Review of Various Scheduling Techniques in Cloud Computing

Harpreet Kaur & Maninder Singh
Department of I.T, L.P.U., Department of C.S.E., L.P.U.
Jalandhar, India

Abstract
Cloud computing technology virtualizes and offers many services across the network. It mainly aims at scalability, availability, throughput, and resource utilization. Cloud model promotes availability and is composed of five essential characteristics, three service models, and four deployment models. We proposed different research papers on cloud computing and scheduling.

Introduction
This chapter explains cloud computing, Service Models, its essential characteristics, Deployment Models and Advantages of Cloud Computing. It also discusses the issues that need to be considered while developing a cloud.

1. Background
Distributed computing is a field of computer science that studies distributed systems. A distributed system consists of multiple autonomous computers that communicate through a computer network. The computers interact with each other in order to achieve a common goal.

Types of Distributed Computing Systems
1. Cluster Computing Systems
2. Grid Computing Systems
3. Peer 2 Peer Computing Systems
4. Cloud Computing Systems

1. Cluster Computing Systems: Cluster computing is not a new area of computing. It is, however, evident that there is a growing interest in its usage in all areas where applications have traditionally used parallel or distributed computing platforms. The growing interest has been fuelled in part by the availability of powerful microprocessors and high-speed networks as off-the-shelf commodity components as well as in part by the rapidly maturing software components available to support high performance and high availability applications [12].

2. Grid Computing Systems: Grid computing can mean different things to different individuals. The grand vision is often presented as an analogy to power grids where users (or electrical appliances) get access to electricity through wall sockets with no care or consideration for where or how the electricity is actually generated. In this view of grid computing, computing becomes pervasive and individual users (or client applications) gain access to computing resources (processors, storage, data, applications, and so on) as needed with little or no knowledge of where those resources are located or what the underlying technologies, hardware, operating system, and so on are. Though this vision of grid computing can capture one’s imagination and may indeed someday become a reality, there are many technical, business, political, and social issues that need to be addressed. If we consider this vision as an ultimate goal, there are many smaller steps that need to be taken to achieve it. These smaller steps each have benefits of their own. Therefore, grid computing can be seen as a journey along a path of integrating various technologies and solutions that move us closer to the final goal. Its key values are in the underlying distributed computing infrastructure technologies that are evolving in support of cross-organizational application and resource sharing—in a word, virtualization—virtualization across technologies, platforms, and organizations [13].

3. Peer 2 Peer Computing Systems: The term “peer-to-peer” refers to a class of systems and applications that employ distributed resources to perform a function in a decentralized manner. The resources encompass computing power, data (storage and content), network bandwidth, and presence (computers, human, and other resources). The critical function can be distributed computing, data/content sharing, communication and collaboration, or platform services [14].

4. Cloud Computing Systems: Cloud computing can be defined as a new style of computing in which dynamically scalable and often virtualized resources are provided as a services over the Internet. Cloud computing has become a significant technology trend, and many experts expect that cloud computing will reshape information technology (IT) processes and the IT marketplace. With the cloud computing technology, users use a variety of devices, including PCs, laptops, smartphones, and PDAs to access programs, storage, and application-development platforms over the Internet, via services offered by cloud computing providers. Advantages of the cloud computing technology include cost savings, high availability, and easy scalability. Google, Yahoo, Amazon, and others have built large, purpose-built architectures to support their applications and taught the rest of the world how to do massively scalable architectures to support compute, storage, and application services. Cloud computing is about moving services, computation, and/or data—for cost and business advantage—off-site to an
Cloud computing exhibits various key characteristics like cost is claimed to be reduced and in a public cloud delivery model capital expenditure is converted to operational expenditure. This is purported to lower barriers to entry, as infrastructure is typically provided by a third-party and does not need to be purchased for one-time or infrequent intensive computing tasks. Pricing on a utility computing basis is fine-grained with usage-based options and fewer IT skills are required for implementation (in-house). Device and location independence enable users to access systems using a web browser regardless of their location or what device they are using (e.g., PC, mobile phone). As infrastructure is off-site (typically provided by a third-party) and accessed via the Internet, users can connect from anywhere. Virtualization technology allows servers and storage devices to be shared and utilization be increased. Applications can be easily migrated from one physical server to another. Reliability is improved if multiple redundant sites are used, which makes well-designed cloud computing suitable for business continuity and disaster recovery. Scalability and elasticity via dynamic (“on-demand”) provisioning of resources on a fine-grained, self-service basis near real-time, without users having to engineer for peak loads. Performance is monitored, and consistent and loosely coupled architectures are constructed using web services as the system interface. Maintenance of cloud computing applications is easier, because they do not need to be installed on each user's computer and can be accessed from different places [19].

### Cloud computing possess the following key characteristics [1] :

1. **On-demand self-service**: A consumer can unilaterally provision computing capabilities, such as server time and network storage, as needed automatically without requiring human interaction with each service’s provider.

2. **Broad network access**: Cloud computing provide the users with various capabilities over the network which are accessed through standard mechanisms that promote use by heterogeneous thin or thick client platforms (e.g., mobile phones, laptops etc.)

3. **Resource pooling**: The provider’s computing resources are pooled to serve multiple consumers using a multi-tenant model, with different physical and virtual resources dynamically assigned and reassigned according to consumer demand. Examples of resources include storage, processing, memory, network bandwidth, and virtual machines.

4. **Rapid elasticity**: Capabilities can be rapidly and elastically provisioned, in some cases automatically, to quickly scale out, and rapidly released to quickly scale in. To the consumer, the capabilities available for provisioning often appear to be unlimited and can be purchased in any quantity at any time.
5. **Measured Service** : Cloud systems automatically control and optimize resource use by leveraging a metering capability at some level of abstraction appropriate to the type of service (e.g., storage, processing, bandwidth, and active user accounts). Resource usage can be monitored, controlled, and reported, providing transparency for both the provider and consumer of the utilized service.

1.4 **Types of Cloud** : There are four types of clouds [1]
   - 1.4.1 Private Cloud
   - 1.4.2 Community Cloud
   - 1.4.3 Public Cloud
   - 1.4.4 Hybrid Cloud

1.4.1 **Private cloud** : The cloud infrastructure is operated solely for an organization. It may be managed by the organization or a third party and may exist on premise or off premise.

1.4.2 **Community cloud** : The cloud infrastructure is shared by several organizations and supports a specific community that has shared concerns (e.g., mission, security requirements, policy, and compliance considerations). It may be managed by the organizations or a third party and may exist on premise or off premise.

1.4.3 **Public cloud** : The cloud infrastructure is made available to the general public or a large industry group and is owned by an organization selling cloud services.

1.4.4 **Hybrid cloud** : The cloud infrastructure is a composition of two or more clouds (private, community or public) that remain unique entities but are bound together by standardized or proprietary technology that enables data and application portability.

1.4 **Advantages of Cloud Computing** : Cloud computing is now evolving like never before, with companies of all shapes and sizes adapting to this new technology. Industry experts believe that this trend will only continue to grow and develop even further in the coming few years. While cloud computing is undoubtedly beneficial for mid-size to large companies, it is not without its downsides, especially for smaller businesses [16]. According to [17] following are the various advantages of cloud computing.

1. **Cost Effective** : There is no need for users to invest their time and money into using standalone servers which would be a bit complicating to use compared to the cloud method. It is a cheaper way to maintain the software and it will save time, as the developers keep track of updates and maintain your programs while you use it. There is no need for replacing capital expenditures on a regular basis. The cost of using cloud resources is very economical for resources such as centralized, real estate, bandwidth, and power. Users will also save money on software updates, management costs, and data storage costs.

2. **Speed & Scales** : There is no need to purchase and setup hardware manually when using the cloud computing method. Depending upon their needs the user can quickly scale up or scale down.

3. **Innovation** : Users can now pay closer attention to the innovation process because they don’t have to manually manage other resources. Cloud computing produces a faster development pace for prototype and testing phases. Projects at which users have to watch over for progress on a regular basis will benefit the most because of this advantage.

4. **Convenient** : Since overheads are low when sharing the same infrastructure the services are available to use immediately. Payments are only billed for the times that the service is being utilized. You can easily check the cost of the bill because the service provider will make them available online for you to view.

5. **Location** : Areas that have lower overheads are able to utilize this service and take advantage of the benefits as well. Many different websites are able to be set up in the case of a disaster recovery which helps the companies to cut costs in different ways.

6. **Multiple Users at one time** : Cloud computing is not only cost effective, but utilizing it also helps to cut back on global wastes. It is environmentally friendly since it is shared by multiple users. The down time is cut in half and the resources are stretched.

7. **Flexible** : There is a high rate of flexibility when using cloud computing because people can opt out of using it whenever they want too. This is also one of the main reasons people love to use this method. Service level agreements are what cover the costs in this case. If the correct quality is not provided then has to pay a penalty cost.
8. **Device Diversity** : The cloud computing method can be accessed through various different electronic devices that are able to have access to the internet. These devices would include and iPad, smartphone, Laptop, or desktop computer.

9. **Lots of Storage Space** : When you use the internet with the cloud services then your company will have lots more room to store the files and data that they need to store.

1.5 **Issues in Cloud** : One of the critical issues in implementing cloud computing is taking virtual machines, which contain critical applications and sensitive data to public and shared cloud environment. According to the following are certain issues in cloud computing [19].

1. **Performance** : The major issue in performance can be for some intensive transaction-oriented and other data-intensive applications, in which cloud computing may lack adequate performance. Also, users who are at a long distance from cloud providers may experience high latency and delays.

2. **Security and Privacy** : Customers are worried about the vulnerability to attacks, when information and critical IT resources are outside the firewall.

3. **Control** : Some IT departments are concerned because cloud computing providers have a full control of the platforms. Cloud computing providers typically do not design platforms for specific companies and their business practices.

4. **Bandwidth Costs** : With cloud computing, companies can save money on hardware and software; however they could incur higher network bandwidth charges. Bandwidth cost may be low for smaller Internet-based applications, which are not data intensive, but could significantly grow for data-intensive applications.

5. **Reliability** : Cloud computing still does not always offer round-the-clock reliability. There were cases where cloud computing services suffered few-hours outages.

6. **Security Policy** : It is very difficult to choose whether the user would have same security policy control over their applications and services or the cloud provider will provide its own policies. If so, then the issue of trusting third party vendor arises.

1.6 **Scheduling** : Scheduling is the method by which threads, processes or data flows are given access to system resources (e.g. processor time, communications bandwidth). This is usually done to load balance a system effectively or achieve a target quality of service. The need for a scheduling algorithm arises from the requirement for most modern systems to perform multitasking (execute more than one process at a time) and multiplexing (transmit multiple flows simultaneously). The scheduler is concerned mainly with [18].

1. **Throughput** : The total number of processes that complete their execution per time unit.

2. **Latency, specifically** :
   - **Turnaround time** - total time between submission of a process and its completion.
   - **Response time** - amount of time it takes from when a request was submitted until the first response is produced.

3. **Fairness / Waiting Time** : Equal CPU time to each process (or more generally appropriate times according to each process’ priority). It is the time for which the process remains in the ready queue.

**Review of Literature**

Peter Mell and Timothy Grance (2011) states that Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction. This cloud model promotes availability and is composed of five essential characteristics, three service models, and four deployment models.

AbiramiS.P. ,ShaliniRamanathan (2012) states that Cloud computing technology virtualizes and offers many services across the network. Emerging techniques focus on scalability and availability. According to authors cloud computing must be advanced to focus on resource utilization and resource management. The cloud environment, embedded with the nimbus and cumulus services will contribute more in making the responsibility of resource utilization in Cloud Computing. Considering the processing time, resource utilization based on CPU usage, memory usage and throughput, the cloud environment with the service node to control all clients request, could provide maximum service to all clients. Scheduling the resource and tasks separately involves more waiting time and response time. The cloud infrastructure is formed with the server based services including Nimbus and Cumulus services which are imported to a server node to establish the IaaS cloud environment and virtualization and an effective scheduling algorithm named Linear Scheduling for Tasks and Resources.
The scheduling algorithm is designed which schedules both the task and the resources. The algorithm mainly focuses on eradicating the starvation and deadlock conditions. The virtualization technique along with the scheduling algorithm will yield higher resource utilization, system throughput, thus improving the performance of the cloud resources. The result analysis of the implementation specifies the amount of resource utilized in the cloud environment when requests are satisfied by Linear Scheduling for Tasks and Resources scheduling technique.

Senthil Kumar, S.K. and P. Balasubramanie (2012) proposed a new algorithm using modified linear programming problem transportation based task scheduling and resource allocation for decentralized dynamic cloud computing. The main objective is to improve the reliability of cloud computing environment by considering the resources available and its working status of each cluster periodically and maximizes the profit for the cloud providers by minimizing the total cost for scheduling, allocation and execution cost and minimizing total turn-around, total waiting time and total execution time. Transportation Problem based scheduling algorithm for task scheduling and resource allocation in decentralized and dynamic cloud computing environment, efficiently schedule and allocate the tasks.

Cui Lin and Shiyong Lu (2011) proposed the Scalable Heterogeneous Earliest Finish Time workflow scheduling algorithm to schedule a workflow elastically on a Cloud computing environment. Scalable Heterogeneous Earliest Finish Time workflow scheduling algorithm enables resources to scale elastically during workflow execution.

Shaminder Kaur, Amandeep Verma (2012) states that Cloud computing is recently a booming area and has been emerging as a commercial reality in the information technology domain. Clouds enable the users to use utility services. Users are required to pay for access to the services based on their usage and level of quality of service required. The scheduling of the cloud services to the consumers by service providers influences the cost benefit of this computing paradigm. According to authors Tasks should be scheduled efficiently such that the execution cost and time can be reduced. The work on this paper focuses on the meta-heuristic based scheduling, which minimizes execution time and execution cost as well. In this paper authors proposed a modified genetic algorithm for single user jobs in which the fitness is developed to encourage the formation of solutions to achieve the time minimization and compared it with existing heuristics. Experimental results show that, under the heavy loads, the proposed algorithm exhibits a good performance.

Lei Ying, Siva Theja Maguluri, R. Srikant (2012) states that Cloud computing services are becoming ubiquitous, and are starting to serve as the primary source of computing power for both enterprises and personal computing applications. The authors consider a stochastic model of a cloud computing cluster, where jobs arrive according to a stochastic process and request virtual machines, which are specified in terms of resources such as CPU, memory and storage space. They first define capacity of model of a cloud, i.e., the maximum rates at which jobs can be processed in such a system. Then, they show that the widely-used Best-Fit scheduling algorithm is not throughput-optimal, and present alternatives which achieve any arbitrary fraction of the capacity region of the cloud. Then, they study the delay performance of these alternative algorithms through simulations. A primary contribution is the development of frame-based non-preemptive Virtual machine configuration policies. These policies can be made nearly throughput-optimal by choosing sufficiently long frame durations, whereas the widely used best-fit policy was shown to be not throughput-optimal. Simulations indicate that long frame durations are not only good from a throughput perspective but also seem to provide good delay performance.

Sandeep Tayal (2011) states that Cloud computing is a latest new computing paradigm where applications, data and IT services are provided over the Internet. The Task management is the key role in cloud computing systems. Task scheduling problems are premier which relate to the efficiency of the whole cloud computing facilities. Task scheduling algorithm plays key role in cloud computing. In this paper they proposed an optimized algorithm based on the Fuzzy Genetic Algorithm optimization which makes a scheduling decision by evaluating the entire group of task in the job queue. According to the new features of cloud computing, such as flexibility, virtualization and etc., this paper discusses task scheduling mechanism based on load balancing in cloud computing. This task scheduling mechanism can not only meet user's requirements, but also get high resource utilization.

Lizheng Guo, Shuguang Zhao, Shigen Shen, Changyuan Jiang (2012) in their paper states that Cloud computing is an emerging technology and it allows users to pay as you need and has the high performance. Cloud computing is a heterogeneous system as well and it holds large amount of application data. In the process of scheduling some intensive data or computing an intensive application, it is acknowledged that optimizing the transferring and processing time is crucial to an application program. In many different domains, in order to improve the efficiency the optimizing task scheduling is necessary. In this paper in order to minimize the cost of the processing authors formulate a model for task scheduling and propose a particle swarm optimization algorithm which is based on small position value rule. By virtue of comparing Particle swarm optimization algorithm with the Particle swarm optimization algorithm embedded in crossover and mutation and in the local research, the experiment results show the Particle swarm optimization algorithm not only converges faster but also runs faster than the other two algorithms in a large scale. Experimental result manifests that the Particle swarm optimization algorithm both gains optimal solution and converges faster in large tasks than the other two.

K. Dinesh, G. Poornima, K. Kiruthika (2012) states that Cloud Computing is an emerging technique in recent years that provides computing as a services. In order to maximize resources utilization, many scheduling algorithms were analyzed and implemented. Job scheduling using Berger model is one of the algorithm for scheduling jobs. The combination of Berger model and Neural Network would overcome the disadvantage of Berger Model i.e., incompletion of
task when tasks-resources match is not achieved. In this work, the submitted jobs are classified based on different parameters like bandwidth, memory, Completion time and Resources Utilization. The classified user tasks are passed to the neural network. Neural network consists of input layer, hidden layer and output layer. With the help of hidden layer, the jobs are matched with the resources by adjusting weight. The performance of the system has been improved by means of efficient use of bandwidth, reducing a completion time which in turn improves resources utilization. Neural network is used either for reclassification of unmapped jobs or fast and efficient mapping from past knowledge.

Linan Zhu, Qingshui Li, Lingna He (2012) in this paper make research and elaboration on the cloud computing technology, and analyze the cloud computing system structure and the realization of mechanism, resources scheduling strategy is the key technology in cloud computing. Therefore the use of ant colony algorithm for the basic model, detailed analysis shows that the algorithm for calculating node distribution and load balancing has good performance and the algorithm has better scheduling performance.

Pardeep Kumar, Amandeep Verma (2012) states that Scheduling is a critical problem in Cloud computing, because a cloud provider has to serve many users in Cloud computing system. So scheduling is the major issue in establishing Cloud computing systems. A good scheduling technique also helps in proper and efficient utilization of the resources. Many scheduling techniques have been developed by the researchers like Genetic Algorithm, Particle Swarm Optimization, Min-Min, Max-Min etc. This paper proposes a new scheduling algorithm which is an improved version of Genetic Algorithm. In the proposed scheduling algorithm the Min-Min and Max-Min scheduling methods are merged in standard Genetic Algorithm. They have designed and tested an algorithm which is able to schedule multiple jobs on multiple machines in an efficient manner such that the jobs take the minimum time for completion.

Shaobin Zhan, Hongying Huo (2012) states that Job scheduling system problem 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. For characteristics of particle swarm optimization algorithm in solving the large-scale combination optimization problem easy to fall into the search speed slowly and partially the most superior, the global fast convergence of simulated annealing algorithm is utilized to combine particle swarm optimization algorithm in each iteration, which enhances the convergence rate and improves the efficiency. This paper proposed the improve particle swarm optimization algorithm in resources scheduling strategy of the cloud computing which increases convergence speed of Particle swarm optimization. The experimental results indicate that, improved particle swarm optimization algorithm shortens the average operation time of tasks, supplies proper resources to user task efficiently in the environment, increases utilization ratio of resources.

Gap Study
Available scheduling algorithms are more complex, time consuming an does not consider reliability and availability of the cloud computing environment. Therefore there is a need to implement a scheduling algorithm that can improve the availability and reliability in cloud environment.

Conclusion
Cloud computing technology virtualizes and offers many services purely a dynamic environment across the network. The existing task scheduling algorithms are mostly consider various parameters like time, cost, make span, speed, scalability, throughput, resource utilization, scheduling success rate and so on. But, available scheduling algorithms are more complex, time consuming and does not consider reliability and availability of the cloud computing environment. Therefore there is a need to implement a scheduling algorithm that can improve the availability and reliability in cloud environment.

References


