Chapter 2

Literature Review

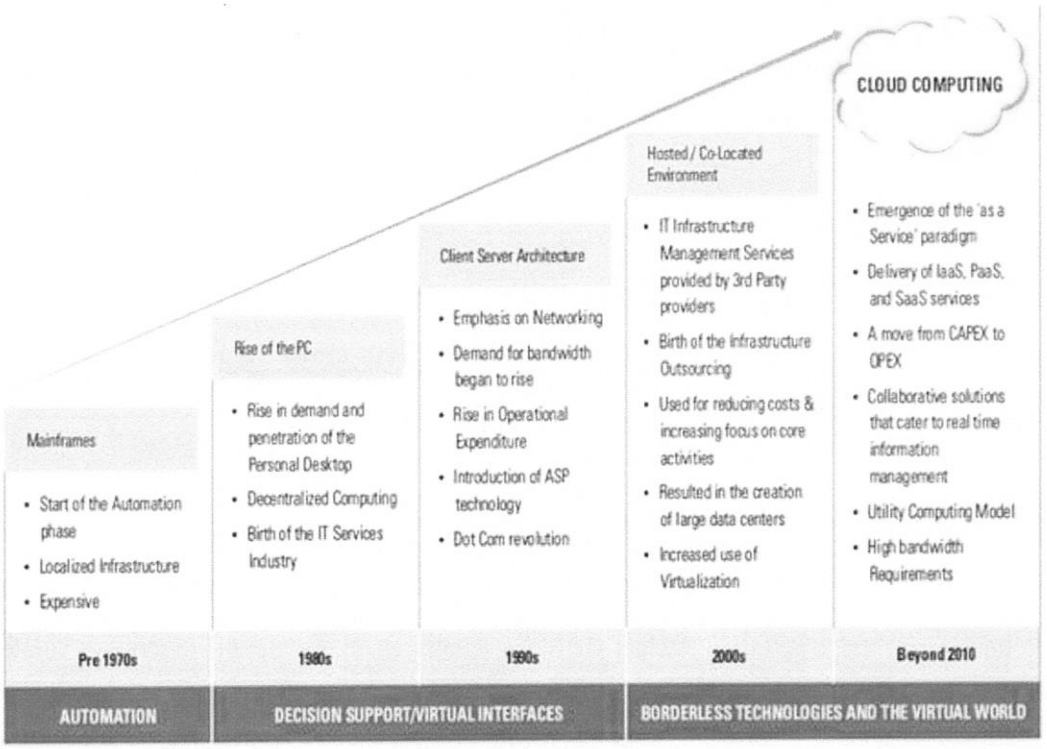
2.1 Introduction

The study primarily centres on the core theme of review of cloud computing ecosystem in India with a focused approach to understand the requirements of SMEs with respect to the Service Level Agreement. However, to fully understand the ecosystem in Indian context, it was deemed necessary to first understand the evolution of cloud computing from the traditional IT processes, related literature discussions, key issues and ongoing works with special reference to India. There is large number of literature available on cloud computing but the same is not collated at one point in the Indian context. This study attempts to present the knowledge together with a view to propose measures for faster adoption of cloud by SMEs.

2.2 Evolution of Cloud Computing

The Cloud computing literature is more than 50 years old, as the concepts were recognized as early as the 1950s in the work done by AT&T in the area of telephony networking. At that time, AT&T had already begun to develop an architecture and system where data would be located centrally and accessed by businesses through redesigned telephones and an updated telephone network. While the service did not materialize, the concepts and advantages were understood and relentlessly pursued through to this day.

The pursuit of centralized, abstracted IT services progressed over the decades with the advent and adoption of technologies such as Internet Service Providers (ISP –where servers were located at the Internet access point), and Application Service/Infrastructure Providers (ASP – where infrastructure was rented to a customer at an offsite location, but used most of the time by the one, paying customer). Other IT services historically offered include Time Sharing Systems, Co-Location, Hosting, and Outsourcing. A diagrammatic evolution towards Cloud computing is given below.



(Source: KPMG's The Cloud: Changing the Business Ecosystem, 2011)

Figure 2.1: Cloud Computing Evolution

As with any evolution, the step from ASP to Cloud computing is subtle yet disruptively important. While ASPs managed the offsite infrastructure for a customer, they were bound to the concept that the infrastructure capacity was predetermined and inflexible; ASP customers were required to declare the quantity of compute and storage capacity needed up front. If the customer's computing needs grew or contracted, the hardware had to be scaled up or down with an associated delay and up-front investment.

IT outsourcing was an option being exercised by organizations and is still an alternative to Cloud migration and adoption. Dibbern, Goles, Hirschheim and Jayatilaka (2004) studied the impact of outsourcing and found that although it was beneficial to the organization at the beginning, outsourcing projects performed unsatisfactorily after going through several rounds of contracts. This led some organizations to take previously outsourced IT systems and services back in house as a result of unsatisfactory service levels, change in strategic direction or cost saving failure (Overby, 2003).

Pallis (2010) states that the year of birth of commercial Cloud Computing is 2009 when the cloud service providers started introducing browser-based enterprise applications.

Cloud computing has created the same paradigm shift as replacement of individual generators by the centralised electricity grid (Etro, 2011). This is exactly how the service providers and consumers do not necessarily reside within same proximity.

Some organizations use Cloud as an alternative to outsourcing their resources. However, Khajeh-Hosseini, Sommerville, Bogaerts and Teregowda (2011) explain that there is a key difference between Cloud Computing and IT outsourcing: self-service, scalability and pay-as-you-go model give clients more flexibility and control than traditional IT outsourcing.

Mahesh, Landry, Sridhar and Walsh (2011) emphasize that the framework provided by cloud computing is in the form of high quality leased IT resources instead of building the IT infrastructure from the scratch. Thus, the in-house versus cloud computing comparison is, analogous to the make or buy decision faced by organizations and is equivalent to outsourcing data centre operations.

2.3 Definitions and characteristics of Cloud computing

Several authors have given comprehensive definitions of cloud computing (Geelan 2009; Stanoevska-Slabeva, Wozniak and Ristol, 2010; Vaquero, Rodero-Merino, Caceres and Lindner, 2009; Grandison, Maximilien, Thorpe and Alba, 2010), however, Cloud computing generally refers to a computing paradigm whose foundation is the delivery of services and ICT assets, often denoted as XaaS (Everything as a Service). The term refers to the cloud-based resources and services provided over the Internet, with the most common examples, following the SPI model, being Software (SaaS), Platform (PaaS) and Infrastructure (IaaS) as a service.

As per the definition provided by Meel and Grance (2011) of the National Institute of Standards and Technology (NIST):

> *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 (On-demand self-service, Broad network access, Resource pooling, Rapid elasticity, and Measured service), three service models (Software as a Service, Platform as a Service, and Infrastructure as a Service), and four deployment models (Private Cloud, Community Cloud, Public Cloud, and Hybrid Cloud).*

Cloud computing is similar to an electricity grid, where resources like hardware, software, information are pooled and shared with the end-user via the internet, which is used as a medium of exchange (Li, Wang, Wu, Li and Wang, 2011).

Marston, Li, Bandyopadhyay, Zhang and Ghalsasi (2009) suggest that Cloud computing is an online service model by which hardware and software services are delivered to customers depending upon their requirements and pay as an operating expense without incurring high cost.

Wyld (2009) explains that cloud computing platforms are based on utility model that enhances the reliability, scalability, performance and need based configurability and all these capabilities are provided at relatively low costs as compared to the dedicated infrastructures.

Chen, Wills, Gilbert and Bacigalupo (2010) define Cloud Computing as a tower architecture where the virtualization layer sits directly on top of hardware resources and sustains high-level cloud services. It goes onto the IaaS, PaaS and SaaS layers. The IaaS layer provides an infrastructural abstraction for self-provisioning, controlling and management of virtualized resources. In PaaS,

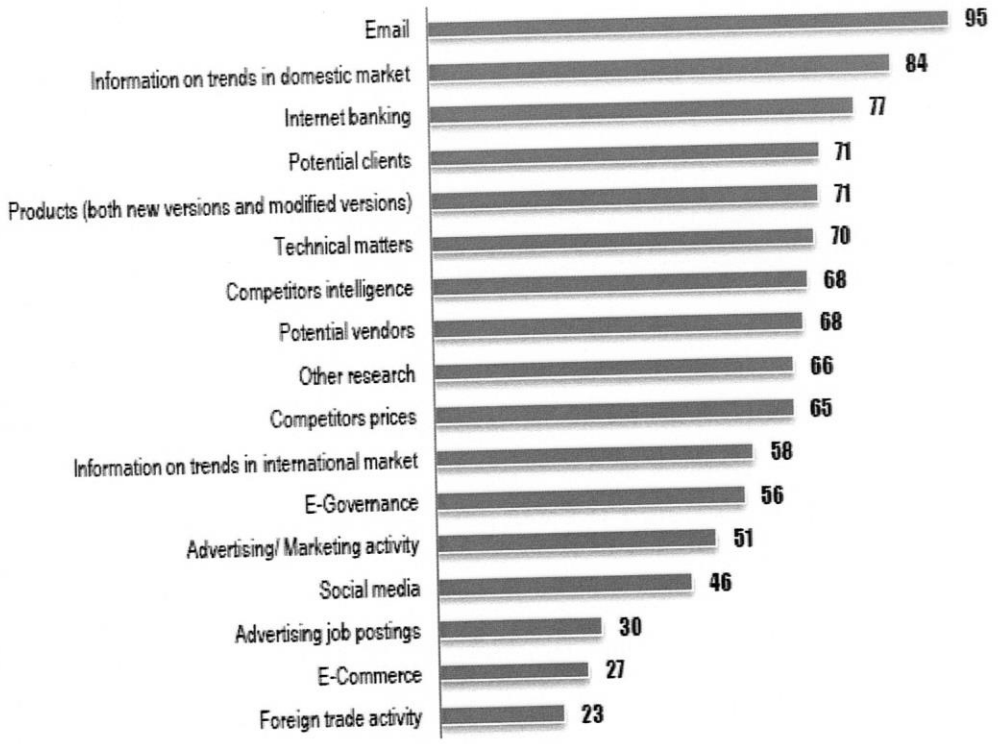
consumers may leverage the development platform to design, develop, build, and deploy cloud applications. The SaaS layer is the top of the cloud architectural tower and delivers specific applications as a service to end users. There is a self-managing cloud system for dynamic capacity planning which is underpinned by monitoring and accounting services. Capacity planning hides complex infrastructural management tasks from users by automatically scaling in and out virtualized resource instances in order to enforce established SLA commitments.

Computing Clouds are commonly classified into Public Clouds, Private Clouds and Hybrid Clouds (Boss, Malladi, Quan, Legregni and Hall, 2007; Marston, Li, Bandyopadhyay, Zhang and Ghalsasi, 2011). Cloud adoption is dependent on the type of Clouds and the intended use for the deployment. For small organizations that aim to save cost and test their software products before release, using public clouds is a good option (Khajeh-Hosseini, Greenwood and Sommerville, 2010). For organizations that have sensitive data and have data ownership and privacy concern, hosting private clouds is more suitable.

2.4 Advantages of cloud adoption for SMEs

Popli and Rao (2009) state that SMEs are said to be the lifeblood of any vibrant economy. They are known to be the silent drivers of a nation's economy. SMEs are leading the way for entering new global markets and for innovations in the emerging economic order. In India 95 percent of the industrial units are SMEs which give over 50 percent of the industrial output. Thus, SMEs form the backbone of the Indian economy.

A study by Nathan India, the Federation of Indian Chambers of Commerce and Industry (FICCI) and Google India (2013) points out the uses of Internet by SMEs as given below:



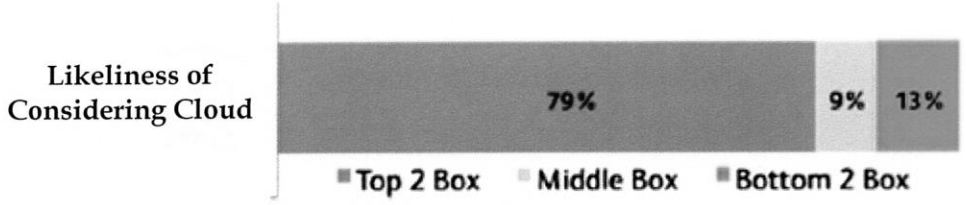
(Source: A study by Nathan India, FICCI, and Google India, 2013)

Figure 2.2: Uses of Internet (in percentage of SMEs)

According to a paper published by the Associated Chambers of Commerce and Industry of India SMEs sector is growing at a rate of 35 percent per annum and it will increase to 40 percent in the coming years (Assocham, 2009). As per the ASSOCHAM report, 60 percent of SMEs are moving towards the technology based infrastructure to increase their productivity with the reduction in their input cost. SMEs are one of the growing sectors and lucrative market places for the implementation of enterprise solutions. As the traditional in-house implementation of ERP solutions incurs high cost for the SMEs so it becomes a major constraint for them. By using and accessing services through the cloud, the companies can buy components relevant to their business on pay per basis instead of buying whole ERP suite (Sharif, 2009).

An Intuit study in collaboration with the Government of India's Ministry of Micro, Small and Medium Enterprises, the National Institute of Entrepreneurship and Small Business Development and the National Small Industries Corporation conducted on a five box pattern indicates that 79 percent SMEs are likely to consider cloud adoption. In the figure given below, 79 percent

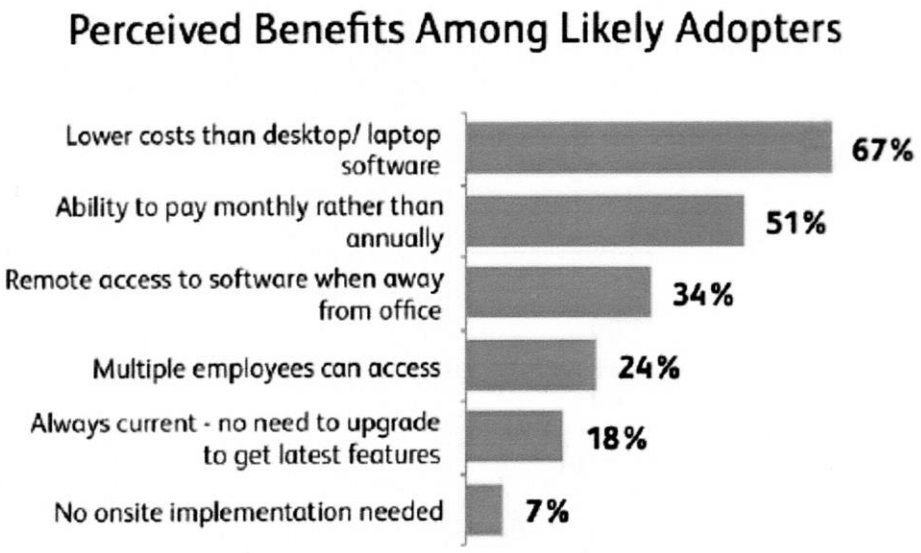
of the respondents indicated their interest for 'very likely' and 'likely' for considering cloud adoption which is shown as the total of top two boxes.



(Source: An Intuit study, 2013)

Figure 2.3: Likeliness of considering Cloud by SMEs

The perceived benefits among likely adopters were also diagnosed and are depicted below.



(Source: An Intuit study, 2013)

Figure 2.4: Perceived benefits among likely cloud adopters

Sharma, Mehra, Jola, Kumar, Misra and Tiwari (2010) state that traditional ERP systems involve higher level of difficulty in terms of adaptability than the cloud computing services.

Online services are better suited for small industries whereas large enterprises face more problems in implementation because of their complex functionalities and data security concerns (Dubey and Wagle, 2007). The requirements of SMEs and their budgets are generally small as compared to the offerings by the service providers of the traditional ERP solution providers (including the cost component) which lead to a wastage of resources on part of at least one of the stakeholder.

Rittinghouse and Ransome (2009) argue that by using the cloud computing environment the SMEs will not have to own the infrastructure so they can abstain from any capital expenditure and instead they can utilize the resources as a service and pay as per their usage of the resources provided by the cloud service provider.

Aggarwal and Barnes (2010) suggest that Cloud based services help the industries to reduce their cost that are involved in on premise ERP solutions such as hardware, software, up-gradation, training and licensing costs. Moreover long implementation cycles with regular maintenance costs adds to the total cost of traditional ERP which is saved by using cloud services.

Marston, Li, Bandyopadhyay, Zhang and Ghalsasi (2011) suggest that due to cloud computing, innovation is nurtured as the entry barrier in terms of cost gets lowered resulting in to that start-ups and SMEs can use cloud computing to introduce new online applications and social media services.

Creegar (2009) argues that one of the biggest advantages of moving to cloud computing is the opportunity cost of freeing up some of the IT administrative time, which can now be applied to the core business activity by the SMEs.

Durkee (2010) suggests that due to cloud computing adoption resulting in to stiff competition and availability of open resource software, the trend of software becoming a commodity like hardware has started. Downward pricing pressures have resulted in cloud services being used as a commodity now, hence large scale adoption of cloud computing has to be ensured, similar to volume sales but at a lower price. Elasticity in ramping up and disposing off cloud capacity when not needed and being extremely budget friendly, is an added advantage.

Krell (2011) suggests that SMEs can now afford applications such as ERP (Electronic Resource Planning), CRM (Customer Relationship Management), SFA (Sales Force Automation) and SCM (Supply Chain Management) due to economical subscription fees without huge upfront fees. Proliferation of smart phones has also improved collaboration within the companies using cloud computing.

Ankeny (2011) suggests that SMEs can move their business components to cloud through step implementation instead of instantaneous shift and growth in the cloud may happen at the pace of business growth resulting in to cost effective scaling and elasticity. Further, small business employees often work outside the actual office location and hence having ubiquitous access to their data using even the mobile devices is a big advantage.

Laura Abrar (2013) suggests that one of the most important advantages of cloud is that it enables employees to work from anywhere without being tied to an office desk, particular system, or office servers. In cloud model, data is stored in the data centres managed by cloud service providers, which can be accessed by even using a mobile device. It enables employees to work from anywhere and mine for real-time data, thus increasing productivity across the board. Most importantly, they can access customer contacts and details through mobile CRM applications and respond with agility.

Gupta, Seetharaman and Raj (2013) argue that for SMEs 'Ease of use and convenience' and 'Security and privacy' are considered to be the top two priorities for adoption of cloud followed by cost reduction or cost savings. A catalyst variable which needs immediate attention and can result in to faster adoption of cloud by SMEs is 'Reliability' which basically defines the data protection, processing and privacy issues.

2.5 Service Level Agreement: Requirement and Evolution

Goo (2010) argues that the role of a contract in Information Technology Outsourcing (ITO) is fulfilled by the Service Level Agreement (SLA), which describes the products or services to be delivered, sets service provider's and service recipient's expectations, identifies contacts for end-user problems, and specifies the metrics by which the effectiveness of various contracted services and lower-level activities, functions, and processes will be measured, examined, changed and controlled.

The literature on ITO often considers SLAs only as one of the constructs, besides for instance security and performance (Hofmann and Woods, 2010; Patel, Ranabahu and Sheth, 2009), size of outsourced IT (Misra and Mondal, 2011), or IT failure and performance degradation (Marques, Sauve, Antao and Moura, 2009), determining the success of outsourcing relationships.

Paquette, Jaeger and Wilson (2010) point out that in a period of less than 60 days, Apple MobilMe, Google Gmail, Citrix, and Amazon S3 all reported outages or periods of unavailability from 2 to 14 hours; in March 2009, Microsoft Windows Azure was down for 22 hours. This necessitates framing of terms and conditions to ensure services to the customers as per the agreed upon terms which are recorded in SLA.

Khaddaj (2014) explains that Service Level Agreement is a type of contract between providers and consumers, where providers guarantee to deliver service with a certain quality of services which are essential to users' business operations. SLAs management provides a facility to agree upon QoS between the users and providers, and to define user requirements and providers guarantees, thus assuring users that they are receiving the services they have requested and paid for, within a defined time frame.

TM Forum (2012) argues that a service level agreement is a formal negotiated agreement between the Service Provider (SP) and the Customer. It is

designed to create a common understanding about quality of service of cloud computing, priorities, responsibilities, etc. SLAs can cover many aspects of the relationship between the customer and the SP, such as performance of services, customer care, billing, service provisioning, etc. However, although a SLA can cover such aspects, agreement on the level of service is the primary purpose of a SLA.

A process oriented definition of service level agreement (ITIL, 2011) defines it as a formal, negotiated document that defines (or attempts to define) in quantitative (and perhaps qualitative) terms the service being offered to a Customer. Any metrics included in a SLA should be capable of being measured on a regular basis and the SLA should record by whom.

Service level agreement (SLA) is also explained as a contract between the service provider and the service consumer in which the expectations of the service provisioning is specified, including penalties that should be imposed as a remedial measure when a violation occurs (Begnum, Burgess, Jonassen and Fagerness, 2006; Greenwood, Vitaglione, Keller and Calisti, 2006; Kandukuri, Paturi and Rakshit, 2009; Yan, Zhang, Lin, Chhetri, Goh and Lowalczyk, 2006; Patel, Ranabahu and Sheth, 2009; Mahbub and Spanoudakis, 2010). It contains certain service level objectives (SLOs) and responsibilities and obligations of the parties that define objectively measurable conditions for the service, e.g. availability, throughput and response time. The terms and condition of SLA will vary depending on the requirements and applications or the data that are outsourced. SLA is also commonly used to address issues of problem management, legal compliance, resolution of disputes, customer duties, security and confidential information and termination. SLA is established by a negotiation process between both parties prior to service provisioning (Pichot, Waldrich, Ziegler and Wiedler, 2007) and can be re-negotiated if the requirements vary in the course of the currency of the agreement. Negotiation in this area is commonly defined as the process by which the involved parties come to a mutually acceptable agreement on some matter (Yaqub, Wieder, Kotsokalis, Mazza,

Pasquale, Rueda, Gomez and Chimeno, 2011). During negotiation, the service is designed to the consumer's needs and provider's capabilities to provide the requisite services (Chhetri, Lin, Goh, Zhang, Kowalczyk and Yan, 2006). Consumers usually want to obtain a high-quality service at low costs. Likewise, providers try to achieve the highest possible profit in line with demand, given their currently available Quality of Service (QoS) levels and capacities. The cloud provider will also need to evaluate its relationships and SLAs with vendors, enterprise data centres, network providers, and content providers. Finally, consumers and providers have to decide on a promising negotiation strategy respectively. Since both parties try to achieve the highest possible utility and, due to the business context, do not want to disclose too much private information (e.g., business goals, cost factors), a negotiation on the QoS parameters is necessary.

Cloud Standards Customer Council (2013) puts forth that SLAs are weighted heavily in the provider's favour, leading to the vendor's liability being limited. The burden is usually more likely on the consumers to recognize breaches of the SLA, notify their service provider and request a credit.

Nurika, Paputungan and Hassan (2014) argue that in addition, there are multiple competing providers in the market, which offer the same type of service, but with different properties. This situation is complicated for new or inexperienced users who just started to try exploiting this emerging technology. Therefore, this type of consumers wishes to have a helpful mechanism in negotiating SLA terms at ease. Likewise, it is also required to conduct concurrent negotiations with multiple providers from a customer's point of view in order to determine the highest possible profit. Further challenges arise in relation to the self-understanding of how good QoS values the customer should need. While most customers are willing to have the maximum service level, they in turn must acquire the necessary and acceptable QoS level or resources needed. Hence, besides the issue of negotiation, the baseline measurement on performance and resource usage prior to negotiation also needs to be considered. Due to the large

number of cloud providers and potential cloud consumers, the information exchange between the parties involved is very complex. Thus, a dynamic, scalable and automated approach is required for negotiating SLAs with multiple providers.

Garg, Versteeg and Buyya (2013) argue that because of the diversity of cloud service offerings, it becomes a challenge for the customers to identify the 'right' cloud provider who meets the requirements as generally there are trade-offs between different functional and non-functional requirements thereby resulting in to different service levels of different service providers.

Collins and Lam (2014) argue that the evaluation of services offered by competing cloud service providers is not trivial. Marketing information published by providers highlights potential benefits but offers little information which might assist potential customers in exercising informed judgements.

Comuzzi, Jacobs and Grefen (2013) argue that presently in Public Cloud solutions it is nearly impossible to change the characteristics of the service, as the services are standard and the corresponding SLAs are fixed and only to be agreed upon by the clients, who cannot negotiate them, thereby rendering them as unilateral documents.

The need for Cloud benchmarking is well established (Luszczek, 2011) and the Transaction Processing Council is working on a new benchmark for assessing transaction based applications on Cloud infrastructure and platforms (Nambier, 2013). Although there have been some experimental tools (Calheiros, 2010), these are complex and hence systematic evaluation of cloud service providers are beyond the reach of most of the SMEs.

McDonald, CEO CloudOne says that though 'Much good work has been completed on SLAs and the entire business model around the cloud, but much remains'.

Given the novelty and rapid evolution of cloud technology, contracting aspects in Cloud Computing have not yet received the same level of attention as in traditional ITO. The fragmentation of related work on SLA specification in cloud computing calls for more research in the area of SLA contract design (Marston, Li, Bandyopadhyay, Zhang and Ghalsasi, 2011).

Jin, Machiraju and Sahai (2002) provide an overview of ten components that an SLA should have in traditional outsourcing: purpose, parties, validity period, scope, service-level objectives, penalties, optional services, exclusions, restrictions, and administration.

Goo (2010) provides a more elaborate overview of eleven elements that should be included in a SLA divided among three categories: foundation elements, change management, and governance characteristics.

Buyya, Yeo, Venugopal, Broberg and Brandic (2009) introduced Service Level Agreement (SLA) led cost-saving models and explained how to calculate savings in detail. Buyya, Ranjan and Calheiros (2010) also demonstrate applications and services developed for Cloud Computing, and these services are helpful for start-up firms to generate additional revenues. Further to their work, Buyya, Beloglazov and Abawajy (2010) introduced a Return on Investment (ROI) power model which can calculate power cost-saving and present it using 3D visualization.

Khaddaj (2010) argues that the major requirements for supporting QoS Cloud are typical software quality factors and are mainly inherited from Web services. These can be listed as: availability, accessibility, reliability, integrity, performance, security and regulatory requirements.

Chung, Jeon and Seo (2014) suggest that quality evaluation items for cloud system have been deduced in five items, system performance, security performance, service performance, network performance and SLA compliance.

SLA compliance can be deducted as the item for measuring the status of SLA compliance provided to users by cloud service provider and may include uptime/downtime, service availability and backup (Jeon and Seo, 2013).