

A Review of Selected Techniques in Job Scheduling in Cloud Computing

Sreenu Banoth

Department of Computer Science Engineering, Ganapathy Engineering College, Warangal

Abstract: *Cloud computing is a general term used to describe a new class of network based computing that takes place over the internet. Cloud computing is the new paradigm for delivering on demand services over internet and can be described as internet amiable software. Job scheduling is one of the major activities performed in all the computing environments. Cloud computing is one the upcoming latest technology. To increase the working of cloud computing environments, job scheduling is one the tasks performed in order to gain maximum profit. In this paper, we introduce multiple scheduling research works to address this area. This review work provides acumen sight on the introduction of cloud computing to understand it and then focus on the various strategies for job scheduling in cloud computing. This paper also provides the base to the new researchers in cloud computing. This paper also takes a few techniques. Our priority bases upon some factors including same number of jobs for execution at each node. This paper provides the survey on scheduling algorithms. There working with respect to the resource sharing. We systemize the scheduling problem in cloud computing, and present a cloud scheduling hierarchy.*

Keywords: Cloud computing, Cloud Architecture Load balancing, queuing system, scheduling, Workflows.

1. Introduction

Cloud computing is rising as the next generation platform for computation. May be in future Cloud computing will be the main platform to save the world; this makes people can have everything they need on it. Main advantages of the Cloud computing is used for on-demand gathering of information, technology services and products. The name Cloud has come from the Internet, based on how it is depicted in computer network diagrams, and is an abstraction for the complex infrastructure it conceals.

In general we can define cloud computing is style of computing in which IT-related capabilities are provided “as a service”, allowing users to access technology -enabled services from the Internet without knowledge of, expertise with, or control over the technology infrastructure that supports them.

Email application was probably the first service on the “cloud”. As now a day many computing industry shifting toward providing Platform as a Service (PaaS), Infrastructure as a Service (IaaS) and Software as a Service (SaaS) for consumers and enterprises to access on demand regardless of time and location, this helps in avoiding the over- supplying of the resource when used with utility pricing, which meets the demands of millions of users. There will be an increase in the number of Cloud platforms available. To get cloud computing to work, you need three things: thin clients (or clients with a thick-thin switch), grid computing, and utility computing. Grid computing links disparate computers to form one large infrastructure, harnessing unused resources. Utility computing is paying for what you use on shared servers like you pay for a public utility (such as electricity, gas, and so on) [1].

But it seems that Cloud computing cannot be suitable for all the kind of companies, because there may be need that company has very secured data they may not proceed with cloud computing mainly for the reasons, stored data might

not be secure. With cloud computing, all your data is stored on the cloud. How secure is the cloud? Can unauthorized users gain access to your confidential data? Cloud computing companies say that data is secure, but it's too early in the game to be completely sure of that. One more negative of cloud computing it, it needs the constant high speed internet to access your own data, dead Internet connection means no work, when you are offline cloud computing simply does not work. Cloud computing is a very specific type of computing that has very specific benefits. But it has specific negatives as well. One thing that is hoped of what the cloud does extremely well and a brief overview of them, but also a short survey on their criteria and challenges ahead of them.

Cloud Computing,” to put it simply, means “Internet Computing.” The Internet is commonly visualized as clouds; hence the term “cloud computing” for computation done through the Internet. With Cloud Computing users can access database resources via the Internet from anywhere, for as long as they need, without worrying about any maintenance or management of actual resources. Besides, databases in cloud are very dynamic and scalable. Cloud computing is unlike grid computing, utility computing, or autonomic computing. In fact, it is a very independent platform in terms of computing. The best example of cloud computing is Google Apps where any application can be accessed using a browser and it can be deployed on thousands of computer through the Internet. [2]

Cloud computing has captured popular imagination with the immense possibilities this computing paradigm seems to offer. The term cloud computing itself is variously defined — it encompasses a broad range of IT capabilities that are accessible irrespective of the user’s geographic location.

These capabilities are offered as a high-availability service by a third party on a pay-as-you-go basis, and the end user interacts with them over the Internet cloud. [3]

You can distinguish among the cloud computing services different vendors offer on the basis of the level of abstractions they provide: hardware-as-a-service or infrastructure cloud (e.g., Amazon Elastic Compute Cloud), platform (e.g., Google App Engine), and applications (e.g., Web-based e-mail services) [2].

2. What is Cloud Computing?

Cloud computing provides the facility to access shared resources and common infrastructure, offering services on demand over the network to perform operations that meet changing business needs. The location of physical resources and devices being accessed are typically not known to the end user. It also provides facilities for users to develop, deploy and manage their applications „on the cloud“, which entails virtualization of resources that maintains and manages itself.



Figure 1: Cloud Computing

Some Generic examples include:

- 1) Amazon's Elastic Computing Cloud (EC2) offering computational services that enable people to use CPU cycles without buying more computers.
- 2) Storage services such as those provided by Amazon's Simple Storage Service (S3)
- 3) Companies like Nirvanix allowing organizations to store data and documents without adding a single on-site server.
- 4) SaaS companies like Salesforce.com delivering CRM services, so clients can manage customer information without installing specialized software.

2.1 Why Cloud Computing

The main advantages of cloud computing over the parallel, grid and utility computing are Cloud Computing is a form Distributed computing vs. Parallel Computing:

Distributed Operating systems are also referred to as Loosely Coupled systems whereas parallel processing systems are referred to as tightly coupled systems.

A Loosely coupled system is one in which the processors do not share memory and each processor has its own local

memory whereas in a tightly coupled system there is a single system wide primary memory shared by all the processors. The processors of distributed operating systems can be placed far away from each other to cover a wider geographic area which is not the case with parallel processing systems.

The no. of processors that can be usefully deployed is very small in a parallel processing operating system whereas for a distributed operating system a larger no. of processors can be usefully deployed.

Global clock is used for controlling SIMD and MIMD in parallel computing. But in distributed computing no any global clock present in this synchronization algorithms are used.

In the distributed operating system there is an unpredictable communication delays between processors whereas the processors in the parallel processing system share over an interconnection network.

To get cloud computing to work, you need three things: thin clients (or clients with a thick-thin switch), grid computing, and utility computing. Cloud computing has the facility of on-demand resource providing or supplying which is not possible in utility computing, which helps in avoiding the over-supplying the resource when used with utility pricing, which meets the demands of millions of users. As the Grid computing act as the back end of the cloud computing.

Compute-intensive Application: Compute-intensive is a term that applies to any computer application that demands a lot of computation, such as meteorology programs and other scientific applications. A similar but distinct term, computer intensive, refers to applications that require a lot of computers, such as Grid computing. The two types of applications are not necessarily mutually exclusive: some applications are both compute- and computer-intensive.

2.2. Cloud –Types

Public Cloud

Public cloud or external cloud describes cloud computing in the traditional mainstream. Public clouds are run by third parties, and applications from different customers are likely to be mixed together on the cloud's servers, storage systems, and networks. A public cloud provides services to multiple customers.

Hybrid Cloud

Hybrid clouds combine both public and private cloud models. This is most often seen with the use of storage clouds to support Web 2.0 applications.

Private Cloud

Private clouds are built for the exclusive use of one client, providing the utmost control over data, security, and quality of service. The company owns the infrastructure and has control over how applications are deployed on it. Private clouds can be built and managed by a company's own IT organization or by a cloud provider.

3. The Architecture of Cloud

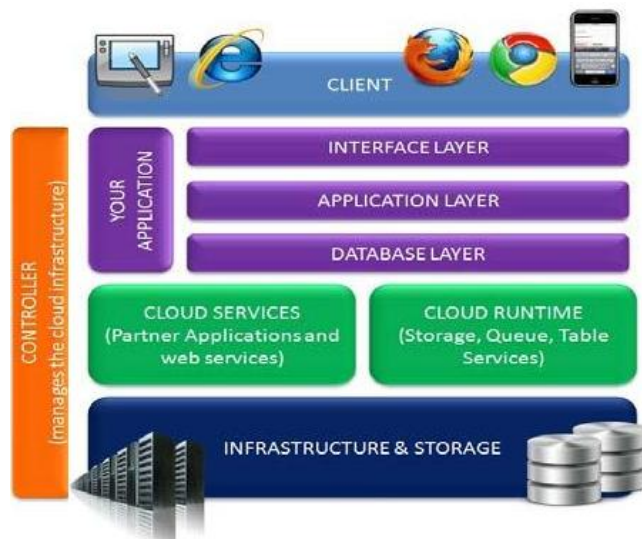


Figure 2: Layered View of Cloud Architecture

Cloud computing architecture has two most significant components that are basically known as front and back end. Front end is that part of cloud which is visible to the user of cloud. This include applications and computer that user uses to access the cloud. Back end of cloud computing is nothing but the cloud itself, which mainly comprises of computers, storage devices.

Cloud Environment basically consists of software applications that are accessed via internet as services when wishes to use them. Applications that are built on Cloud Architectures are such that the underlying infrastructure of computing is used only when it is actually needed, draw the necessary resources when someone make a demand for those, perform a particular job, then relinquish the unneeded resources and often dispose them after the job is completed. During their operation the applications scale up or down elastically based on need of resources.

Cloud Architectures address key difficulties that mainly related to processing of large amount of data. In traditional way of processing the data it is quite difficult to get as many machines that an application needs for its operation to complete. It is one of the most difficult things to get the machines when one/particular needs them. It is really a tough work to distribute and co-ordinate a large-scale job on different-different machines, run processes on them, and provisions another machine to recover if one machine fails during the operation. It is also difficult to auto-scale up and down based on changing nature of workloads. It is difficult to get rid of all those machines when the job is completed.

Applications that built over Cloud Architectures run in-the-cloud where the actual physical location of the infrastructure is determined by the cloud providers. They simply take advantage of simple APIs of Internet-accessible services that are able to scale on-demand, where the complex logic of scalability and reliability of the services remains implemented and hidden inside the vast cloud.

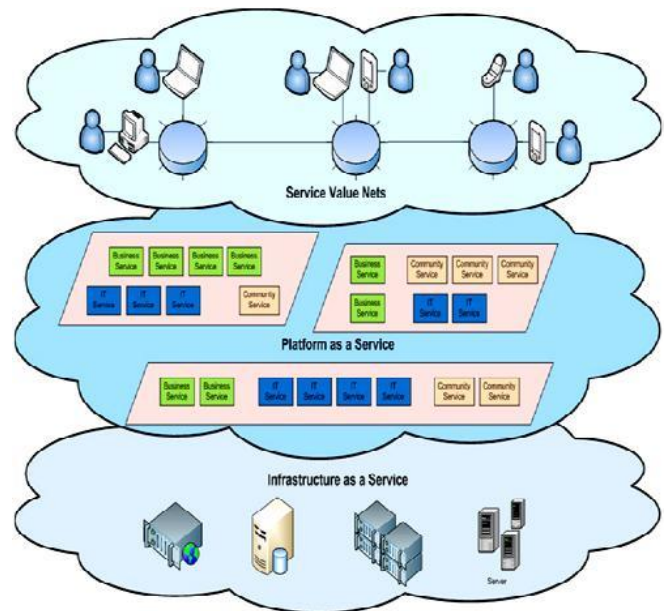


Figure 3: Service Layers of Cloud

IaaS is the lower most layers of the cloud computing systems and it basically provides virtualized resources ex: computation, communication and storage that are available on demand. PaaS makes the cloud easily programmable. SaaS can be term as the software delivery model.

4. Workflow Scheduling

Workflow Scheduling compacts with the credentials of suitable resources and allocation of the tasks on those resources. In workflow scheduling, bigger task are divided into different-different small co-tasks these co-task or sub-tasks are allocated to resources in such a way that achieve some pre-define objective. There is various problems in bioinformatics, astronomy and business enterprise in which set of sub-tasks are executed in particular sequence and ordered to carry out the bigger task[6]. In general, a workflow application requires series of steps to be executed in a particular fashion. These steps have the parent-child relationship [7]. The parent task is associated to child-task according to some rules and must be executed before its child-task.

A workflow is represented by Directed Acyclic Graph (DAG). The number of tasks is denoted by vertex (V) and data dependencies among tasks are denoted by an edge (E).

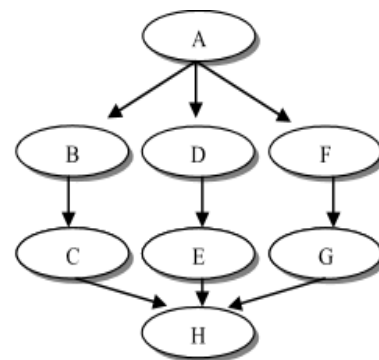


Figure 4: Workflow represented in the form of a graph

4.1 Nature of Workflow Algorithm

The workflow scheduling algorithms can be two type:

4.1.1 Heuristic in Nature

The heuristic algorithms are priority based and mainly problem centric. The maker can use his individual knowledge to allocate priority for workflow applications and cloud resources.

4.1.2 Meta Heuristic in Nature

Meta-heuristic scheduling algorithms are the ones that do not need the human interface and provide the over-all solution to workflow applications.

4.2 Workflow Scheduling Parameters

Workflow scheduling parameters which may consider in different algorithms as per developer requirements.

- **Execution Time (ET):** It refers to the time taken by CPU to execute the task or job.
- **Response Time(RT):** The response time is the time at which system start responding for the submitted task.

$$RT = \text{Service Time} + \text{Wait Time}$$

- **Makespan (M-s):** Time difference between the submission of request and completion time of the last task of workflow is known as makespan time.

$$M-s = \text{Completion Time} - \text{Start}$$

- **Energy Consumption:** Energy Consumption is referred as consumption of power or energy by resources during the service.
- **Throughput:** The rate of production or how much task are completed in the particular unit of time is known as the throughput of the system.
- **Scalability:** The ability of system grow itself according to increasing the demand or increasing the data.
- **Resource Utilization (RU):** Keep the resources as busy as possible. It covers the ranges from 0 to 100% but in practice 100% is not possible.
- **Load Balancing:** It is a process that keep the resources or servers in balance by pooling or pushing the load from one resource to another.
- **Fault Tolerance:** To give the surety of availability and reliability of critical services is known as fault tolerance.
- **Reliability:** Reliability is define as the trust value of the resources, services and Cloud service provider (CSP). It defines user must be received continue service without failures and quality must be meet the SLA agreement.

5. Workflow Scheduling Algorithms

In the current scenario, many scientist/ researcher are working for improving the proper utilization of cloud resources. This problem is similar with some other environment like the grid or distributed. In the cloud computing environment, we can improve QoS with the help of workflow scheduling algorithms. We studied various

algorithm and some of them will be discussed here.

5.1 Cost-Based Scheduling Of Scientific Workflow Applications On Utility Grids [8] is based on Nimrog-G [9] and A Market-Based Workflow Management System [10] which schedules independent tasks and assign deadline of individual task respectively. In this algorithm author solved the task independency problem by considering the tasks with certain dependencies. Through this, author minimize the execution cost while meeting the deadline constrain and solve optimally the scheduling problem for sequential tasks by modeling the branch partition as a Markov Decision Process(MDP)[3] that cut down the execution cost and improve the QoS in Utility Grids environment.

5.2 Scheduling Service Workflows for Cost Optimization in Hybrid Clouds [14]. This algorithm work in Hybrid cloud scenario. When cloud user demands new resources from the public cloud to fulfill their need at that time public cloud resources should be schedule in a proper way. The author using PCH [4] strategy that separate the public cloud resources and private cloud resources are requested by cloud user in order to minimize the cost and meet the deadline. After the comparison of proposed work give the better result in compare to the individual only private cloud and only public cloud environment.

5.3 Scheduling Scientific Workflows Elastically For Cloud Computing [15] in this paper Cui Lin et al,

optimizing the execution time as well as resources to scale elastically at run time. This was done by formulize the computing environment and scientific workflow.

5.4 Scheduling Technique of Data-Intensive Application Workflows In Cloud Computing [16] The main goal of this paper is the selection of suitable hosts for accessing resources and creating a virtual machine to execute applications for more effective execution. The selection of suitable data centers that have minimum transfer time to access replica among other data centers and create VM for this data center. These procedures undersized the access cost that minimize the broadcast time of input and output data and make more efficient execution.

5.5 A Truthful Dynamic Workflow Scheduling Mechanism for Commercial Multicloud Environments [19]. It is Multicloud environment workflow scheduling algorithm which has objected to minimize the completion time and monetary cost while meeting the deadline for delivering the result. Algorithm based on game theory and mathematical approach to analyze the decision of agents in problem for trust accuracy of the cloud resource [20, 21].

5.6 Energy Efficient Workflow Job Scheduling for Green Cloud [22] was presenting a realistic model that consume less energy and minimize the CO2 emission. Each task of scientific workflow run without affecting the performance based on DVFS strategy. Resource utilization is maximized in this algorithm, through VM allocation strategy that enhance the system performance. When this algorithm was simulated then its save the energy

consumption up to 30% and maximize the resource utilization 25% and reduce emission of CO₂.

5.7 Cluster based Scheduling of Workflow Applications in Cloud [24] in this paper author propose a new type of workflow algorithm that is based on the cluster of the task. In this algorithm reducing the overall makespan is the main objective of the author. Cluster are calculated by obtaining the execution time, based on it author put the task in some cluster

6. Literature Review

The following section provides the review of already done literature work on scheduling in cloud computing. The review work presented here provides base to the new researchers in cloud computing scheduling.

A). Praveen K. Gupta, Nitin Rakesh [5] presents, Cloud computing has come out to be an interesting and beneficial way of changing the whole computing world. In this paper, we deal with the various methodologies adopted to handle all the processes and jobs concurrently executing and waiting into the web application and web server housed into the same system or different systems. Also, these different methods will be compared taking into account the same number of jobs, but varied environmental conditions and hence, the result would be formulated. Various issues like virtual resources, queuing strategies, resource managers etc. has been discussed here apart from the main coverage points. All these aspects will be closely studied, observed and proved with proper explanations.

B). Meng Xu, Lizhen Cui, Haiyang Wang, Yanbing Bi [1] presents that cloud computing has gained popularity in recent times. As a cloud must provide services to many users at the same time and different users have different QoS requirements, the scheduling strategy should be developed for multiple workflows with different QoS requirements. In this paper, we introduce a Multiple QoS Constrained Scheduling Strategy of Multi-Workflows (MQMW) to address this problem. The strategy can schedule multiple workflows which are started at any time and the QoS requirements are taken into account. Experimentation shows that our strategy is able to increase the scheduling success rate significantly.

C). Luqun Li [2] presents that job scheduling system problem which is a core and challenging issue in Cloud Computing. How to use Cloud computing resources efficiently and gain the maximum profits with job scheduling system is one of the Cloud computing service providers' ultimate goals. In this paper, firstly, by analysis the differentiated QoS requirements of Cloud computing resources users' jobs, build the corresponding non-preemptive priority M/G/1 queuing model for the jobs.

D). HUANG Qiyi, HUANG Ting-lei [3] discusses cloud computing refers to constructed data center or "super computer" by virtualization technology and provide computing and storage resources, as well as the application container environment of software running, to software developers in a manner of free or hiring. The Cloud

Computing is promoted by the business rather than academic which determines its focus on user applications. Different users have different QoS Requirements. So according to the given deadline and budget, this article conducts research on scheduling model from the user's perspective.

E). Yujia Ge, Guiyi Wei [4] discusses task scheduling problems that of paramount importance which relate to the efficiency of the whole cloud computing facilities. In Hadoop, the open-source implementation of Map Reduce, scheduling policies, such as FIFO or delay scheduling in FAIR scheduler is used by the master node to distribute waiting tasks to computing nodes (slaves) in response to the status messages of these nodes it receives.

F). S.Selvarani¹, Dr. G.Sudha Sadhasivam [6] concludes cloud computing has been build upon the development of distributed computing, grid computing and virtualization. Since cost of each task in cloud resources is different with one another, scheduling of user tasks in cloud is not the same as in traditional scheduling methods. The objective of this paper is to schedule task groups in cloud computing platform, where resources have different resource costs and computation performance.

G). Jorge Ejarque et al [7] presents about the license management which is one of the main concerns when Independent Software Vendors (ISV) try to distribute their software in computing platforms such as Clouds. They want to be sure that customers use their software according to their license terms. The work presented in this paper tries to solve part of this problem extending a semantic resource allocation approach for supporting the scheduling of job taking into account software licenses.

H). Baomin Xu et al [8] provides and considers the commercialization and the virtualization characteristics of cloud computing, the Paper proposed for the first time an algorithm of job scheduling based on Berger model. In the job scheduling process, the algorithm establishes dual fairness constraint. The first constraint is to classify user tasks by QoS preferences, and establish the general expectation function in accordance with the classification of tasks to restrain the fairness of the resources in selection process. The second constraint is to define resource fairness justice function to judge the fairness of the resources allocation. We have expanded simulation platform CloudSim, and have implemented the job scheduling algorithm proposed in this paper. The experimental results show that the algorithm can effectively execute the user tasks and manifests better fairness.

I). Ramendu Bikash Guin et al [9] presents and discusses cloud computing focuses about delivery of reliable, fault tolerant and scalable infrastructure for hosting Internet based application services. Our work presents the implementation of an efficient Quality of Service (QoS) based smart-scheduler along with Backfill strategy based light weight Virtual Machine Scheduler for dispatching jobs. The user centric smart-scheduler deals with selection of proper resources to execute high level jobs. The system centric Virtual Machine (VM) scheduler optimally dispatches the

jobs to processors for better resource utilization.

J). Srinivasa Rao V, Nageswara Rao N K , E Kusuma Kumari [10] gives overview about the basic concept, defines the terms used in the industry, and outlines the general architecture and applications of Cloud computing. It gives a summary of Cloud Computing and provides a good foundation for understanding.

K). Norman W. Paton et al [11] presents and discusses cloud computing services to potentially numerous remote users with diverse requirements. Although predictable performance can be obtained through the provision of carefully delimited services, it is straightforward to identify applications in which a cloud might usefully host services that support the composition of more primitive analysis services or the evaluation of complex data analysis requests. In such settings, a service provider must manage complex and unpredictable workloads. This paper describes how utility functions can be used to make explicit the desirability of different workload evaluation strategies, and how optimization can be used to select between such alternatives. The approach is illustrated for workloads consisting of workflows or queries.

7. Conclusion

We presented a review of detailed working of different job scheduling algorithms behind the web application on a web server on the same system or on different systems in a virtual cloud computing environment [1]. The various terminologies and concepts were thoroughly looked and thus explained.

This work will give way to more future findings regarding the scheduling techniques in a cloud environment. More efficient and faster ways to schedule jobs and increase CPU throughput [2] needs to be discovered. Also, this will fuel the greater knowledge and popularity of cloud environment among the peoples.

References

- [1] "A Multiple QoS Constrained Scheduling Strategy of Multiple Workflows for Cloud Computing" Meng Xu, Lizhen Cui, Haiyang Wang, Yanbing Bi School of Computer Science and Technology Shandong University Jinan, China. 2009
- [2] "An Optimistic Differentiated Service Job Scheduling System for Cloud Computing Service Users and Providers" Luqun Li Department of Computer Science and Technology Shanghai Normal University 100 #Guilin Road, Shanghai, China . 2009
- [3] "An Optimistic Job Scheduling Strategy based on QoS for Cloud Computing" HUANG Qiyi, HUANG Ting-lei School of Computer Science and Engineering Guilin University of Electronic Technology Guilin, China. 2010
- [4] "GA-Based Task Scheduler for the Cloud Computing Systems" Yujia Ge, Guiyi Wei, School of Computer and Information Engineering Zhejiang Gongshang University Hangzhou, china. 2010
- [5] "Different Job Scheduling Methodologies for Web

Application and Web Server in a Cloud Computing Environment" Praveen K. Gupta, Nitin Rakesh, Member, 2010

- [6] "IMPROVED COST-BASED ALGORITHM FOR TASK SCHEDULING IN CLOUD COMPUTING" Mrs.S.Selvarani1, Dr.G.Sudha Sadhasivam Department of Information Technology, Tamilnadu College of Engineering, Coimbatore, India. 2010
- [7] "Job Scheduling with License Reservation: A Semantic Approach" Jorge Ejarque, Andras Micsik, Raül Sirvent, Peter Pallinger , Laszlo Kovacs and Rosa M. Badia Grid Computing and Clusters Group - Barcelona Supercomputing Center (BSC), Barcelona, Spain. 2011
- [8] "Job scheduling algorithm based on Berger model in cloud environment" Baomin Xu, Chunyan Zhao , Enzhao Hua, Bin Hu. 2011
- [9] "ScheduleSuite Modelling & Simulation of a Smarter Job Scheduling System for Cloud Computing Service Providers and Users" Ramendu Bikash Guin, Sayan Chakrabarti, Chinmoy Tarafdar Computer Science & Engineering Department Kalyani Government Engineering College Kalyani, Nadia- 741235, West Bengal (INDIA). IEMCON 2011
- [10] Cloud Computing: An Overview" Srinivasa Rao V, Nageswara Rao N K, E Kusuma Kumari.2011
- [11] "Optimizing Utility in Cloud Computing through Autonomic Workload Execution" Norman W. Paton, Marcelo A. T. de Aragão, Kevin Lee, Alvaro A. A. Fernandes, Rizos Sakellariou School of Computer Science, University of Manchester, U.K. 2009

Author Profile



Mr. Sreenu Banoth received the M.Tech degree in Medical imaging and Image analysis from IIT KHARAGPUR University, India in 2009 and had 7 years of teaching and one year industry experience. Presently working as Asst.Professor, Department of CSE, Ganapathy Engineering College, Warangal. Research Interest include Cloud computing, Data Mining and Image Processing.

