARE MODEL FOR INTELLIGENT SCHEDULING IN FEDERATED CLOUD

Shishira S R[1], A. Kandasamy[2]
Research Scholar[1], Professor[2]
Department of MACS, NITK Surathkal [1, 2]
shishirasr@gmail.com[1], kandyi@nitk.ac.in[2]

Abstract

Cloud computing helps in serving the client request in a large manner. It serves n number of client requests. Services provided by the cloud would be servers, networks, storage and other applications. These services are proved by a cloud service provider which in turn given to a broker in a federated architecture. For fasten the execution and manage the workloads, it is very necessary to predict the workloads. Also, predicted workloads can be smoothly optimized for better usage without waving off the SLA that are agreed between the provider and clients. Thus, in the present paper, a context-aware model is provided which contains different system properties involved for effectively managing the workload in the federated clouds. This also can be used for the researchers for research purposes in the cloud domain.

Keywords: *Federated cloud, Conceptual-Framework, Prediction, Optimization*

1. Introduction

Cloud computing is a paradigm that serves clients request. It provides services to the consumer's on-demand basis, that is pay per usage [8][9]. Cloud computing consists of various number of services such as providing network, storage databases, servers, virtual machines, and various applications. Figure 1 shows the primer on cloud computing.

It also consists of various deployment models such as Private cloud-which is owned by a particular organization for its own purpose, Public cloud- which consists of resources which are shared publicly on-demand and Hybrid cloud- which is a combination of private and public cloud, where an organization or an institution shall store data on private cloud and offload some of the data to the public cloud[12]. But there exists a security risk in public cloud as it is a third-party cloud and it can be accessed by any single client. Recent advancement in cloud computing is a Federated cloud, which consists of more than one cloud service provider located geographically. Properties of various federated cloud advancements are described in the upcoming sections in the paper.

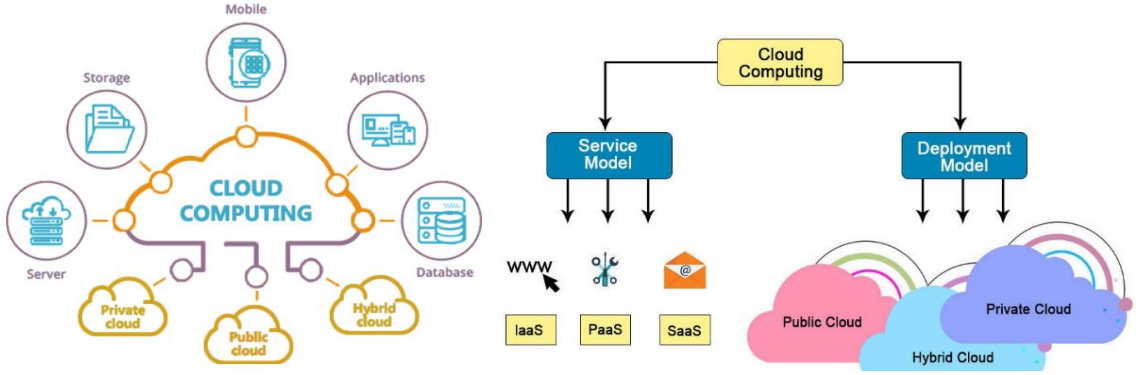

Figure1: Cloud computing system and deployment models

ISSN: 2233-7857 IJFGCN

Figure 2 describes the type of services provided by Cloud computing. IaaS-Infrastructure as a service deals with the storage, servers and network type of services. Examples include Microsoft Azure, Amazon Web service, Google computes engines. Similarly, Paas- Platform as a Service, provides various platforms such as Google app engines, database runtime environments, and SaaS- Software as a Service deal with google apps, Dropbox, etc.

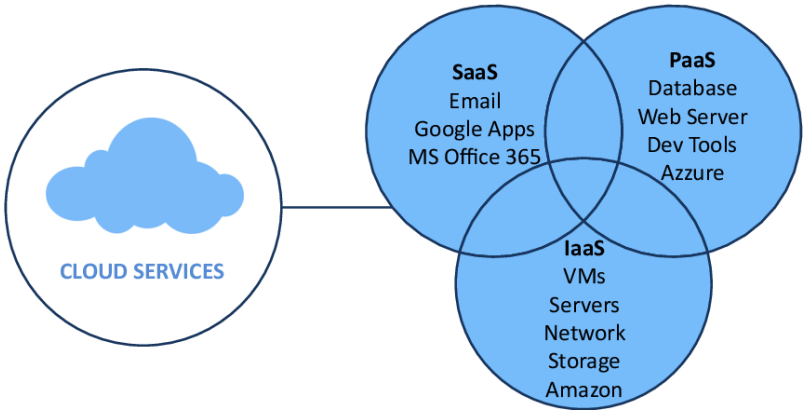

Figure 2: Types of cloud services [1]

This paper gives the context-aware model for Intelligent Management of workloads in the federated cloud. In cloud computing, initializing the cloud resources is not instantaneous, and involves a certain time of delay. And hence load prediction is very much necessary. Decisions can be made in advance to efficiently initiate the resources in the cloud environment. Also, optimizing certain cloud parameters help in reducing the delay in providing the service to the clients also profitable for the providers.[10][11] This the main motivation to do a conceptual framework that shall be helpful for the future researchers to work on the architecture provided in this paper.

Present paper deals with the following contributions described as follows,

- Provided a set of concepts which give the properties and relationships between the different modules in the cloud system.
- Provided a system architecture for cloud computing prediction and optimization.
- Provided an abstract design for the prediction module.
- Provided a different optimization algorithm for future research purposes.

2. Related works

Federated cloud environment refers to the n number of cloud service providers that are distributed geographically. It is a platform, where all the cloud service providers join together or share their resources to serve the client's request. Clients or consumers shall not be aware of the background process or setup involved in the federation. Cloud Infrastructure service is provided in the form of virtual machines during the user request [2]. This type of federation is helpful during the natural disaster while there exists a loss of data [3]. Cloud broker helps in processing the client request [4][5]. The broker will be choosing the best provider resource based on the client request [6].
Cloud workloads refer to the load produced by one or more number of infrastructures. The workload is basically a combination of jobs and tasks which are produced by a various number of client machines [7]. It’s very necessary to know the workload patterns to efficiently characterize and schedule a

ISSN: 2233-7857 IJFGCN

particular resource [15]. Workloads are classified into different processing models, computing environment, based on the generation, and also based on the applications [8].
Prediction of workloads helps in making the decisions in advance. This is because initializing the cloud instance is not instantaneous and it takes several minutes of delay [13]. And hence it is very necessary to decide the resources for a particular request for efficiently managing the client request.

From the literature, we observe that the workloads are scheduled based on the different types of workload patterns [17]. There exists optimized scheduling, green aware scheduling, delay-constraint, and cost-aware scheduling by optimizing certain cloud parameters which are helpful in maintaining the Quality of Service [7].

The rest of the paper is organized as follows. Section 3 briefs about the proposed system architecture, while section 4 gives the characteristics of federated cloud computing. section 5 explains the properties of workloads and section 6 provides the abstract design of the prediction module. Finally, section 7 gives the conclusion.

3. Proposed System Architecture

This section provides the conceptual framework for the management of workloads in the federated environment. We have provided this framework to understand the properties of each module in the system and the inter-relation between them. Figure 3 shows the framework which consists of a workload generator which are real or synthetic workloads from the client-side. The workloads are then analyzed and features are extracted for testing and training during the prediction. The broker helps in finding the best cloud service provider by optimizing the cloud parameters such as response time, delay, execution time, etc.

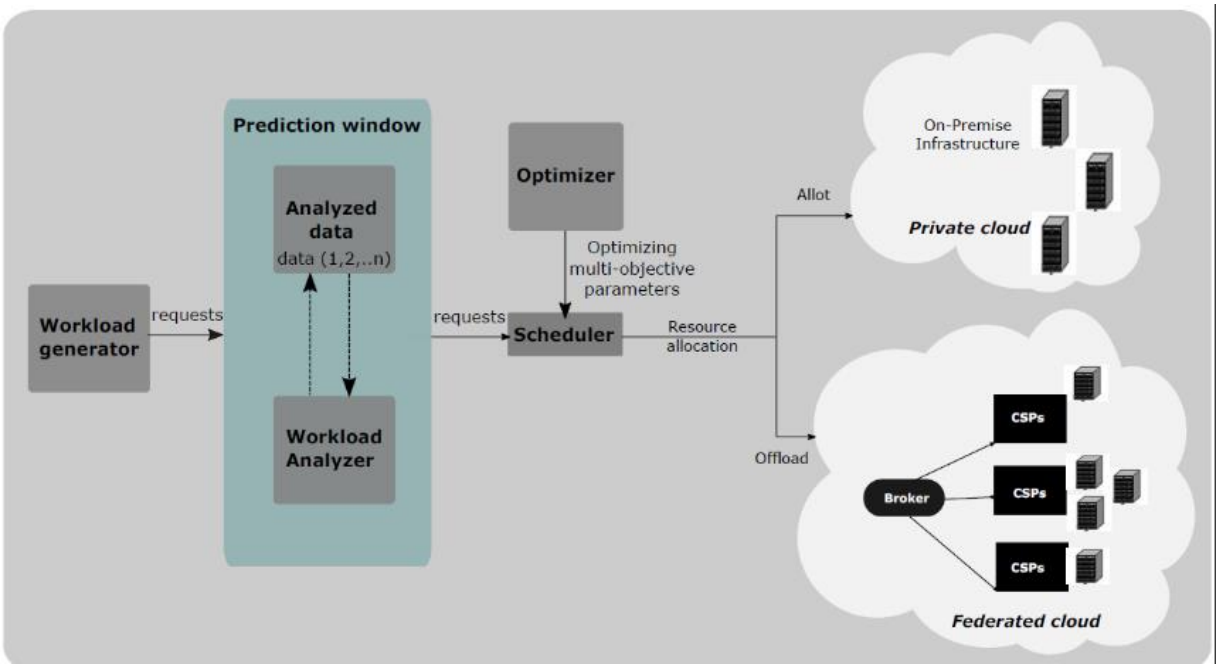

Figure 3: Conceptual framework for management of workloads

These properties help the scheduler to have the information in advance intelligent and helps in efficiently allocate the resources to the particular client. Once the scheduler allows the requests if

ISSN: 2233-7857 IJFGCN

storage capacity is not entitling in the on-premise infrastructure, it shall be offloaded into the federated cloud.

4. Characteristics of Federated cloud computing

Different types of entity involved in a cloud federation. The main characteristics of cloud federation involve scalability, fault tolerance and increased performance [14]. Also, the different coupling is involved namely loosely coupled, partially coupled and tightly coupled based on the levels in the cloud environment [16]. We have classified cloud federation architectures into various types,

4.1 Cloud bursting Architecture

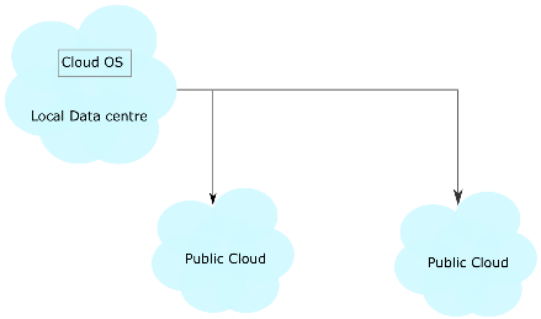

Figure 4: Bursting architecture

During the resource shutdown, the client can demand or bursts the data for the service from the third-party cloud which is available publicly.

4.2 Cloud Broker architecture:

Brokering helps in sending and serving the client request between consumer and provider.

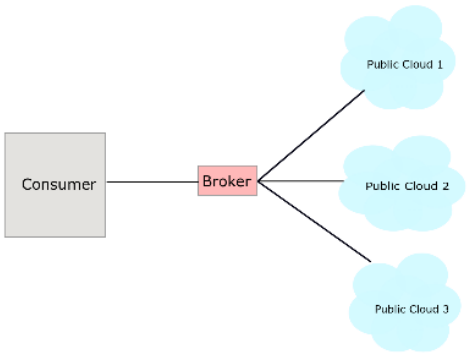

Figure 5: Broker architecture

4.3 Aggregated cloud architecture:

As explained earlier, if there exists a shut down in resources or there exists a resource constraint, cloud providers aggregate their resources to serve the request.

ISSN: 2233-7857 IJFGCN

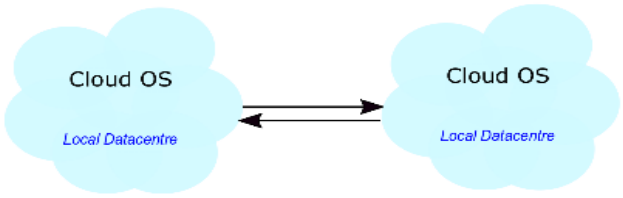

Figure 6: Aggregated architecture

5. Properties of workload

During the literature review, we observed that there doesn't exists a particular definition for workloads as it usually defined various applications. Hence, we have defined the workload in cloud application as the combination of jobs, which in turn distributed as tasks at the granular level.

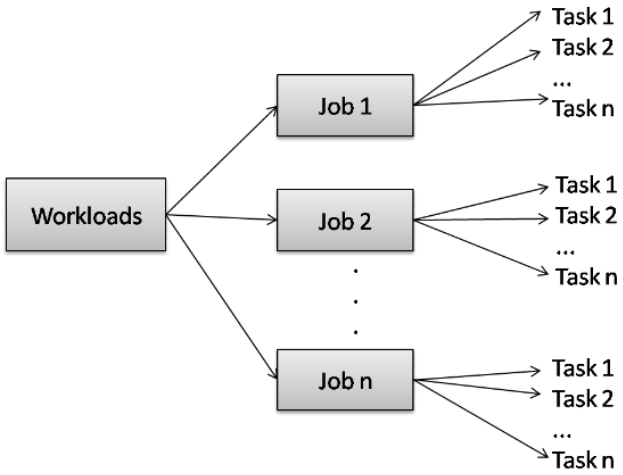

Figure 7: Distribution of workloads

6. System module: Abstract design

6.1 Forecasting properties

Figure 8 shows the abstract design of the Prediction module. Historical workload information is analyzed and fed into the load predictor, which outputs the future demands and the estimated number of resources. Thus, when a user sends the request also called workloads, Application Provisioner will communicate with an incoming estimated number of resources from the predictor and requests for the scheduling of resources.

- Workload analyzer: User historical data is processed, analyzed and fed into the load predictor.
- Admission control: User submitted task is processed and sent to Provisioner for further process.
- Load Predictor: Outputs the future demands and the estimated number of resources.

ISSN: 2233-7857 IJFGCN

- Application Provisioner: When a user sends the request also called workloads, the Application Provisioner will communicate with an incoming estimated number of resources from the predictor and requests for the scheduling of resources.

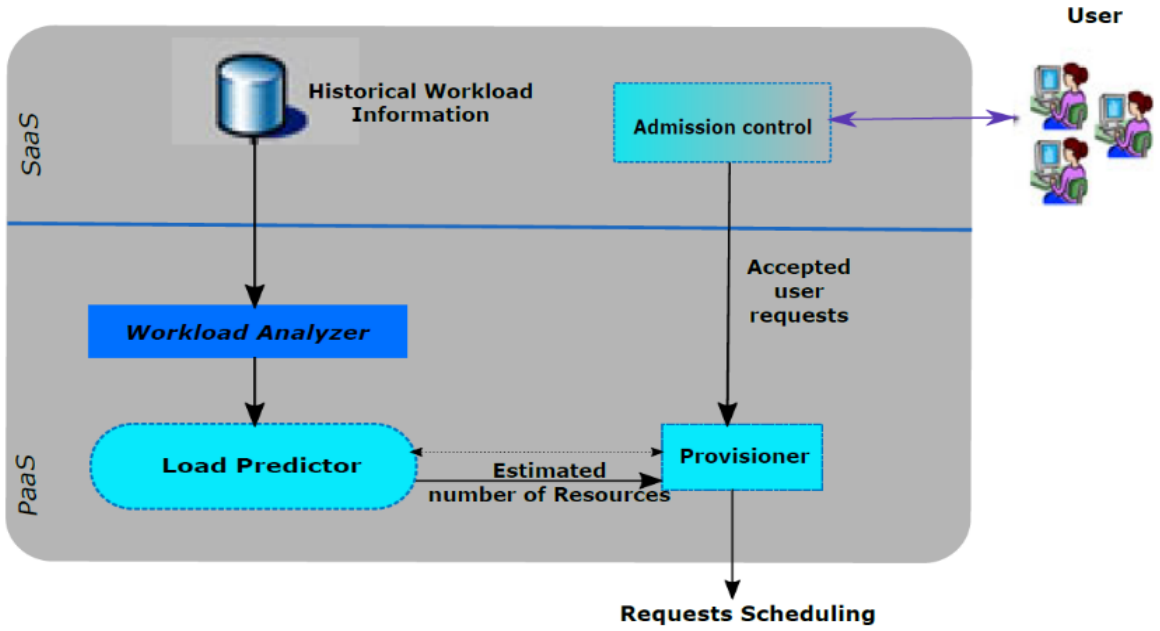

Figure 8: Abstract design of Prediction module

Workload prediction problem is resolved effectively by including deep learning techniques. Having a historical workload in the database for a specific time periods, one can efficiently predict the incoming workloads. Machine learning helps in learning the patterns and extract the features to make the decisions. Historical workload data is considered in a training window which is used to predict the future workloads. Testing data which is included in a prediction window which is after prediction.
Figure 9 shows the training and prediction windows.

Figure 9: Prediction and training window

Algorithm1 gives the steps involved in the prediction of workloads.
Algorithm1
//Task: Prediction based on historical workload

1. Input1<-Workload sample
2. Input2<- Historical data
3. Include pattern matching technique and extract features.
4. For extracted feature<-use data mining techniques for efficient prediction.

6.2 Parameter optimization and scheduling

ISSN: 2233-7857 IJFGCN

Figure 10 gives the classification of optimization algorithms. Optimization parameters considered in Ant Colony Optimization (ACO) and Particle Swarm Optimization (PSO) are Task scheduling, SLA aware, Energy Aware. Genetic Algorithm (GA) considers Load balancing, SLA and energy-aware. Task scheduling is prioritized in League Championship Optimization (LCO).

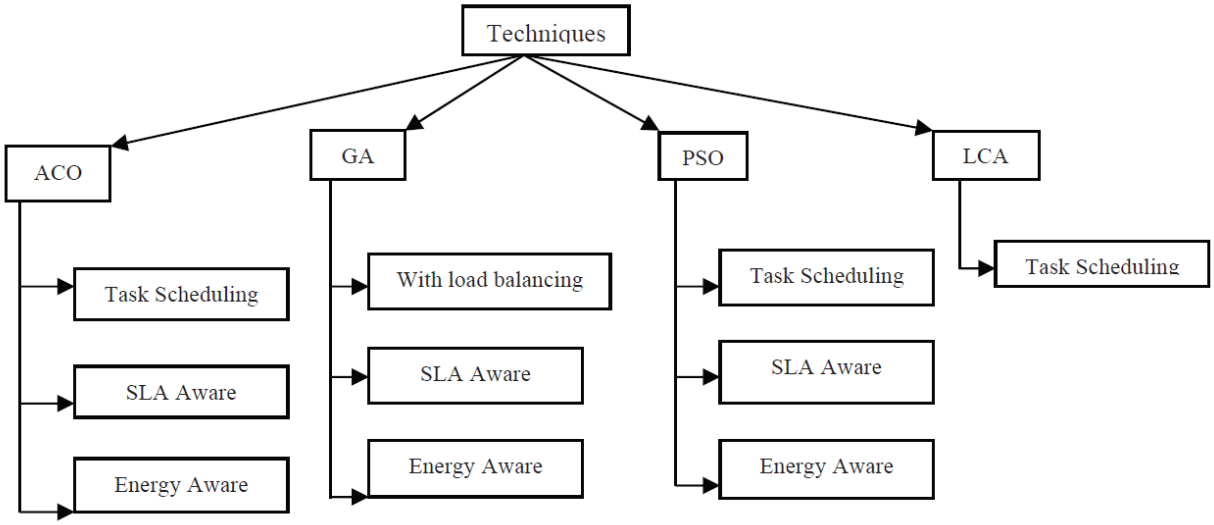

Figure 10: Classification of optimization algorithms

Optimization parameters are divided based on the provider-based consumer desire. Table gives provider constraint parameters.

Table 1: Provider constraint optimization parameters

Criteria	Provider-constraint
Resources	Maximize the utilization
Throughput	Maximize the execution per unit time
Priority	Based on the deadline of the task
Dependency	Precedence order among the task
Budget	Restricted on the total cost on the task

Table 2: Consumer-constraint optimization parameters

Criteria	Consumer-constraint
Make span	Minimization for faster execution
Tardiness	Must be zero for an optimal scheduling
Fairness	Each job to share equal number of resources
Waiting Time	Time between submission and execution time must be minimal
Turnaround Time	Complete execution time must be minimal

Steps involved in optimizing different service parameters are given in Algorithm 2.

Algorithm2

//Task: Optimizing QoS parameters

1. Input<- predicted workloads
2. Require<- Define quality of service parameters

ISSN: 2233-7857 IJFGCN

3. Process request<- Use meta-heuristic optimization technique for efficient processing of the workloads.
4. Information fed into the scheduler to allot the resources.

7. Conclusion

A Federated cloud environment is a recent advancement in the cloud deployment model. When there is a resource shutdown, one can use a federated cloud which consists of n number of service providers who provide services whose data centers are located geographically. Load prediction also helps in efficiently manage the workloads and maximize the execution time which is helpful for both provider and consumer. Hence, this paper provides a conceptual framework that contains system modules, each explaining its properties. Abstract design of each muddle is also provided with the future research directions which shall be helpful for the researchers in the area of cloud computing.

References

1. Fatima, A., Javaid, N., Sultana, T., Hussain, W., Bilal, M., Shabbir, S., & Ilahi, M. (2018). Virtual machine placement via bin packing in cloud data centers. *Electronics*, *7*(12), 389.
2. Bohn, R. B., Messina, J., Liu, F., Tong, J., & Mao, J. (2011, July). NIST cloud computing reference architecture. In *2011 IEEE World Congress on Services* (pp. 594-596). IEEE.
3. Shishira, S. R., Kandasamy, A., & Chandrasekaran, K. (2017, January). Workload scheduling in cloud: A comprehensive survey and future research directions. In *2017 7th International Conference on Cloud Computing, Data Science & Engineering-Confluence* (pp. 269-275). IEEE.
4. Rochwerger, B., Breitgand, D., Levy, E., Galis, A., Nagin, K., Llorente, I. M., ... & Ben-Yehuda, M. (2009). The reservoir model and architecture for open federated cloud computing. *IBM Journal of Research and Development*, *53*(4), 4-1.
5. Najm, M., & Tamarapalli, V. (2019, January). A cost-aware algorithm for placement of enterprise applications in federated cloud data center. In *Proceedings of the 20th International Conference on Distributed Computing and Networking* (pp. 510-510).
6. Badger, L., Grance, T., Patt-Corner, R., & Voas, J. (2012). *Cloud computing synopsis and recommendations: Recommendations of the national institute of standards and technology*. CreateSpace Independent Publishing Platform.
7. Shishira, S. R., Kandasamy, A., & Chandrasekaran, K. (2017, November). Workload Characterization: Survey of Current Approaches and Research Challenges. In *Proceedings of the 7th International Conference on Computer and Communication Technology* (pp. 151-156).
8. Shishira, S. R., Kandasamy, A., & Chandrasekaran, K. (2016, September). Survey on meta heuristic optimization techniques in cloud computing. In *2016 International Conference on Advances in Computing, Communications and Informatics (ICACCI)* (pp. 1434-1440). IEEE.
9. Javadi, B., Abawajy, J., & Buyya, R. (2012). Failure-aware resource provisioning for hybrid Cloud infrastructure. *Journal of parallel and distributed computing*, *72*(10), 1318-1331.
10. Selvanathan, N., Jayakody, D., & Damjanovic-Behrendt, V. (2019, August). Federated Identity Management and Interoperability for Heterogeneous Cloud Platform Ecosystems. In *Proceedings of the 14th International Conference on Availability, Reliability and Security* (pp. 1-7).
11. Buyya, R., Ranjan, R., & Calheiros, R. N. (2010, May). Intercloud: Utility-oriented federation of cloud computing environments for scaling of application services. In *International*

ISSN: 2233-7857 IJFGCN

Conference on Algorithms and Architectures for Parallel Processing (pp. 13-31). Springer, Berlin, Heidelberg.

12. Lucas-Simarro, J. L., Moreno-Vozmediano, R., Montero, R. S., & Llorente, I. M. (2013). Scheduling strategies for optimal service deployment across multiple clouds. *Future Generation Computer Systems*, *29*(6), 1431-1441.
13. Rosa, M. J., Araüjo, A. P., & Mendes, F. L. (2018, December). Cost and Time Prediction for Efficient Execution of Bioinformatics Workflows in Federated Cloud. In *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)* (pp. 1703-1710). IEEE.
14. Najm, M., & Tamarapalli, V. (2019, January). A cost-aware algorithm for placement of enterprise applications in federated cloud data center. In *Proceedings of the 20th International Conference on Distributed Computing and Networking* (pp. 510-510).
15. Comden, J., Yao, S., Chen, N., Xing, H., & Liu, Z. (2019). Online Optimization in Cloud Resource Provisioning: Predictions, Regrets, and Algorithms. *Proceedings of the ACM on Measurement and Analysis of Computing Systems*, *3*(1), 1-30.
16. Chaufournier, L., Ali-Eldin, A., Sharma, P., Shenoy, P., & Towsley, D. (2019, April). Performance evaluation of Multi-Path TCP for data center and cloud workloads. In *Proceedings of the 2019 ACM/SPEC International Conference on Performance Engineering* (pp. 13-24).
17. Vecchiola, C., Calheiros, R. N., Karunamoorthy, D., & Buyya, R. (2012). Deadline-driven provisioning of resources for scientific applications in hybrid clouds with Aneka. *Future Generation Computer Systems*, *28*(1), 58-65.

ISSN: 2233-7857 IJFGCN