Applying Autonomic Techniques To Cloud Computing For Resource Monitoring And Prediction.

Submitted By

Kedar Pandya

16MCEN08

DEPARTMENT OF INFORMATION TECNOLOGY

INSTITUTE OF TECHNOLOGY

NIRMA UNIVERSITY

AHMEDABAD-382481

MAY 2018

Applying Autonomic Techniques To Cloud Computing For Resource Monitoring And Prediction.

Major Project

Submitted in partial fulfillment of the requirements

for the degree of

Master of Technology in Computer Science and Engineering(Network Technologies)

Submitted By

Kedar Pandya

(16MCEN08)

Guided By

Prof. Vivek. K. Prasad

DEPARTMENT OF INFORMATION TECNOLOGY

INSTITUTE OF TECHNOLOGY

NIRMA UNIVERSITY

AHMEDABAD-382481

MAY 2018

Certificate

This is to certify that the Major Project entitled **"Applying autonomic techniques to cloud computing for resource monitoring and prediction."** submitted by **Kedar Pandya (Roll No: 16MCEN08)**, towards the fulfillment of the requirements for the award of degree of Master of Technology in Computer Science and Engineering(Network Technology) of Nirma University, Ahmedabad, is the record of work carried out by him under my supervision and guidance. In my opinion, the submitted work has reached a level required for being accepted for examination. The results embodied in this Research semina project, to the best of my knowledge, haven't been submitted to any other university or institution for award of any master, degree or diploma.

Dr. Gaurang Raval
Associate Professor,
Coordinator M.Tech - Network Technologies,
Institute of Technology,
Nirma University, Ahmedabad.

Prof. Vivek. K. Prasad
Guide & Assistant Professor,
CSE Department,
Institute of Technology,
Nirma University, Ahmedabad.

Dr. Madhuri Bhavsar
Professor and Head,
Information Technology Department,
Institute of Technology,
Nirma University, Ahmedabad.

Dr. Alka Mahajan
Director,
Institute of Technology,
Nirma University, Ahmedabad.

Statement of Originality

I, **Kedar Pandya**, Roll. No. **16MCEN08**, give undertaking that the Major Project entitled "**Applying autonomic techniques to cloud computing for resource monitoring and prediction.**" submitted by me, towards the partial fulfillment of the requirements for the degree of Master of Technology in **Computer Science & Engineering(Network Technologies)** of Institute of Technology, Nirma University, Ahmedabad, contains no material that has been awarded for any degree or diploma in any university or school in any territory to the best of my knowledge. It is the original work carried out by me and I give assurance that no attempt of plagiarism has been made. It contains no material that is previously published or written, except where reference has been made. I understand that in the event of any similarity found subsequently with any published work or any dissertation work elsewhere; it will result in severe disciplinary action.

_____

Signature of Student

Date:

Place:

Endorsed by

Prof. Vivek. K. Prasad

(Signature of Guide)

Acknowledgements

It gives me immense pleasure in expressing thanks and profound gratitude to **Prof. Vivek. K. Prasad**, Assistant Professor, Computer Science Department, Institute of Technology, Nirma University, Ahmedabad for his valuable guidance and continual encouragement throughout this work. The appreciation and continual support he has imparted has been a great motivation to me in reaching a higher goal. His guidance has triggered and nourished my intellectual maturity that I will benefit from, for a long time to come.

It gives me an immense pleasure to thank **Dr. Gaurang Raval**,PG Coordinator of M.Tech. Computer Science and Engineering (Network Technologies), Institute of Technology, Nirma University, Ahmedabad for his kind support and providing basic infrastructure and healthy research environment.

A special thank you is expressed wholeheartedly to **Dr. Madhuri Bhavsar**, Hon'ble Head, Institute of Technology, Nirma University, Ahmedabad for the unmentionable motivation he has extended throughout course of this work.

I would also thank the Institution, all faculty members of Information Technology Department, Nirma University, Ahmedabad for their special attention and suggestions towards the project work.

- Kedar Pandya

16MCEN08

Abstract

Today we are in the era of distributed system and cloud computing is the best example for the same. Resource management is an important issue in cloud computing, we need to keep track of the available resources in cloud, so that we can give services to user to fulfill their requirement. Which leads in helping to generate maximum revenue, lower powerconsumption, carbon emission and ultimately leads to green computing. So with the help of resource monitoring we can get the data about how, when, in what amount of resources for a particular cloud and the user. With effective resource monitoring, by minimizing some monitoring units we can reduce the cost of monitoring in terms of computation and power consumption. Our main objective is to reduce the monitoring technique, so that the amount of computation and power consumption can be saved which will lead to smart and green computing. In this project we had proposed an algorithm for reducing the monitoring overhead for cloud computing. Resource prediction can make resource management much more easier. Prediction techniques like machine learning or neural networks can be very healpfull for predicting our cloud ressources. In thhis project Long Short Term Memory (LSTM) is being applied for the resource prediction in cloud computing. So with the help of resource monitoring and predition in cloud we can manage the cloud resources very effectively.

Abbreviations

CC	Cloud Computing.
IaaS	Infrastructure as a Service
PaaS	Platform as a Service
SaaS	Software as a Service
ANN	Artificial Neural Network
RNN	Recurrent Neural Network
LSTM	Long Short Term Memory
RMSE	Root Mean Squared Error

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Cloud Computing

Cloud Computing is the technology which provides infrastructure for computing resources such as hardware, networking resources, storage and various applications are available on demand. Cloud are built on the infrastructure on the top of which services are developed and provided to customers. Cloud classification is majorly done in 4 types:- Public, Private, Hybrid, Community [1] . Figure below shows the cloud types.

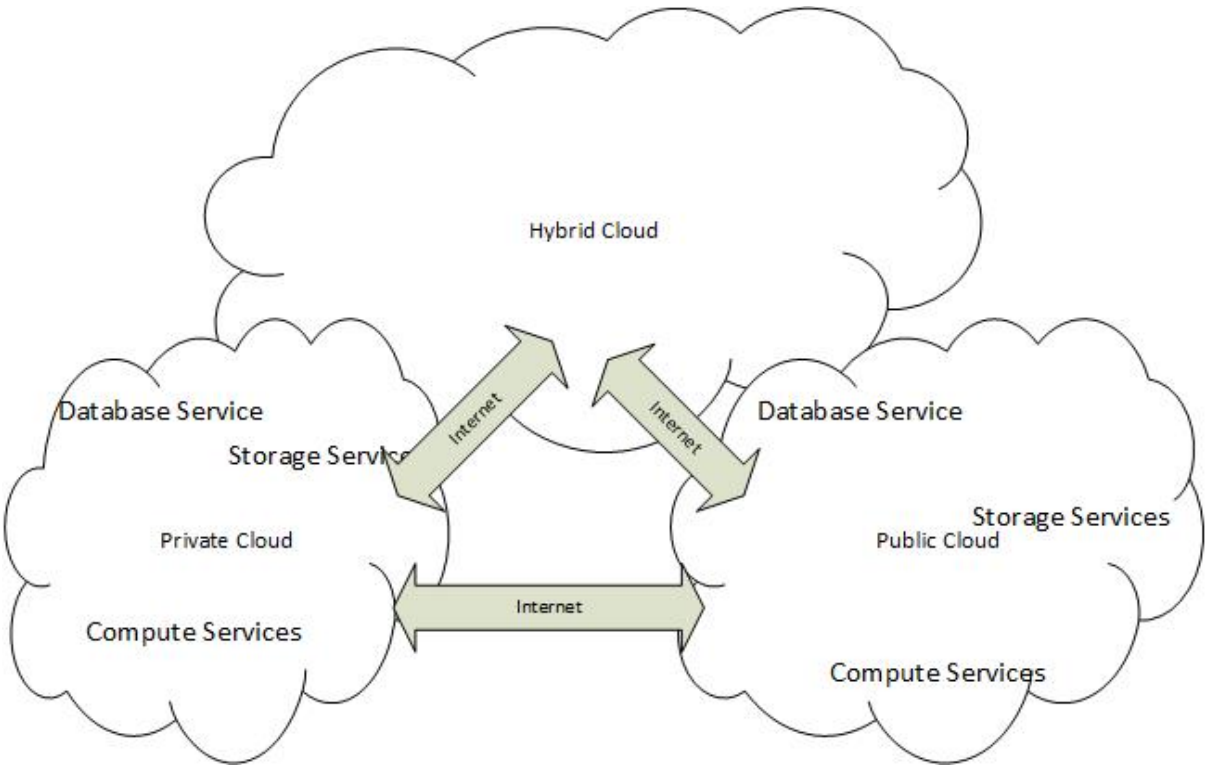

Figure 1.1: Cloud Computing Services

- Public Cloud:- In this type of cloud services of cloud are open to all users or cloud is wider to all public.
- Private Cloud:- In this type cloud is implemented to certain institute, company, organization. It fulfills the requirements of that organization only not all public user and users of particular organization can only access it.
- Hybrid Cloud:- This type of cloud is combination of public and private type of clouds. For example if a public cloud service provider provides a private cloud or private service for a particular organization only then it is called Hybrid cloud.
- Community Cloud:- This type of cloud is characterized by a multiple domain containing models like as mentioned above public, private, hybrid. This tye of cloud is designed for specific industry and to reach out specific demands only.

Cloud Computing Services are mainly classified as:-Infrastructure as a service (IaaS), Platform as a service (PaaS), Software as a service (SaaS) . Figure below show the typical cloud computing services diagram.

These Cloud Computing Resources are mainly classified as :-

- Infrastructure as a service (Iaas)
- Platform as a service (Paas)
- Software as a service (Saas)

Here IaaS is referred to provisioning of cloud services based on the Infrastructure point of view; such as providing user with Hardware such a Processing Unit, RAM, Storage devices etc. Platform as a service (PaaS) provides platform such as computational platform, storage platform on which cloud user can make and deploy its application for examples programming APIs, frameworks etc [1]. to build applications. In Software as a Service (SaaS) customers are provided with the software application that can be accessed by user anytime and anywhere. As we saw, these are three main services in cloud computing the key highlighting thing is that all 3 services are based on resources (processing, ram, memory etc.) of cloud. For any cloud provider resources are the biggest investment and through which they can generate revenue. So handling resource very effectively is very important for cloud provider. Resources for any user should not be

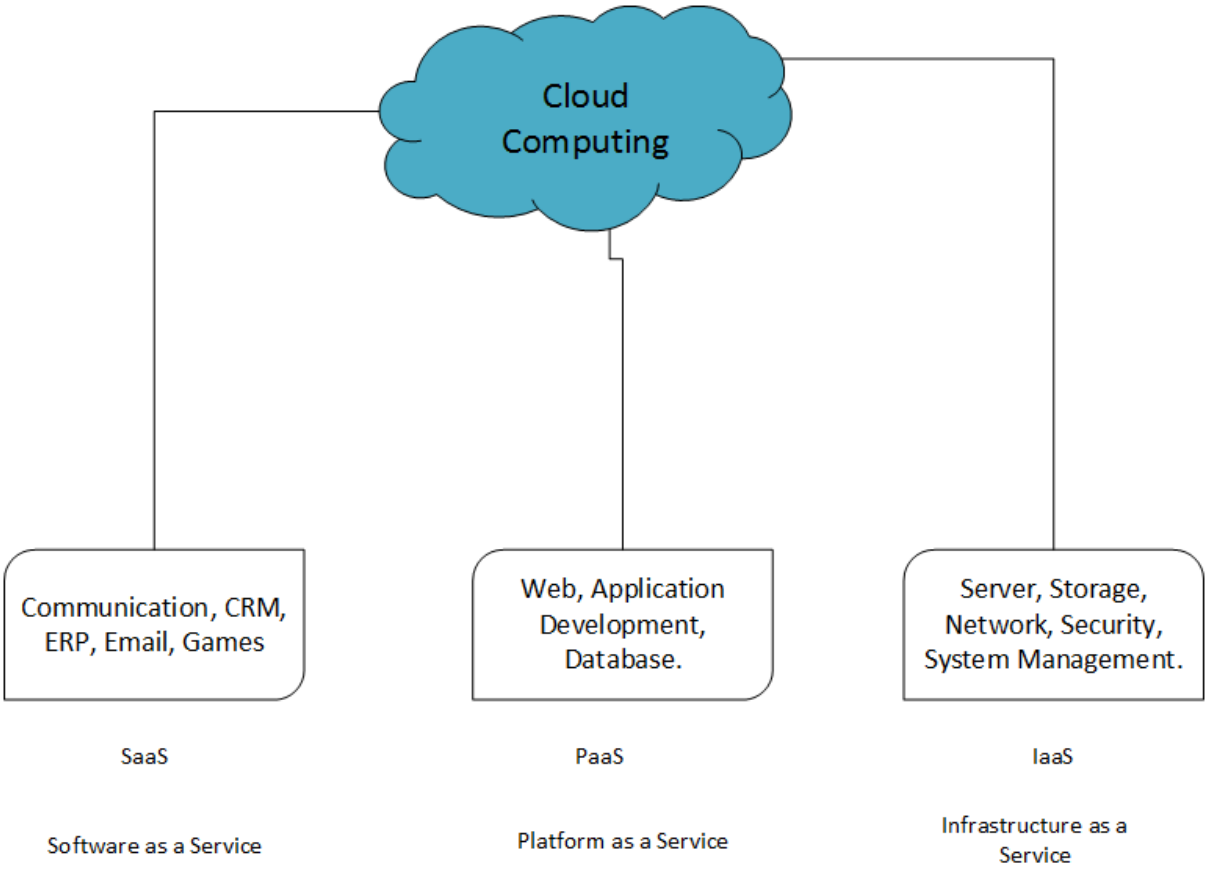

Figure 1.2: Cloud Computing Services

under-provisioned and over-provisioned [2]. Resource monitoring is the technique which helps us to solve this problem. In resource monitoring technique we can get data which particular resource is being consumed in what amount. So with the help of these data we can get the information of which user uses which resources in how much of time; due to which we can solve the problem of resource under and over provisioning at Cloud Service Provider side.

So with the help of these Resource monitoring data we can get the series of data from which we can get usefull information for future. Resorce Prediction can be done in the of particaular user of particular job with the help of various Resource Prediction methods like Machine Learning, Data mining and many more. So with the help of Monitoring data and Prediction techniques our target is predict the particular amount of resource to be used for particular job in cloud computing due to which we can get the best out of every resource which will able to lead the solution of under provsion & over provisioning.

1.2 Deep Learning - ANN, RNN, LSTM networks

1.2.1 Artificial Neural Network (ANN)

Artificial Neural Network (ANN) are inspired by the biological neural network system of human brain. ANN is on collection of collected neurons. As shown in figure 1.3 we can see typical Ann collection and connection of neurons.

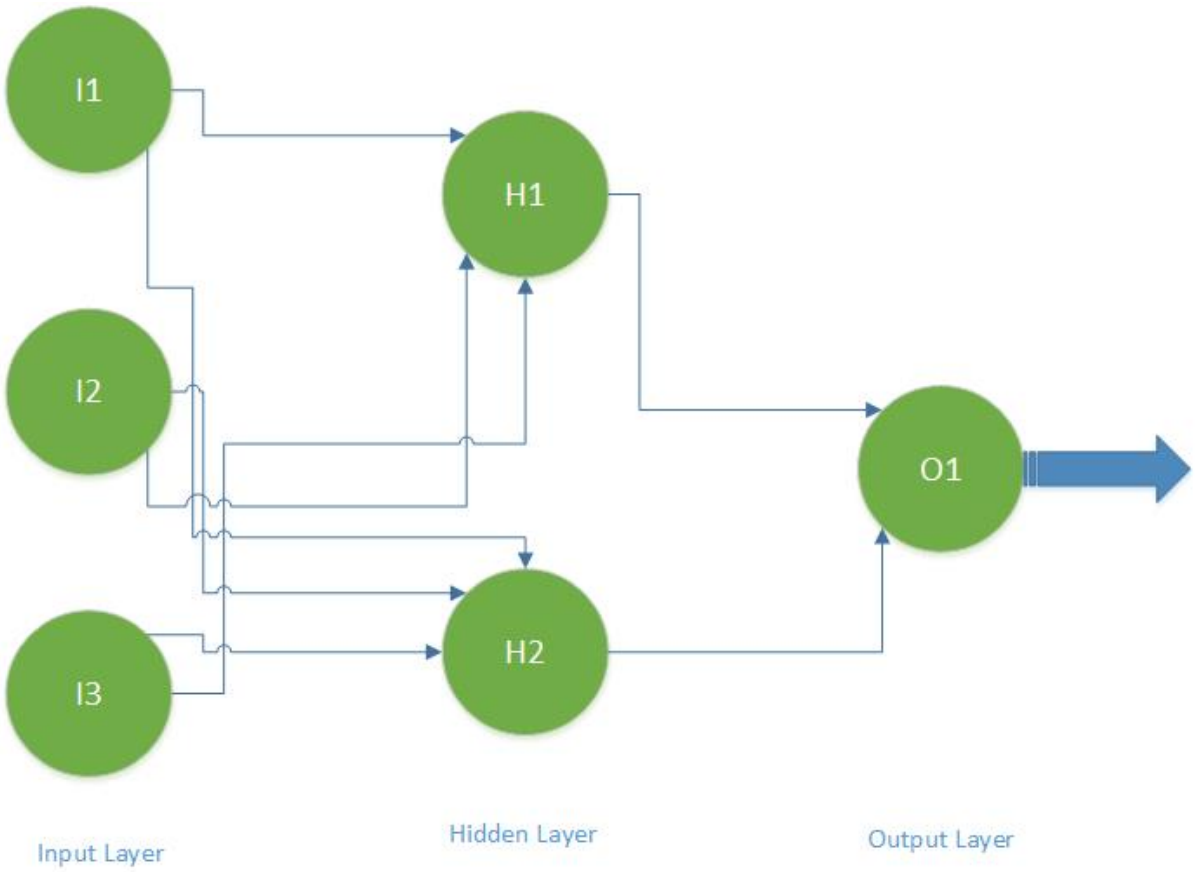

Figure 1.3: Artificial Neural Network

ANN is feed-forward network. It allows signals to transfer in one way from input layer to hidden layer and to output layer; after that we get the output. Feed-forward network ANNs are straight-forward networks associated to inputs and outputs. There is no loop back in the network so output of the any previous layer does not effect on the next layer. Feedforward neural networks are ideally suitable for modeling relationships between a set of predictor or input variables and one or more response or output variables.

1.2.2 Recurrent Neural Network (RNN)

RNN is a type of ANN, which has a recurrent connection to its own network. This type of recurrent connection to its own helps to learn the neural network the information of previous input i(time-1) along with its current input i(time) and it will predict the output at time t of o(t).Figure below show the typical diagram of RNN.

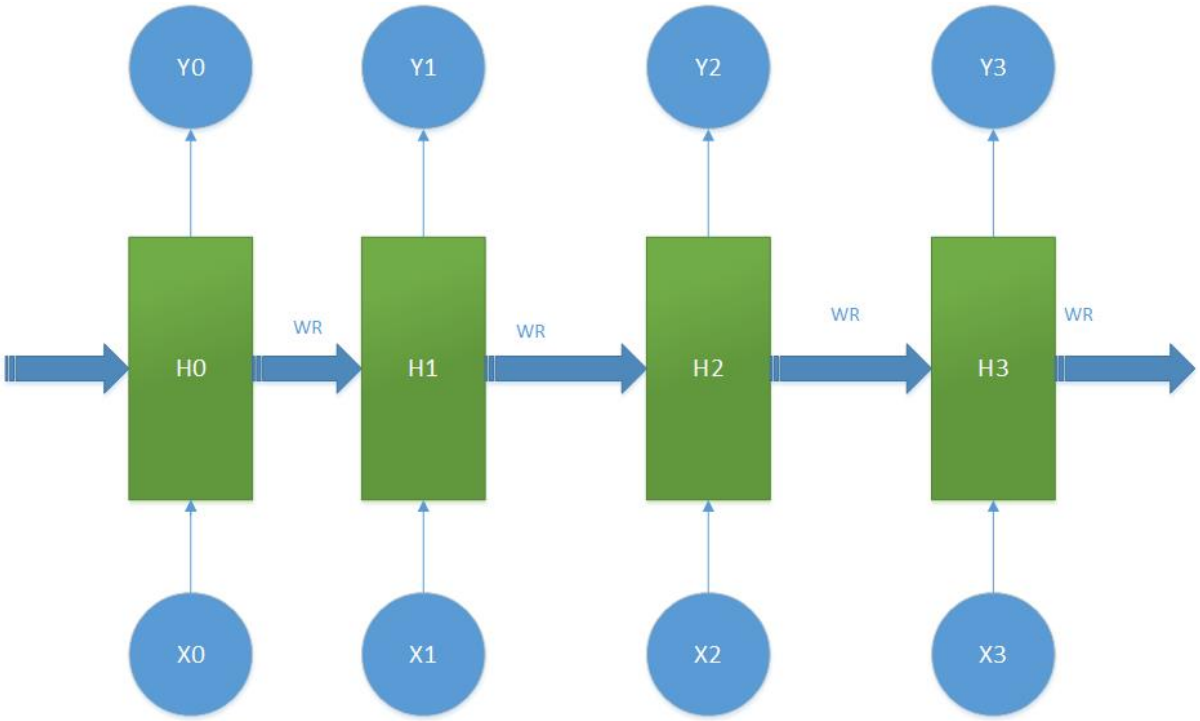

Figure 1.4: Recurrent Neural Network

RNN and LSTM are typical Feedback Networks. Recurrent or feed-forward networks can data travelling in both directions by inccluding loops in the network. These type of neural network are very powerfull and complicated. Data or Information from earlier input is feeded into next netowrk. Feed-forward networks are highly dynamic and their states changes continuously until they reach equillibrium.

Problem in RNN is :-

- Vanishing gradient problem:- It is one of the biggets problems with RNN networks. The gradients gets vanished to 0 when backpropagation done. This situation is able of noticed because of the derivative of the functions (activation) called tanh or sigmoid() are less than 0.25 and 1. After that these derivatives are multiplied then the gradient becomes to 0. Due to this earlier layers learns slowly than later layers.

1.2.3 Long Short Term Memory (LSTM)

LSTM model is used because it solves the problem of vanishing gradients. It does by the introduction of a new state called cell state. Also having a CEC which helps to allow the error to back-propagate and not being vanished. Figure below show the typical diagram of LSTM.

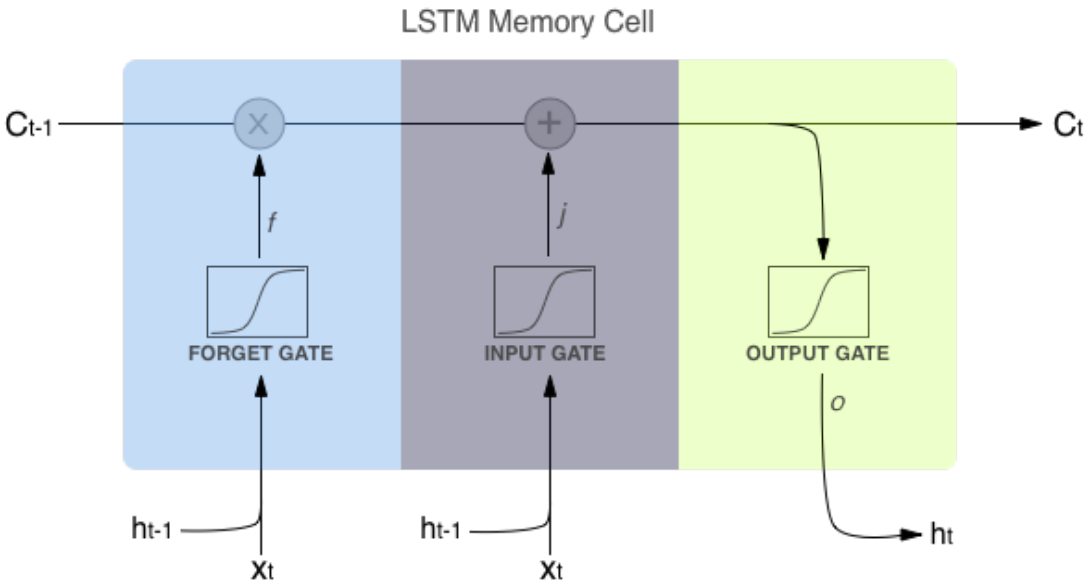

Figure 1.5: Long Short Term Memory

With the help of Forget gate model can learn on which time to clear the information of the cell-state. It gives model the information when to remember the data or when to clear the data when the content or information starting to get irrelevant. In the model there is obvious takes the input and output gates; input-gates controls in what amount of new information has to be added in the cell-state. Output gates(o) controls when to retrieve the data in the cell state for producing ht. LSTMs are used over RNN because can hold the information in to its memory for a longer time period.

Chapter 2

Literature Survey

2.1 Literature Survey for Resource Monitoring

This main aim of survey is for finding the current technology and work in cloud resource monitoring; also to find the drawbacks or future research directions so that work can done in the correct direction.

Paper 1 :-

- Title of paper :- Resource management for Infrastructure as a Service (IaaS) in cloud computing: A survey [3].
- Type of work :-In this paper classifying the resource management, drawing out all the elements in work towards resource mapping, provisioning, allocation.
- Methodology :- In this paper they had compared performance metrics on basis of Reliability, Ease of deployment, QoS, Delay, Control overhead.
- Network queuing model, SLA oriented methods, Dynamic and automated framework, Resource Pricing, Adaptive resource provisioning, Optimal cloud resource provisioning (OCRP) etc. methods under Resource Provisioning schemes.
- Comparison of methods like Market Oriented Allocation, Dynamic resource allocation method, Real time resource allocation mechanism, Congestion control method under Resource Allocation Schemes
- Comparison of methods like Load-aware mapping, Impatient task mapping, Opportunistic resource, Mapping a virtual network onto a substrate network etc. under

resource mapping scheme.

Paper 2 :-

- Title of paper :- Dynamic Resource Allocation Using Virtual Machines for Cloud Computing Environment [1].
- Type of work :- In this paper they have introduce the concept of Skewness to measure the uneven usage of multidimensional resources & by minimizing it we can improve over-all utilization of servers due to which work over-load on Physical Machines is avoided degradation in performance of VM.
- Methodology :- Skewness Algorithm of using resources in uneven manner. This algorithm executes periodically to evaluate the resource allocation status based on the predicted future resource demands of VMs.

Paper 3 :-

- Title of paper :- Load Prediction and Hot Spot Detection Models for Autonomic Cloud Computing [2].
- Type of work :-This paper is being written for the use of load prediction algorithms in cloud computing. This approach is used by different load decision system cloud platforms computational complexity compatible to run-time environment.
- Methodology :- Load prediction of the algorithms for cloud computing using 2 step approach of load-tracking followed by load prediction, by cubic spline interpolation & by hotspot detection algorithms.

Paper 4 :-

- Title of paper :- Resource Consumption Prediction Using Neuro-Fuzzy Modeling [4].
- Type of work :- In this paper they had presented that that resource consumption prediction is better by with the help of the neuro-fuzzy models rather than the machine learning algorithms.

- Methodology:- Adaptive Neuro fuzzy Inference System(ANFIS) & machine learning models is applied on the RAxML & BLAST bio-informatics applications and then these algorithms are compared by Relative Squared Error (RSSE) method.

Paper 5 :-

- Title of paper :- Autonomic Cloud Computing: Open Challenges and Architectural Elements [5].
- Type of work :- In this paper identifies open issues on autonomic resource provisioning and presents many new management techniques for supporting SaaS applications hosted on Clouds. They presented conceptual model able to achieve the different goals like Quality of Service (QoS), Energy efficiency,Security and present results that gives the advantages of autonomic management of Cloud infrastructures.
- Future work:- Implementation of more dynamic provisioning algorithms that are QoS and security-aware and energy efficient, and will demonstrate their effectiveness with real applications from domains.

Paper 6 :-

- Title of paper :- Resource monitoring and prediction in cloud computing environments [6].
- Type of work :- This paper addresses the resource monitoring and prediction problem in cloud computing environment, designs and implements an adaptive resource monitoring framework for cloud computing.
- Methodology :- A resource prediction mechanism based on Vector Auto Regression (VAR) by the correlation between various resources.
- Future work:- The main direction of their future work is how to optimize the performance & finding the algorithm that has lower time complexity.

Paper 7 :-

- Title of paper :- Resource Prediction Based on Double Exponential Smoothing in Cloud Computing [7].

- Type of work :- In this paper, they had presented a resource prediction model based on double exponential smoothing technique, on which it considers the current state of resources but also the history of resorces.
- Methodology :- In this paper they had used double exponential smoothing as resource prediction method.
- Future work:- In the future works, further research will be done on energy efficiency which is also useful to save money for the customers and cloud providers.

Paper 8 :-

- Title of paper :- Accurate Resource Prediction for Hybrid IaaS Clouds Using Workload-Tailored Elastic Compute Units [8].
- Type of work :- This paper has written, to get perfect performance prediction and cost-optimal resource management for hybrid clouds.
- Methodology :- In this paper they had developed Resource Configuration Algorithm to support the cost-optimal throughput based resource management.
- Its steps are:-
- Compute the target throughput and set a corresponding computational power requirement.
- Check if the private cloud could satisfy the computational need.
- If Step 2 donot satisfy the requirement, use extra resources in the cloud.
- Assignment to workers to the instances allocated in Steps 2 & 3,

Paper 9:-

- Title of paper :- DARGOS: A highly adaptable and scalable monitoring architecture for multi-tenant Clouds [9].
- Type of work :-DARGOS is monitoring architecture and is distributed and very efficient which spreads information of monitoring of resource in the cloud. It accurately measures resources in cloud physical and virtual resources in cloud.

- Methodology :- DARGOS adapts and allows to monitor new monitoring metrics easily. Architecture and related tools are integrated in the OpenStack. With monitoring any resource, hosts, Virtual machine with very limited and scalable monitoring overhead.

- Future work:- Future on they want to work on the optimizing techniques such as to reduce the monitoring overhead.

Paper 10:-

- Title of paper :- GMonE: A complete approach to cloud monitoring [10].

- Type of work :-GMonE is general purpose tool which does all the aspects of the monitoring. This tool has made on their proposed architecture. It can mainly deployed on the in high-scale model.

- Methodology :- GMonE had evaluated performance, scalability and overhead using OpenNebula cloud middleware; results surpasses the performances of Amazon EC2 and OpenNebula. Proposed study has used publish/subscribe model for message communication.

- Future work:- To test the behavior of GMonE in different scenarios specially heterogeneous and federated cloud system.

Paper 11:-

- Title of paper :- Towards an Architecture for Monitoring Private Clouds [11].

- Type of work :-PCMONS is an open-source cloud monitoring architecture for the private cloud.

- Methodology :- PCMONS is of three layered architecture which is integrated with tools like Nagios. PCMONS supports two approaches agents and central monitoring which is very adaptable, flexible and extensible.

- Future work:- To provide monitoring in other type of cloud also like public, hybrid etc.

Paper 12:-

- Title of paper :- Touchless and always-on cloud analytics as a service [12].
- Type of work :-Cloud monitoring is another module in cloud but not as a part of cloud operating system. NFM (Near Field Monitoring) proposes a new way of monitoring the cloud architecture.
- Methodology :- NFM does not monitor cloud giving interrupts to VMs or resources. NFM has its own logic of collecting monitoring data irrespective to the health of VMs. Author has used the kernel data for monitoring a particular metric of cloud.
- Future work:- To implement the NFM in highly scalable and heterogeneous architecture.

Paper 13:-

- Title of paper :- Towards Multi-Tenant and Interoperable Monitoring of Virtual Machines in Cloud [13].
- Type of work :- This paper enlightens and meets the new challenge like multi-tenancy and complexity of cloud for monitoring. For this they had proposed a system called New-Generation-MONitoring (Ngmon). Ngmons is event based monitoring.
- Methodology :- Ngmon collects monitoring information from logs, notifications and activities which are being described event based. After that processing of that data is done. Next step to distribute the data to interested parties. After that visualization of data takes place.
- Future work:- Inter-Cloud monitoring is one future aspect and use of CEP (Complex Event Processing) technique and algorithm for pattern detection and correlation in monitoring.

Paper 14:-

- Title of paper :- MonPaaS: An Adaptive Monitoring Platform as a Service for Cloud Computing Infrastructures and Services [14].
- Type of work :- MonPaas is the monitoring tool that is open source. MonPass increases the metrics of information of monitoring. It does monitoring in two mode as cloud provider monitoring and user monitoring.

- Methodology :- For monitoring MonPaas uses another monitoring VM (MVM) for any new user. MonPass uses Nconf, DNX, Nagios as monitoring support and graphical user interface for showing the data.
- Future work:- As future work they want to test MonPass in bigger scenarios. Extend the monitoring architecture by using workload balancing for cloud monitoring.

Paper 15:-

- Title of paper :- Application level interface for a Cloud Monitoring service [15].
- Type of work :- Monitoring as a services system is provided which can be used in all clouds.
- Methodology :- Monitoring as a service is implemented using OOCI (Open Cloud Computing Interface) API. Agents that collect monitoring information of different metric which is defined Mixins. Mixins have 3 different features containing metric, aggregator and publisher.
- Future work:- Docker based prototype for monitoring is made which is open source and complex. So to implement in different situation is the main future work.

Paper 16:-

- Title of paper :- A Framework for Consumer- Centric SLA Management of Cloud-Hosted Databases [16].
- Type of work :- This paper proposes the framework of the SLA based monitoring. Monitoring is done on the requirements of the user.
- Methodology :- Framework uses the database of the services of the cloud. It monitors the cloud in such a way that SLA violation is prevented.
- Future work:- It does monitoring to satisfy the SLA, so it may not monitor all the aspects of SLA.

Paper 17:-

- Title of paper :- A Publish/Subscribe Middleware for Dependable and Real-time Resource Monitoring in the Cloud [17].

- Type of work :- This paper present solution of monitoring cloud in real and updating the cloud information with least amount of data loss in real time called as SQRT-C.
- Methodology :- Solution is based on the OMG Data Distribution Service (DDS) real-time publish/subscribe paradigm for real-time scalable monitoring in cloud.
- Future work:- SQRT-C operation in fault tolerant orchestrator. Testing this approach in different cloud platforms.

Paper 18:-

- Title of paper :- CloudSense: Continuous Fine-Grain Cloud Monitoring With Compressive Sensing [18].
- Type of work :- This paper presents data centre monitoring tool which provides fine grain status monitoring of a cloud data centre.
- Methodology :- A new Switch design which performs all the in-networking compression via low complexity encoding and compressive sensing.
- Future work:- A general framework for performance monitoring of hosted application in cloud.

2.2 Literature Survey for Resource Prediction

This main aim of survey is for finding the current technology and work in cloud resource prediction; also to find the drawbacks or future research directions so that work can done in the correct direction.

Paper 1:-

- Title of paper :- Function points-based resource prediction in cloud computing [19].
- Type of work :- In this paper they had used the linear regression based resource prediction model to predict the resource in the cloud based on the function point used by the users request.
- Methodology :- In proposed model first function point is calculated to find complexity and size of user request. After that linear regression model is used to predict the usage of resource. ANN using linear regression resource, current and previous resource usage predict for future resources prediction.
- Future work:- Apply and testing this model in different environment of cloud computing.

Paper 2:-

- Title of paper :- Virtual Resource Prediction in Cloud Environment: A Bayesian Approach [20].
- Type of work :- In this paper they had proposed Bayesian model to estimate the long term and short term usage of virtual resources in cloud on basis of several workload patterns on different cloud specifically for CPU utilization.
- Methodology :- Model is simulated on the SamIam Bayesian network simulator and workload traces of Amazon EC2 and Google CE data.
- Future work:- Future work suggests about SLAs, they want to do: (i) shift the workloads of weekdays to weekends, and (ii) share the workloads of a CSP among multiple DCs.

Paper 3:-

- Title of paper :- Survey on prediction models of applications for resources provisioning in Cloud [21].
- Type of work :- This papers review the stand out prediction methods of resources in different aspects. And also states the issues in prediction for future work.
- Methodology :- Bayesian Theory, K-Nearest Neighbor Random forest, Neural Network, Support Vector Machine, Markov Model, Classification/Clustering, Dimension Reduction Reinforcement Learning, String Matching, Fuzzy Logic, Hurst Exponent, Qn Theory, Control Theory, Smoothing Filters etc. methods and techniques used for prediction.
- Future work:- Challenges for prediction of resources are Choosing the pattern length, modern transparency and adaptability, robustness, hybrid prediction approaches etc.

Paper 4:-

- Title of paper :- Workload Prediction for Cloud Cluster using A Recurrent Neural Network [22].
- Type of work :- In this paper, we propose an approach using recurrent neural networks (RNN) to realize workload prediction, where CPU and RAM metrics are used to evaluate the performance of the proposed approach.
- Future work:- Try with another method like LSTM (Long short-term memory).

Chapter 3

Problem Statement and Proposed Model & Solutions

3.1 Problem Statement

Below table 3.1 shows the survey of the monitoring system.

Resources are the biggest investment for Cloud Service Provider to manage these resources (CPU, memory, RAM) of cloud are very important and critical to manage it. Cloud Service Provider have to solve the two main problems related to resources utilization which are called under provisioning and over provisioning of resources. Underprovisioning of resources occurs when the application is assigned with fewer numbers of resources than the demand. Over-provisioning of resources arises when the application gets surplus resources than the demanded one. According to survey of:- Right Scale 2017 State of the Cloud Report we can see that which are the key challenges of 2017, it is shown in the figure 3.1.

With the help of monitoring and prediction techniques we can reduce these challenges of under and over provisioning. Monitoring is one of the key technique in which every cloud service provider is investing lot of money on it. In an another survey of Right Scale 2017 State of the Cloud Report ,we can see that how companies are optimizing their cloud costs section that in monitoring the cloud instances the companies are investing more time and money on monitoring the instance rather than any other technique in cloud computing for optimizing the cloud cost or money. We can see this in figure 3.2.

Montoring System	Scalable	Cloud aware	Fault Tolerant	Multi Granual	Compr-ehensive	Time Sensitive	Auto nomic
Cloudinit		Yes	Yes				
CloudSense	Yes					Yes	
DARGOS	Yes	Yes		Yes		Yes	
Dingra		Yes		Yes			
GMonE	Yes	Yes					
Konig	Yes	Yes			Yes		
Logstash	Yes			Yes		Yes	
OpenNebula		Yes					
PCMONS	Yes	Yes					
Sensu	Yes	Yes					Yes
SQRT-C	Yes	Yes				Yes	
Varanus	Yes	Yes				Yes	

Table 3.1: Monitoring System Table

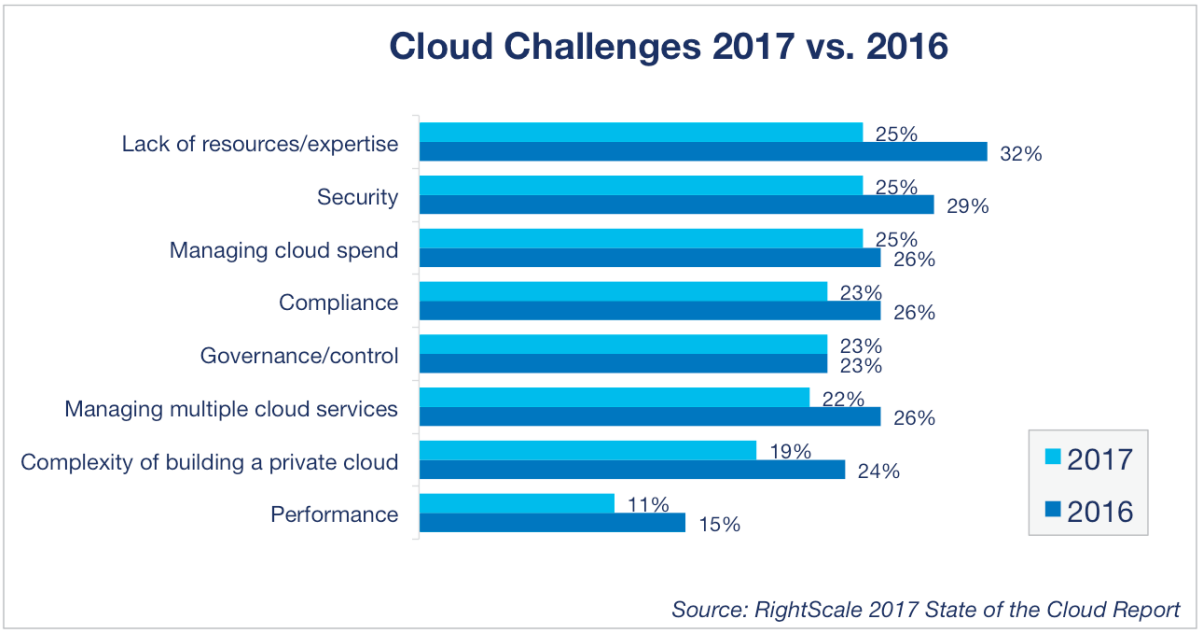

Figure 3.1: Cloud Computing Challenges

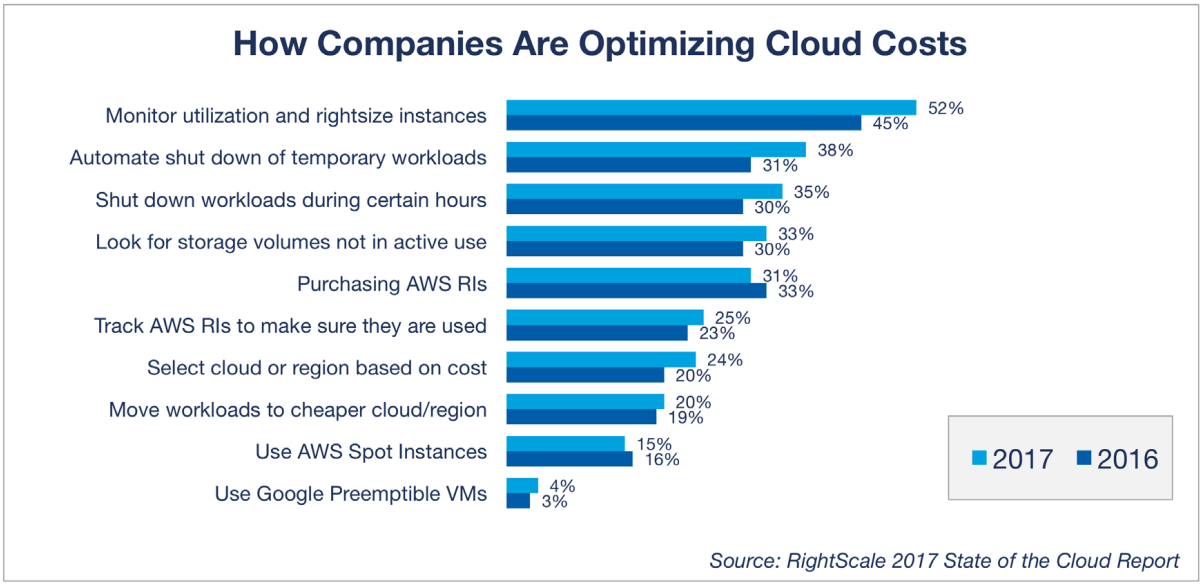

Figure 3.2: Optimization in cost cloud computing

3.2 Proposed Model

So from above images we can conclude that resources management is important aspect for the cloud service provider and to do this, monitoring technique is an important task for resource management. So our one of the main objective of the research is effective resource monitoring in cloud computing. But monitoring in all the aspect and every time will be burden on the cloud provider, so we have proposed the algorithms, where the monitoring technique will be called whenever the systems performance will be going down , so as to precautionary measures can be taken well in advance.

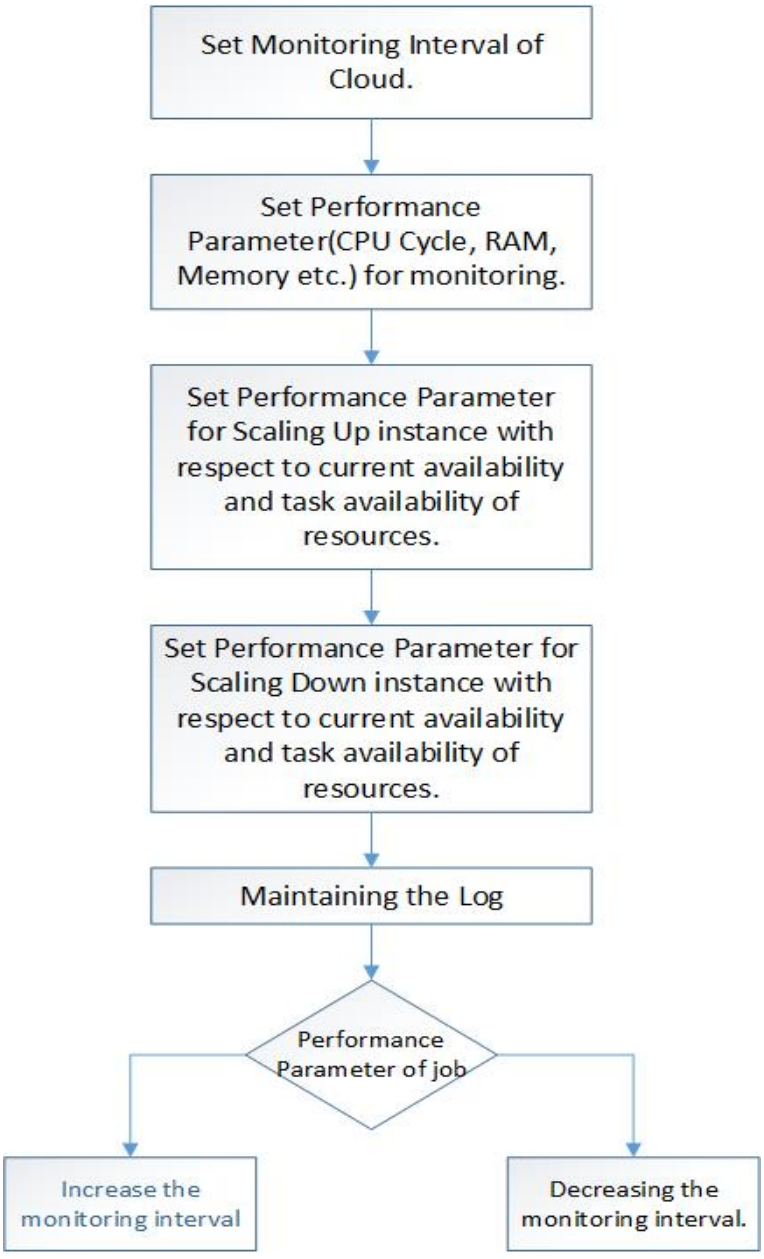

Figure 3.3: Flow Diagram of proposed Model

Our one of the main aim of the research is to effectively monitor the resources when need. If user has its own product which he/she has launched that product in cloud. Now Cloud Service Provider(CSP) has to continuously monitor that cloud instance for checking that the instance is healthy or not. The health check of the instance depends on the values of the various resource parameters like CPU utilization, RAM utilization, Memory usage, Network usage etc. Every CSP can set its own parameters or can set its own logic of healthy instance. CSP decide to monitor every instance on let say every t seconds. So at every t seconds it check that instance is healthy or not.Every application in cloud have their own necessity of resources for example if cloud has application of the social networking website then main requirement of this application is memory, network latency and RAM .

Here CPU utilization is moderate in normal condition but other parameters like memory and RAM have much more use. But if the user has product which is based on computation of machine learning; then here CPU cycles utilization, RAM utilization and memory is much more than social networking website. But in machine learning based product network latency is less used in normal condition. Our proposed idea of effective monitoring is that every application uses a particular resource in particular amount. Every resource uses particular resource in particular pattern now we had proposed the algorithm about to effectively monitor the instance so that minimum amount of monitoring and performance also not being degraded.

The real need of monitoring technique is needed when the value of performance parameter of any resource is crossing threshold or the value of performance parameter is near about threshold because if the performance parameter crosses the threshold than new instance has to be created. The flow diagram of our proposed algorithm is shown below in figure 3.3. So more monitoring has to be done near the threshold only; because necessary steps(scaling up instance or scaling down instance or killing the processes etc.) has to be taken so that instance does not gets overloads and performance does not degrades.

3.3 Proposed Algorithm

Algorithm 1 Effective Resource Monitoring :-

Step 1: Set a monitoring interval for any instance at which,
it is being monitored for any interval = m1 seconds.
Step 2: Let performance parameters be CPU Cycle Usage be P1,
Memory Usage be P2, RAM Usage be P3,..etc.
Step 3: Let threshold of any performance parameter for
Scaling Up instance be t1.
Step 4: Let threshold of any performance parameter for
Scaling Down instance be t2.
Step 5: Create a log of performance parameter monitoring of
every job in cloud train the log with prediction like
Markov Modeling.
Step 6: Let P1 or P2 or P3 of new job be nt1
if $nt1 > t1$ or $nt1 \leq t2$ **then**
m1=m1-(x%m1)
end if
if $t1 > nt1 > t2$ **then**
m1=m1+(x%m1)
end if

Above algorithm 1 is our proposed algorithm of monitoring of the instances in effective manner. In step 1 we have to set the monitoring intervals at which at that interval cloud monitors the instance. In Step 2 we set the parameters to which we want to monitor performance of it; for example CPU utilization, RAM, memory etc. In Step 3 and 4 user have to decide the thresholds for scaling up and scaling down the instance. In Step 5 CSP has record all the information of jobs and its performance parameters. This data is used in train the machine learningl or any other prediction technique. Step 6 adjusts the monitoring of instance according to the performance of the instance.In IF$nt1 > t1$ or $nt1 \leq t2$ statement states that if the performance of job is beyond any of two scaling up and down threshold then in decrease the interval of monitoring which increases the number of monitoring operations. In IF$t1 > nt1 > t2$ statement states that if the performance of job is in between two scaling up and down threshold then in increase the interval of monitoring which decreases the number of monitoring operations.

3.4 Proposed Resource Prediction

For resource prediction data set from Delft University of Technology is taken. From that dataset of 1750 VMs are being taken. Dataset contains all :- Timestamp, CPU cores, CPU usage (in MHZ), CPU usage (in percentage), Memory provisioned (memory requested of the VM in terms of KB), Memory usage (in KB), Disk read throughput (in KB/s), Disk write throughput (in KB/s), Network received throughput (in KB/s), Network transmitted throughput (in KB/s).

LSTM model is being applied on this dataset and accuracy is taken on it.

Chapter 4

Experimental Results

4.1 Resource Monitoring Results

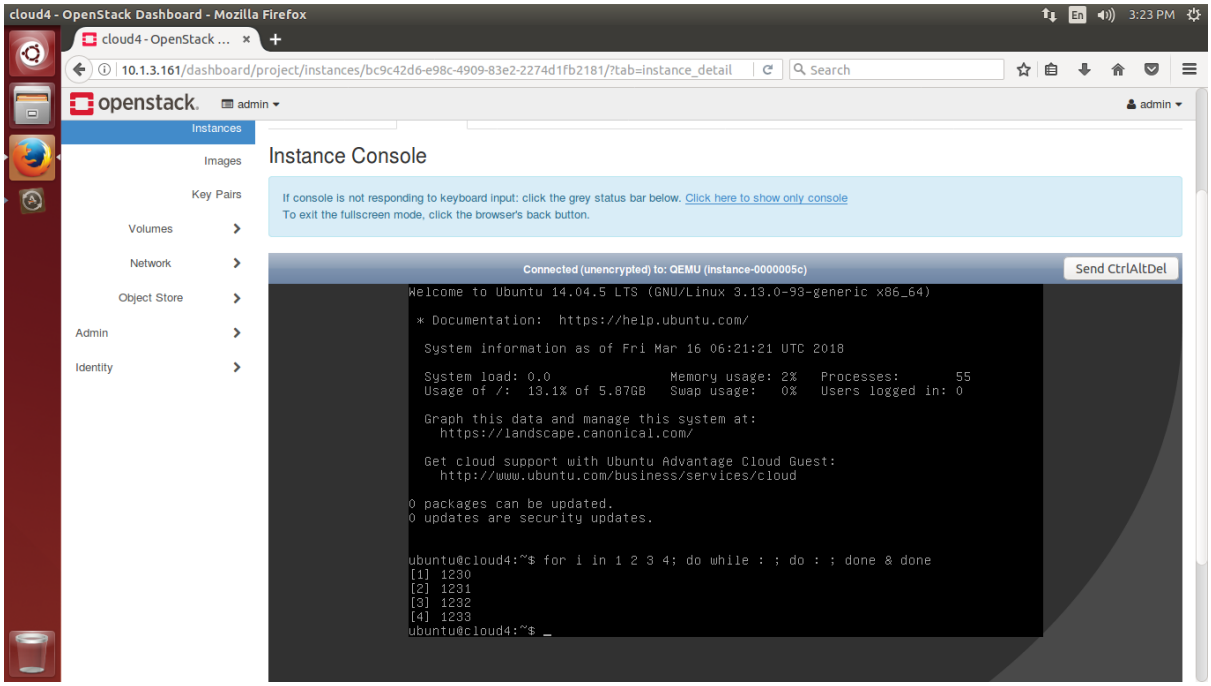

Figure 4.1: Process forking in cloud instance

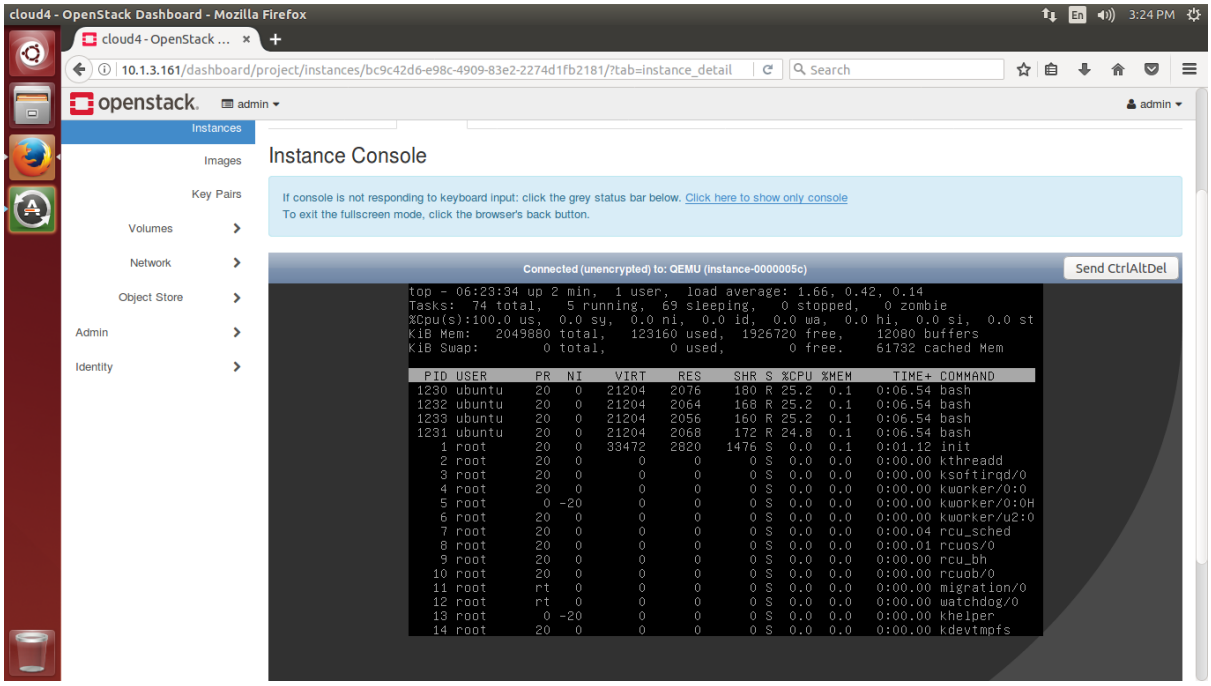

Figure 4.2: CPU Utilization for Process forking in cloud instance

- Screenshots above are the monitoring data of CPU utlization of process forking.
- Below is the graph of nature of the process fork cpu utilization.

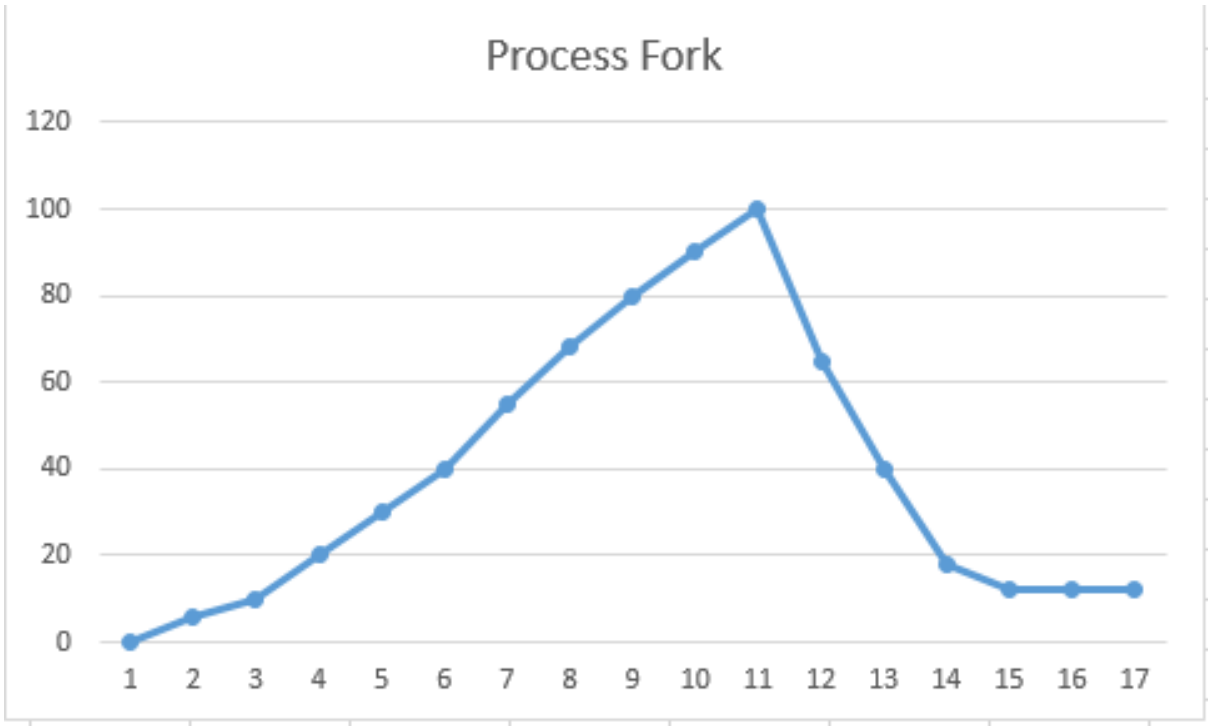

Figure 4.3: Process fork graph of sever

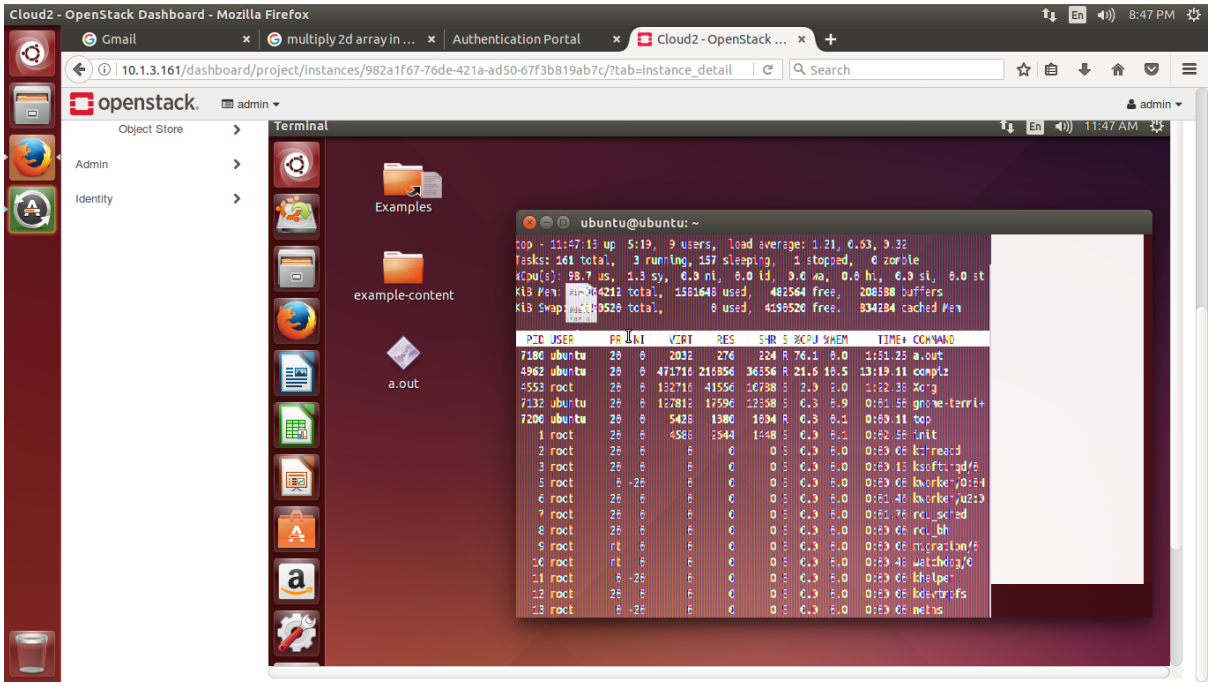

Figure 4.4: Matrix multiplication CPU utilization

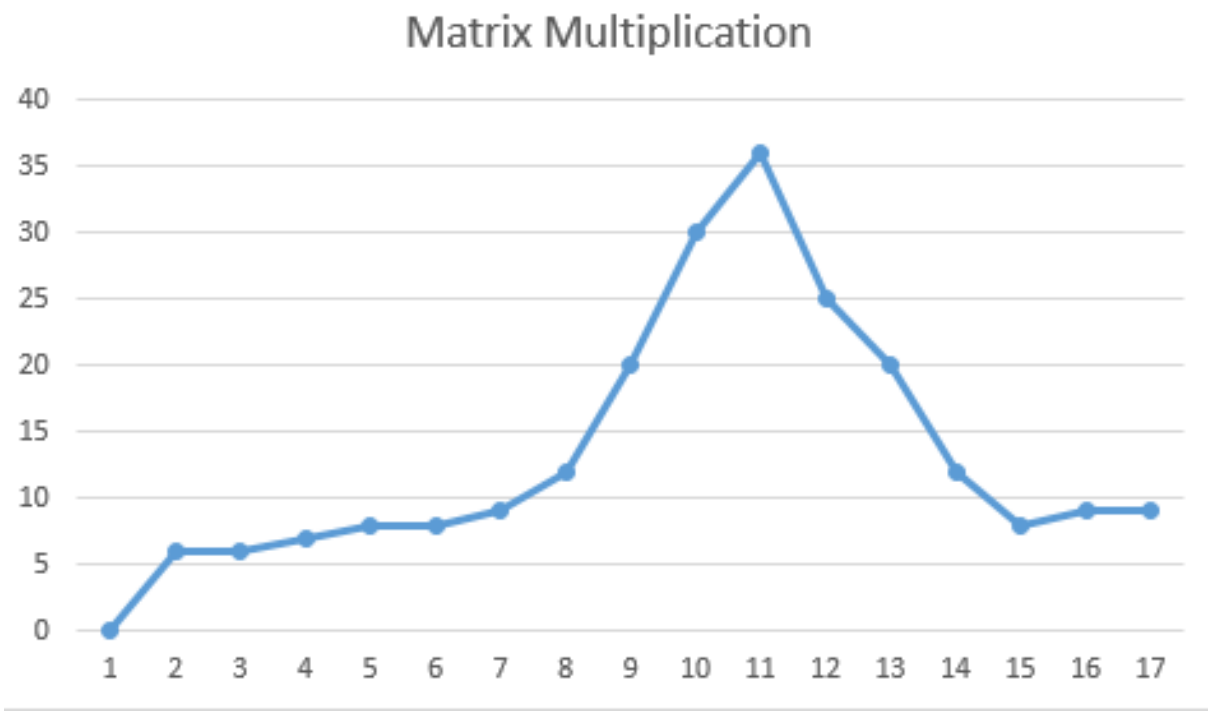

Figure 4.5: Matrix multiplication CPU utilization graph

4.2 Resource Prediction Results

- Cloud resource prediction is done on the basis of the virtual machine performance prediction.
- Resorce prediction is done with the help of neural network technique called LSTM.
- The experiments were performed on Intel CoreTM I3-M520 processor of 2.40GHz clock speed having 8 GB of memory.
- We had taken Python along with Keras (tensorflow at backhand) library as a tool for implementation.
- We performed the experiments on public cloud dataset.
- Dataset contains all :- Timestamp, CPU cores, CPU usage (in MHZ), CPU usage (in percentage), Memory provisioned (memory requested of the VM in terms of KB), Memory usage (in KB), Disk read throughput (in KB/s), Disk write throughput (in KB/s), Network received throughput (in KB/s), Network transmitted throughput (in KB/s).
- But for this experiment CPU usage (in MHZ) is taken into consideration.
- Figure 4.6 is graph of CPU utilization of the dataset.
- Figure 4.7 is graph of predicted CPU utilization of the dataset.
- We had taken CPU utilization applied LSTM on it.
- From that we had predicted train set and test set on it.
- On that we had got Root Mean Squared Error(RMSE) for train set is 3.23 and test score is 3.48.
- In this way of predicting the resources; dataset is splitted into training set and test set into 70-30
- For this 70% of data is taken for training and 30% of data is used for testing.
- In figure 4.8 CPU utilization is predicted by training data into every 10 timestep and predicting next 10 timestep of data.

- RMSE of this way of predicting the CPU utilization is 4.143.

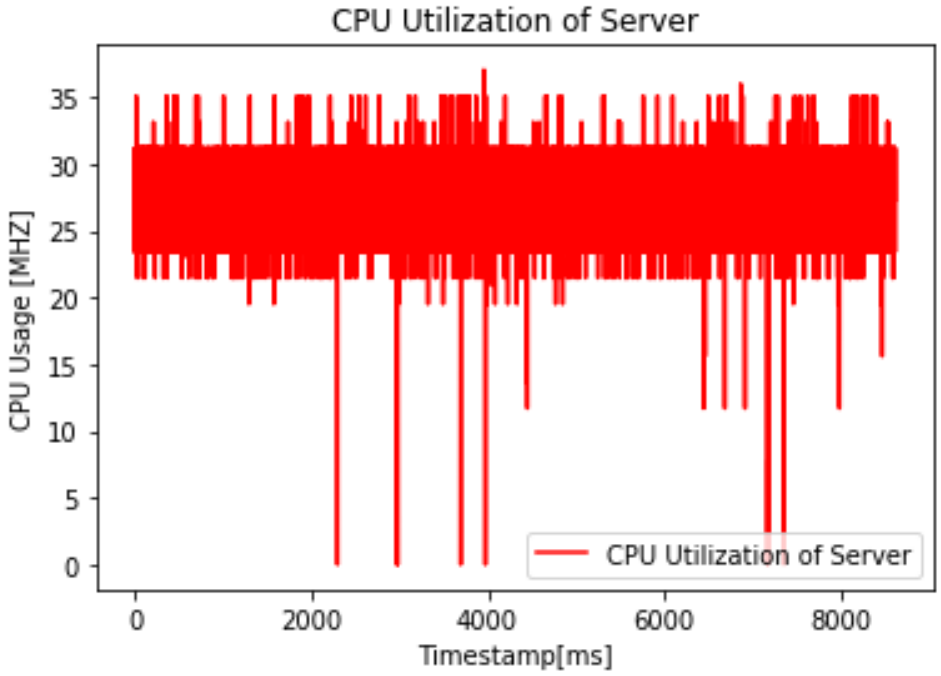

Figure 4.6: Graph of CPU utilization of server

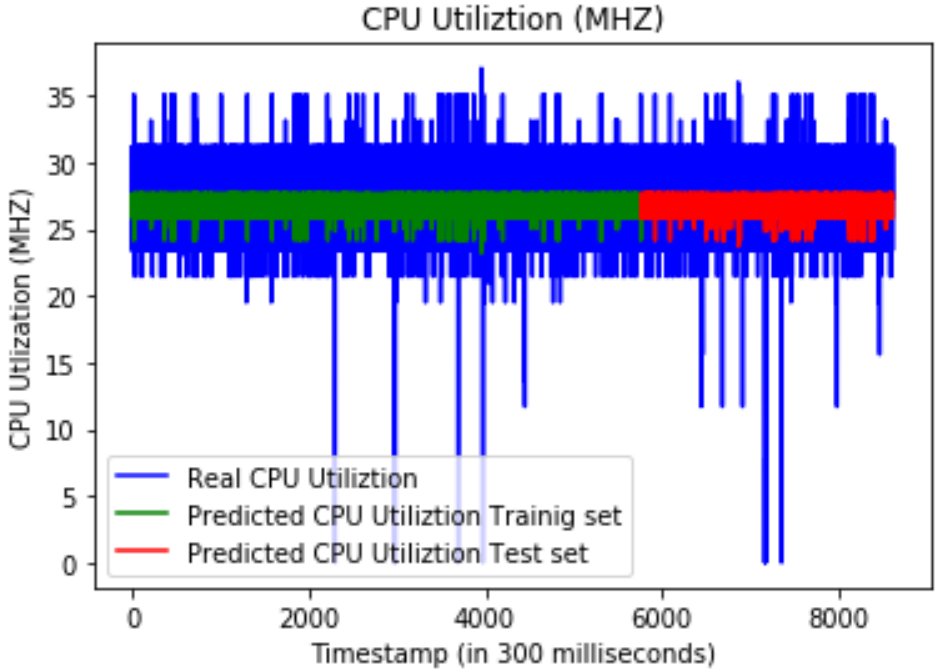

Figure 4.7: Graph of CPU utilization prediction of server

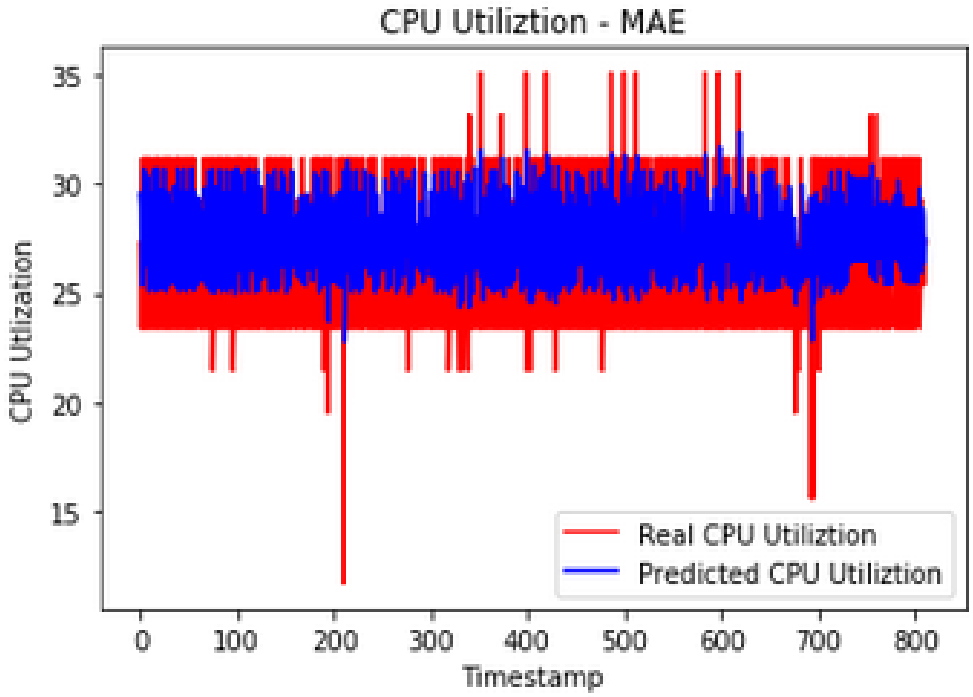

Figure 4.8: Graph of CPU utilization prediction of server

Chapter 5

Conclusion and Future Work

5.1 Conclusion

As cloud computing resources are highly elastic, scalable and dynamic in nature; we have to cope with challenges like resource availability, high throughput of resource, we have understood the significance of resource monitoring and prediction in cloud computing. In this project we have proposed the algorithm for monitoring of resources in effective way and used the LSTM (neural network) for the resource prediction purpose which leads to green computing and generation of revenue for CSP. Here we have only implemented on the one performance parameter which is CPU utlization(MHZ). Our future work is to perform on all the performance parameter of cloud computing such as latency, network bandwidth and throughput etc. Other mechanism such as Incremental Neural Network in LSTM can be implemented.

Due to better way of monitoring and prediction we can reduce the monitoring overhead due to which less computation is done on monitoring and effective utilization of resource is done. This all scenarios leads to energy saving which an important aspect of green computing. This ultimately is able help CSP to generate more revenue.

5.2 Future Work

In future we will implement incremental neural network in LSTM for solving the randomness problem which LSTM is unable to predict.

Bibliography

[1] Z. Xiao, W. Song, and Q. Chen, "Dynamic resource allocation using virtual machines for cloud computing environment," *IEEE transactions on parallel and distributed systems*, vol. 24, no. 6, pp. 1107–1117, 2013.

[2] P. Saripalli, G. Kiran, R. R. Shankar, H. Narware, and N. Bindal, "Load prediction and hot spot detection models for autonomic cloud computing," in *Utility and Cloud Computing (UCC), 2011 Fourth IEEE International Conference on*, pp. 397–402, IEEE, 2011.

[3] S. S. Manvi and G. K. Shyam, "Resource management for infrastructure as a service (iaas) in cloud computing: A survey," *Journal of Network and Computer Applications*, vol. 41, pp. 424–440, 2014.

[4] R. C. Barranco and P. J. Teller, "Resource consumption prediction using neuro-fuzzy modeling," in *Fuzzy Information Processing Society (NAFIPS), 2016 Annual Conference of the North American*, pp. 1–6, IEEE, 2016.

[5] R. Buyya, R. N. Calheiros, and X. Li, "Autonomic cloud computing: Open challenges and architectural elements," in *Emerging Applications of Information Technology (EAIT), 2012 Third International Conference on*, pp. 3–10, IEEE, 2012.

[6] H. Chen, X. Fu, Z. Tang, and X. Zhu, "Resource monitoring and prediction in cloud computing environments," in *Applied Computing and Information Technology/2nd International Conference on Computational Science and Intelligence (ACIT-CSI), 2015 3rd International Conference on*, pp. 288–292, IEEE, 2015.

[7] J. Huang, C. Li, and J. Yu, "Resource prediction based on double exponential smoothing in cloud computing," in *Consumer Electronics, Communications and*

Networks (CECNet), 2012 2nd International Conference on, pp. 2056–2060, IEEE, 2012.

[8] S. Imai, T. Chestna, and C. A. Varela, "Accurate resource prediction for hybrid iaas clouds using workload-tailored elastic compute units," in *Utility and Cloud Computing (UCC), 2013 IEEE/ACM 6th International Conference on*, pp. 171–178, IEEE, 2013.

[9] J. Povedano-Molina, J. M. Lopez-Vega, J. M. Lopez-Soler, A. Corradi, and L. Foschini, "Dargos: A highly adaptable and scalable monitoring architecture for multi-tenant clouds," *Future Generation Computer Systems*, vol. 29, no. 8, pp. 2041–2056, 2013.

[10] J. Montes, A. Sánchez, B. Memishi, M. S. Pérez, and G. Antoniu, "Gmone: A complete approach to cloud monitoring," *Future Generation Computer Systems*, vol. 29, no. 8, pp. 2026–2040, 2013.

[11] S. A. De Chaves, R. B. Uriarte, and C. B. Westphall, "Toward an architecture for monitoring private clouds," *IEEE Communications Magazine*, vol. 49, no. 12, pp. 130–137, 2011.

[12] S. Suneja, C. Isci, R. Koller, and E. de Lara, "Touchless and always-on cloud analytics as a service," *IBM Journal of Research and Development*, vol. 60, no. 2-3, pp. 11–1, 2016.

[13] D. Tovarnak and T. Pitner, "Towards multi-tenant and interoperable monitoring of virtual machines in cloud," in *Symbolic and Numeric Algorithms for Scientific Computing (SYNASC), 2012 14th International Symposium on*, pp. 436–442, IEEE, 2012.

[14] J. M. A. Calero and J. G. Aguado, "Monpaas: an adaptive monitoring platformas a service for cloud computing infrastructures and services," *IEEE Transactions on Services Computing*, vol. 8, no. 1, pp. 65–78, 2015.

[15] A. Ciuffoletti, "Application level interface for a cloud monitoring service," *Computer Standards & Interfaces*, vol. 46, pp. 15–22, 2016.

[16] L. Zhao, S. Sakr, and A. Liu, "A framework for consumer-centric sla management of cloud-hosted databases," *IEEE Transactions on Services Computing*, vol. 8, no. 4, pp. 534–549, 2015.

[17] K. An, S. Pradhan, F. Caglar, and A. Gokhale, "A publish/subscribe middleware for dependable and real-time resource monitoring in the cloud," in *Proceedings of the Workshop on Secure and Dependable Middleware for Cloud Monitoring and Management*, p. 3, ACM, 2012.

[18] H. Kung, C.-K. Lin, and D. Vlah, "Cloudsense: Continuous fine-grain cloud monitoring with compressive sensing.," in *HotCloud*, 2011.

[19] S. K. Sood, "Function points-based resource prediction in cloud computing," *Concurrency and Computation: Practice and Experience*, vol. 28, no. 10, pp. 2781–2794, 2016.

[20] G. K. Shyam and S. S. Manvi, "Virtual resource prediction in cloud environment: a bayesian approach," *Journal of Network and Computer Applications*, vol. 65, pp. 144–154, 2016.

[21] M. Amiri and L. Mohammad-Khanli, "Survey on prediction models of applications for resources provisioning in cloud," *Journal of Network and Computer Applications*, vol. 82, pp. 93–113, 2017.

[22] W. Zhang, B. Li, D. Zhao, F. Gong, and Q. Lu, "Workload prediction for cloud cluster using a recurrent neural network," in *Identification, Information and Knowledge in the Internet of Things (IIKI), 2016 International Conference on*, pp. 104–109, IEEE, 2016.