CHAPTER 1

INTRODUCTION

Internet-based, on-demand computing where shared resources, data, information and other devices are available to user on-demand is known as Cloud Computing. The computing resources from shared pool are accessed by users on the basis of demand and store their data in third-party data centers at distant locations. At the establishment of cloud platform infrastructure and resources are the wider concept to achieve coherence and economies of scale. A model of cloud computing can be quickly adapted provisioned and quashed with least effort for enabling omnipresent, suitable, on-demand network based access to a shared pool of configurable computing resources. To improve the effectiveness of the cloud resources are dynamically reallocated as per demand and shared by multiple users [1]. With enhanced manageability and less maintenance cloud computing allows enterprises to acquire their applications up and consecutively more rapidly without purchasing licenses for different applications.

Cloud computing ensures the access to single server by multiple users for retrieval and update of their data from cloud computing. Under the pay-as-you-go model (customers pay for services on pay-per-use basis) Cloud computing delivers infrastructure, platform, and software (applications) as subscription based services [2], which are provided to customers and supports hosting of pervasive applications from domestic, research and enterprise domains.

1.1 Cloud Computing: An Overview

Cloud computing is recent development and has its roots in Grid Computing, or initial form of Parallel & Distributed Computing. Grid computing was modeled as volunteer ensemble of computing, storage and data resources. Respectively, such collaborative
computing was named as computing grid, storage grid and data grid. As an electrical grid provides power and client pays back for the power drawn from grids, so was with grid computing framework. Clients may access grids for the services offered by grid and adopts pay-as-you-use model. Grids computing also named utility computing as otherwise, unused resources are converted into a utility.

Cloud computing framework evolved from grid computing in the sense that clouds are also ensemble of software and hardware resources. Evolution though is multi-dimensional, but still it is client and server model. Cloud is complex form of utility computing and presents a pure form of distributed and parallel computing (Figure 1.1).

1.1.1 Evolution of Cloud Computing

A distributed system is distributed in the sense that it’s a collection of independent machines situated local or at remote locations. This ensemble of computers acts as a single large machines. Users of distributed systems are in illusion that they have one united machine at their disposal. Cloud Computing has evolved from distributed computing and evolved through multiple stages. The pictorial representation has been shown as in (Figure 1.2). Distributed computing gives an illusion to its users of one
complete machine for his usage ensuring availability and scalability. Evolving from thin client to grid computing and evolving finally as cloud computing through grid computing, an on-demand service model has taken its final shape. Many definitions of Cloud Computing prevails in literature and one most widely accepted definition was given by National Institute of Standards and Technology (NIST).

“Cloud computing is a model for enabling 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” [4].

![Figure 1.2: Evolution of Cloud Computing](image)

### 1.1.2 Service Models

Cloud computing can be realized by its three service and four deployment models (Figure 1.3). Cloud computing presents/implements a service-driven business model that advocates “everything as a service” offer their “services” according to various models. Infrastructure and resources are provided as services on-demand basis. However,
in practice, services offered by cloud computing environment can be classified into three classes: Software as a service (SaaS), Platform as a Service (PaaS) and Infrastructure as a Service (IaaS) [4].

- **Infrastructure as a Service:** IaaS providers offers the infrastructural or virtualized resources like storage, network infrastructure to cloud users as on-demand, usually in terms of VMs. From their large pools of equipment installed in data centers, IaaS-cloud providers supply additional resources like virtual-machine disk-image library,
firewalls, load balancers, IP addresses and VLANs resources on-demand basis. Examples of IaaS providers include Amazon EC2[5], GoGrid [6] and Flexi scale[7].

- **Platform as a Service**: PaaS provides a framework consisting of layers of software and platforms on which user can develop applications and run, test and host them. PaaS provider develops tools and conformance standards for product and application development by customers. In all PaaS provides Operating Systems, development and execution environment, popular database, and web server for application developers. On cloud platforms, software developers can build up and run their software applications without the price and complication of buying and organizing the underlying hardware and software layers. Cloud user doesn’t have to allocate resources manually because resources are scaled automatically according to user’s application demand. Examples of PaaS providers include IBM, Google App Engine, Force.com [8]. There are two main types of PaaS: Public PaaS and Enterprise PaaS [9]. Services provider is delivered Public PaaS for building applications. For examples Salesforce Heroku [10], AWS Elastic Beanstalk [11], Microsoft Azure [12], and Engine Yard [13]. Central IT within an organization is delivered Enterprise PaaS to developers and possibly partners and business customers. Examples include Gartner which uses the term “Cloud-Enabled Application Platform” or CEAP. Other example includes Apprenda [14], VMware [15] and EMC owned Pivotal [16], and Red Hat OpenShift [17].

- **Software as a Service (SaaS)**: SaaS refers to providing on-demand applications over the web to be used by end user. Cloud user’s right to use software from cloud clients and providers, mount and run application software in SaaS model. Users are unaware of how the applications are run on cloud infrastructure, which eases management and support. Cloud applications are run on several VMs at run-time to meet dynamic work requirements. Task or Job Scheduler schedules these tasks on different VMs and load balancer balance the work over VMs. The cloud users don’t know this process as they only experience a single access point. Cloud providers give the ability to use multitenant applications, where any machine may serve
multiple organizations. Examples of SaaS providers include Salesforce.com [18], Rack space [19].

1.1.3 Deployment Models

There are different ways by which cloud services can be deployed, depending on the organizational structure and the provisioning location. Precisely there are four different models of deployment: Public, Private, Community and Hybrid Cloud [20].

- **Public Cloud**: This model is wholly owned and managed by third party for various management, maintenance and support functions. Clients of public clouds experience affordable computing framework, network resources, databases and supporting infrastructure without spending huge capitals. Access is metered and any scale up and scale down is best experienced at lowest cost. Examples includes Amazon’s and Google’s Cloud.

- **Private Cloud**: Private clouds make services accessible for one designated organization. The specific organization’s employee can get access to it and is accessible within organization premises by means of authentication. It allows the organization to implement and maintain control over security, privacy and authentication. Private cloud may be on premises or off-premises. Yet another aspect is that private cloud may be managed by owner or third party. Private clouds are not the real example of cloud computing. Extra security and control in private cloud is paid in terms of huge capital investment in infrastructure and IT supports.

- **Community Cloud**: This deployment model is conglomerate model where several organizations with similar requirements may pool the infrastructure. It is generalization of private cloud. Data security and integrity is also major problem in this model. Operations of community clouds may be in house or with a third party on the premises.

- **Hybrid Cloud**: Joining together two or more types of deployment model like private and public clouds generally results into hybrid clouds. Mission critical and
business line data and information may be managed in private instance where as non-critical information and data may be managed through and on public clouds. Seamless integration of external processes with that of in-house processes is the key to success of hybrid clouds. To further add to the benefits of hybrid clouds is the scenario called Cloud bursting where in organization may access the public cloud infrastructure in high load conditions and prefer own resources in low load situations.

1.2 Characteristics of Cloud Computing

Cloud services express the following characteristics [21] that reveal their relation to and differences from conventional computing approaches {Figure 1.4}:

![Figure 1.4 Characteristics of Cloud Computing](image-url)
• **On-Demand Self-Service:** The Cloud services are invoked only when users need them such as server time and network storage, as required automatically without requiring human contact with each service provider.

• **Customization:** A cloud is a dynamically reconfigurable that best adapts to client’s requirements [22].

• **Reliability:** Reliability in terms of virtualized resources and replicated infrastructures is achieved in cloud computing with least efforts. High reliability makes the cloud a perfect solution for disaster recovery and business critical tasks.

• **Resource Pooling:** Resource pooling is a cloud computing environment to depict a situation in which cloud providers serve multiple clients with conditional and scalable services. Different physical and virtual resources are animatedly assigned and reassigned according to user demand. There is a sense of location liberty in that the client generally has no control or knowledge over the exact location of the provided resources.

• **Virtualization:** Virtualization refers to the abstraction of computer resources, i.e., a technique for hiding physical aspect of resources from user’s applications and interactions. It is a technique which multiplexes the hardware resources and enables the users to run multiple operating systems on same physical hardware. This technology helps in slicing a single data centre or high power server to act as multiple machines.

• **Rapid Elasticity:** Rapid Elasticity is the ability to provide scalable services. Capabilities can be elastically provisioned and released according to user specific service requirement [23].

• **Measured/Metered Services:** Due to affordable nature of cloud, the cloud services can be billed as pay-per-use model. The cloud provider can calculate storage levels, processing, bandwidth and the various user accounts which can be billed accordingly [24].
• **Multi-tenancy**: This feature allows cloud providers to save energy requirements by means of hosting applications from multiple clients on to the same infrastructure by sharing of resources [25].

### 1.3 Benefits of Cloud Computing

Cloud computing has been most advanced form of distributed and parallel computing. The applications of cloud computing in various domains offer unique benefits which are not available in other platforms. This section briefly presents the benefits of cloud computing [26].

**Reduction of Cost**

Cloud platform offers hosting services for business entities. Businesses may choose cloud platform for hosting their information and websites etc. Cloud platform allows users to choose from varied degrees of QoSs and thus pay according to subscriptions. The hardware costs of businesses get reduced as there is no need to buy hardware to host the information. It may however seem difficult scenario from user’s data privacy and security, but security provision in cloud is being worked upon and will improve to expected levels.

**Universal Access**

Access to resources and information by employees requires only an internet connection. With universal access to contents on clouds, employee may enjoy work from home and helps employers to extract best from their employees.

**Software Updates**

A Cloud service provider may seek to update on its software on the basis of feedbacks provided by users for existing or old releases. This will help users to work in latest environments with incurring any extra costs on updates and new releases.
Choosing an Application

Any business or enterprise which is using cloud platform for hosting their websites or other applications for development may choose different platform or software from what is available from cloud service providers. The fast implementation by cloud provider and pro data billing strategy allows users to go with their choices instantly.

Green and Mean Computing

The onsite deployment of infrastructure and hardware requires much more energy than needed in clouds. Clouds resources are shared by users and thus energy costing is reduced. This not only helps in reduction of carbon footprints but also reduces energy expensases.

Flexibility

Option to switch from one application to another or one software platform to another as may seem appropriate from user’s perspective is best exemplified by cloud platform. This is best form of flexibility example useable and available.

Elasticity

The scaling of resources up or down as per user requirement is desirable characteristic of clouds. The sudden rise in resource demand is provided by cloud provider without any investment from user. The use of cloud as SaaS by Taxation Department for example, may see rise in resources near the end of financial years and reduced demands during remaining times. This can be easily handled in cloud environment without up-front investment by enterprise.

Pay-As-You-Go

The requirements of users may be scaled up or scaled down. The billing of the user depends upon the duration and scale. Thus users are charged as per their usage. Scaling down of resources will reduce charging rates immediately. Further as any business may
see altogether different hardware and software requirement during their journey, so user will be able to resources as per their needs. Investment by enterprises in resources in prevented and that too at much smaller subscription charge.

1.4 Application of Cloud Computing

- **Test and Development**: The best use case of cloud is its capability for being used as development environment. Further, testing of developed apps can also be done in real-time. It opens up altogether a new platform where developing tools for a wide variety of applications are made available. Setting up hardware resources, employing manpower and investing time and money, all curtailed with use of online accessible development framework. To further ease the users, cloud platforms provided tailor made environments which can be further customized at one’s fingertips. Figure 1.5 represents one scenario for use of cloud as development platform [27].

![Figure 1.5: Applications of Cloud computing [28-33]](image)

- **Big data analytics**: Another major benefit experienced by enterprises is leveraging cloud computing for tapping tremendous volumes of structured and unstructured
data and extract information of business value. M-commerce and E-Commerce companies’ rope in cloud computing platform for migrating huge data volumes and estimating future trends and behaviour of potential buyers and using it customize business policies and procedures[27] {Figure 1.5}.

- **File Storage**: Users can store their large number of files and retrieve them through web enabled interface. Simplifies web enabled interfaces ensured customized experience for cloud user by offering different views and protection as per client’s requirements. Only enough storage is allocated to accommodate client’s requirements. Storage management procedures and consolidation related responsibilities are delegated to service providers. Users are relieved from worries related to storage management, security and privacy of his data, besides high availability ensured by cloud supported storage services [27].

- **Disaster Recovery**: Another indirect benefit experienced through cloud environment is the assurance of fast recovery and improved Disaster Recovery (DR) due interconnected resources at several distinct sites than traditional DR sites having fixed assets and rigid procedures [34].

- **Backup**: Traditional backing up of data using tapes and derives and shifting them from point of origination of data to some remote physical location exposes media to any kind of disastrous acts like robbery, natural calamity or accident etc. Data stored at some physical location is again prone to similar kind of accidental issues. Availability of cloud resources for storage of huge data with ease of transfer through wire and that too with assured level of safety and isolation besides preventing accidental corruption of data. Cloud based backup is certainly a solution of current and future vision and offered most feasible solution for both personal and business and enterprises [34].

- **Online Software Library**: Licensing of each instance of software is practice in big corporations. It requires huge capital investments besides hardware and other infrastructure. Cloud computing offered metered service to charge the usage
organization wide instead of license wise pay out. It saved organization from getting strangulated in legal complexities of pirated software and helped in realization of online software library

- **Online Hardware**: Company’s infrastructure investments have been minimized as huge servers and data bases subscribed through cloud based web services. Physical space requirements have been reduced besides hardware or software upgrade investments.

- **Cloud Database**: Backend of web applications rely on data bases. Traditionally web developers managed and tuned their databases besides huge capital investments in Databases. Management and tuning of data bases requires expertise and involves a lot of complexities. Cloud databases offered a viable solution to web developers by offering a chance to access databases, managed and tuned for almost every requirement they specify. This whole is part of services. Rackspace is one popular example in this category of services [35].

- **CRM**: Customer Relationship Management (CRM) is core business activity related to customer satisfaction and revenue. Mission critical activities like CRM are popularly being migrated to cloud environments as most of the organizations don’t want to invest in CRM activities. Cloud environment offered a viable solution consider such critical activities which otherwise ignored or considered in forcing circumstances. CRM may become a blueprint for moving mission critical activities to cloud environment [35].

- **Email**: Moving personal and organization’s employee’s emails to cloud is another mission critical activity. Gmail and Hotmail are representative example in this scenario. Anytime and anywhere access to mails and even you are on move. One needs not to worry about the space usage storage scalability, security and privacy of the emails. Clients are free from the worries of up-time or downtime of the servers. Default fault tolerance through multiple copies of databases further supports the candidature of email on clouds.
- **Web Site Hosting**: Hosting websites is huge IT related activity, but diverts the efforts of developers from development to management related activities. At present, blogs, corporate websites, social sites all have been hosted on the cloud and allows developers effort to be directed in improving customer’s experience.

- **E-Commerce**: E-commerce and M-commerce pose spiked requirement due sudden increase in demands may be seasonal, festive or holiday based demand surge. Such traffic requires that website should remain available and that too with tolerable delays. Long delays while accessing websites may lead to customer denial to portal and may permanent customer migration to another seller. Cloud environment provides scalable environment in terms of resources which scale at peak times and quashed later automatically. Without much of capital investment and using metered services cloud environment is heaven for e-commerce and m-commerce.

### 1.5 Challenges in Cloud Computing

Feature oriented cloud environment offers services at user discretion. Such a huge ensemble of resources gives way to a large number of problems and challenges. In this section, some of the challenging issues [36] in cloud computing are summarized as:

- **Quality of Service (QoS)**: Cloud Service Providers (CSPs) need to ensure that adequate amount of resources are made available to fulfill the QoS requirements such as deadline, response time and budget constraints. These QoS requirements form the basis for SLAs (Service Level Agreements) and any violation will lead to penalty. Therefore, CSPs need to ensure that these violations are avoided or minimized by dynamically provisioning the right amount of resources in a timely manner [36].

- **Energy Efficiency**: Cloud runs on infrastructure that includes huge servers and communication and storage devices. This requires huge power expenses and emits carbon footprints. Energy aware scheduling and server consolidation are common approaches to address energy efficiency [37].
- **Security:** The sharing of resources is inherent property of cloud environments. Such environment may expose the data and increase the vulnerability of cloud based solution. Authentication, confidentiality and integrity are most common challenges that may interfere with the acceptability potential of cloud computing paradigm. To further add to the scenario is instances of Denial of service attacks due to explosive arrival of authenticated access requests. This may be an attack and ay be treated as an instance of scale up of resources. The cost of running the application and wastage of energy will increase [38].

- **VM Migration:** Virtual Machines are instantiated to address client’s requests. VMs may have to be migrated in order to achieve server consolidation but may have traded off with performance degradation due to migration [39].

- **Resource Allocation:** The requirement of any customer is documented in the form if Service Level Agreement (SLAs). Once an SLA has been agreed between client and service provider, it becomes provider’s responsibility to allocate the resources to meet the agreed level of service levels. This requires dynamic scale up or scale down of the resources to requests [40].

- **Resource Scheduling:** It becomes cloud provider responsibility to schedule the requests in such a manner that maximizes the utilization and reduce makespan, in order to extract maximum economic gains [41].

- **Management Capabilities:** The cloud management is most crucial and important. This area still poses potential for improvement though there are many cloud service provides available. The sudden scalability requirement need to be addressed as this area is still in its infancy [42].

- **Regulatory and Compliances:** The strict regulation on privacy of data and information about the citizen and businesses in the country requires that cloud service providers must set-up data centers and storage capability inside country boundaries. Such strict compliances and restrictions poses a potential threat in the
growth of cloud computing as most natural form of distributed and parallel computing [42].

1.6 Performance Improvement

The expectation of users in terms of performance and financially viable alternate can be met only if the internal management of resources ensures performance oriented approach. Cloud providers must therefore consider, resource allocation, Task Scheduling, Energy Efficiency and Load balancing through the use of capable approaches for scheduling, allocation, load balancing and Server consolidation through VM migration.

1.6.1. Resource Allocation and Task Scheduling

When a user submits his applications or tasks to clouds for processing and other requirements over internet, cloud management system allocates resources as per specified SLA and task requirements. This process is termed as Resource Allocation (RA) and it uses underline implementation called Resource Allocation Strategy (RAS). RAS is responsible for performance within cloud’s limits and specified user QoS requirements. The RAS deals with keeping the track of resources in cloud along with efficient mapping of tasks into VM and then mapping VMs onto physical resources through virtualization. This very process requires that RAS is able to meet current demands and keeps reserves of resources as per future estimations [43]. A lot of research work in the field of Resource Allocation is done in recent past [44-63] with goals like energy efficiency, cost optimization, makespan, utilization efficiency etc. In [47], authors proposed a user framework called Nephele to exploit the dynamic resource allocation mechanism and was compared with Hadoop, a well-established data processing framework. An application of Map-reduce to processing of heterogeneous datasets was explored in [48], especially w.r.t. relational and join operations. A Rule based Resource Manager was proposed in [49] to address the scalability of on-demand cloud service and reduce the operational costs. The performance of resource manager was evaluated. A new search algorithm was proposed in optimization module of Virtual
Design Advisor (VDA) [50]. The algorithm improved resource portioning and resource allocation characteristics in clouds. Algorithm used a hybrid of Greedy and PSO and thus named GPSO. The evaluation established that GPSO was able to recover from local minima much better than greedy approach. A Multi-Dimensional Resource Allocation (MDRA) was proposed in [63] to dynamically allocate the resources among running cloud applications in an effort to reduce the cost by reducing the number of nodes.

Task scheduling had been primary measure for performance improvements in cloud computing environments. Several representative works in recent past has established the importance of task scheduling in clouds. Min-Min, Max-Min and Max-Max techniques represent historic work in scheduling proposals. Several new combinations like Min-Max [64] were proposed and validated by comparative analysis with Min-Min and Max-Min. Such proposals improved load imbalance and results improved the utilization time and reduced average response time. Such solution may be classified as heuristic based solutions. Metaheuristic like PSO, ACO, Fire-Fly, and Swarm inspired approaches offered an improved class of algorithms. Several work based on PSO [65-75], ACO [76-82], Artificial Bee Colony (ABC) [83-89] and BAT [90-92] explored the application of metaheuristic in task-scheduling. Performance of metaheuristics in task scheduling established the superiority of such solutions. Genetic Algorithm (GA) [93-98] and Bio-inspired computing opens another dimension in task scheduling. GAs emulates biological function and generates new population; during each iteration. After few iterations GAs converge. The performance of GAs was challenged by metaheuristic in terms of convergence time and quality of solution. The solutions in this class strive to reduce makespan, improve utilization and energy consumption characteristics of cloud environments.

1.6.2 VM Migrations

VM migration offers yet another dimension in performance improvement in cloud computing. Proposals in [99-104] proposed several strategies for VM migration and selection of best VM in given set of VMs. In general VM migration helps in server/host
consolidation and improvement of energy characteristics in cloud environments. VM migration results in performance degredation. VM migration helps to reduce SLA violations and allows cloud customers service providers to meet their objectives. An ACS based VM Consolidation (ACSVMC) [99] approach finds a near-optimal solution that reduces energy consumption and number of VM migrations. Decentralized VM migration [100] based on two-threshold, resource aware VM Migration[101] to meet resource scarcity by admission control and scheduling schemes, heuristic based VM allocation for Minimization of Migrations [102], metaheuristic combined with uniform random initialization, the binary search and Boltzmann selection policy (PS-ABC) [103] to achieve an improved ABC-based approach for power saving and novel adaptive heuristics inspired scheme for dynamic consolidation of VMs based on an analysis of historical data from the resource usage by VMs are representative examples in VM migration scope. Most of the approaches results in server or host consolidation and meeting the SLA requirements.

1.6.3 Energy Efficiency

Energy efficiency is misleading term in itself as it portrays a big canvas including VM migration, task scheduling, load balancing, server consolidation through switching off under-utilized servers etc. For example Information and Communication Technologies (ICT) is being implemented on cloud platforms, but it has energy efficient implementation internally in terms of execution, scheduling etc.

As it is clear from context that servers consume a lot of energy and emits carbon footprints. To reduce the energy consumption and reduce environments hazards is another deriving force in the arena of performance tuning. Several solutions [105-110] in recent times explored new approaches for enhancing the energy efficiency. Combining task allocation with resource provisioning and energy efficient task scheduling [105] and using cloud’s elasticity with task’s real time guarantee, VM packing based approach like Low Perturbation aware Bin Packing (LPBP) & Power and Computing capacity Aware- Best Fit Decreasing (PCA-BFD) algorithms [106] to improve upon energy efficiency and network load balancing, energy aware task
scheduling for heterogeneous environments (EHEFT) [107], Three methodologies VM reconfiguration, VM migration, PM power management to minimize energy consumption [108], GA based energy efficient approaches, Shuffled Frog-Leaping Algorithm (SFLA), Energy and budget constraints based approaches [109], and data replication in geographically distributed cloud computing data centers [110] presents few most promising energy efficient solutions.

1.7 Contributions

Cloud computing offered a new paradigm of computation for growing needs of rented computation, storage and networking requirements. Citing the challenges in cloud computing and evaluation of existing works exposes gaps in performance tuning measures. Considering that Scheduling of tasks, Migration of VMs and Energy efficient solutions have direct impact on the performance of clouds; we addressed several issues related to performance tuning in clouds.

1.7.1 Motivations

1. **Scheduling of Independent Tasks:** Most of the task scheduling schemes relies on mathematical solution which may result in maximizing and minimizing one or more parameters but leave several issues unaddressed.

2. **Workflow Scheduling:** As most of the real time scheduling in industrial environment occurs in inter-task dependency way. Such environment is best represented through workflows. Any scheduling scheme may apply only to independent tasks or workflows. No solution simultaneously handles both independent tasks and workflow scheduling.

3. **Task Types based Solutions:** Most of the scheduling schemes consider one kind of tasks. In majority only CPU bound tasks or I/O based tasks are considered. No single proposal addressed scheduling with an objective of multiple task types.
4. Planning Algorithm: Planning algorithm are similar to scheduling algorithm with global optimization. Only few well known planning algorithms are proposed in literature. Most of planning algorithms achieved global optimization and failed to exploit local optimization. Another major failure comes from the fact that planning algorithm can’t approximate ready queue to better precision.

1.7.2 Overview of Contributions

- A review on task scheduling taxonomies was proposed. This work identifies a new taxonomy based on Goal Oriented Task Scheduling (GOTS) and Constrained Oriented Task Scheduling (COTS). A thorough review was presented with features, limitation and scope of latest work in task scheduling.

- A VM migration scheme based on minimizing variance of remaining loads on hosts after VM migration was proposed. Scheme converges on best VM Selection (Var_Sel) for migration. Var_Sel was compared and proved betters than several existing VM migration policies under various parameters besides minimizing VM migration and reduced instances of SLA violations. Also an extended version of this work was carried by roping in more performance parameters. Var_Sel proved to be an energy efficient procedure in cloud computing.

- A variance aware scheduling scheme for workflows was proposed called Nearest Neighbor (NN). NN was then used to obtain hybrid variants of several exiting independent tasks scheduling schemes. This work proposed hybrid variants of Minimum Completion Time (MCT), NN_MaxMin, NN_MinMin. The hybrid versions of legacy scheduling schemes were improved in performance than their original variants.

- A cost aware hybrid meta-heuristic based scheduling was proposed. Considering PSO [65] and Cost Aware PSO [67] as base variants, NNCA_PSO was proposed. Comparing PSO, CA_PSO and NNCA_PSO on makespan, energy
efficiency and utilization etc. NNCA_PSO proved better than existing hybrid PSO variant.

- A planning algorithm called Robust HEFT (RHEFT) was proposed for workflows by using an Interior Scheduling (IS) approach. Tasks in workflow were ranked and grouped into sets of independent tasks. Set of independent tasks were then subjected to scheduling using IS approach. The comparison of RHEFT with HEFT and DHEFT was performed and proved better than both in terms of energy, makespan and utilization etc.

1.8 Organization of Thesis

Research work of this thesis has been organized into 7 chapters. Chapter 1 introduces cloud computing and its challenges and applications. A brief background on research work in performance tuning of clouds is also presented. Chapter 2 presents a detailed literature Review on scheduling, VM Migration and Energy Efficiency. Chapter 3 presents a new taxonomy on Task Scheduling. Chapter 4 presents VM migration approach proposed during this research work. Chapter 5 presents a Hybrid metaheuristic based scheduling schemes for set of independent tasks. Chapter 6 describes workflow scheduling implementations. Chapter 7 explores in Novel Robust Planning algorithm carried out during this research work. Finally Chapter 8 presents the conclusion and future aspects of this research work.

1.9 Summary

In this chapter, a brief overview of cloud computing, its application and challenges have been presented. This chapter elaborates the brief of literature in open challenges of cloud computing. Further Chapter presents contributions of the work and presents the organization of the thesis. Chapter 2 presents related work in the field of VM Migrations, Load Balancing and Task Scheduling with a brief list of gaps identified.