A Comparative Study on Load Balancing and Energy Efficiency Techniques in Cloud Paradigm

Shivani Bajaj [1], Manisha Malhotra[2]

University Institute of Computing, Chandigarh University, Mohali, Punjab, India

Abstract: Cloud computing is a materialise technology now-a-days that provides computing, storage resources and communication resources and it is that kind of technology which attracts the ICT (Information and communication Technology) services that provides huge distribution of online services. It provides a great infrastructure which involves large centres including the large amount of servers dealing with the requests given by the clients. There is rapid growth of demand for the power which computes the creation of large-scale data centres. The data centres consume a huge amount of electricity which results in carbon dioxide emission and high operational costs. This paper discusses the load balancing approach in cloud computing environment. Numerous loads are concerned with cloud such as memory load, network load and processor load. Load balancing is the task of distributing the loads across the nosed in the network which provides efficient resource utilization in the case of overloaded node. How these all loads are under loaded to make the network work easily in the cloud environment is the main issue which is resolved with different prescribed algorithms such as round robin, ant colony and opportunistic load balancing. The comparison of the algorithms and the load balancing techniques are done to show which is better in distinct network cases.

Keywords: Cloud Computing, Energy Efficiency, Load Manager, Power Consumption, Resource Scheduling, Security.

I. INTRODUCTION

The cloud computing is that kind of concept which gives an immediate action of many well researched domains like grid computing, cluster computing, distributed computing and virtualization. The data centers of cloud computing give many virtualization technologies that may allow the scheduling of workload on less number of servers which keeps the better utilization as the different kind of workload may give different kind of footprints. There are many companies which are offering cloud computing services for the expansion of cloud infrastructures scuh as Microsoft, Amazon, Gogrid, vCloud Express, Layered Technologies, ENKI Prima Cloud. From the Google's point of view, there are five characteristics of cloud computing user centric, task centric, powerfulness, programmability and intelligence. Cloud computing gradually writes off the initial cost over the diversity of workloads, shared system operators and the distributed servers' offers different kind of services based on computation and operational tasks.

There are many unique issues based on the cloud computing infrastructure such as standardization, dynamic scalability, debugging, reduction of operational costs, reduction of carbon emission and the most important privacy and security of Information and Communicational Technology resources. Now-a-days it has become the major concern of research that "How to reduce carbon emission". This is all because of the energy which is required by the large-scale data centers for its power supply, cooling, and operations. Therefore reduction of power consumption has become a major issue in this present time. [6]

Cloud computing is defined as a type of computing that relies on sharing computing resources rather than having local servers or personal devices to handle applications. Cloud computing is comparable to grid computing, a type of computing where unused processing cycles of all computers in a network are harnesses to solve problems too intensive for any stand-alone machine.

In cloud computing, the word cloud (also phrased as "the cloud") is used as a metaphor for "the Internet," so the phrase cloud computing means "a type of Internet-based computing," where different services —such as servers, storage and applications — are delivered to an organization's computers and devices through the Internet.

The Cloud model promotes availability and is composed of three essential characteristics, three service models, and four deployment models.

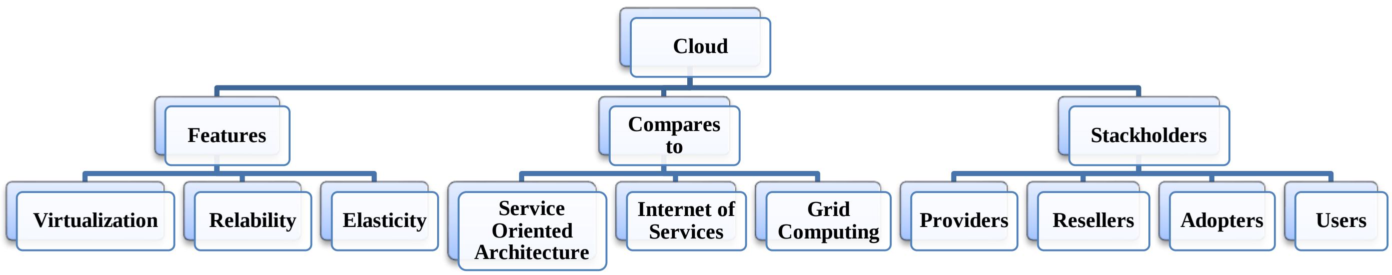

Fig.1: Cloud Computing Model

1.1 Services of Cloud Computing:

Services over the internet are offered under some deployment models that are: infrastructure as a service (IaaS), software as a service (SaaS), platform as a service (PaaS) and network as a service (NaaS). Operators or providers give many service portfolios which vary in service specification. This comprises of configuration of the VMs, the nature of hardware/software security services the programmer's degree of control, portability guarantees, network service configuration and storage scalability.

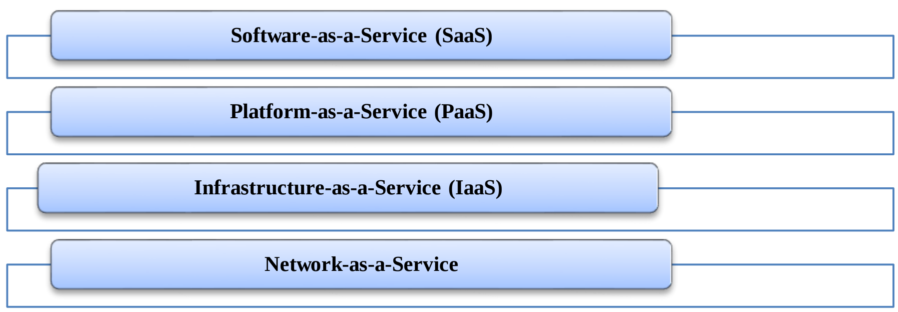

Fig. 2: Architecture of Cloud computing

A. *Infrastructure as a Service (IaaS):* This layer delivers hardware components (like server and storage) and software as services.
B. *Platform as a Service (PaaS):* Cloud application developers are the users of this layer. Automatic scaling, load balancing and integration with other services (e.g. email services) are the major benefits to cloud application developer.
C. *Software as a Service (SaaS):* This layer hosts the software and provide to the customer through Internet. It reduces the purchase and maintenance cost of the customer.

Cloud consists of a number of resources that are different with one other via some means and cost of performing tasks in cloud using resources of cloud is different so scheduling of tasks in cloud is different from the traditional methods of scheduling and so scheduling of tasks in cloud need better attention to be paid because services of cloud depends on them.

1.2 Benefits of Cloud Computing

There are several benefits of Cloud computing which are described below:

Open Access: Cloud service provider can be accessed with the help of fair internet connection.

Improved economies of scale: On the user side, by decrease in the investment and running costs and on the provider side, by providing higher productivity in arranging infrastructure services with high survivability and flexibility.

Capacity for on-demand infrastructure and computational power: Users can demand for the computational power, storage and other infrastructure according to their need according to pay per use model.

Improved resource utilization: Resources are utilized properly because whenever users don't need a resource then they return it back to the cloud provider. So in this way elasticity and flexibility can be increased.

Reduced information technology (IT) infrastructure needs: Cloud computing provides infrastructure-as-a-service on demand for user. So there is no permanent need to purchase the infrastructure for the IT. The user can purchase it from cloud provider whenever needed.

Resource pooling: The consumer generally has no information about the locality of the service provider. Thus the provider serves multiple consumers by assigning resources dynamically and virtually.

Control systems with abstract policies: There is no need to provide unnecessary detail about the core business component to the user.

Organizations focus on their core competencies: Non IT user can contact to IT service provider for their business activity needs.

1.3 Goal of Cloud Computing

The goal of cloud computing is to apply traditional supercomputing, or high-performance computing power, normally used by military and research facilities, to perform tens of trillions of computations per second, in consumer-oriented applications such as financial portfolios, to deliver personalized information, to provide data storage or to power large, immersive online computer games.

To do this, cloud computing uses networks of large groups of servers typically running low-cost consumer PC technology with specialized connections to spread data-processing chores across them. This shared IT infrastructure contains large pools of systems that are linked together. Often, virtualization techniques are used to maximize the power of cloud computing.

1.4 Deployment Model of Cloud Computing

Basically, there are four modes of deployment in cloud computing, or four kinds of cloud, the private cloud, the community cloud, the public cloud and the hybrid could.

1. Private cloud: The cloud infrastructure is built separately for a customer, and provides the most effective control of the data, the safety and quality of service. Private cloud can be deployed in the enterprise data center, also can be deployed on a hosting site, by a single organization owned or leased.
2. Community cloud: Infrastructure is shared by several organizations, and is a common concern of community service (such as task, security requirements, policies and compliance considerations).
3. Public cloud: Infrastructure is a sales of cloud computing services organizations have, the organization will cloud computing services to the general public or to a large industry group, often in public cloud apart from customer where a building hosting, and by providing a like enterprise infrastructure of flexible even temporary extension, provides a method to reduce the risk and cost of the customers.
4. Hybrid cloud: Infrastructure is composed of two or more than two kinds of cloud (private, community and public), each cloud still remain independent, but combine them with the standard or proprietary technology, with data and the portability of applications (for example, can be used to treatment of sudden load), hybrid cloud helps to provide on-demand and external supply expansion.

1.5 Load Balancing:

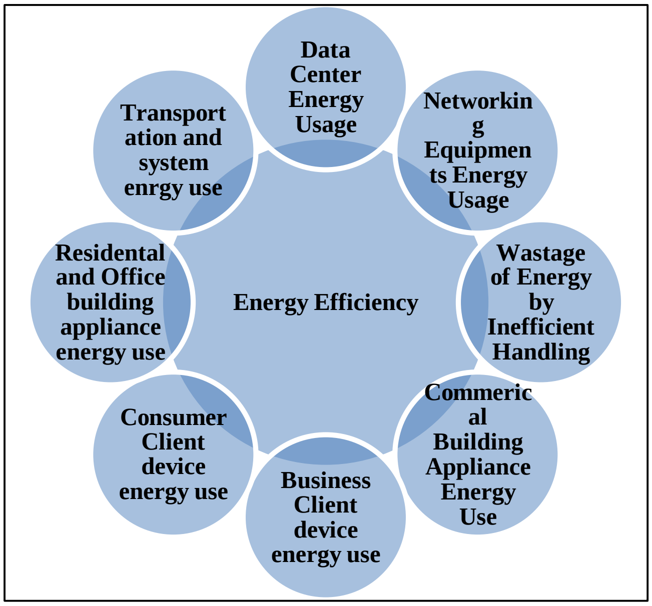

Fig. 3: Energy Efficiency

In the field of cloud computing load balancing plays a vital role as it allocate and balance the load of various resources among the distinct components and nodes on the network. The main purpose of load balancing is to optimize resource use, minimizing the time of response, maximizing the throughput and the main key point is to avoid the key load on the any single node. Load balancer is the main component which balances the load over the nodes. It is a software program that listens to the external clients on the port which request for services. The work of load balancer is to forward request to the server at backend, which again replies back to load balancer. This permit the load balancer to answer the request of the user.

1.6 Energy Efficiency in Cloud Computing

Energy Efficiency can be defined as reduction of energy used by a given service or level of activity[14]. In spite of scale and complexity of data center equipment it can be highly difficult to define the proper activity that could be examined for the efficiency of energy. So there can be four scenarios which may define within the system where the energy is not utilized in an efficient manner. The basic research on energy efficiency originated from the grid computing and the environments of data centers. Energy efficiency is a major concern in the large-scale data centers. In US the data centers consumes about 1.5 percent of electricity in earlier time which is quite equal to the annual ebergy consumption. It costs a lot to around 4.5 billion per year, so the owners of data centers are searching the better way to save energy and low down the power consumption. Apart from this there are also many environmental issues that are also generalized to make call for the reduction of energy in large data centers. The carbon footprints and energy costs are increasing rapidly in data centers. All these factors are made on high priority to find out the solutions to low down energy consumption in data centers and this practice has been given to the academia, industries and also the government agencies. There are many large data centers examples like the cloud data centers which offer the various services on certain devices.

To calculate the amount of energy consumed by data centers, two metrics were established by Green Grid, an international consortium. The metrics are Power Usage Effectiveness (PUE) and Data Centre Infrastructure Efficiency (DCiE)as defined below:

$$\text{PUE} = \text{Total Facility} \frac{\text{Power}}{\text{IT}} \text{Equipment Power} \qquad (1)$$

$$\text{DCiE} = \frac{1}{\text{PUE}} = \left(\text{IT Equipment} \frac{\text{Power}}{\text{Total}} \text{Facility Power}\right) \text{x } 100\% \qquad (2)$$

The IT equipment power is the load delivered to all computing hardware resources, while the total facility power includes other energy facilities, specifically, the energy consumed by everything that supports IT equipment load. In cloud infrastructure, a node refers to general multicore server along with its parallel processing units, network topology, and power supply unit and storage capacity.

1.7 Energy consumption of computing servers

The server power consumption depends upon the utilization of CPU. A server which is ideal consumes 2/3 rd of the highest power consumption. This energy consumption by ideal server is because sever must keep the disks, modules of memory, I/O resources and other exterior operations even if not in use or no computations are performed on that. So when the load comes on the network the power consumption is increased [26].

$$P_s(l) = P_{fixed} + \frac{(P_{peak} - P_{fixed})}{2} (1 + l - e^{-\frac{l}{a}}) \qquad (1)$$

In equation 1, the following variables depict the meaning as follows:

P_{fixed} = Idle power consumption

P_{peak} = Power consumption at its peak load

l = Load of server

a = Utilization point at which server reaches linear consumption of power than the load offered.

Processor consumption of power is proportional to the $V^2 f$ in which

V = Voltage

f = Operating frequency

When the frequency downshift takes place the voltage is reduces at that time and infers a cubic relation from f. The equation inferred from the relation is as follows:

$$Pb2_s(l) = P_{fixed} + \frac{(P_{peak} - P_{fixed})}{2} (1 + l^3 - e^{-\frac{l^3}{a}}) \qquad (2)$$

On the basis of equation 2, the power management can adjust the frequency of operation during the ideal routine of server and leads to operational power consumption.

II. LITERATURE SURVEY

Manisha Malhotra et.al, has presented the secure agent based framework for optimal resource scheduling. This paper has integrated the entire previously proposed framework and explored the real world applicability. The work has been evaluated basis on the five main metric that are response time, cost, workload, encryption time and decryption time etc. For future work trust can be considered as the other parameter.

Shikha Gupta et al, aimed at A quick growth is happening in the development of clouds and its management through the cloud computing has increased the speed of the research in this field. After seeing the growth it can say that the future of internet technology is totally based on cloud computing. It provide “as a service” on demand of user. It can be a software, platforms or infrastructure. The cloud owner’s relationship with the consumer highly depends upon how efficiently the consumers are able to use the cloud resources, which in turn depend upon the effective cloud management. Many resources, big data and high demand, may deteriorate the service due to heavy loading of the server. This calls for the balance load on server by distributing the task to the appropriate node in the server. This paper presents a critical review and comparison of the existing techniques for load balancing.

Jayant Baliga et. al, has presented a study of the consumption of energy in cloud computing. Both public and private clouds are taken and contains consumption of energy in transmission of data, switching, data storage and data processing. It has been shown that the energy consumed in switching and transportation of data is much amount of the total energy consumed in cloud computing. The energy consumption is mainly associated with the three services of cloud computing that are storage as service, processing as a service and software as a service. In future merging of these three can be included for analysis of energy consumption.

Geethu Gopinath P P, has reviewed about the load balancing algorithms in cloud computing which as min –min and max-min algorithms. High study is done on how to implement the different load balancing algorithms. The result described that Max-Min is better than the Min-Min algorithm. But in some situation Min-Min has worked better than Max-Min algorithm. When the job of small execution time are more than heavy execution time than the Max-min performs better but if the heavy jobs are more than Min-Min performance is better.

M.Lawanyashri, proposed a multi-objective hybrid fruit fly optimization technique which is based on simulated annealing to increase rate of convergence and accuracy of optimization. This approach is needed for proper utilization of resource, reduction in energy consumption and cost in environment of cloud computing. The simulation results of hybrid FOA approach performs better for convergence rate and datacenter.

ChaimaGhribi, presented algorithms for energy efficient scheduling. The allocation and migration algorithm was combined in this paper to achieve saving of energy. The problem of placement of VM in cloud data centers were explored. The main focus of paper is on accurately identifying the algorithms for placement and consolidation that reduce energy consumption and cost of migration.

Arindam Banerjee et al, presented Cloud computing is an "evolving paradigm" that has redefined the way Information Technology based services can be offered. It has changed the model of storing and managing data for scalable, real time, internet based applications and resources satisfying end users' needs. More and more remote host machines are built for cloud services causing more power dissipation and energy consumption. Over the decades, power consumption has become an important cost factor for computing resources. In this paper we will investigate all possible areas in a typical cloud infrastructure that are responsible for substantial amount of energy consumption and we will address the methodologies by which power utilization can be decreased without compromising Quality of Services (QoS) and overall performance. We also plan to define the scope for further extension of research from the findings we would have from this paper.

Andreas Beri et al, focused on energy efficiency is increasingly important for future information and communication technologies (ICT), because the increased usage of ICT, both with increasing energy costs and the need to reduce green house gas emissions call for energy-efficient technologies that decrease the overall energy consumption of computation, storage and communications. Cloud computing has recently received considerable attention, as a promising approach for delivering ICT services by improving the utilization of data centre resources. In principle, cloud computing can be an inherently energy-efficient technology for ICT provided that its potential for significant energy saving that have so far focused on hardware aspects, can be fully explored with respect to system operation and networking aspects. Thus this paper, in the context of cloud computing, reviews the usage of methods and technologies currently used for energy-efficient operation of computer hardware and network infrastructure. After surveying some of the current best practice and relevant literature in this area, this paper identifies some of the remaining key research challenges that arise when such energy-saving techniques are extended for use in cloud computing environments.

Renuka M. Dhanwate et al, aimed at Mobile devices of today offload their compute intensive application to Cloud, but also consume huge energy while communicating using mobile network services like 3G/4G. Cloudlets can be used to provide such services with wireless LANs. Worldwide Android becomes the fastest-growing mobile OS. Millions of new Android devices are activated worldwide every single day. Even so, it is the fact that Android smart phones have limited resources, such as battery charge capacity, network bandwidth utilization, storage capacity, and processor performance. These restrictions may be relieved by computation offloading: sending heavy computation to resourceful servers and receiving the results from these servers. Several issues related to offloading have been investigated. The energy consumption of under-utilized resources, particularly in a cloud environment, accounts for a substantial amount of the actual energy use. Inherently, a resource allocation strategy that takes into account resource utilization would lead to better energy efficiency; this, in clouds.

III. CONCLUSION

In this paper we can conclude that Lots of research work has been conducted and many techniques were proposed in order to enhance the efficiency and also minimize the consumption of energy for the servers and virtual machines. But there is a gap present to fill in order to reduce the consumption of energy for servers. Another related issue is Quality of Service and power Consumption because after implementing the techniques for load balancing and energy efficiency the parameters of quality gradually decreased. The parameters those comes under these issues of the system are dependent on the parameters as delay, throughput etc. An efficient load balancing algorithm is needed and resource utilization which works on each virtual machine so that the work can be distributed fairly. A procedure of dissimlar task on the basis of priority or order of job execution which can be analysed as load balancing task scheduling. Since the further enhancements are still required for maintaining the quality of the services by providing less delay factor or high throughput, modules that are provided by the cloud computing.

IV. REFERENCES

[1]. Foram F Kherani et al, "Load Balancing in Cloud Computing", IJEDR, Vol 2, Issue 1, Pp 907-912, 2014

[2]. Shikha Gupta et al, "Load Balancing in Cloud Computing: A Review", IJSETR, Vol 4, Issue 6, Pp 1912-1916, 2015

[3]. P.P Geethu Gopinath et al, "An In depth Analysis and study of Load Balancing Techniques in the Cloud Computing Environment", ELSEVIER, Vol 50, Pp 427-432, 2015

[4]. Arindam Banerjee, "Energy Efficiency Model for Cloud Computing", IJEIC, Vol 4, Issue 6, Pp 29-42, 2013

[5]. Andreas Beri et al, "Energy Efficient Cloud Computing", Research Gate, Pp 1-8, 2009

[6]. Renuka M. Dhanwate et al, "A Literature Review on improving Energy Efficiency on Android using Cloud computing based Service", IJARCSMS, Vol 2, issue 12, Pp 209-213, 2014

[7]. Kalange pooja R, "Application of Green Cloud Computing in Energy Efficiency and Environment Sustainability", IOSR-JCE, Pp 25-33

[8]. M. Lawanyashri, BalamuruganBalusamy, and S. Subha. "Energy-aware hybrid fruitfly optimization for load balancing in cloud environments for EHR applications." Informatics in Medicine Unlocked (2017).
[9]. Ibrahim AbakerTargioHashem, IbrarYaqoob, NorBadrulAnuar, SalimahMokhtar, Abdullah Gani, and SameeUllah Khan. "The rise of "big data" on cloud computing: Review and open research issues." Information Systems 47 (2015): 98-115.
[10]. Niroshinie Fernando, Seng W. Loke, and WennyRahayu. "Mobile cloud computing: A survey." Future generation computer systems 29, no. 1 (2013): 84-106.
[11]. XiaofengGao, "Traffic load balancing schemes for devolved controllers in mega data centers." IEEE Transactions on Parallel and Distributed Systems 28.2 (2017): 572-585.
[12]. ChaimaGhribi, MakhloufHadji, and DjamalZeghlache. "Energy efficient vm scheduling for cloud data centers: Exact allocation and migration algorithms." Cluster, Cloud and Grid Computing (CCGrid), 2013 13th IEEE/ACM International Symposium on. IEEE, 2013.
[13]. Namboothiri, Aswathy B., and R. Joshua Samuel Raj. "A comparative study on job scheduling algorithm augmenting load balancing in cloud." In Science Technology Engineering and Management (ICONSTEM), Second International Conference on, pp. 582-588. IEEE, 2016.
[14]. Rafael Moreno-Vozmediano, Rubén S. Montero, and Ignacio M. Llorente. "Key challenges in cloud computing: Enabling the future internet of services." IEEE Internet Computing 17.4 (2013): 18-25.
[15]. KekeGai, "Dynamic energy-aware cloudlet-based mobile cloud computing model for green computing." Journal of Network and Computer Applications 59 (2016): 46-54.
[16]. Shang-Liang Chen, Yun-Yao Chen, and Suang-Hong Kuo. "CLB: A novel load balancing architecture and algorithm for cloud services." Computers & Electrical Engineering 58 (2017): 154-160.
[17]. Ranesh Kumar Naha, and Mohamed Othman. "Cost-aware service brokering and performance sentient load balancing algorithms in the cloud." Journal of Network and Computer Applications 75 (2016): 47-57.
[18]. MehiarDabbagh, "Toward energy-efficient cloud computing: Prediction, consolidation, and overcommitment." IEEE network 29.2 (2015): 56-61.
[19]. KousikDasgupta, BrototiMandal, ParamarthaDutta, Jyotsna Kumar Mandal, and Santanu Dam. "A genetic algorithm (ga) based load balancing strategy for cloud computing." Procedia Technology 10 (2013): 340-347.
[20]. AlirezaSadeghiMilani, and NimaJafariNavimipour. "Load balancing mechanisms and techniques in the cloud environments: Systematic literature review and future trends." Journal of Network and Computer Applications 71 (2016): 86-98.
[21]. DhineshBabuLD, and P. Venkata Krishna. "Honey bee behavior inspired load balancing of tasks in cloud computing environments." Applied Soft Computing 13, no. 5 (2013): 2292-2303.
[22]. Saurabh Kumar Garg, and RajkumarBuyya. "Green cloud computing and environmental sustainability." Harnessing Green IT: Principles and Practices (2012): 315-340.
[23]. JayantBaliga, "Green cloud computing: Balancing energy in processing, storage, and transport." Proceedings of the IEEE 99.1 (2011): 149-167.
[24]. PP GeethuGopinath, and Shriram K. Vasudevan. "An in-depth analysis and study of Load balancing techniques in the cloud computing environment." Procedia Computer Science 50 (2015): 427-432.
[25]. R.Raju, J. Amudhavel, NevedhaKannan, and M. Monisha. "A bio inspired Energy-Aware Multi objective Chiropteran Algorithm (EAMOCA) for hybrid cloud computing environment." In Green Computing Communication and Electrical Engineering (ICGCCEE), 2014 International Conference on, pp. 1-5. IEEE, 2014.
[26]. OpeyemiOsanaiye, Shuo Chen, Zheng Yan, Rongxing Lu, Kim Choo, and MqheleDlodlo. "From cloud to fog computing: A review and a conceptual live VM migration framework." IEEE Access (2017).

AUTHORS

Ms. Shivani Bajaj is pursuing her Ph.D from Chandigarh University in the area of cloud computing. She has the degree of MCA and BCA. She has attended the national conference. She has worked on a major project Akhar 2016. Her task was on the making of transliteration system which deals with Gurmukhi to Roman Transliteration System.

Dr. Manisha Malhotra is presently working as Associate Professor in Chandigarh University Gharuan. She has credible academic record with degrees like Ph.D (computer Applications), MCA & B.Sc (Computer Science). She is the member of many professional bodies like IEEE, IEANG. She has published more than 20 research papers in National and International conferences & in many reputed journals. She has attended many national & international conferences. She is awardee of young faculty in the field of cloud computing. Her area of research is Cloud computing and Agent technology.