Journal of Global Research in Computer Sciences

2021

Vol.12 No.1

Numerous Zones of Cloud Computing in Information Technology – Survey

C. Edward Jaya Singh*[1], E. Baburaj[2] and W.R. Sam Emmanuel[1]
[1]Nesamony Memorial Christian College, Marthandam, India
[2]Narayanaguru College of Engineering and Technology, Manjalumoodu, India

Information Technology changes the style of the people every day in the organizations. The increasing use of information technology has brought with it overheads in the implementation and maintenance of in-house computer systems. The amount of time and finances spent in managing it has increased exponentially; each decade since the 1970s has seen the evolution of it into new phenomenon. In the early 1990s, the Internet transformed the way businesses communicate. By the mid- 90s, e-commerce virtualized purchasing for customers and business partners.

Today, the cloud computing plays a vital role and undertaking broad changes in the way IT services are designed, delivered, consumed, and managed. The boom in cloud computing over the past few years has led to a situation that is common to many innovations and new technologies: “Cloud computing” was coined for what happens when applications and facilities are moved into the “cloud” World. The cloud computing allows users, wherever they are, to obtain computing capabilities through the Internet from a remote network of servers. Cloud computing is not something that suddenly appeared overnight; in some form it may trace back to a time when computer systems remotely time-shared computing resources and applications. More currently though, cloud computing refers to the many different types of services and applications being delivered in the internet cloud, and the fact that, in many cases, the devices used to access these services and applications do not require any special applications. Characteristics of cloud computing are shared infrastructure1, dynamic provisioning2, network access1, managed metering2 and so on.

These services are over and above the support of service deployments of Voice Over Internet Protocol (VOIP) systems3, collaboration systems, and conferencing systems for both voice and video. They can be accessed from any location and linked into current services to extend their capabilities, as well as standalone as service offerings. In terms of social networking, using cloud-based communications provides click-to-call capabilities from social networking sites, access to Instant Messaging systems and video communications, broadening the interlinking of people within the social circle. The Public cloud2, Private cloud2, Hybrid cloud2 and Community cloud2 are the different types of cloud deployment models.

Some of the possible benefits for those who offer cloud computing-based services and applications are Cost saving, Scalability/Flexibility, reliability1, maintenance and mobile accessibility.

The rest of the paper is organized as follows. Chapter 2 explains the common structure of the Cloud environment and its application. The chapter 3 elaborates the working methodology of the image processing applications in clouds. The chapter 4 deals the security mechanism followed to get high confidence of data in cloud environment. It also discusses the data placement methods of clouds in chapter 5. The different allocation models of the resources through the cloud are discussed in chapter 6. The chapter 7 gives the scheduling applications of Cloud. The chapter 8 explains the mining algorithms and its applications in cloud model followed by the conclusion in chapter 9.

General construction of cloud computing See figure 1.

Cloud based image processing

The image processing area plays important role in cloud computing technology. The Outsourced Image Recovery Service (OIRS) feats different domain technologies and takes security, efficiency and design complexity into consideration from the very beginning of the service flow. Data owners only need to outsource compressed image samples to cloud for reduced storage ahead. In OIRS, data users can harness the cloud to securely reconstruct images without revealing information from either the compressed image samples or the underlying image content. The OIRS design is the emblematic application scenario for compressed sensing, and then show its natural extension to the general data for meaningful tradeoffs between efficiency and accuracy. It is thoroughly analyze the privacy-protection of OIRS and conduct extensive experiments to demonstrate the system effectiveness and efficiency4.

Image processing algorithms related to remote sensing have been tested and utilized on the Hadoop MapReduce parallel platform. Although there has been considerable research utilizing the Hadoop platform for image processing rather than for its original purpose of text processing, it had never been proved that Hadoop can be

successfully utilized for high-volume image files. There are eight practical image processing algorithms are researched for image processing to know the advantage of using Hadoop. It is extend the file approach in Hadoop to regard the whole TIFF image file as a unit by expanding the file format that Hadoop uses. The experiments have shown that the method is scalable and efficient in processing multiple large images used mostly for remote sensing applications, and the difference between the single PC runtime and the Hadoop runtime is clearly noticeable5.
Dynamic Switch of Reduce Function (DSRF) algorithm is proposed for MapReduce to switch dynamically to the next task according to the achieved percentage of tasks and reduce the idle time of reduce. The DSRF scheme efficiently improves MapReduce performance in running 2-D to 3-D applications6.
Energy conservation is accomplished by mobile device computation to the cloud using image processing, enabling the mobile device to save energy in the idle mode. The local execution energy consumption is replaced by additional transmissions between the mobile device and the cloud. Cloud computing is defined as applications delivered as services over the Internet and the hardware and software in data centers providing those services. Cloud computing allows quick utilization of cheap, scalable services7. Cloud computing, and subsequently cloud robotics have risen as an alternative to offer solutions to the expanding needs in robotics. Rather than limiting the robot to use only the onboard software, cloud robotics offers access to vast resources; in most cases through wireless internet to complete computational needs remotely. A local network and an external camera are used to control the robot and perform image recognition utilizing the cloud system. Cloud robotics minimizes the need of hardware, which traditionally translates to lower prices for products that are highly technologically advanced. Cloud robotics is additionally appealing due to the simple way in which it can be continuously improved and easily shared8.
Images and Videos are the major new generated big datas today which are generated by various sources. Besides the algorithms become much more complex, which poses great demands to data storage and computation power. This image processing cloud project targets to support the image processing research by leveraging the cloud computing and big data analysis technology. A new design is developed for image processing cloud architecture, and big data processing engine based on Hadoop. It is also reported the performance scalability and analysis on the cloud using several widely used image processing algorithms9. The management of medical image data in a cloud computing environment with access to mobile users through a mobile application to help the patients and doctors to view patient health records and prescriptions on their handheld devices that supports Android OS. The application that has been effectively implemented allows a flexible medium for patients to access vital health records at their convenience, without a need to visit the hospital to view the same. The application is an advantage to people residing in remote areas and who cannot access hospitals in cities at tease10.

Cloud security
Cloud computing permits users to use applications without install software and access their personal files and application at any computer with internet or intranet access. Many users place their data in the cloud, so correctness of data and security is key concern. The software enabled work to change the software working environment is done by Cloud Computing for the next generation. It is interconnecting the large- scale computing resources to effectively integrate, and to computing resources as a service to users. To ensure the correctness of data, it is considered the task of allowing a Third Party Auditor (TPA), on behalf of the cloud client, to verify the integrity of the

data stored in the cloud. The auditing process should bring in no new vulnerabilities towards user data privacy, and introduce no additional online burden to the user. It is proposed a secure cloud storage system supporting privacy- preserving public auditing. It is further extend the result to enable the TPA to perform audits for multiple users simultaneously and efficiently with RC5 encryption algorithm. It shows the proposed scheme is highly efficient and data modification attacks, and even server colluding attacks11.
One of the most important aspect refers to security: while some cloud computing security issues are inherited from the solutions adopted to create such services, many new security questions that are particular to these solutions also arise, including those related to how the services are organized and which kind of service/data can be placed in the cloud. Aiming to give a better understanding of this complex scenario, it is necessary to identify and classify the main security concerns and solutions in cloud computing, and propose taxonomy of security in cloud

computing, giving an overview of the current status of security in this emerging technology12.

Guaranteeing the security of corporate data in the cloud is difficult. Each service has their own security issues. The different aspects of security issues related with cloud computing, and its possible solution are presented here13. Security issues for cloud computing and present a layered framework for secure clouds and then focus on two of the layers, i.e., the storage layer and the data layer. A pattern for secure third party publications of documents in a cloud will converse secure federated query processing with map Reduce and Hadoop, and discuss the use of secure co-processors for cloud computing. The XACML implementation for Hadoop that building trusted applications from untrusted components will be a major aspect of secure cloud computing14. Successful implementation of cloud computing in an enterprise requires proper planning and understanding of emerging risks, threats, vulnerabilities, and possible counter-measures. It is believed enterprise should analyze the organization security risks, threats, and available Counter measures before adopting this technology. The security risks and concerns in cloud computing and enlightened steps an enterprise can take to reduce security risks and protect their resources15. A detailed analysis of the cloud computing security issues and challenges focusing on the cloud computing types and the service delivery types16. The security for Cloud Computing is emerging area to provide security topic in terms of cloud computing based on analysis of Cloud Security treats and Technical Components of Cloud Computing17.

It is deliberated the task of allowing a Third Party Auditor (TPA), on behalf of the cloud client, to confirm the integrity of the dynamic data or information which is stored in the cloud. The introduction of TPA eliminates the involvement of the client through the auditing of whether his data stored in the cloud is indeed intact, which can be important in achieving economies of scale for Cloud Computing. The support for data dynamics via the most general forms of data operation, such as block modification, insertion and deletion, is also a significant step toward practicality, since services in Cloud Computing are not limited to archive or backup data only. The potential security problems of direct extensions with fully dynamic data updates are identified and the elegant verification scheme for the seamless integration of these is constructed by the design. To achieve efficient data dynamics, the storage models by manipulating the classic Merkle Hash Tree construction for block tag authentication is improved. To support efficient handling of multiple auditing tasks, the technique of bilinear aggregate signature is used to extend the main result into a multi-user setting18.

Data Protection Application Security Privacy is an important security issues that have to be included in cloud computing. A model system in which cloud computing system is combined with Cluster Load balancing, SSL over AES and secure session. In this model, some important security services, including authentication, confidentiality and integrity, are provided in cloud computing system19.

Cloud computing has great potential of providing robust computational power to the society at reduced cost. It empowers customers with limited computational resources to outsource their large computation workloads to the cloud, and economically enjoy the massive computational power, bandwidth, storage and even appropriate software that can be shared in a pay-per-use manner. Despite the tremendous benefits, security is the primary obstacle that prevents the wide adoption of this promising computing model, especially for customers when their confidential data are consumed and produced during the computation. Treating the cloud as an intrinsically insecure computing platform from the viewpoint of the cloud customers, it must design mechanisms that not only protect sensitive information by enabling computations with encrypted data, but also protect customers from malicious behaviors by enabling the validation of the computation result. Such a mechanism of general secure computation outsourcing was recently shown to be feasible in theory, but to design mechanisms that are practically efficient remains a very challenging problem. Focusing on engineering computing and optimization tasks, this paper investigates secure outsourcing of widely applicable linear programming (LP) computations. In order to achieve practical efficiency, our mechanism design explicitly decomposes the LP computation outsourcing into public LP solvers running on the cloud and private LP parameters owned by the customer. The resulting flexibility allows us to explore appropriate security or efficiency tradeoff via higher level abstraction of LP computations than the general circuit representation. In particular, by formulating private data owned by the customer for LP problem as a set of matrices and vectors, we are able to develop a set of efficient privacy- preserving problem transformation techniques, which allow customers to transform original LP problem into some arbitrary one while protecting sensitive input or output information. To validate the computation result, it is further explore the fundamental duality theorem of LP computation and derive the necessary and sufficient conditions that correct result must satisfy. Such result verification mechanism is extremely efficient and incurs close-to- zero additional cost on both cloud server and customers. Extensive security analysis and experiment results show the immediate practicability of the mechanism design20. The research into the use of multi-cloud providers to maintain security has received less attention from the research community than has the use of single clouds. The uses of multi-clouds due to its facility to diminish the safety threats which affect the cloud computing user in various ways21.

Storing data in a third party's cloud system causes serious concern over data confidentiality. General encryption schemes protect data confidentiality, but also limit the functionality of the storage system because a few operations are supported over encrypted data. Constructing a secure storage system that supports multiple functions is challenging when the storage systems is distributed and has no central authority. It is proposed a threshold proxy re-encryption scheme and integrate it with a decentralized erasure code such that a secure distributed storage system is formulated. The distributed storage system not only supports secure and robust and data storage and retrieval, but also lets a user forward his data in the storage servers to another user without retrieving the data back. The main technical contribution is that the proxy re- encryption scheme supports encoding operations over encrypted messages as well as forwarding operations over encoded and encrypted messages. This method fully integrates encrypting, encoding, and forwarding. It is analyzed and suggested suitable parameters for the number of copies of a message dispatched to storage servers and the number of storage servers queries by a key server. These parameters allow more flexible adjustment between the number of storage servers and robustness22.

Data placements in cloud

A secure multi owner data sharing scheme, named Mona, for dynamic groups in the cloud. By leveraging group signature and dynamic broadcast encryption techniques, any cloud user can anonymously share the data with others. Meanwhile, the storage overhead and encryption computation cost of this scheme are self- governing with the number of revoked users. In addition, it is analyzed the security of this scheme with rigorous proofs, and demonstrate the efficiency of this scheme in experiments23. Data centres are used by the Data manager to store the data effectively. When one task needs several datasets located in different data centres, the movement of large volumes of data becomes a challenge. A matrix based k-means clustering strategy for data placement in scientific cloud workflows contains two algorithms that group the existing datasets in k data centres during the workflow build-time stage, and dynamically clusters newly generated datasets to the most appropriate data centres based on dependencies during the runtime stage24.

Many scientific workflows are data intensive: large volumes of intermediate datasets are generated during their execution. Some valuable intermediate datasets need to be stored for sharing or reuse. Traditionally, they are selectively stored according to the system storage capacity, determined manually. As doing science on clouds has become popular nowadays, more intermediate datasets in scientific cloud workflows can be stored by different storage strategies based on a pay- as-you-go model. To build an intermediate data dependency graph (IDG) from the data provenances in scientific workflows deleted intermediate datasets can be regenerated, and as such develop a novel algorithm that can find a minimum cost storage strategy for the intermediate datasets in scientific cloud workflow systems. The strategy achieves the best trade-off of computation cost and storage cost by automatically storing the most appropriate intermediate datasets in the cloud storage. This strategy can be utilized on demand as a minimum cost benchmark for all other intermediate dataset storage strategies in the cloud. The Amazon clouds' cost model is utilized and applies the algorithm to general random as well as specific astrophysics pulsar searching scientific workflows for evaluation25. With continued research advances in trusted computing and computation-supporting encryption, life in the cloud can be advantageous from a business intelligence standpoint over the isolated alternative that is more common today26.

Infrastructure as a Service (IaaS) cloud computing has revolutionized the way of thinking of acquiring resources by introducing a simple change: allowing users to lease computational resources from the cloud provider's datacenter for a short time by deploying virtual machines (VMs) on these re-sources. This new model raises new challenges in the design and development of IaaS middleware. One of those challenges is the need to deploy a large number (hundreds or even thousands) of VM instances simultaneously. Once the VM instances are deployed, another challenge is to simultaneously take a snapshot of many images and transfer them to persistent storage to support management tasks, such as suspend-resume and migration. It is important to enable efficient concurrent deployment and snapshotting that are at the same time hypervisor independent and ensure a maximum compatibility with different configurations. The challenges by proposing a virtual file system specifically optimized for virtual ma-chine image storage. It is based on a lazy transfer scheme coupled with object versioning that handles snapshotting transparently in a hypervisor-independent fashion, ensuring high portability for different configurations27.

Enterprises can save storage infrastructure investments and reduce administration costs tremendously while outsourcing their data to third-party cloud storage providers. However data privacy and integrity are hindering their data migration steps to flexible and cost effective cloud storage. And traditional security schemes can provide data protection for this complicated cloud environment by sacrificing convenient operations, such as searching and sharing, which conflicts the flexibility and availability of cloud environment with multi-services for multi- tenants. This raises the demand for innovative and secure searching schemes to protect data privacy, data storage and retrieval process, and to access data flexibly on cloud storage. A privacy preserved data sharing scheme on cloud storage is used to protect data privacy without sacrificing the cloud flexibility and accessibility: 1) Privacy Preserving Data Searching scheme to provide secure and efficient searching and sharing on privacy hidden data on cloud storage; 2) Policy based access control provides flexible yet data access ability on the privacy protected data28.

Prognostics and Health Management (PHM) has been extensively studied in recent years, with many sophisticated techniques and intelligent algorithms developed for machinery data analysis, health assessment and decision making. PHM solutions for typical components such as roller bearing and machine tool have grown and proved to be reliable enough for industrial application. However utilization of PHM solutions in industry is still much limited due to high research, development and implementation costs. Such limitations are more severe in smaller scale factories and workshops where no ample resource is available for implementing PHM systems. With advantages that can be delivered with the emerging cloud computing paradigm, a cloud-based prognostics and health management system for manufacturing industry has been developed based on Watchdog Agent® tools and the ideology of PHM as a Service. In addition to traditional data acquisition and management functions in a machine condition monitoring system, the cloud based PHM platform is able to further provide on-demand, customizable and low-cost data analysis service. Machinery data accumulated within the cloud system further enables more advanced services such as machine-to-machine comparison, data mining and knowledge discovery29.

A storage management framework for Web 2.0 services places users back in control of their data. Current Web services complicate data management due to data lock-in and lack usable protection mechanisms, which makes cross-service sharing risky. Our framework allows multiple Web services shared access to a single copy of data that resides on a personal storage repository, which the user acquires from a cloud storage provider. Access control is based on hierarchically, filtered views, which simplify cross-cutting policies, and enable least privilege management30. The concept of deniable cloud storage that guarantees privacy of data even when one's communication and storage can be opened by an adversary. It clearly shows that existing techniques and systems do not adequately solve this problem. It is designed the first sender and receiver deniable public- key encryption scheme that is both practical and is built from standard tools. The practical aspect of user collaboration provides an implementation of a deniable shared file system, DenFS31.

Resource allocation through cloud

Clouds can be used to deliver extra resources whenever necessary for the enterprises or organizations. For this vision to be achieved, requires both policies defining when and how cloud resources are allocated to applications and a platform implementing not only these policies but also the whole software stack supporting management of applications and resources. Aneka is a cloud application platform capable of provisioning resources obtained from a variety of sources, including private and public clouds, clusters, grids, and desktops grids. The Aneka's deadline driven provisioning mechanism is responsible for supporting quality of service (QoS) aware execution of scientific applications in hybrid clouds composed of resources obtained from a variety of sources. Aneka is able to efficiently allocate resources from different sources in order to reduce application execution times32.

Many of the touted gains in the cloud model come from resource multiplexing through virtualization technology. A system for resource multiplexing through technology used to allocate data center resources dynamically based on application demands and support green computing by optimizing the number of servers in use. The concept of “skewness” used to measure the unevenness in the multi-dimensional resource utilization of a server. By minimizing skewness, we can combine different types of workloads nicely and improve the overall utilization of server resources33. With the increased demand for delivering services to a large number of users, they need to offer differentiated services to users and meet their quality expectations. Existing resource management systems in data centers are yet to support Service Level Agreement (SLA)-oriented resource allocation, and thus need to be enhanced to realize cloud computing and utility computing. In addition, no work has been done to collectively incorporate customer-driven service management, computational risk management, and autonomic resource management into a market-based resource management system to target the rapidly changing enterprise requirements of Cloud computing. SLA-oriented resource management architecture supports integration of market based provisioning policies and virtualization technologies for flexible allocation of resources to applications. The performance results obtained from our working prototype system shows the feasibility and effectiveness of SLA-based resource provisioning in Clouds34.

A comprehensive solution for resource allocation is fundamental to any cloud computing service provider. Any resource allocation model has to consider computational resources as well as network resources to accurately reflect practical demands. Another aspect that should be considered while provisioning resources is energy consumption. This aspect is getting more attention from industrial and government parties. Calls for the support of green clouds are gaining momentum. With that in mind, resource allocation algorithms aim to accomplish the task of scheduling virtual machines on the servers residing in data centers and consequently scheduling network resources while complying with the problem constraints. Several external and internal factors that affect the performance of resource allocation models are introduced. Design challenges are discussed with the aim of providing a reference to be used when designing a comprehensive energy-aware resource allocation model for cloud computing data centers35.

To build a management framework dedicated to automatic resource allocation in virtualized applications, identify from experiments the sources of instabilities in the controlled systems. Two types of policies were analyzed, threshold-based and reinforcement learning techniques to dynamically scale resources. Both approaches are tricky and that trying to implement a controller without looking at the way the controlled system reacts to actions, both in time and in amplitude, is doomed to fail. To build good resource management policies, and longer term issues on which are currently working to manage contracts and reinforcement learning efficiently in cloud controllers36.

User communities are rapidly transitioning their ‘‘traditional desktops’’ that have dedicated hardware and software installations into ‘‘virtual desktop clouds’’ (VDCs) that are accessible via thin-clients. To allocate and manage VDC resources for Internet-scale desktop delivery, existing works focus mainly on managing server-side resources based on utility functions of CPU and memory loads, and do not consider network health and thin-client user experience. Resource allocations without combined utility-directed information of system loads, network health and thin-client user experience in VDC platforms inevitably results in costly guesswork and over-provisioning of resources. In this paper, we develop an analytical model viz., ‘‘Utility- Directed Resource Allocation Model (U- RAM)’’ to solve the combined utility- directed resource allocation problem within VDCs. The solution involves an iterative algorithm that leverages utility functions of system, network and human components obtained using a novel virtual desktop performance benchmarking toolkit viz., ‘‘VDBench’’ that was developed. The combined utility functions are used to direct decision schemes based on Kuhn–Tucker optimality conditions for creating user desktop pools and determining optimal resource allocation size/location. It is deployed VDBench in a VDC testbed featuring: (a) popular user applications (Spreadsheet Calculator, Internet Browser, Media Player, Interactive Visualization), and (b) TCP/UDP based thinclient protocols (RDP, RGS, PCoIP) under a variety of user load and network health conditions. Simulation results based on the utility functions obtained from the testbed demonstrate that our solution maximizes VDC scalability i.e., ‘VDs per core density’, and ‘user connections quantity', while delivering satisfactory thin-client user experience37.

Service demanders intend to solve sophisticated parallel computing problem by requesting the usage of resources across a cloud-based network, and a cost of each computational service depends on the amount of computation. Game theory is used to solve the problem of resource allocation. A practical approximated solution with the two steps is proposed. First, each participant solves its optimal problem independently, without consideration of the multiplexing of resource assignments. A Binary Integer Programming method is proposed to solve the independent optimization. Second, an evolutionary mechanism is designed, which changes multiplexed strategies of the initial optimal solutions of different participants with minimizing their efficiency losses. The algorithms in the evolutionary mechanism take both optimization and fairness into account. It is demonstrated that Nash equilibrium always exists if the resource allocation game has feasible solutions38.

Ad-hoc parallel data processing has emerged to be one of the most imperative applications for Infrastructure-as-a-Service (IaaS). Major Cloud computing companies have started to integrate frameworks for parallel data processing in their product portfolio, making it easy for customers to access these services and to deploy their programs. The processing frameworks are currently used have been designed for static, homogeneous cluster setups and disregard the particular nature of a cloud. The allocated compute resources may be inadequate for big parts of the submitted job and unnecessarily increase processing time and cost. Nephel's architecture offers for efficient parallel data processing in clouds. It is the first data processing framework for the dynamic resource allocation offered by today's IaaS clouds for both, task scheduling and execution. Particular tasks of a processing job can be assigned to different types of virtual machines which are automatically instantiated and terminated during the job execution39.

Purlieus is a MapReduce resource allocation system aimed at enhancing the performance of MapReduce jobs in the cloud. Purlieus provisions virtual MapReduce clusters in a locality-aware manner enabling MapReduce virtual machines (VMs) access to input data and importantly, intermediate data from local or close-by physical machines. The locality-awareness during both map and reduce phases of the job not only improves runtime performance of individual jobs but also has an additional advantage of reducing network traffic generated in the cloud data center40. The emerging cloud computing paradigm provides administrators and IT organizations with tremendous freedom to dynamically migrate virtualized computing services between physical servers in cloud data centers. Virtualization and VM migration capabilities enable the data center to consolidate their computing services and use minimal number of physical servers. VM migration offers great benefits such as load balancing, server consolidation, online maintenance and proactive fault tolerance. In cloud computing environments the cost of VM migration requires thorough consideration. Each VM migration may result in SLA violation, hence it is essential to minimize the number of migrations to the extent possible. Failure to do so will result in performance degradation and the cloud provider will have to incur the cost in monetary terms. This will play a major role in avoiding the performance degradation encountered by a migrating VM41.

Resource allocation is an integral, evolving part of many data center management problems such as virtual machine placement in data centers, network virtualization, and multi-path network routing. Since the problems are inherently NP-Hard, most existing systems use custom-designed heuristics to find a suitable solution. Such heuristics are often rigid, making it difficult to extend them as requirements change. Wrasse is a generic and extensible tool that cloud environments can use to solve their specific allocation problem. Wrasse provides a simple yet expressive specification language that captures a wide range of resource allocation problems. At the back-end, it leverages the power of GPUs to provide solutions to the allocation problems in a fast and timely manner. The extensibility of Wrasse by expressing several allocation problems in its specification language42. One of the major pitfalls in cloud computing is related to optimizing the resources being allocated. Because of the uniqueness of the model, resource allocation is performed with the objective of minimizing the costs associated with it. The other challenges of resource allocation are meeting customer demands and application requirements43.

Major Cloud computing companies have started to integrate frameworks for parallel data processing in their product portfolio, making it easy for customers to access these services and to deploy their programs. The processing frameworks which are currently used have been designed for static, homogeneous cluster setups and disregard the particular nature of a cloud. The allocated compute resources may be inadequate for big parts of the submitted job and unnecessarily increase processing time and cost. Nephele is the first data processing framework to explicitly exploit the dynamic resource allocation offered by today's IaaS clouds for both, task scheduling and execution. Particular tasks of a processing job can be assigned to different types of virtual machines which are automatically instantiated and terminated during the job execution. Based on this new framework extended evaluations of MapReduce- inspired processing jobs on an IaaS cloud system is performed and the results to the popular data processing framework Hadoop is compared44. A cloud environment consists of multiple customers requesting for resources in a dynamic environment with possible constraints. In the existing economy based models of cloud computing, allocating the resource efficiently is a challenging job. The developed resource allocation algorithm is based on different parameters like time, cost, No of processor request etc. The developed priority algorithm is used for a better resource allocation of jobs in the cloud environment used for the simulation of different models or jobs in an efficient way. After the efficient resource allocation of various jobs, an evaluation is being carried out which illustrates the better performance of cloud computing with profit45.

Scheduling

One of the most important aspects which differentiate a cloud workflow system from its other counterparts is the market- oriented business model. This is a significant innovation which brings many challenges to conventional workflow scheduling approaches and polices. To investigate such issue, it is proposed a market-oriented hierarchical scheduling strategy in cloud workflow systems. In particular, the service-level scheduling deals with the Task-to-Service assignment where tasks of individual workflow instances are mapped to cloud services in the global cloud markets based on their functional and non- functional QoS requirements; the task-level scheduling deals with the optimization of the Task-to-VM (virtual machine) assignment in local cloud data centres where the overall running cost of cloud workflow systems will be minimized given the satisfaction of QoS constraints for individual tasks. Based on our hierarchical scheduling strategy, a package based random scheduling algorithm is presented as the candidate service-level scheduling algorithm and three representative metaheuristic based scheduling algorithms including genetic algorithm (GA), ant colony optimization (ACO), and particle swarm optimization (PSO) are adapted, implemented and analyzed as the candidate task-level scheduling algorithms. The hierarchical scheduling strategy is being implemented in our SwinDeW-C cloud workflow system and demonstrating satisfactory performance. The experimental results shows that the overall performance of ACO based scheduling algorithm is better than others on three basic measurements: the optimization rate on makespan, the optimization rate on cost and the CPU time46.

One of the major contribution of cloud computing is to avail all the resources at one place in the form a cluster and to perform the resource allocation based on request performed by different users. It is defined the user request in the form of requirement query. Cloud Computing devices being able to exchange data such as text files as well as business information with the help of internet. Technically, it is completely distinct from an infrared using a new super-sensitive optical sensor (SSOS). The transmission and storage of large amounts of information, and become propulsion of fiber-optic accelerating towards 40G/100G. Its foreground is to provide secure, quick, convenient data storage and net computing service centered by internet47.

Scheduling is a critical problem in Cloud computing, because a cloud provider has to serve many users in Cloud computing system. A good scheduling technique also helps in proper and efficient utilization of the resources. Many scheduling techniques have been developed by the researchers like GA (Genetic Algorithm), PSO (Particle Swarm Optimization), Min-Min, Max-Min, X-Sufferage etc. It is proposed a new scheduling algorithm which is an improved version of Genetic Algorithm. This scheduling algorithm, the Min-Min and Max-Min scheduling methods are merged in standard Genetic Algorithm. Min-Min, Max-Min and Genetic Scheduling techniques are discussed and in the last the performance of the standard Genetic Algorithm and proposed improved Genetic Algorithm is compared48. The users' perspective of efficient scheduling may be based on parameters like task completion time or task execution cost. Service providers like to ensure that resources are utilized efficiently and to their best capacity. A scheduling algorithm addresses the major challenges of task scheduling in cloud. The incoming tasks are grouped on the basis of task requirement like minimum execution time or minimum cost and prioritized. Resource selection is done on the basis of task constraints using a greedy approach. This model is implemented and tested on simulation toolkit. The results validate the correctness of the framework and show a significant improvement over sequential scheduling49.

The number of cloud users has been growing exponentially and apparently scheduling of virtual machines in the cloud becomes an important issue to analyze. It is proposed a new algorithm that combines the advantages of all the existing algorithms and overcomes their disadvantages50.

Conventional scheduling methodo- logy encounters a number of challenges. During the tasks scheduling in cloud systems, how to make full use of resources and how to effectively select resources are also important factors. At the same time, communication delay also plays an important role in cloud scheduling, which not only leads to waiting between tasks but also results in much idle interval time between processing units. A fuzzy clustering method is used to effectively preprocess the cloud resources. Combining the list scheduling with the task duplication scheduling scheme, a new directed acyclic graph based scheduling algorithm called earliest finish time duplication algorithm for heterogeneous cloud systems is presented. Earliest finish time duplication attempts to insert suitable immediate parent nodes of the current selected node in order to reduce its waiting time on the processor51.

The problem of online task scheduling of jobs such as MapReduce jobs, Monte Carlo simulations and generating search index from web documents, on cloud computing infrastructures are investigated. It is considered the virtualized cloud computing setup comprising machines that host multiple identical virtual machines (VMs) under pay-as-you-go charging, and that booting a VM requires a constant setup time. The cost of job computation depends on the number of VMs activated, and the VMs can be activated and shutdown on demand. It is introduced a new bi-objective algorithm to minimize the maximum task delay, and the total cost of the computation52.

Cloud computing is known as a provider of dynamic services using very large scalable and virtualized resources over the Internet. Due to novelty of cloud computing field, there is no many standard task scheduling algorithm used in cloud environment. Especially that in cloud, there is a high communication cost that prevents well known task schedulers to be applied in large scale distributed environment. Today, researchers attempt to build job scheduling algorithms that are compatible and applicable in Cloud Computing environment Job scheduling is most important task in cloud computing environment because user have to pay for resources used based upon time. Hence efficient utilization of resources must be important and for that scheduling plays a vital role to get maximum benefit from the resources53. Bee Swarm optimization algorithm called Bees Life Algorithm (BLA) applied to efficiently schedule computation jobs among processing resources onto the cloud datacenters. It is considered as NP-Complete problem and it aims at spreading the workloads among the processing resources in an optimal fashion to reduce the total execution time of jobs and then, to improve the effectiveness of the whole cloud computing services. BLA has been inspired by bees' life in nature represented in their most important behaviors which are reproduction and food source searching. A set of experimental tests has been conducted to evaluate the effectiveness and the performance of this algorithm54.

Data mining

Popularity of cloud computing is increasing day by day in distributed computing environment. There is a growing trend of using cloud environments for storage and data processing needs. To use the full potential of cloud computing, data is transferred, processed, retrieved and stored by external cloud providers. Data owners are very skeptical to place their data outside their own control sphere. Their main concerns are the confidentiality, integrity, security and methods of mining the data from the cloud. The efforts directed to which degree this skepticism is justified, by proposing to model Cloud Computing Confidentiality Archetype and Data Mining 3CADM. The 3CADM is a step-by-step framework that creates mapping from data sensitivity onto the most suitable cloud computing architecture and process very large datasets over commodity clusters with the use of right programming model. To achieve this, the 3CADM determines the security mechanisms required for each data sensitivity level, which of these security controls may not be supported in certain computing environments, which solutions can be used to cope with the identified security limitations of cloud computing. The model achieves data confidentiality while still keeping the harmonizing relations intact in the cloud. It also achieves an algorithm to mine the data from the cloud using sector/sphere framework with association rules55. Data Mining is useful for extracting useful data from raw data. Data Mining

Techniques are used commonly in our day- to-day lives to extract useful information and to improve businesses by reducing cost. The use of Data Mining techniques through cloud computing will help end users to retrieve meaningful information from virtually integrated data that reduces the cost of setting up the infrastructure and storage place56.

A "Collaborative approach" tells how component based systems are used with data mining as well as cloud computing. Data Mining helps for extracting potentially useful information from the raw data. The function of association rules are an important data mining technique used to find the interesting relationship with the other related objects. The logical significance to study the concept of dependency in component based systems with data mining is to show the relation between one or more products where a change of one season to another season leads to a potential for a change of one product offer to another product offer. These association rules can either be use at a single level or in multiple levels. Sometimes, association rules becomes important because of output of one rule may be act as an input for other rules. The limited numbers of association rules are generated at different levels. Data mining techniques help businesses to become more efficient by reducing costs. Here, we uses two types of data mining techniques called "Feature Selection & feature extraction" and "Attribute importance" which helps us for "Product Mining". Data mining techniques and applications are very much needed in the cloud computing paradigm. While implementing data mining techniques with cloud computing allows the users to retrieve meaningful information from virtually integrated data warehouse that reduces the costs of infrastructure. The major benefit to combine component based systems with cloud computing is easy accessing with "platform as a service". The other benefits to combine these concepts are making systems more reliable, well maintained and also cost effective57.

An algorithm is announced to mine the data from the cloud using sector/sphere framework with association rules. Data mining is the process of analyzing data from different perspectives and summarizing it into useful information. Mining association rules is one of the most important phases in data mining. Association rules are dependency rules which predict occurrence of an item based on occurrences of other items. Apriori is the best-known algorithm to mine association rules. Cloud can be meant as an infrastructure that provides resources and/or service over the internet. A cloud can be a storage cloud that provides block or file based storage service or it can be a compute cloud that provides computational services. The design and implementation of sector storage cloud and sphere compute cloud are reviewed. Sector is the distributed file system, while sphere is the parallel in-storage data processing framework that can be used to process data stored in sector58.

The integration of data mining techniques with Cloud computing allows the users to extract useful information from a data warehouse that reduces the costs of infrastructure and storage. Security and privacy of user's data is a big concern when data mining is used with cloud computing. An important security concern is privacy attacks based on data mining involving analyzing data over a long period to extract valuable information. A single cloud provider stores the entire client data on a single cloud. This gives the provider and outside attackers an opportunity to gain unauthorized access to cloud, allowing them to analyze client data over a long period to extract sensitive information causing violation of privacy of clients. This is of major concern for many clients of cloud. A cryptography-based scheme can be used for mining the cloud data in a secure way without loss of accuracy. The problem of Knearest neighbor (KNN) classification over horizontally distributed databases is addressed using order preserving symmetric encryption (OPSE) without revealing any unnecessary information. It is very important to use an effective data mining strategy in the cloud to extract interesting patterns that may be forecasting or predictions to be used by the companies in near future to increase their sales. These prediction should be mined securely so as to protect them from interception, thus using a secure cloud mining architecture59.

Data security and access control are the most challenging research work going on, at present, in cloud computing. This is because of the users sending their sensitive data to the cloud providers for acquiring their services. In cloud computing, the data is going to be stored in storage area provided by the service providers. The service providers must have a suitable way to protect their client's sensitive data, especially to protect the data from unauthorized access. A common method of information privacy protection is to store the client's data in encrypted form. If the cloud system is responsible for both storage and encryption/decryption of the data, the system administrators may simultaneously obtain encrypted data and the decryption keys. This allows them to access the information of the client without any authorization. This leads to the risk of sensitive information leak and the method involved of storage and encryption/ decryption is costly. To overcome these problems, a model (cloud server) has been proposed which accepts only those data which are required in an encoded form, performs the service opted by the client and sends the result in the encoded format to be understood by the respective client60.

Every day people are confronted with targeted advertising, and data mining techniques help businesses to become more efficient by reducing costs. Data mining techniques and applications are very much needed in the cloud computing paradigm. The implementation of data mining techniques through Cloud computing will allow the users to retrieve meaningful information from virtually integrated data warehouse that reduces the costs of infrastructure and storage61.

A novel time series data mining and analysis framework inspired by ancient Chinese culture I-Ching. The proposed method converts the time series into symbol spaces by employing the concepts and principles of I-Ching. Algorithms are addressed to explore and identify temporal patterns in the resulting symbol spaces. Using the analysis framework, major topics of time series data mining regarding time series clustering, association rules of temporal patterns, and transition of hidden Markov process can be analyzed. Dynamic patterns are derived and adopted to investigate the occurrence of special events existing in the time series62.

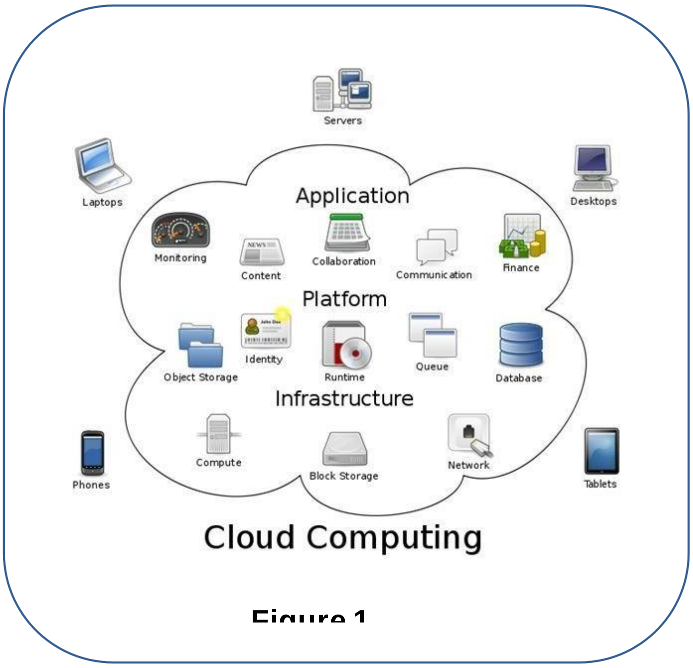

Figure 1

Exploiting Data Mining Techniques for Improving the Efficiency of Time Series Data using SPSSCLEMENTINE are addressed. It is helpful for an organization or individual when choosing Right software to meet their mining needs. It utilizes the famous data mining software SPSS Clementine to mine the factors that affect information from various vantage points and analyze that information. The purpose is to review the selected software for data mining for improving efficiency of time series data. Data mining techniques is the exploration and analysis of data in order to discover useful information from huge databases. So it is used to analyze a large audit data efficiently for Improving the Efficiency of Time Series Data. SPSS-Clementine is object-oriented, extended module interface, which allows users to add their own algorithms and utilities to Clementine’s visual programming environment. The overall objective is to develop high performance data mining algorithms and tools that will provide support required to analyze the massive data sets generated by various processes that is used for predicting time series data using SPSS- Clementine63.

CONCLUSION

Cloud computing plays an effective role in all grounds of information technology. Some of the key factors of cloud are discussed and reported here. Cloud computing technology supports small, medium and large level organization, institutions and business people all over the world. The challenges of cloud computing also increase day by day in every aspects. Thus, to overcoming the problems is very much essential, since large numbers of peoples are migrated to cloud because of its benefits.