Cost Effective Resource Provisioning Approach for Cloud Environments

Rajesh Agrawal (Technical Architect, Technowings Pune)

Shlok Sabarwal (Research Intern, Technowings Pune)

Abstract*— *Cost Effective Resource Provisioning Approach is the processes of reducing your overall cloud spend by identifying mismanaged resources, eliminating waste, reserving capacity for higher discounts, and Right Sizing computing services to scale. The main aim of this paper is to implement a tool that can able to monitor and analyse cost patterns on cloud accounts(like Amazon Web Services) and capable of giving suggestions about cost optimization and cost containment.

It can guide the businesses to understand how much they are spending on cloud computing services by providing regular emails; alerts and a dashboard showing monthly spend to date, and a monthly estimate. This helps companies to manage all their cloud spend and understand when cost overages are happening (not after the fact). It also offers an optimizer service to identify other ways to save money on the cloud and evaluate the performance of workloads on EC2 instances.

***Keywords: Cloud Computing, Cloud Resource Optimizer, Cost Advantages, Private Clouds, Performance Evaluation, Price Reduction,* EC2 instances.**

I. INTRODUCTION

Cloud computing is dynamically evolving, nonetheless creating important impact in numerous areas of life. Additional enterprises are migrating to the cloud and new technological developments are going down. Cloud computing offers many edges to organizations, enterprises and little businesses alike. Corporations have to be compelled to look on the far side this seriously consider the important price of incorporating the Cloud in their businesses. This can be aimed toward serving to corporations analyse many characteristics of their own business furthermore as pre-existing IT resources to spot their favourability within the migration to the Cloud design.

Cost optimisation [3] in cloud computing is that the most up-to-date topic studied by the investigator. Just like the author of [4] projected an evaluation algorithmic rule for cloud computing resources and also the Author [5] projected a system named Dyna to cut back the expected financial cost [2]. Once more to characterize the cost-performance of cloud resources, the author of [6] used the Performance Cost Ratio (PCR) metric. By extending the Gustafson's fixed-time scaling within the context of cloud, and, investigate fixed-cost-time scaling of applications and show that victimisation resources with higher PCR yields higher cost-time performance. Each cloud supplier has totally different evaluation methods for computing resources.

Subhas Chandra Misra et al. [7] offers a framework for serving to corporations analyse many characteristics of their own business furthermore as pre-existing IT resources to spot their favourability within the migration to the Cloud design. The System provides Associate in nursing in-depth analysis of the money perspective of CC in a very lucid and straightforward manner alongside the subjective decision-making tool to seek out the suitableness of an organization for adopting CC.

In the course of a cloud, implementation users have the flexibility to settle on the EC2 instance sort that has the acceptable mixture of resources for the target application and employment. They apply charges supported resource utilization, however it's terribly high. the most purpose of the system is to form a non-public cloud (testbed) by victimisation (Amazon Account) alongside watching essential resources like RAM, CPU, memory, bandwidth, partition data, running method data and utilization and swap usages, etc. Also, advocate the value reduction strategy. The projected technique offers Associate in nursing economical task of workloads to servers to decrease price by exploiting the resource utilization. The system will able to monitors VMs (EC2 Instances) on personal clouds like Amazon or Google to supply low infrastructure prices.

Types of cloud services: IaaS, PaaS, serverless and SaaS.

- **Infrastructure as a service (IaaS)-**The most basic category of cloud computing services. With IaaS, customer rent IT infrastructure—servers and virtual machines (VMs), storage, networks, operating systems—from a cloud provider on a pay-as-you-go basis.
- **Platform as a service (PaaS)** - Platform as a service refers to cloud computing services that supply an on-demand environment for developing, testing, delivering and managing software applications. PaaS is designed to make it easier for developers to quickly create web or mobile apps, without worrying about setting up or managing the underlying infrastructure of servers, storage, network and databases needed for development.
- **Software as a service (SaaS)**- Software as a service is a method for delivering software applications over the Internet, on demand and typically on a subscription basis. With SaaS, cloud providers host and manage the software application and underlying infrastructure and handle any maintenance, like software upgrades and security patching. Users connect to the application over the Internet, usually with a web browser on their phone, tablet or PC.

II. LITERATURE SURVEY

Subhas Chandra Misra et al. [1] gives a framework for helping companies analyze distinctiveness of their business in addition to previous IT resources to recognize their favorability in the relocation to the Cloud Architecture. A common Return on Investment (ROI) representation has also been explained for consideration of the different indefinable effect of Cloud Computing, despite the cost. The analysis presented herein provides a much broader viewpoint and approaching into Cloud Computing to its likely adopters.

A] Title- Monetary cost optimizations for hosting Workflow-as-a-Service in IaaS Clouds.

Description - Recently, we've got witnessed workflows from science and alternative data-intensive applications rising on Infrastructure-as-a-Service (IaaS) clouds, and plenty of progress service suppliers giving Workflow-as-a-service (WaaS). The most important concern of WaaS suppliers is to attenuate the financial value of death penalty workflows within the IaaS clouds. The choice of virtual machines (instances) sorts considerably affects the financial value and performance of running a progress. Moreover, IaaS cloud surroundings is dynamic, with high performance dynamics caused by the interference from coinciding executions and value dynamics like spot costs offered by Amazon EC2. Therefore, we tend to argue that WaaS suppliers ought to have the notion of giving probabilistic performance guarantees for individual workflows to expressly expose the performance and value dynamics of IaaS clouds to users. It tends to develop a planning system known as Dyna to attenuate the expected financial value given the user-specified probabilistic point guarantees. Dyna includes AN A-based instance configuration methodology for performance dynamics, and a hybrid instance configuration refinement for victimization spot instances. Experimental results with 3 scientific progress applications on Amazon EC2 and a cloud machine demonstrates:

(1) The flexibility of Dyna on satisfying the probabilistic point guarantees needed by the users
(2) The effectiveness on reducing financial value compared with the prevailing approaches.

What they have proposed?

1) They need projected Workflow-as-a-service conception during a Software-as-a-Service branch of cloud computing.
2) It guarantees probabilistic performance that is nothing however dynamic deadlines.

What we have referred?

1) Settled deadlines are used for static task execution however given the dynamic nature of elastic cloud computing we tend to need additional rigorous notion for deadlines.
2) Value improvement by selecting the acceptable form of instance and dynamicity of progress.

B] Minimum-Cost Cloud Storage Service Across Multiple Cloud Providers [IEEE 2017]

Liu et al. [3] provides a model to decrease the payment value of shoppers and at constant time is guarantee their SLOs (service level objective) with the globally distributed information canters belonging to totally different CSPs with different resource unit costs. The price minimisation drawback will be solved by victimization whole number programming.

C] Title- Cost-aware cloud profiling, prediction, and provisioning as a service.

Description - Here we tend to Scalable Cost-Aware Cloud Infrastructure Management and Provisioning (SCRIMP) - a service-based system that permits application developers and users to dependably source the task of provisioning cloud infrastructure. we tend to show that by understanding application necessities, predicting dynamic market conditions, And mechanically provisioning infrastructure in keeping with user-defined policies and time period conditions that our approaches will scale back prices by an order of magnitude once victimization industrial clouds whereas additionally up execution performance.

What they have proposed?

1) Harnessing the Cloud for Research Computing.
2) Identification Application Performance on Heterogeneous Infrastructure.
3) Predicting Market Conditions and Establishing Probabilistic Guarantees.
4) Provisioning and Managing Instances during a Multi-cloud Environment.

What we have referred?

1) Given the character of complexities in varied cloud computing services furthermore as among a particular cloud computing service for instance AWS, 3 areas particularly identification, prediction and provisioning are integrated to deliver a versatile cloud computing expertise.
2) we'd like to use a middleware but, effective and economical use of cloud infrastructure is technically difficult, and these challenges are displeased during a multi cloud surroundings.

Ryan Chard et al. [2] proposes a Scalable Cost-Aware Cloud Infrastructure Management and Provisioning (SCRIMP), a service-dependent scheme that enables application developers and users to consistently deployed the job of provisioning cloud infrastructure. It shows that by accepting application necessities, forecasting dynamic market situation, and repeatedly provisioning infrastructure along with user defined policies and real-time conditions that this method be able to decrease costs by an order of magnitude when using commercial clouds while also improving execution performance and efficiency.

D] A Pricing Algorithm for Cloud Computing Resources

G. Tang et al.[7] Projected an evaluation algorithmic rule for cloud computing resources. The authors projected the cloud agent model as a resource agency from a world perspective that provides analysis and steerage for all members.

Chi Chow et al. [8] presents a planning system referred to as Dyna to attenuate the expected financial value given the user-specified probabilistic point guarantees. Dyna includes AN A-based instance configuration methodology for performance dynamics and a hybrid instance configuration refinement for victimization spot instances. Experimental results with 3 scientific progress applications on Amazon EC2 and a cloud machine show (1) the capability of Dyna on satisfying the probabilistic point guarantees needed by the users; (2) the potency of reducing financial value compared with the prevailing approaches.

E] Identification of a company's suitability for the adoption of cloud computing and modelling its corresponding Return on Investment

Chandra Misra et al. [9] gives a framework for helping companies analyze several characteristics of their own business as well as pre-existing IT resources to identify their favourability in the migration to the Cloud Architecture. A general Return on Investment (ROI) model considers various intangible impacts of Cloud Computing, apart from the cost. The analysis presented herein provides a much broader perspective and insight into Cloud Computing to its prospective adopters.

Advantages

- The system provides an in-depth analysis of the financial perspective of CC in a very lucid and simple manner.
- It provides both the objective as well as the subjective decision-making tool to find the suitability of a company for adopting CC.

Xinhui Li et al. [3] fills the gap in between cost computation and examination in Cloud environment using suits of metrics and formulas for the computation of Cloud entire Cost of Ownership (TCO) and consumption Cost. The elastic characteristic of Cloud infrastructure and extensively used virtualization technology in the Cloud are taken into consideration. This gives a base for assessing financial effectiveness of Cloud and offers the suggestion for cost optimization of Cloud. This calculation and analysis strategy used in the interior Cloud environment and show firstly its investigation ability on the cost allocation and use imbalance factor.

The exploration in the performance of cloud computing services for technical computing workloads is explained in [5]. They measure the occurrence in actual technical computing workloads of Many-Task Computing (MTC) users. Again users who use insecurely coupled applications including lots of jobs to achieve their scientific goals. The outcome of the proposed system shows that the present clouds required a sequence of magnitude in performance, development to be helpful to the scientific community, and show which development should be measured first to deal with this inconsistency between offer and demand.

Krishnadas Nath et al. [6] gives a structure in studying the cost - benefits to make a decision upon the adaptability of cloud computing. A variety of factors are taken into consideration an organization likes the number of servers, power requirements and other computational/non - computational resources. A three-layer strategy is used for the cost-benefit learning and draws imminent on the productivity when an organization transfer to cloud computing in each layer. Base cost estimation, data pattern based cost estimation and project specific cost estimation are the three mentioned layers. These layers are intended to give different stages of decision making to aid managers in their attempt to find out the prospects of using cloud computing in their organization.

III. PROPOSED SYSTEM

In today's competitive market, measuring application success as a "user interface" alone is no longer enough. Poor availability costs revenue, loyalty and brand image. Application leaders are shifting business-centric metrics to service level management (SLM) to bring IT closer to business. The main intention is to build up a scalable CLOUD explanation which is able to carry requirements of Stock Broking firm with no negotiation on performance, scalability, and cost. Following figure shows the architecture of the proposed system which can monitor VMs (EC2 Instances) on private clouds like Amazon or Google and provide solutions to reduce infrastructure costs.

In our proposed model, we will be doing

1. **Cloud Setup -** Creating private cloud (test bed) by using (Amazon Account)

2. **Resource Monitoring -** monitoring critical resources like RAM, CPU, memory, bandwidth, partition information, running process information and utilization and swap usages etc.

3. **Authentication and authorization –** we need to connect to existing user's amazon account using user id and password and fetch all the performance matrix like CPU, RAM, storage etc.

4. **Testing -** In order to evaluate the performance of complete setup, need to deploy resource monitoring and load balancing tools on test bed and evaluate need of available resources.

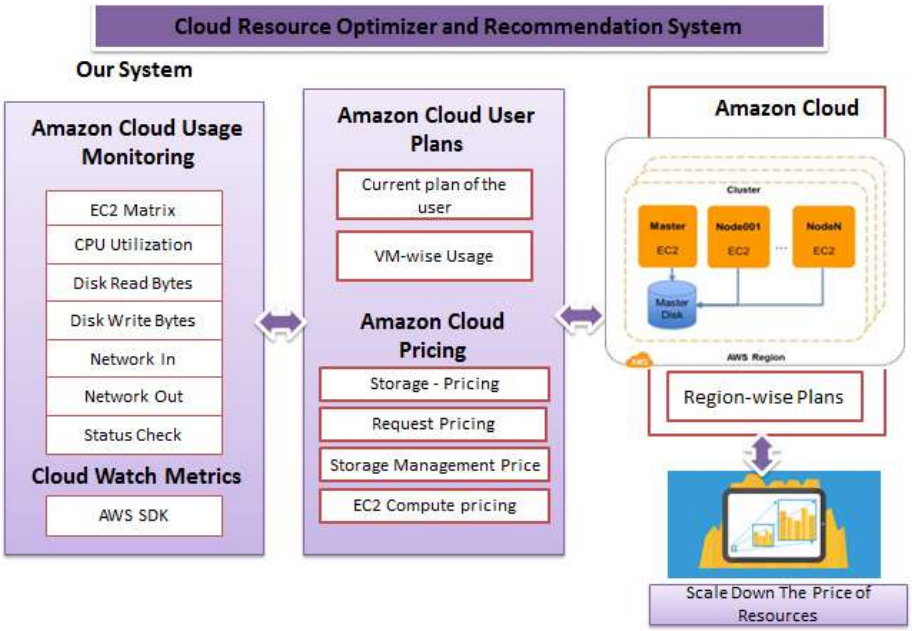

Figure: - System Architecture

Modules:

1. Resource Monitoring of Cloud Nodes:

a. User should be able to view CPU and RAM usage utilization of amazon ec2 nodes.

b. CPU and RAM utilization statistics should be dynamic and should refresh every second.

2. Select Cloud Plans for popular clouds like amazon. Cost of service depends on region of server, memory usage, CPU etc. Cloud service providers charge for following services which need to be added in system.

a. Storage – Pricing
b. Request Pricing
c. Storage Management Price
d. CPU pricing

3. Monitor account wise VM Usage of following parameters

a. CPUUtilization
b. DiskReadBytes
c. DiskWriteBytes
d. NetworkIn
e. NetworkOut
f. StatusCheck

4. Propose efficient resource utilization

1. by suggesting memory cutdown,
2. by suggesting cpu cutdown,
3. by suggesting storage cut down.

The figure 2 shows the exact flow of the proposed system for cost optimization

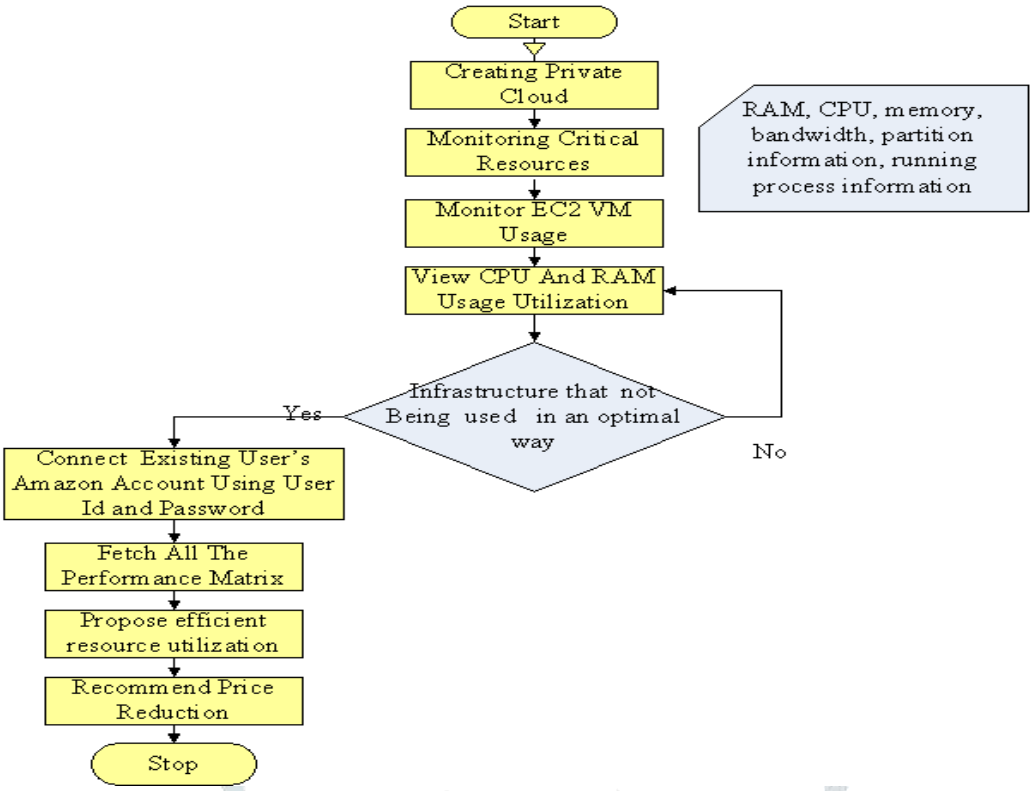

Figure: - System Flow

IV. ALGORITHM USED

A. AES Algorithm

AES is widely used for large size data encryption. AES is one of the symmetric key block cipher algorithm used worldwide for data encryption. It's particular structure of encrypting and decrypting data make it more secure so that it cannot be hacked. AES can deal with different key sizes such as AES 128, 192 and 256 bit and each of these ciphers has 128 bit block size.

The key size used for an AES cipher specifies the number of repetitions of transformation rounds that convert the input, called the plaintext, into the final output, called the cipher text. The number of cycles of repetition is as follows:

- 10 cycles of repetition for 128-bit keys.
- 12 cycles of repetition for 192-bit keys.
- 14 cycles of repetition for 256-bit keys.

Each round consists of several processing steps, each containing four similar but different stages, including one that depends on the encryption key itself. A set of reverse rounds are applied to transform cipher text back into the original plaintext using the same encryption key.

B. Support Vector Machine (SVM)

- SVM is a powerful classifier that is able to distinguish two classes. SVM classifies the test image in to the class with highest distance up to the neighboring point in the training.
- SVM training algorithm built a model that predict whether the test image fall into this class or another.
- SVM necessitate a vast training data to decide a decision boundary and computing cost is very high although we are using single pose (frontal) detection.
- The SVM is a learning algorithm for classification which attempt to discover the finest distinguishing hyper plane which minimize the error for unseen patterns.

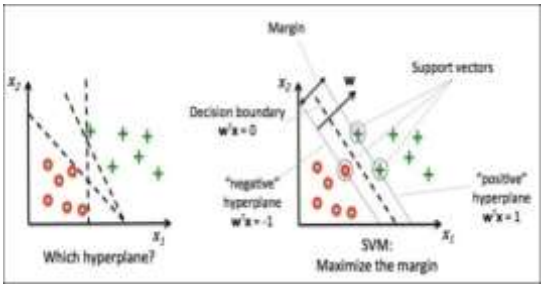

Figure: Distinguishing Hyper Plane to Minimize the Error

- The data which cannot be distinguished the input is mapped to high-dimensional attribute space where they can be separated by a hyper plane. This projection is well performed by means of kernels.

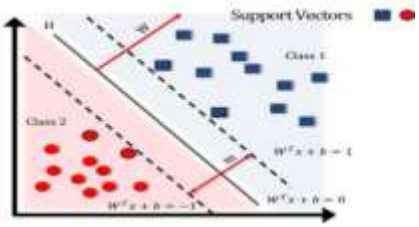

Figure: Separating Hyper Plane by Equation

- If training set of samples and the equivalent resultant values {-1, 1}. So SVM intend to get the best separating hyper plane specified by the equation WTx+b that make use of the distance between the two classes as shown in above figure.

V. EXPERIMENTAL RESULT

The AWS EC2 Nodes are created using AWS administrator account. The usage data of each node is fetched from the AWS monitoring tables and stored into local database. Application table name AWS statistics stores the values of below data parameters for each EC2 node created.

CPUUtilization, DiskReadBytes, DiskWriteBytes, NetworkIn, NetworkOut, StatusCheck.

The existing systems have fixed threshold values. Additional user load is shifted to next node for load balancing whatever be the load of that node. Below figure 8 gives the load balancing structure of existing system.

Table I: Existing System Load Balancing Table

Existing System				
	Userload	**Additional UserLoad**	**Fixed Threshold**	**Next Node Load Balancing**
Node1	50	10	50	Node2
Node2	10	50	50	Node3
Node3	20	10	50	Node4
Node4	20	10	50	Waits

Below figure shows the plot of the same.

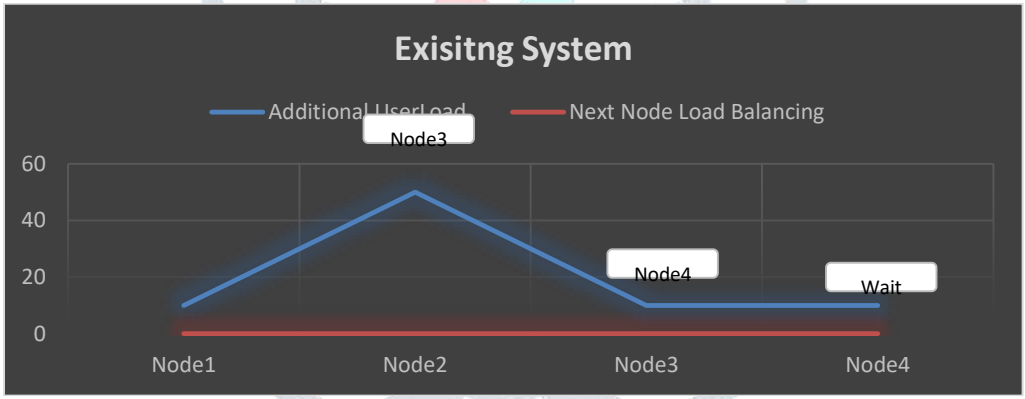

Figure: - Plot of Load Balancing Table for Existing System

The proposed method can monitor EC2 Instances on private clouds with reduced infrastructure cost. The system can also help to optimal utilization of cloud resources. In proposed system threshold values are dynamically updated and Inactive User count is decreased.

Table II: Next Load Balancing Table for Proposed System

Cloud Cost Analyzer						
	Userload	**Additional UserLoad**	**Fixed Threshold**	**Next Node Load Balancing**	**% RAM Fr**	**% CPU Free**
Node1	50	10	Threshold Values are	Node3	20	45
Node2	10	50	dynamically updated	Node3	10	45
Node3	20	10	and Inactive User	Node3	50	45
Node4	20	10	count is decreased	Node3	30	45

Here, node 3 is selected as it contains large amount of free space and CPU utilization. The offloading can done at 2 levels i.e. Resource Level and User Level. The below figure gives percentage of free RAM and CPU usage for shifting the load to next node.

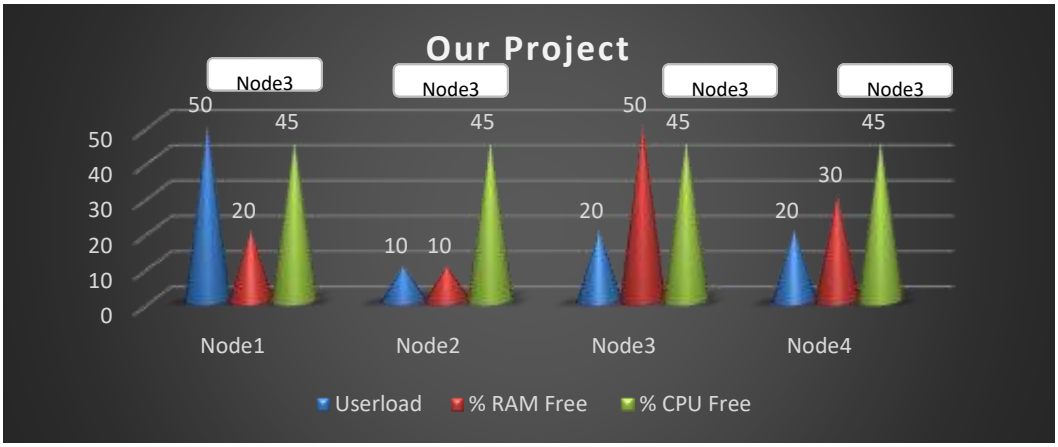

Figure: - Next Load Balancing Table for Proposed System

VI. CONCLUSION

Cloud computing refers to a paradigm for accessing computing resources which is becoming increasingly popular. Despite the very fact that having a cloud infrastructure is typically cheaper than maintaining a physical information center, owners massive and sophisticated IT infrastructure would possibly incur large prices. Therefore, the matter of value optimization in cloud computing is changing into more and more vital.

This system analyses the matter of price optimization in cloud computing. We tend to additionally evaluate the performance of the resource observance. This method monitors the VM node on the non-public cloud to cut back infrastructure prices from the customer's prospective.

REFERENCES

[1] Keith R. Jackson,Krishna Muriki,Shane Canon, Shreyas Cholia, John Shalf Harvey J. Wasserman, and Nicholas J. Wrig,"Performance Analysis of High Performance Computing Applications on the Amazon Web Services Cloud",2nd IEEE International Conference on Cloud Computing Technology and Science.

[2] Amelie Chi Zhou, Bingsheng He and Cheng Liu Nanyang Technological University"Monetary Cost Optimizations for Hosting Workflow-as-a- Service in IaaS Clouds",IEEE TRANSACTIONS ON CLOUD COMPUTING, VOL. X, NO. X, AUGUST 2014.

[3] https://cloudcheckr.com/cloud-cost-management/cloud-cost-management-cost-optimization-best-practices-for-reducing-cloud-bills/

[4] H. Li, J. Liu and G. Tang, A Pricing Algorithm for Cloud Computing Resources, Proc. Int. Conference on Network Computing and Inform. Security, 2011.

[5] Amelie Chi Zhou, Bingsheng He and Cheng Liu Nanyang Technological University,"Monetary Cost Optimizations for Hosting Workflow-as-a-Service in IaaS Clouds",IEEE TRANSACTIONS ON CLOUD COMPUTING, VOL. X, NO. X, AUGUST 2014.

[6] Sunimal Rathnayake ; Lavanya Ramapantulu ; Yong Meng Teo , "Cost-Time Performance of Scaling Applications on the Cloud" , 2018 IEEE International Conference on Cloud Computing Technology and Science (CloudCom).

[7] Subhas Chandra Misra ∗, Arka Mondal1 "Identification of a company's suitability for the adoption of cloud computing and modelling its corresponding Return on Investment.

[8] Xinhui Li, Ying Li, Tiancheng Liu, Jie Qiu, Fengchun Wang ,"The Method and Tool of Cost Analysis for Cloud Computing", 2009 IEEE International Conference on Cloud Computing.

[9] Ryan Chard,Kyle Chard,Rich Wolski,Ravi Madduri,Bryan Ng and Kris Bubendorfer,Ian Foster,"Cost-Aware Cloud Profiling, Prediction, and Provisioning as a Service",PUBLISHED BY THE IEEE COMPUTER SOCIET, 2017.

[10] Keith R. Jackson,Krishna Muriki,Shane Canon, Shreyas Cholia, John Shalf Harvey J.

[11] Wasserman, and Nicholas J. Wrig,"Performance Analysis of High Performance Computing Applications on the Amazon Web Services Cloud",2nd IEEE International Conference on Cloud Computing Technology and Science.

[12] Alexandru Iosup, Member, IEEE, Simon Ostermann,Nezih Yigitbasi, Member, IEEE, Radu Prodan, Member, IEEE, Thomas Fahringer, Member, IEEE, and Dick Epema, Member, IEEE"Performance Analysis of Cloud Computing Services for Many-Tasks Scientific Computing",IEEE TPDS, MANY-TASK COMPUTING, NOVEMBER 2010.

[13] Guoxin Liu and Haiying Shen, Senior Member, IEEE, Member, ACM"MinimumCost Cloud Storage Service Across Multiple Cloud Providers",IEEE/ACM TRANSACTIONS ON NETWORKING,2017.