Prediction and Monitoring Agents using Weblogs for improved Disaster Recovery in Cloud

Rushba Javed[1], Sidra Anwar[2], Khadija Bibi[3], M. Usman Ashraf[4], Samia Siddique[5]
GC Women University Sialkot/ Department of Computer Science & Information Technology, Pakistan
E-mail: [1]rushbajaved@gmail.com,[2]engrsid.es@gmail.com,[3]info.tec15018@gcwus.edu.pk,[4]usman.ashraf@gcwus.edu.pk,
[5]Samiasaddique53@gmail.com,

Abstract—Disaster recovery is a continuous dilemma in cloud platform. Though sudden scaling up and scaling down of user's resource requests is available, the problem of servers down still persists getting users locked at vendor's end. This requires such a monitoring agent which will reduce the chances of disaster occurrence and server downtime. To come up with an efficient approach, previous researchers' techniques are analyzed and compared regarding prediction and monitoring of outages in cloud computing. A dual functionality Prediction and Monitoring Agent is proposed to intelligently monitor users' resources requests and to predict coming surges in web traffic using Linear Regression algorithm. This solution will help to predict the user's future requests' behavior, to monitor current progress of resources' usage, server virtualization and to improve overall disaster recovery process in Cloud Computing.

Index Terms—Cloud Computing, Disaster, Prediction, Monitoring, Disaster Recovery, Prediction Agent, Monitoring Agent, Monitoring and Prediction using Weblog, Server virtualization, Monitor Downtime, Optimize Recovery time.

I. INTRODUCTION

Cloud computing is a prototypical for approving universal, suitable, on-need and worldwide network access to a communal pool of settable computing assets (e.g., webs, servers, storage, applications, and services) that can be swiftly provisioned and liberate with slight managing struggle or facility provider's collaboration. Cloud Computing is fetching one of the greatest well-liked and capable technologies quickly. It deals with numerous opportunities that help the organizations to make well their businesses and practice the expertise in an effective way. Also, it has fruitfully gained the interest from organizations because it deals with a wide range of results and benefits to businesses such as improved elasticity, scalability, accessibility, quickness, and decreases expenses and higher competences.

Despite of growing technological usage, some management concerns still remain. The foremost impending problem in cloud computing is effective disaster management [1]. In cloud, whenever a system smashes or power failure arises, there is a chance of data loss [2] and eventually it results in economic loss [3]. This system crashing and other difficulties happens due to natural Disasters or by the humans from causing costly service interruptions or increasing and decreasing resources demands [4]. A Disaster is a surprising incident that happens in the cloud environment during its active lifespan [5]. However, there are systems for recovering data after a Disaster but these systems take much time to recover the data and sometimes system becomes down/slow which cause trouble for the users. As your growing business is becoming more IT oriented and Internet dependent, you open yourself up to particular risks. Then your technical infrastructure has become more complicated and it needs appropriate systems, backup processes, and recovery plans are used to help reduce the high-cost danger of downtime. Downtime is intensely predominant, with 90 percent of businesses unpredictably losing access to their crucial systems, and one third of its big business are dealing with downtime every month [6]. The regular downtime in the U.S. is 7.9 hours per day while in Europe, it is 10.3 hours. However, several business and management service station use Disaster Recovery (DR) systems to lessen the downtime suffered by catastrophic system disasters.

Here we are going to demonstrate some examples regarding Disaster in cloud computing that brings the system down. Cloud providing businesses such as Amazon, Microsoft, and Google etc. have been experiencing cloud disasters with huge devastation of data and servers. For example, Clients of Virgin Blue were really upset when they couldn't board their booked flights, during an outage that continued up to 11 days [7]. The outage fired up a lot of undesirable press, well as costing the organization in billions. Virgin Blue's airline's registration and online booking systems went down in September 2010 [8]. The outage severely disturbed the Virgin Blue business for a period of 11 days, disturbing around 50,000 customers and 400 flights, and was restored to normal on October 6.

The Cloud Computing IaaS supplier guarantees to money back the users [8] on experiencing consistent downtime. But refund the clients is not the solution. Some infrastructure as service providers (IaaS) also defines flexibility in terms of the redundancy of their

structure. In other words, a distinctive IaaS SLA does not observe disaster resilience. On the other hand, an IaaS supplier could be disaster resistant, ensuring a given QoR after a disaster happening.

To overcome these disasters, Disaster recovery (DR) tools are set in place to support monitor variations that happen in data storage locations and be able to inform consumers of these variations before a disaster happens. Because they monitor changes, DR monitoring tools can frequently improve data security and shorten disaster recovery procedures and as well as reduces downtime. Many researchers, lately, have been using different predictions techniques and machine learning algorithms for disaster prediction. And these techniques and algorithm are presented in form of analysis table in the literature review.

In this paper, a solution is proposed that we will focus on predicting and monitoring by controlling long-term disaster recovery improvement in Cloud Computing. Therefore, the web logs data is taken as a sample while enacting prediction algorithms on it to predict disaster in the cloud. Additionally, these applications would be continuously monitored to classify surges and resources under small to high category of demand. For the purpose of knowing about everything is in actual statistics, we set periodic audits and schedule updates to check that the system is working properly or there is a chance of disaster occurrence in the future. This will let users to freely interact with the technical support employees and will improve the reliability of data availability at the time of disaster recovery. The system will minimize the immediate loss of data by minimizing downtime at the time of disaster. It enables effective management of recovery tasks and lessens the complication of recovery struggles.

The rest of this paper is planned as follows: In the section 2 disaster recovery's background and related work have been discussed. In section the related work, contributions, and limitations of disaster predictions and monitoring techniques are analyzed in comparison. In section 4, the methodology is presented and enacted on a scenario. In section 5, the expected results and outcomes are examined as per that application in disaster. Finally, the paper ends with the discussion and conclusion.

II. Background

Disaster recovery actually came into play in the late 1970s and become a common conversation for companies all over the world [10]. Disaster recovery has turn into a mutual discussion for businesses all over the world because of high-profile cyber-attacks and data losses [21, 22]. In the early days of disaster recovery, the majority of businesses keeps paper records and didn't depend on credible [21]. They converted extra attentive of the possible disturbance triggered by technology downtime when their companies began to clasp the flexibility and storage benefits of digital tech [10, 11]. In 1978, SunGard Information Systems for an organization located in Philadelphia developed the first hot site [10, 22]. Equally

consumer consciousness and the organization itself began to grow rapidly due to the arrival of open systems and real-time handling in the late 1980s and early 1990s. Open systems and real-time handling expands the amount an organization is influenced by their IT systems. Since 1990s, there has been a rapid evolution on the internet [10, 11,21]. Continuous availability of their IT systems depends on the swift growth that leads to organizations of all sizes. Some organizations even set objectives 99.9 percent of the accessibility of their critical systems in the occurrence of a disaster [10,22,23]. This made information maintenance and backup an easier process but downtime was still a problem.

In the era of 2000s, the 11th September attacks on the World Trade Centers have a reflective impact on disaster recovery approach both in the US and abroad. Server virtualization makes the recovery from a disaster a much faster process [11]. With server virtualization, the capability to control processes to a terminated or backup server when the major asset flops is also an efficient way for moderating interruption [10,11,21]. Now recovery is not about back-up and standby servers, but about Virtual Technologies and data sets, that may have been repeated within minutes of the production systems and can be running as the live system within minutes [21,23]. The acknowledgement by all organizations of the significance of keeping Business Stability and having a trustworthy Disaster Recovery plan is returned in the evolution forecast for the sector, with the DRaaS market expect to be worth $6.4 billion by 2020 [10,22].

Cloud Computing faces many problems and challenges such as Security, Data Replication System Monitoring, Maintenance, Disaster Recovery, Server Virtualization and many more. Because you do not know where cloud vendor is storing your data and you have a fear of data loss. And recover your data from the disaster is a big issue, also resources on the cloud are not managed properly. Above all mentioned problem we will work on Disaster recovery, Monitoring, and server virtualization.

III. Related Works

C.-S. M. S. L. S. &. T. D. Pernggives emphasis to use monitoring agents with regular monitoring of resources used or resource utilization [12]. To fix bivariate (includes two variables, essentially independent of one another) the authors encourage and develop a novel technique, starting point for a bivariate time sequence [24]. This technique is the best beginnings that split two-time sequences in such a way that the related facts among them are maximal.

Knobbe et al. implement data mining procedures to information for conduct experiments composed by network monitoring agents. The concern of their study is with real-life information of tracing claim for aircraft and also an affordable portion tracking [13]. Their major goal behind this was to recognize the reason that disturbs the actions of performance metrics. The data values are saved by monitoring agents on 250 limits at orderly time span. CPU capacity permitted memory space, stored readings

and NFS activates were the composed limits. After every 15 minutes, the agents completes the reading for two months, 875000 records with 3500 time-slice of 250 limits subsequently stored in a data metric table as a result. Decision tree algorithm, top n algorithm, rule induction algorithm, and inductive logic programming procedures were used [24]. Authors were able to recognize numerously unpredicted and real complications, for example, all the selected methodologies with performance blockage.

A circulated data controlling scheme called Astrolabe was presented by R. V. B. K & V. W. A. Renesse Astrolabe allows fast updates, gathers significant system state, able to manage user increasing or decreasing request for using server and providing on-the-fly quality collection [14]. Later this ability authorize an application to trace a resource request and also deal with an accessible method to track system state as it changes over a period of time. To resolve an extensive variety of management and monitoring problems is possible by the arrangement of the structures [24]. The attention plan of the author in his proposed system is on its scalability. Later for finding resource request, publish-subscribe, scattered management in large systems, Astrolabe service and use of Astrolabe was described. Astrolabe used a confined form of mobile code and also used point-to-point protocols based on the SQL query language for collection. The protocol they used to give rise to a different correspondence model. Some security deliberations which use a built-in PKI were addressed by Astrolabe. The system scalability is tested by using both simulation and experiments. Astrolabe could measures to thousands of resource requests, with information broadcast interruptions and also system becomes down sometimes.

Vrishali P. Sonanane suggested the prediction of consumer navigation designs via the LCS algorithm [15]. The authors describe the consumer prediction behavior, this contains the following stagesthat are: Information pretreatment, navigation design forming, navigation design mining, assembling a prediction engine. Firstly, in the information treatment phases all the data of weblogs is scrubbed and cleaned. After that for displaying the pages accesses data as an undirected graph, the navigational design mining on resulting consumer access pastern and also model navigational patterns algorithms are applied. In the assembling phase, they attempt to catch a collection of intensely interrelated pages. In the end, in the prediction engine, for the categorized user navigation patterns and predicts the user coming demand, the LCS algorithm is used. To use this LCS algorithm [19] we are able to predict customer coming demands more specifically than estimation of web usage mining.

VedpriyaDongre, JagdishRaikwal [16] presented a system design for locating the hidden navigational patterns via weblog information. In this system design, the server connected too many users and produced a log and the entire information is sent to the pre-processing stage where it is wiped and cleaned. Then, the

information goes to assembling stage where to discover related information from the large database, i.e. web pages retrieved by the focused customer and other consumers; K-means algorithm is applied [19]. At last for prediction the entire cluster records are sent for regression analysis, in the regression analysis cluster records are used to precise records over the numeric values after approximating the load for a target consumer.

R.khanchana and M.Punithanali [17] presentedthe web usage mining method for predicting the user's browsing activities by FPCM clustering. The authors considered the two-stages of prediction model by combing the Bayesian theorem and Markov model. But both models are used for common objective. The authors practice Fuzzy Possibilistic Algorithm for clustering for diversity in consumer's actions to overcome this problem. Tested consequences demonstrate that the percentage of FPCM cluster is 63.6 percent [19], on the contrary, the worldwide and hierarchical percentage is 57.8 percent and 59.9 percent correspondingly. And the authors perceived that FPCM has the talent for improved prediction.

MehradadJalali, Narwat Mustapha, Md. Nasir Sulaiman, Ali Mannat [18] progressive their earlier effort and re-titled their structural design as Web PUM. For the cruising pattern mining phase, they applied the consumer cruising patterns. To predict consumer near future movement, the authors used the LCS algorithm for categorizing all the consumer actions. They over sighted two key tests for cruising pattern mining and their scheme has been verified on CTI and MSNBC.

Dilpreetkaur, A.P.sukhpreetkaur[19]suggested a system framework with Fuzzy Clustering i.e. fuzzy c-means and kernelized fuzzy c-means algorithm. Firstly, the weblog information is picked up and then it sends to the preprocessing stage. After this stage, for the consumer future demand prediction, the fuzzy clustering algorithm is allocated, and the consequences were taken. Researchers decided that KFCM is not only extra strong than FCM but also produces well clusters for prediction.

K. R. Suneetha, R. Krishnamoorthi [20] absorbed the collection of the regularly retrieved outlines of involved consumers. It supports the website engineers in the enhancement of the execution of the web by granting choices to the outlines traversed by the consistently involved consumers [25]. In the first step, from web log information needless as well as unwanted items are displaced. In the second step, for the association of involved consumers from web server log queue, the novel type of the decision tree C4.5 algorithm is applied. The outcomes displayed the enhancement equally for the time as well as memory consumption.

IV. ANALYSIS

The methodology, contributions and limitations of all the above discussed techniques are described in the Table 1.

Table 1.

Ref	Methodology	Contribution	Limitations
[12]	Mutual information	Server system monitoring and reporting	Threshold setting for bivariate time series
[13]	Decision trees, top n algorithm, rule induction algorithm	Data collected by monitoring agents	Performance bottlenecks
[14]	Peer-to-peer	Server system monitoring	Difficult to handle, Time complexity self-configuration
[15]	Use LCS Algorithm	Architecture of Prediction Systems	We can use Web usage mining algorithm more accurately for predicting rather than LSM algorithm.
[16]	K-means and Regression analysis.	Prediction system architecture.	The writer didnot arguesthe factors like time taken for prediction and as well as memory used to store the information.
[17]	FPCM clustering.	Predicting user'sbrowsing behavior.	This technique cannot deals with heterogeneity in user's Behavior
[18]	Online and Offline architecture phases, LCS Algorithm, clustering.	Web PUM.	The core problem with this technique is that the websites made up from dynamically created pages cannot be easily handled.
[19]	Fuzzy Clustering i.e. fuzzy c-means and kernelized fuzzy c-means algorithm.	User future request prediction.	The main problem in this research work is that FCM is less robust and cannot apply on the enormous sets.
[20]	Decision Tree C4.5 Algorithm (Enhanced version)	Web Performance improvement	This research cannot easily manage dynamic website.

There are different techniques and methods analyzied to understand prediction and monitoring of user behavior. To start with first monitoring technique in which C.-S. M. S. L. S. & T. D. Perng [12] used mutual information to emphasize on server monitoring, reporting and utilization but threshold setting for bivariate time series was difficult to implement.

For another researcher, A. V. d. W. D. & L. L. Knobbe [13] used decision trees, top n algorithm, rule induction algorithm for collecting data from monitoring agents to manage user resource requests but this creates several unexpected problems like minimize performnce and requests bottleneck.Additionally, K. R. Suneetha, R. Krishnamoorthi [20] continued the work of and used Decision tree C4.5 Algorithm (Enhanced version) to improve and monitor cloud performace but this research couldn't easily manage dynamic website.

Moreover, R. V. B. K. &. V. W. A. Renesse [14] propsed distributed information management system called Astrolabe which uses peer-to-peer technique for server monitoring. This system permits rapid updates and handle thousands of user's resource requests. But the major problem in this system is that it is difficult to handle and also takes time to proceed user resource demand.

In the first prediction technique, Vrishali P. Sonanane [15] used LCS algorithm to predict user behavior which used weblog data to predict user behavior for future resource demands. The problem is minimal web usag instead of LCS algorithms which is easy to use and handle.

Secondly, VedpriyaDongre, JagdishRaikwal [16] Proposed Prediction system architecture which used K-means and Regression algorithm for finding hidden user navigational behavior from weblog data but the limitations in this system is that writer did not discuss the factors like time taken for prediction and also memory used to store the information.

Additionally, R.khanchana and M.Punithanali [17] continued the previous work and used FPCM algorithm for predicting user behavior in cloud. This technique minimizes the previous limitations and the two levels of prediction model gives improved consequences for common cases but the problem is with the heterogeneity of user's behavior.

Thirdly, MehradadJalali, Narwat Mustapha, Md. Nasir Sulaiman [18] propsed Web PUM which uses Online and Offline architecture phases, LCS Procedure, clusters to predict consumer activities and future movements. The main disadvantage of this system is the websites made up from dynamically created pages that are hard to be handled.

Moreover, Dilpreetkaur, A.P.sukhpreetkaur [19] continued the previous work and used Fuzzy clustering for user future request prediction. This solved the previous problem but in this research work FCM is not as much of robust and cannot be implemented on the large sets.

By discussing them all, the major problem discovered in previous systems is their inability to manage and control users increasing or decreasing resource demands and also if thousands of people using server at the same time and all users generate requests, the practiced system do not handle the situation well which in result happened to reduce system performance and sometimes system got stuck. Also these systems are failed to predict the future resource requests in dynamic websites and takes time in prediction. To alleviate these consequences, a support is needed through which vendor can predict and monitor user resource requests and can minimize downtime by improving performance and resource utilization.

V. METHODOLOGY

The Monitoring Agent for disaster recovery in cloud is proposed to intelligently monitor users' resources requests and to predict coming surges in web traffic. Performing the auto-scaling on cloud services will help to enhance resource utilization and to minimize downtime. There are three Machine learning algorithms in use for predicting requests in website traffic that are: Linear Regression, Multilayer perception and Support vector Machine. Out of aforementioned algorithms, Linear Regression is preferred for the prediction of user resource requests because it outperforms SVM and Multilayer Perceptron on whole [10,13], while there is no significant difference in performance of the other two techniques.

A. Prediction

a) Auto Scaling Algorithm

In cloud, auto scaling is a method that is used to automatically adjust computing resources to ensure the service availability. Firstly, it requires finding the capacity needed for resource demand then to compute plan and compare CPU acknowledgement and number of requests. At the end selected plan and requests would be returned as a result.

Here is the auto-scaling algorithm (in pseudo code) for prediction of future user resource requests.

1. **Begin** {
   Capacity needed = current - required capacity
   Compute scaling plan for Capacity Needed
2. **If** *(scaling plan>1) {*
3. **For** *(each scaling plan)*
   Compute CPU credit and number of instances
   Compare CPU credit and Number of instances
4. **End if{**
   Selected plan = plan with max CPU credit && max number of instances
5. **Else return**
   Number and type of instances of selected plan
   }
6. **End**
   We apply following input to receive required output.
   }
   }

b) Down scaling Algorithm

Downscaling is a method to derive higher resolution from lower resolution variables. It is based on dynamical or statistical approaches. Initially, the capacity needed for resource demands would be provided then a plan would be computed and compare remaining time of requests. Eventually, the selected plan and requests would generate results.

Here is the down-scaling algorithm (in pseudo code) for prediction of future user resource requests.

1. **Begin{**

Capacity needed = current - required capacity
Compute scaling plan for Capacity Needed
2. **If** *(scaling plan>1) {*
3. **For** *(each scaling plan)*
   Compute average remaining time
   Compare remaining time of instances
4. **End if {**
   Selected plan = plan with minimum remaining time
5. **Else return**
   Number and type of instances of selected plan
   }
6. **End**
   Received input and output.
   }
   }

B. Monitoring

To monitor the current metrics of resources such as CPU utilization, Request Rate, Network in/out, a conceptual workflow is designed in Fig.1.
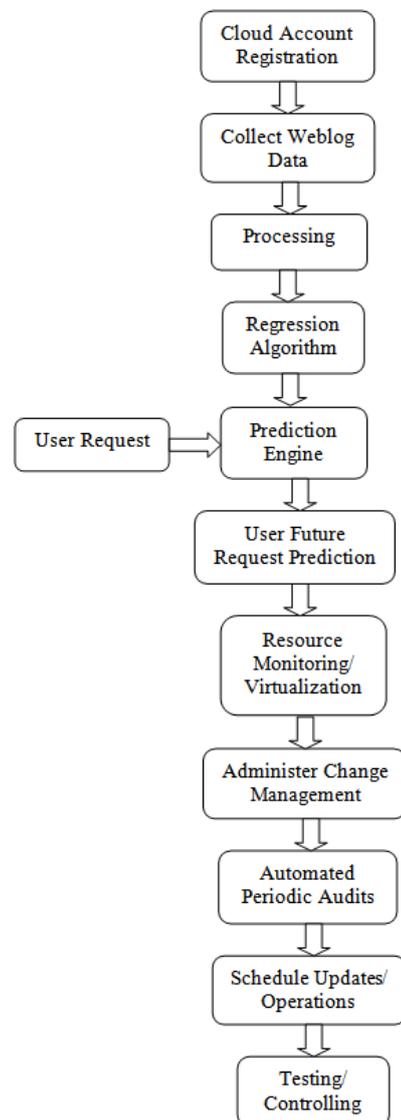

Fig.1.

VI. Anticipated Results

Take an application which is experiencing sudden surge of instances when total capacity of infrastructure is less than the current workload of the application for specific time duration. Additionally, Total capacity of the infrastructure is the workload that it can endure without exhausting. Whereas, generating workload that results in 100% CPU Utilization of VMs is practically challenging for experimental environment. However, enacting the proposed methodologies would be generating following solutions in result:

1. Threshold for maximum CPU Utilization ~= 45% to 35% for different type of instances
2. Total Capacity= Workload that system can endure without exceeding its CPU Utilization threshold
3. Data from near future would be used as only most recent history produces better prediction results as compared to using large history data sets.
4. Time window for which you are making predictions is very important specifically in case of surges so smaller the time window the better would be the prediction results.
5. Linear regression would be practiced as it is examined to outperform the SMO and Multilayer Perceptron on whole, while there is no significant difference in performance of the later two techniques.
6. Auto-scaling buffer is studied and expected to minimize the gap between current and required capacity of the system significantly.

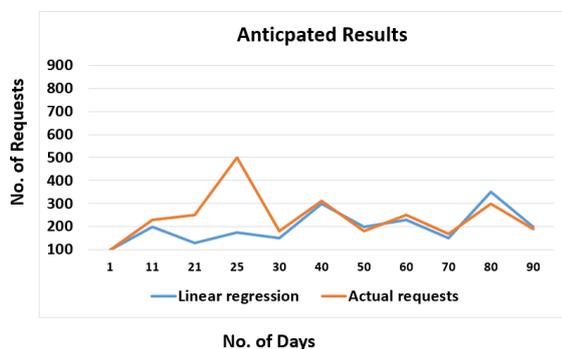Based on aforementioned strategies, expected results are:

Fig.2.

The graph shows the trend at which user's requests for checking the result on a web that is ilmkiduniya.com over three month's period from October-December 2018. The figure shows the No. of days in the horizontal axis and No. of requests in the vertical axis.

Overall, the trend of the website is steady in its regular working days. But if we look at the trend overtime, we can see that the trend gradually increases as the number of requests increases from 400-500 in the time span of 21-25 October. Except a sudden surge was observed from 21-25 October, when result of B.A Sargodha University

declared on the website and No. of requests spiked up to check result.

On the horizontal axis, the trend of graph is steady from 1-20 days and sharply spiked up from 21-25 days and then again trend is regular in the remaining days.

On the vertical axis, the trend of graph is steady when No. of requests are 1-200 and it dramatically increase when No. of requests are increases from 400-500 and becomes steady afterward.

By comparing the website ilmkiduniya.com and other techniques in the analysis of related work, one may observe that by using linear Regression Algorithm to the near past weblog data could predict 70% of the user future requests. And it is considered that adding monitoring activities that are presented in the form of conceptual model in Fig.1 in the experiment can be seen that results of prediction are raised from 70% to 83% as well as downtime is minimized because system is continuously monitored. It is therefore clear that by using algorithms like decision tree algorithm and others mention in analysis table does not produce fully same results of prediction because it requires continuous monitoring of resources in it. However, complete expert analysis and measurements of data are not conducted which prevents us from saying that which things are actually problematic and requires expert analysis. Anyhow, it seems that resource monitoring could benefits better weblog data prediction for future demands also minimizes downtime and can increase the performance of the system.

A. Auto Scaling Output

Input:
1. Predicted workload
2. Required capacity for surge
3. Current capacity

Output:
1. Number and type of instances to launch

B. Down Scaling Output

Input:
1. Predicted workload
2. Required capacity for surge
3. Current capacity

Output:
1. Number and type of instances to terminate

C. Post-Predictions Actions

An instance buffer would be maintained to ensure timely provisioning of instances to handle sudden surges. A variable sized buffer would be used that contains instances in ready state.

The instance buffer will be following these three rules:

1. Update every 20 minutes
2. Minimum 2 instances
3. Maximum 4 instances

D. Monitoring Results

Surges in web traffic monitoring would be classified

into three categories based on the nature of the application usage.

1. **Small Surge:** The surge in traffic for a web application where totals numbers of requests is very small such as small business.
2. **Medium Surge:** The traffic spike of a web application where total no. of requests is average such as university websites.
3. **Large Surge:** The traffic surge of millions of users for very popular websites such as Facebook and Amazon.

VII. DISCUSSION

To examine the prediction as well as monitoring outages in cloud have been argued in a way to highlight the problem domains and formulating the desired strategies to alleviate the consequences. To start with description, a brief overview on disasters, recovery need and cloud outages are given and further, discussed how surges have been arisen and monitored so far while which techniques are under use in cloud computing. By analyzing these techniques, the limitations are extracted and presented through a comparison Table 1: Analysis of Techniques. Therefore, a solution is proposed that is Monitoring Agent for disaster recovery in which predictions are made as well as monitoring at one platform by applying linear regression algorithm. The results are analyzed and presented taking a situation in sudden demands to explain the workflow model. Conclusively, using this cogent plan can help in achieving following outcomes:

- By using monitoring model, there will more appropriate and efficient time optimization.
- Predictions will be made before any disaster occurrence and eventual bottleneck.
- There will be accurate minimization of time as well as proper monitoring of resources.
- Using linear regression algorithm can help in controlled web usage.
- Minimized downtime and concurrent change management.
- Regular conduction of periodic audits would help review the current state.
- Substantial help by automating the system updating and also suitable for dynamic nature of web application.

VIII. CONCLUSION

In daily life, cloud computing becomes more important and almost every company depends on it. By the passage of time disaster has become very problematic in the field of IT. Almost every company is facing big loss of data and also financial loss just because of disasters and it is difficult to recover data in less time, so the system slows down. . After reviewing and examining the loopholes of

current researcher's techniques a dual functionality prediction and monitoring agent is proposed to alleviate the consequences. This Agent will intelligently predict user's coming requests in near future by applying regression algorithm and would also monitor the server utilization to avoid the potential disasters' occurrence. Consequently, an instance buffer would be maintained to ensure timely provisioning of instances to handle such surges.

REFERENCES

[1] T. G. Peter Mell (NIST), "The NIST Definition of Cloud Computing," September 2011.
[2] M. G. Avram, "Advantages and Challenges of Adopting Cloud Computing from Enterprise," 2014.
[3] S. e. a. Jafar, "Identifying Benefits and Risks Associated with Utilizing Cloud Computing," 2013.
[4] F. T. W. X. Colin Ting Si Xue, "BENEFITS AND CHALLENGES OF THE ADOPTION OF," International Journal on Cloud Computing: Services and Architecture (IJCCSA), December 2016.
[5] T. Harris, "CLOUD COMPUTING – An Overview," 2011.
[6] M. PERLIN, "Downtime, Outages and Failures - Understanding Their True Costs," 17 September 2012. [Online]. Available: https://www.evolven.com/blog/downtime-outages-and-failures-understanding-their-true-costs.html. [Accessed 19 November 2018].
[7] S. S. M. E. M. C. H. M. K. C. Rodrigo de Souza Couto, "Network Design Requirements for Disaster Resilience in IaaS Clouds," IEEE Communications Magazine, vol. 52, no. 10, p. 52 – 58 , October 2014.
[8] K. Morrison, "90% of Companies Have Experienced Unexpected Downtime [Infographic]," ADWEEK, 17 October 2014. [Online]. Available: https://www.adweek.com/digital/90-percent-companies-experienced-unexpected-downtime-infographic/. [Accessed 19 November 2018].
[9] M. PERLIN, "Downtime, Outages and Failures - Understanding Their True Costs," 17 September 2012. [Online]. Available: https://www.evolven.com/blog/downtime-outages-and-failures-understanding-their-true-costs.html. [Accessed 19 November 2018].
[10] R. Troutman, "The History of Disaster Recovery," 23 November 2015. [Online]. Available: https://www.engadget.com/2015/11/23/the-history-of-disaster-recovery/. [Accessed 15 October 2018].
[11] N. Cornish, "A brief history of disaster recovery," 9 Junary 2016. [Online]. Available: https://www.comparethecloud.net/articles/a-brief-history-of-disaster-recovery/. [Accessed 15 October 2018].
[12] C.-S. M. S. L. S. &. T. D. Perng, "Data-driven Monitoring Design of Service Level," 9th IFIP/IEEE International Symposium on Integrated Network, 2005.
[13] A. V. d. W. D. &. L. L. Knobbe, "Experiments with data

mining in enterprise," Proceedings of the Sixth IFIP/IEEE International Symposium on, 1999.

[14] R. V. B. K. &. V. W. A. Renesse, "A robust and scalable technology for distributed system monitoring, management, and data mining," ACM Transactions on Computer Systems (TOCS), vol. 21, no. 2, p. 164–206, May 2003.

[15] Vrishali P. Sonavane,"Study And Implementation Of LCS Algorithm For Web Mining", International Journal of Computer Science Issues, Vol. 9, Issue 2, No 3, March 2012.

[16] VedpriyaDongre, JagdishRaikwal ,"An improved user browsing behavior prediction using web log analysis", International Journal of Advanced Research in Computer Engineering and technology (IJARCET), Vol. 4, Issue 5, , May 2015.

[17] R. Khanchana and M. Punithavalli, "Web Usage Mining for Predicting Users' Browsing Behaviors by using FPCM Clustering", IACSIT International Journal of Engineering and Technology, Vol. 3, No. 5, October 2011.

[18] M. Jalali, N. Mustapha et al," WebPUM: A Web-based recommendation system to predict user future movements", in international journal Expert Systems with Applications 37 (2010) 6201–6212.

[19] S. T. Abhishek Chauhan, "Prediction of User Browsing Behavior Using Web Log Data," Gautam Buddha University, Uttar Pradesh, India, vol. 2, 2016.

[20] ShailyG.Langhnoja , Mehul P. Barot , Darshak B. Mehta," Web Usage Mining Using Association Rule Mining on Clustered Data for Pattern Discovery, International Journal of Data Mining Techniques and Applications, Vol 02, Issue 01, June 2013.

[21] N. R. Information, "History of Disaster Recovery: Four Key Facts," 3 March 2015. [Online]. Available: https://rfa.com/blog/2015/03/03/history-disaster-recovery-four-key-facts/. [Accessed 14 December 2018].

[22] P. Croetti, "IT business continuity and disaster recovery: Past to future," 26 Febrary 2018. [Online]. Available: https://searchdisasterrecovery.techtarget.com/podcast/IT-business-continuity-and-disaster-recovery-Past-to-future. [Accessed 14 December 2018].

[23] R. DeVos, "50% of disaster recovery strategies will fail – Here's how you can save yours," 10 October 2017. [Online]. Available: https://www.cloudsecuretech.com/50-disaster-recovery-strategies-will-fail-heres-can-save/. [Accessed 14 December 2018].

[24] J. K. P. S. &. M. H. Sami Nousiainen, "Anomaly detection from server log data A case study," VTT RESEARCH NOTES 2480, 2009.

[25] P. M. B. Megha P. Jarkad, "Improved Web Prediction Algorithm Using Web Log Data," International Journal of Innovative Research in Computer and Communication Engineering, vol. III, no. 5, May 2015.

Authors' Profiles

Rushba Javed has lived in Sialkot since 2000. She did her Matric in Science subjects from Government Higher secondary school Model Town Ugoki and FSC (pre-Engineering) from Superior College Sialkot. Now she is student of BS (IT) at GC Women University Sialkot. She currently works on research as final year project in cloud Computing. She

also attended ethical hacking workshop at University of Gujarat and also certified as Microsoft Office specialist. Her research and publication interests include Cloud architect, Computer science and robotics.

Sidra Anwar was born in Sialkot, Pakistan in 1992. She is a lecturer in Computer Science at the Govt. College Women University, Sialkot, Pakistan and member of the ORIC Steering committee. She did her BS in Software Engineering from Fatima Jinnah Women University, the Mall, Rawalpindi, Pakistan and MS in Project Management from COMSATS Institute of Information Technology, Islamabad, Pakistan. She is also a Registered Engineer under Pakistan Engineering Council Act 1976 with Registration# COMP/14954. She has personally worked and supervised many projects regarding decision support systems serving multidisciplinary domains. Currently, she is working on Outage Prediction and Monitoring, and Efficient Disaster Recovery in Cloud Computing.

Khadija Aashiq born in Sialkot in 1997. She did her matriculation level in 2013 from Government Girls Higher Secondary School, Ugoki Sialkot in Science Subjects and her intermediate level (I.C.S) in 2015 from Government Girls Higher Secondary College Ugoki Sialkot. Now she is doing her BS (Hons) in Information Technology (IT) from GC WOMEN UNIVERSITY, Sialkot. She attended ethical hacking workshop at University of Gujarat and also certified as Microsoft Office specialist. Currently she has been working on research as final year project in Cloud Computing topic. Her main areas of research interest are Cloud Computing, primary dangerous computer viruses, Data Base and Programming Languages.

Dr. M. Usman Ashraf was born in Sialkot, Pakistan in 1988. He received the B.Sc. degree in Mathematics from The University of Punjab, Pakistan in 2007, M.S. degrees in Computer Science from the University of Lahore, Pakistan, in 2014 and currently doing Ph.D. degree in Computer Science from King Abdulaziz University, Jeddah, Saudi Arabia. He is Assistant Professor in the Departmentof Computer Science, GC Women University, Sialkot, Pakistan. He is also a member of Software Engineering group at King Abdulaziz University Jeddah, Saudi Arabia. From 2010 to 2014, he was a Senior Software Engineer (SSE) at Coeus Software solutions, GmbH. His research interests include High Performance Computing (HPC), Parallel Computing, Exascale Computing, Ubiquitous computing, Software Engineering, Location Based Service Systems and Recommender Systems.

Samia Siddique is a student at the Department of Information technology at the government college women university Sialkot. She is Bachelor candidate at the government college Women University Sialkot .where she furthers her research about Information technology, cloud

computing from Chinese and Australia.