

# Exhausting Autonomic Techniques for Meticulous Consumption of Resources at an IaaS Layer of Cloud Computing

Vivek Kumar Prasad<sup>(✉)</sup> and Madhuri Bhavsar

Nirma University, Ahmedabad 382481, Gujarat, India  
{vivek.prasad,madhuri.bhavsar}@nirmauni.ac.in  
<http://www.nirmauni.ac.in/>

**Abstract.** Internet based computing has provided lots of flexibility with respect to the usages of resources, as per the current demand of the users, and granting them the said resources has its own benefits, if given in proper manner i.e. exactly what the user has asked. In this paper the autonomic computing concepts has been discussed which will be very useful for the better utilisation of the resources at an IaaS (Infrastructure as a Service) level of the cloud computing. As Cloud Computing is highly scalable and virtualisation has become an important means for the efficient utilisation of the resources. Seeking the right amount of the resources at right time should be the goal of any CSP (Cloud service provider), On the other hand the CSPs has to deal with the situation of over provisioning and under provisioning, there should be some self-managing scheme through which the resources should be made available to the requesting user in an efficient manner to satisfy the need of their requirement with an improved resource utilisation. We have discussed the usage of autonomic computing to enhance the resource utilisation in the IaaS of cloud computing through various ways.

**Keywords:** Autonomic computing · Cloud computing  
Over provisioning · Under provisioning · Resource utilisation  
Infrastructure as a Service

## 1 Introduction

Cloud [22] is basic necessity today for the organisation because of its dynamic benefits and even though the organisation is 100% in cloud, it still requires the skilled professionals and managers, who can understand the right solutions where the cloud can be well suited to their enterprises. The reports indicates that the cloud will emerge in the future for betterment of the enterprise stakeholders, for handling the services [17] such as, IaaS (Infrastructure as a Service), PaaS (Platform as a service) and SaaS (software as a Service).

The resources at an IaaS level of the cloud the will give the illusion to the end user that, the an infinite pool of resources are present for the end user because of the term called as virtualisation.

Every VMs should be associated to certain amount of resources available at an IaaS of cloud computing. Our aim should be to efficiently use these VMs so that there will be minimum wastage of the resources. In certain situations, their will be the wastage of resources [3].

The autonomic computing [12] can be classified into four parts i.e. self-configuration means the system should be adaptable with the changing environment, Self-healing which means discovering and diagnosing the problem in advance or heal the problem and prevent disruption, Self-Optimisation, i.e. tune the resources and workload to maximise the utilisation, Self-protecting i.e. to anticipate, detect and identify the problem and protect itself.

## 2 Objective

- Growing demand of the infrastructure has also increased the energy consumption. Our objectives will be to reduce the energy consumption using efficient resource usages. The resources has to be used in such a manner, so the even the last resource available at the data centre has to be used efficiently. To make use of the Autonomic Techniques to utilise the resources available at an IaaS level of the cloud computing.

## 3 Methodology

The Fig. 1 indicates the impact of autonomic computing at various levels/areas where, if the resources are used efficiently then, the energy will be reduced and the systems will have the properties of go green concepts and an increased revenue. The security, QoS Dynamic resource allocation, Iterative optimisation, Root cause analysis and energy efficiency can be combined with autonomic computing to efficiently use the resources available at IaaS.

### 3.1 Future Direction for Research in the Field of Resource Optimisation at the IaaS Level with the Help of Autonomous Computing

The autonomic computing [12], have four parts known as self-configuration, which means the system should be adaptable to the changing environment, Self-Healing, means the act to heal itself from the upcoming problem, self-optimisation indicates the handling of the system in such a way so that the resources should be maximally utilized, followed by self-protection, it handles the identification of security issues and protect itself.

**Quality of Service:** Cloud service providers (CSPs) need to ensure that sufficient amount of resources are available as per the requirement of the end users, the QoS [2] requirements Such as limit (deadline), response time (latency), and budget constraints should be encountered with the current requirement of the



**Fig. 1.** The taxonomy of the autonomic computing and efficient resource management in an IaaS level of cloud computing.

end user. These QoS foundation is derived from SLAs (Service Level Agreements), and if any violation happens w.r.t. the SLA, will lead to penalty and the QoS will be affected.

The DeSVi [8] architecture has been proposed for monitoring and detecting of any SLAs violation in cloud computing, the main components are VM deployer, which is responsible for allocation of tasks and its mapping to the available resources and another one is application deployer, which is responsible for the execution of the application and its metrics.

**Security:** Security [13] is an important aspects of cloud computing because of its distributed nature, as Differentiate between steady request verses DDOS attack, as if a coordinated attack is launched against the SaaS provider, the unforeseen rise in traffic might be erroneously assumed to be from the original/legitimate requests and resources would be scaled up to handle them. This in turn will increase running cost of application and also the wastage of the resources. Here again the resources (data storage and network hardware) will be misused.

**Performance Anomalies [7]:** In this paper, the concept of UBL (Unsupervised Behaviour Learning) has been discussed, which is a black-box unsupervised behaviour learning and is used to find out for the anomaly prediction of the system at IaaS clouds. The unsupervised Behaviour Learning influences the SOM (Self-Organizing Map) which captures the changing behaviour of cloud instances with human intervention. Based upon this learning techniques the UBL predict

previously unknown performance anomalies and delivers clues for any anomaly causes. This also target the scalable behaviour learning by making use of virtualizing and distributing the learning task to the distributed hosts.

**Energy Efficiency [4]:** As cloud will give the illusion to the end user that cloud provides infinite pool of resources and on the other side the cloud needs to look for efficient usages of the energy also, example include as over provisioning of the resources, minimising the carbon footprint and server consolidation. Applications need to be scheduled in such a way, so that their total energy consumption is minimized by the effective usages of the resources at IaaS level. There are various research has been done on this field, some of them has been discussed below:-

In paper [9] considers the SLA with effective VM placement is used to minimize the operational cost in the cloud computing system which in turn reduces the energy consumption by considering the general queuing theory models for real world workloads.

In this research paper [5] the energy has been reduced and there is also a decrease in resource wastage by fixing up the predefined SLAs (Service Level Agreements) i.e. the SLA violation = 0% and the mechanism that has been used here are as Monte Carlo and random methods.

Here the authors [6] has mentions about the advantages of using a good architecture for energy efficient cloud computing and also discussed about the algorithms that are energy aware and should work within the limits of the SLAs.

**Dynamic Resource Allocation:** Scaling in/Scaling out [28] (that is the expanding/shrinking of resources, also called as elasticity and is one of the important property of the cloud computing), has to be carried out w.r.t. changing demands of the end users. The dimensions of the resources that has to be considered are, number of CPUs, amount of memory and size of the virtual disk at IaaS level of cloud computing. In the research paper [14], the CometCloud autonomic cloud engine concept has been highlighted that works on policy based mechanism, The keywords Autonomic cloudbridging and cloudbursting has been described, the cloudbridging merges the computation of local environment (i.e. grids and datacenters) and public cloud services (i.e. Eucalyptus and Amazone EC2) as well, the autonomic cloudbursting allows spikes in demand and dynamic application scale out for dynamic workloads.

**Prediction for Resource Selection:** The total cost of the resources depends on the type of the resources made available to the end user or resources provisioned, a prediction mechanism should be realised, and that takes into account of historical data execution statistics, for the fulfilment of the resources demand. If the predicted resources estimates are correct, then the wastage of the resources are minimised.

In this paper [25] authors have discussed about the workload forecasting and about the optimal resources allocation, challenged involved in autoscaling, predictive algorithms for autoscaling and its empirical results which satisfies the QoS and less operational cost.

The ecosystem of existing Big data tools [11], the analytics today require the support for big data and its implementation in the cloud computing. As there are various open source technologies related to big data analytics are available such as Spark (analytics), Hive and Pig, Storm (used in stream processing), YARM (MapReduce and other parallel programming)/Hadoop, HDFS (File systems and NOSQL databases), Cassandra and CouchDB.

The main aim is to identify how to use these tools, so that minimum resources has to be used (without the wastage of the resources) and no SLAs violation.

In [1] authors have discussed the state of the art of scalable data management for cloud computing infrastructure for heavy and analytical workloads. The designing of the data management has been highlighted. Apart from this different multitenancy models in the database has also been identified.

**Scalable Decision Making Algorithms:** Will be used for Big data analytics kind of scenarios [20] and it has to be recognised using scalable data management systems which uses machine learning, artificial intelligence, decision making and data mining techniques in clouds.

Here in this research paper the authors [29] have mentioned about the migration scenarios in the cloud, as migrations can be done in seconds or sometime more than this depending on the size, work type and bandwidth of the VMs and physical machines, the migration techniques should be automated through application environment with pre-defined strategy, and less human intervention. The two steps for implementing autonomous concepts are to adopt a distributed architecture where resource management is decomposed and each tasks is performed by Autonomous Node Agents, which are tightly coupled with the physical machines.

**Root Cause Analysis/Identifying the Co-relation:** In the real application scenario of cloud computing, the changes done at one end can affect the other end too, because of coupling hence mining dependency between anomalies of two different application layers are an important promising research direction. If the dependency has been identified, then it can be modelled maintained in the form of knowledge representation languages in the system premises and is known as knowledge-base.

The paper [21] discuss about the formal mathematical based decision model which established a logical chain of services requirements, the basic aim is to determine which cloud provider is well suited w.r.t. the requirement of the users, based upon these analysis, the risk are modelled by keeping view of integrity, availability and confidentiality.

**Multi-resource Anomaly Detection:** If multiple resources are identified for an anomaly detection then its an advantage, as its nice to find out the target resources which contributes more to the anomalies, that has been detected in the application and that affects the QoS, on the other side considering only one resource at a time causes needless delay.

In this paper [24] the Intrusion Detection and Prevention Systems (IDPS) and alarm management techniques has been discussed, the hardware and software

changes constructs the autonomic managers monitoring scheme and it optimise its use of resources without any human intervention.

**Software defined network** have current challenges as efficient utilisation of storage and network- resources utilisation at IaaS level through mathematical and statistical methods.

Here the authors [23] have discussed about the autonomous software defined network which make use of agent based architecture which controls the network, and the network controller generates events every time, when the network changes its state, and these events can be used to update the states or the facts in the knowledge base.

**Failure Management** in cloud computing, which is also known as Fault tolerance for performance optimisation or grace full degradation which in turn results as robustness, stability and reliability of the system, lead to the better utilisation of the resources.

The authors [15] have proposed an efficient fault management mechanism, which is based upon cognitive control loops, which uses alarm dataset, which consists of the data that has been inserted at the time of learning phase of the control loops, the SWRL (Semantic Web Rule Language) and Ontology are used to show the relationship between the alarm and its consequent services/actions, even the alarms are used w.r.t. there priority and is useful for finding the root causes easily by imposing the concept of associated rule miming techniques.

**The Iterative Optimisation**, which entitles the subset of analytic task, the functions which is being repeated, need to be profiled, information should be reserved, so that the cost and execution time will be known in advance and the resources can be utilised properly.

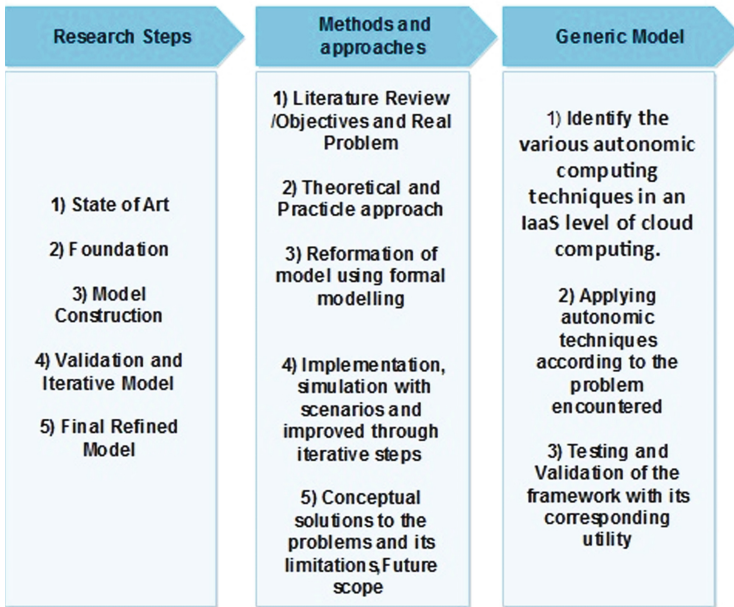
In this paper [10] the authors have proposed the task scheduling optimisation algorithms, with minimised of cost of resources used. They have used the technique of PSO (Particle swarm optimisation) with crossover, mutation and local search algorithms.

The paper [18] discussed about the QoS and its optimisation s design requirements, the application software should be adaptable to different run time situations, the cloud should also have infrastructure for deploying monitoring applications based on the QoS parameters and there should be a cloud feedback loop, that must be able to track the performance model and also to be used for management decisions.

In paper [16] the scheduling of dataflow operations has been considered, as minimum completion time in a given budget, than to minimise the monetary cost given a deadline and the differences between completion time and monetary cost, these values will be used in approximate optimisation framework which deals with the elasticity in the cloud. The practice proposed in the final step of the optimization of the system and its all of the parameters are instantiated automatically using functions or statistics collected during previous executions. System agents [19] (embedded in to system), monitoring CPU, schedulers, optimisation of the dynamic resource handling techniques, here the different agents will interact with each other for Self-Management.

In paper [27] the Multi Agents System concepts has been used, where agents will interact with each other for intelligent behaviour which result into a flexible cloud, autonomic and scalable and it also mentions about the techniques and methodology that account for the changing states and its behaviour. In an another paper [26], the concept of multi agent system has been included in the cloud environment for achieving the goal of implementing high performance complex systems and intelligent applications, the intelligence can be achieved by incorporating dynamic, flexible and autonomous behaviour.

**The Generic Architecture and Its Approach:** The autonomic computing techniques has been used to automate the operations in the cloud with minimal intervention of human interaction. As if multi resources anomaly detection can be handled using autonomic techniques then the resources can be used with its maximum utility, same is the case with SLA management with respect to the QoS parameters and others, same has been discussed above. The correct prediction will enhance our system and will lead to the optimum usages of the resources.



**Fig. 2.** Research approach towards the generic model.

The Fig. 2 discuss about the approach to the generic model, which initially starts from the state of art and move along the various methods and approaches which finally leads to the generic model, that should consists of the following features:-

- The concepts such as QoS metrics, unsupervised behaviour, queuing theory, Monte Carlo methods can be used for energy efficient resource management. The ideas of policy based management, dynamic resource demand, machine learning, artificial intelligence, data mining techniques decision making using pre-defined strategy and autonomous node agents can be used for the process of predicting resource selection and scalable decision making algorithms. Alarm management techniques, SDN agents and raising alarm with its related action decision approaches with monitoring mechanism can add benefits to the automatic anomaly detection mechanism.

Similarly new avenues has to be explored for making the cloud fully autonomous. The pairing of the autonomous techniques to its branch technology has to be analysed and implemented to make worth utilisation of the resources available at an IaaS level.

## 4 Conclusion

The autonomic computing mechanism to cloud computing will be a boon to the current technological scenario where cloud computing concepts are being used, as this will reduce the human intervention and the system itself will take wise decisions based upon the dynamism but still we have lots of open issues and challenges. In this paper we have highlighted a number of questions which can be used as a research agenda for making full utilisation of the resources available at an IaaS level of cloud computing by exhausting autonomic techniques. We have also identified the solution for the above problem and also mentioned some basic solutions which can, if implemented properly will result into the better utilisation of the resources.

## References

1. Agrawal, D., Das, S., El Abbadi, A.: Big data and cloud computing: new wine or just new bottles? Proc. VLDB Endow. **3**(1–2), 1647–1648 (2010)
2. Al-Shehri, S.F.S., Li, C.L.: Quality of service for cloud computing. In: Advanced Materials Research, vol. 905, pp. 683–686. Trans Tech Publications (2014)
3. Armbrust, M., Fox, A., Griffith, R., Joseph, A.D., Katz, R., Konwinski, A., Lee, G., Patterson, D., Rabkin, A., Stoica, I., et al.: A view of cloud computing. Commun. ACM **53**(4), 50–58 (2010)
4. Berl, A., Gelenbe, E., Di Girolamo, M., Giuliani, G., De Meer, H., Dang, M.Q., Pentikousis, K.: Energy-efficient cloud computing. Comput. J. **53**(7), 1045–1051 (2010)
5. Borgetto, D., Maurer, M., Da-Costa, G., Pierson, J.-M., Brandic, I.: Energy-efficient and SLA-aware management of IaaS clouds. In: Proceedings of the 3rd International Conference on Future Energy Systems: Where Energy, Computing and Communication Meet, p. 25 (2012)
6. Buyya, R., Beloglazov, A., Abawajy, J.: Energy-efficient management of data center resources for cloud computing: a vision, architectural elements, and open challenges. arXiv preprint [arXiv:1006.0308](https://arxiv.org/abs/1006.0308) (2010)



7. Dean, D.J., Nguyen, H., Gu, X.: UBL: unsupervised behavior learning for predicting performance anomalies in virtualized cloud systems. In: Proceedings of the 9th International Conference on Autonomic Computing, pp. 191–200. ACM (2012)
8. Emeakaroha, V.C., Netto, M.A.S., Calheiros, R.N., Brandic, I., Buyya, R., De Rose, C.A.F.: Towards autonomic detection of SLA violations in cloud infrastructures. *Future Gener. Comput. Syst.* **28**(7), 1017–1029 (2012)
9. Goudarzi, H., Ghasemazar, M., Pedram, M.: SLA-based optimization of power and migration cost in cloud computing. In: 2012 12th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGrid), pp. 172–179. IEEE (2012)
10. Guo, L., Zhao, S., Shen, S., Jiang, C.: Task scheduling optimization in cloud computing based on heuristic algorithm. *JNW* **7**(3), 547–553 (2012)
11. Hashem, I.A.T., Yaqaob, I., Anuar, N.B., Mokhtar, S., Gani, A., Khan, S.U.: The rise of big data on cloud computing: review and open research issues. *Inf. Syst.* **47**, 98–115 (2015)
12. Huebscher, M.C., McCann, J.A.: A survey of autonomic computing degrees, models, and applications. *ACM Comput. Surv. (CSUR)* **40**(3), 7 (2008)
13. Hwang, K., Li, D.: Trusted cloud computing with secure resources and data coloring. *IEEE Internet Comput.* **14**(5), 14–22 (2010)
14. Kim, H., Parashar, M.: Cometcloud: an autonomic cloud engine. In: *Cloud Computing: Principles and Paradigms*, pp. 275–297 (2011)
15. Kim, S.-S., Seo, S.-S., Kang, J.-M., Hong, J.W.-K.: Autonomic fault management based on cognitive control loops. In: 2012 IEEE Network Operations and Management Symposium (NOMS), pp. 1104–1110. IEEE (2012)
16. Kllapi, H., Sitaridi, E., Tsangaris, M.M., Ioannidis, Y.: Schedule optimization for data processing flows on the cloud. In: Proceedings of the 2011 ACM SIGMOD International Conference on Management of Data, pp. 289–300. ACM (2011)
17. Kundu, A., Banerjee, A., Saha, P.: Introducing new services in cloud computing environment. In: *International Journal of Digital Content Technology and Its Applications*, AICIT. Citeseer (2010)
18. Li, J., Chinneck, J., Woodside, M., Litoiu, M., Iszlai, G.: Performance model driven QoS guarantees and optimization in clouds. In: Proceedings of the 2009 ICSE Workshop on Software Engineering Challenges of Cloud Computing, pp. 15–22. IEEE Computer Society (2009)
19. Lopez-Rodriguez, I., Hernandez-Tejera, M.: Software agents as cloud computing services. In: Demazeau, Y., Pěchouček, M., Corchado, J.M., Pérez, J.B. (eds.) *Advances on Practical Applications of Agents and Multiagent Systems. Advances in Intelligent and Soft Computing*, vol. 88, pp. 271–276. Springer, Heidelberg (2011). [https://doi.org/10.1007/978-3-642-19875-5\\_35](https://doi.org/10.1007/978-3-642-19875-5_35)
20. Low, C., Chen, Y., Mingchang, W.: Understanding the determinants of cloud computing adoption. *Ind. Manag. Data Syst.* **111**(7), 1006–1023 (2011)
21. Martens, B., Teuteberg, F.: Decision-making in cloud computing environments: a cost and risk based approach. *Inf. Syst. Front.* **14**(4), 871–893 (2012)
22. Mell, P., Grance, T., et al.: The NIST definition of cloud computing (2011)
23. Passito, A., Mota, E., Bennesby, R., Fonseca, P.: AgNOS: a framework for autonomous control of software-defined networks. In: 2014 IEEE 28th International Conference on Advanced Information Networking and Applications (AINA), pp. 405–412. IEEE (2014)
24. Patel, A., Taghavi, M., Bakhtiyari, K., JúNior, J.C.: An intrusion detection and prevention system in cloud computing: a systematic review. *J. Netw. Comput. Appl.* **36**(1), 25–41 (2013)

25. Roy, N., Dubey, A., Gokhale, A.: Efficient autoscaling in the cloud using predictive models for workload forecasting. In: 2011 IEEE International Conference on Cloud Computing (CLOUD), pp. 500–507. IEEE (2011)
26. Talia, D.: Cloud computing and software agents: towards cloud intelligent services. In: WOA, vol. 11, pp. 2–6. Citeseer (2011)
27. Talia, D.: Clouds meet agents: toward intelligent cloud services. *IEEE Internet Comput.* **16**(2), 78–81 (2012)
28. Xiao, Z., Song, W., Chen, Q.: Dynamic resource allocation using virtual machines for cloud computing environment. *IEEE Trans. Parallel Distrib. Syst.* **24**(6), 1107–1117 (2013)
29. Yazir, Y.O., Matthews, C., Farahbod, R., Neville, S., Guitouni, A., Ganti, S., Coady, Y.: Dynamic resource allocation in computing clouds using distributed multiple criteria decision analysis. In: 2010 IEEE 3rd International Conference on Cloud Computing (CLOUD), pp. 91–98. IEEE (2010)