ELK ASIA PACIFIC JOURNAL OF COMPUTER SCIENCE AND INFORMATION SYSTEMS

Volume 1 Issue 1, September (2014)

ELK Asia Pacific Journals

www.elkjournals.com

CONCEPTS, PRINCIPLES AND IMPLEMENTATION OF CLOUD COMPUTING

Prof. M. Kamaraj
MES College of Engineering,
Kuttippuram, Kerala

ABSTRACT

Cloud computing is one of the recent emergences for the hosting and delivering services over the internet. It has and is constantly attracting business owners as eliminates the need for users to plan ahead for provisioning or buffer, and rather allows businesses to start from the small giving the flexibility to increase resources only when the demand rises. Despite the fact that cloud computing entails huge opportunities for the IT industry, it is still crawling at infancy, struggling with many unaddressed issues. This paper presents a survey of cloud computing with an attempt to highlight its key concepts, principles, implementation and research challenges.

Keywords: Cloud computing, Data centres, Hosting services, Virtualisation

1. Introduction

Fast improvements in the technologies of processing, followed by storage, and the development of Internet communication, resources of computing being cheaper, there are more powerful and ubiquitously availability noted than before. Such trends of technology enabled realization meant for newer model of computing known as cloud computing, whereby the resources (as storage and CPU) are of general use that get leased and further released by different users by means of internet with on-demand mode. In case of an environment with cloud computing, service provider gets distributed into two determined aspects. These are: *infrastructure providers*, managing cloud platforms and usage-based resources for pricing model; and *service providers* renting resources from single or diversified infrastructural sources in order to serve end users. Need for cloud computing led extreme impact over the Information Technology (or the IT) industry, the recent past, whereby large companies like Amazon, Google, and Microsoft are in the process of offering powerful, cost-efficient and reliable cloud platforms, as well as business enterprises trying to reshape respective business models for the attainment of newer paradigm. As a matter of fact, cloud computing offers many compelling features that are liable to make it attractive towards respective business owners, and noted below.

No up-front investment: the Cloud computing follows the model of pay-as you-go pricing. For a service provider there is no need for investing in infrastructure to get profits from cloud computing. The approach just rents resources from respective cloud as per the determined needs and further pay for usage.

Lowering operating cost: Resources under cloud environment gets rapidly allocated and even de-allocated as per demand. Thus, the service provider is never in the need of any capacities related to the peak load. As a result huge savings for the resources get released in order to save costs for operating as the service demand gets low.

Highly scalable: Respective infrastructure offers pool with larger resources from the centres of data and thus is easily accessible. For a service provider there can be an easy expansion of service towards large scales for handling rapid demands in service (like, flash-crowd effect), also termed as surge computing.

Easy access: Services that are hosted under cloud get web-based provisions in general. Thus, the same remains easily accessible by varied devices that have Internet connections. Such devices are inclusive of desktop as well as laptop computers, and even cell phones and the PDAs.

Reduced expenses on maintenance and business risks: Through the process of outsourcing determined service infrastructure towards clouds, respective service provider is liable to make a shift during business risks (like hardware failures) towards the infrastructure providers, who gain better expertise as well as better equipped aspects for managing risks. Moreover, service provider is open to get cut down maintenance of hardware and costs of staff training.

Still, though cloud computing comprises of considerable scopes towards IT industry; it further brings various challenges to be addressed specifically. The current paper, offers a survey over cloud computing, focussing basic concepts, principles of architecture, implementing state-of-the-art, and challenges faced during research. We aimed to get a better knowledge related to the design challenges meant for cloud computing and further recognise the importance of research directions for the selected topic.

Remainder of current paper gets organized in the following way. In Sect. 2 we offer overview about cloud computing and further compared it with some similar technologies. In Sect. 3, we illustrate cloud computing architecture and design principles. Detailed basic features of cloud computing are noted in Sect. 4. Section 5 initiates the surveys towards the commercial products and current technologies as implied towards cloud computing. Sect. 6 is the summation of current research topics related to cloud computing. Lastly, Sect. 7 is the conclusion.

2. Cloud Computing: An Overview

Here the general overview regarding cloud computing is presented along with definition as well as comparison with necessarily related concepts.

2.1 Definitions

Cloud computing is an old process of computing and dates back of 1960s, when John McCarthy envisioned facilities of computing offered for generalised public such as utility. “Cloud” gets implied for diversified contexts like illustrating large networking of ATM since 1990s. Still, as CEO Google, Eric Schmidt implied it for business model offering services in Internet since2006, and started gaining relevant popularity. From that time onwards, cloud computing gets utilised in marketing domain in diversified contexts for the representation of different ideas. There is a lack of standard definition related to cloud computing that has generated market hypes and remain sceptic as well as confusing. This is the reason that enough standardization towards the definition of cloud computing has been initiated. The work from 20 different definitions attained from different sources towards confirmation of standard definition. This current paper adopts cloud computing as per the definition of the National Institute of Standards and Technology (NIST), since it comprises our opinion, along with important aspects related to cloud computing:

Cloud Computing: NIST Definition:

> *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.*

Basic reason for the different perceptions related to cloud computing is the state of being unlike to other technical terms, not being a new technology, yet being a new operations model with current technologies in order to run business under diversified way. As a matter of fact, various technologies implied by cloud computing, like virtualization as well as pricing based on utility that are not new.

Rather, cloud computing influences current technologies in order to meet demands of technology and economic standards of IT.

2.2 Related Technologies

Cloud computing in general gets compared to the technologies noted below and each of these is liable to share determined features of cloud computing:

Grid Computing: it has been identified as a distributed paradigm of computing, coordinating resources from network for the attainment of common objectives of computation. Development related to this computing originates by the means of scientific applications that remain as computation-intensive. The process of cloud computing is same as grid computing and further employs distributed mode of resources for the attainment of objectives of application. Still, cloud computing adds the act of leveraging virtualization at multiple levels (platform of application and hardware) in order to realize sharing and dynamic provisioning of resources.

Utility Computing: it represents resource model on-demand as well as charging customer-based usage instead of flat rate. Cloud computing is liable to perceive in terms of utility realization over computing and adopts pricing scheme based on utility related to economic reasons. Following resources on-demand, the utility-based pricing, for the service providers gets highly maximized for the use of resource and minimizes costs of operation.

Virtualization: this is a kind of technology, abstracting physical hardware and further offers virtualized resources at the higher application levels. Virtualized server gets termed as virtual machine (or the VM). It forms the base for cloud computing, and offers pooling capabilities to the computing resources from respective clusters of determined servers and further dynamically assigning/reassigning virtual resources for on-demand application.

Autonomic Computing: it has been introduced by IBM, 2001, and creates computing systems with self-management features, that can assist in reacting towards external and internal observations without any intervention of human. The core aim is to overcome complexity of management in the current computer systems. Though cloud computing aims in exhibiting determined autonomic features, like provision for automatic resource, it still aims in lowering resource cost instead of reduction of complexity of the system.

To sum up, leverage of cloud computing is for virtualization technology for the attainment of offering computing resources in terms of utility. It further shares determined features with grid computing and then again with autonomic computing, yet stands away from each other. Thus, there are unique benefits added by imposed challenges for meeting necessary demands.

3. Architecture of Cloud Computing:

In this current section, illustrations about the architectural, operational and business models for cloud computing get initiated.

3.1 Cloud Computing: Layered Model

In general, architecture for the environment of cloud computing is subject to get divided in 4 layers. These are hardware/datacentre layer, infrastructure layer, platform layer and application layer, (see Fig. 1). We illustrate all of these in detail:

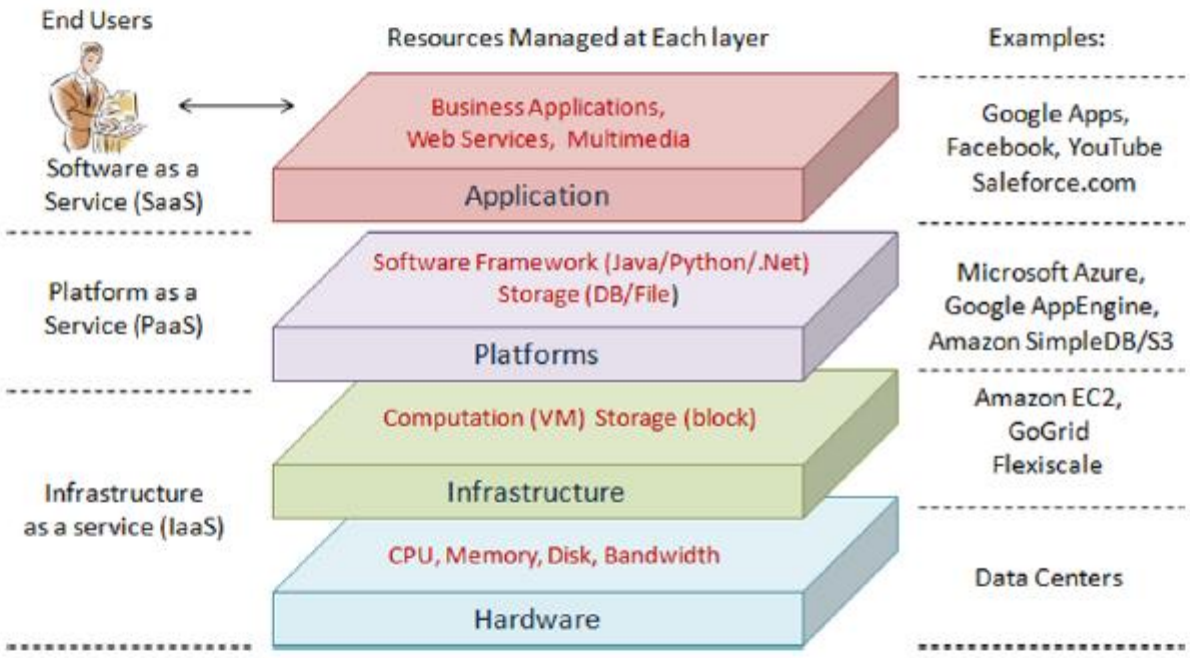
The hardware layer: it manages physical resources of cloud, added by physical servers, power, switches, routers and cooling systems. Practically, this layer gets implemented under data centres. Data centre in general comprises of thousands of servers, organized in interconnectivity by switches, racks, routers or some other fabrics. Typical concerns are related to the configuration of hardware, tolerance of fault, management of traffic, managing cooling and power resources.

The infrastructure layer: this is also called virtualization layer creating a storage pool as well as resource computing through the partitioning of the relevant physical resources through the implication of virtualization technologies like Xen [55], VMware [52] and KVM [30]. This is an important layer with necessary component related to cloud computing, as most of its basic features, like assignment of dynamic resource, are just made available by virtualization technologies.

The platform layer: it is created over infrastructure layer, and comprises operating systems as well as framework of application. The core purpose related to platform layer concentrates in minimizing the burden related to deployment of applications in a direct way into VM containers. As for instance, Google App Engine is liable to operate at platform layer for offering support of API towards the implementation of storage, business logic and database related to typical applications of web.

The application layer: it is noted as the highest level in the hierarchy, whereby the application layer comprises of actual applications of cloud. There are diversification from traditional applications, whereby applications of cloud can leverage feature of automatic-scaling for the attainment of better performance, lower cost of operation and availability. As against the environment of traditional-service hosting, like dedicated modes of server farms, cloud computing architecture is more modular. Every layer gets loosely coupled with the top and bottom layers to evolve separately. This stands same as OSI model design meant for network protocols. Modularity of the architecture permits cloud computing in terms of supporting wide-ranged application of the necessary demands, as reducing overhead maintenance and management activities.

Fig. 1 Architecture for Cloud Computing



3.2 Business model

Cloud computing is liable to employ service-driven model for any business. In a way, the hardware as well as resources from platform-level offer services as per demand. Thus, each of the layers of architecture is related to the former section and can be implied as service to aforementioned layer. As against this, each of the layers get perceived as per customer of layer noted below. Still, practically clouds offer services under three categories. These are –

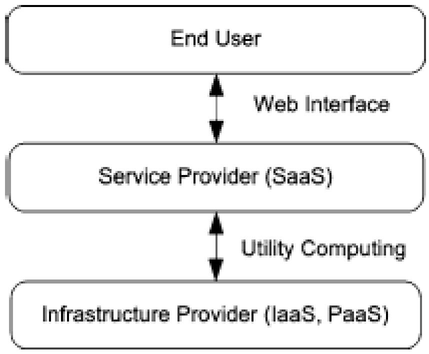
- software as a service (SaaS) : is about on-demand scope for infrastructural resources, especially for VMs. Owner of the cloud offering IaaS is an IaaS provider. As for instance, GoGrid [15], Amazon EC2 [2] and Flexiscale [18].
- platform as a service (PaaS) : is about offering platform layer resources, added by support for operating system support and frameworks for software development. As for instance, Google App Engine [20],

Force.com [41] and Microsoft Windows Azure [53]

- infrastructure as a service (IaaS): is about offering on demand applications related to Internet. As for insatance, Salesforce.com [41], SAP Business By Design [44] and Rackspace [17].

Model for business for cloud computing is in Fig. 2. As per layered architecture meant for cloud computing, there is every possibility that PaaS provider operates cloud over the top of cloud of IaaS provider. Still, current practice notes providers of IaaS and PaaS being integral to same organization (as Google and Salesforce). Thus, providers of PaaS and IaaS are usually termed as the *infrastructure providers* or the *cloud providers* [5].

Fig. 2 Cloud Computing: Model for Business



3.3 Types of clouds

Various issues related to cloud computing are noted as we moved the application into an enterprise level. As for instance, there are some service providers who mostly concentrate in lowering cost of operation, whereas other are into higher modes of security and reliability. Thus, there are various kinds of clouds, with respective benefits as well as drawbacks:

Public clouds: Here the service providers provide resources in terms of services towards public. Public clouds have many benefits towards service providers, and remains inclusive of primitive investment of capital over infrastructure as well as process of shifting infrastructural risks. Still, public clouds are without fine-grained control in terms of data, security and network settings, hampering effectiveness under varied business scenarios.

Private clouds: These are also termed as internal clouds, and are designed for the exclusive application of single organization. Private cloud can be built and further managed through organization or external providers. Private cloud provides highest controlling degree in the maintenance of performance, security and reliability. Still, there are criticisms for being same as traditional proprietary related to the server farms and for not offering scopes for up-front costs of capital.

Hybrid clouds: This is an amalgamation of models for public control and private cloud and attempting to meet the shortcomings in each approach. Hybrid cloud is part of service infrastructure actively related to private clouds, whereas rest of the part are in public clouds. These clouds comprises of more flexibility, against public and private clouds, and so offer tighter security and control in the process of implementing data against public clouds. There still remains the facility of expansion of on-demand service and contraction. Towards down side, design of hybrid cloud needs best split up between the private and public clouds.

Virtual Private Cloud: it has been noted as an alternative solution towards the process of addressing limitations as in the private and the private clouds. Virtual Private Cloud or VPC stands as the base running above public clouds and the basic difference as noted by VPC is the act to leverage the technology of virtual private network (VPN) allowing service providers with the scope to design their topology as well as settings of security like the firewall rules. VPC gets noted as a more holistic design as it virtualizes servers as well as applications, yet underlies communication network. Moreover, many companies get VPC with seamless transition led by proprietary service infrastructure towards cloud-based infrastructure relevantly owing virtualized layer of network. Many service providers, consider right model of cloud as per their business scenario. As for instance, scientific computation-intensive applications remain as the best possible deployment over public clouds to gain cost-effectiveness. On a more logical manner, there are some clouds that turn up more popular against others. There was this prediction about the hybrid clouds that remain dominant to most of the most organizations. Still, the virtual mode of private clouds started gaining more popularity from 2009 that is the year of inception.

4. Cloud Computing: Features

Salient features of cloud computing relies over various traditional computing services that can be summarize as following –

Multi-tenancy: services led by multiple providers in a cloud environment remain co-located under single data centre. Issues related to management and performance for such services gets a shared among the relevant service providers and respective infrastructure provider. Layered architecture related to cloud computing offers natural categorisation of responsibilities; where owner for each layer must concentrate on the goals of the associated layer. Still, multi-tenancy further introduces hurdles in terms of understanding as well as managing the interactions that are among different stakeholders.

Shared resource pooling: Infrastructure provider comes up with computing resources' pool, being dynamically assigned towards multiple resources of the consumers. These kinds of dynamic resources offer more flexibility towards infrastructure providers in the process of managing usage of resource and costs of operation. As for example, an IaaS provider is liable to leverage migration technology of VM in terms of attaining higher degree of server consolidation, thus maximizes usage of resource and minimizes cost related to consumption of power and cooling.

Geo-distribution and omnipresent network access: Clouds remain accessible by Internet connectivity delivered by the network server. Thus any kind of device having Internet connectivity, whether it is a mobile phone, PDA ora laptop, one can attain accessibility to the cloud services. Moreover, in order to attain higher performance by the network and localization, various clouds comprise of data centres from various locations all over the world. Service provider can leverage respective geo-diversity to attain maximum service utility.

Service oriented: aforementioned context of cloud computing notes service-driven model for operation, with serious emphasis over the act of managing service. A cloud having IaaS, PaaS and SaaS provider gives service as per Service Level Agreement (or the SLA) negotiated along with customers. Assurance of SLA thus remains complex aim of every provider.

Provisioning Dynamic Resource: cloud computing computes resources and the same is attained and released over fly. Against traditional model offering resources as per peak demand, provisioning of the dynamic resource permits service providers the way to attain resources in accordance to current demand that can remain lower with cost of operation.

Self-organizing: as resources get allocated/de-allocated on-demand, respective service providers get empowered with the capability to manage determined consumption of resource as per their needs. Moreover, management of automated resource feature yields high agility that enables service providers to respond quickly to rapid changes in service demand such as the flash crowd effect.

Utility-based pricing: Cloud computing is keen in employing pay per-use model for pricing. Exact pricing scheme can remain somewhat different from one service to another. As for instance, SaaS provider might need to rent virtual machine from IaaS provider on the basis of per-hour. Further SaaS provider offering on-demand CRM (customer relationship management) can charge customers as per the clients it is serving (like, Sales force). Pricing based on utility lowers cost of service operation as the same charges customers as per-use. Still, there are complexities in terms of controlling cost for operation. Thus, companies like V. Kernel [51] offer software in order to assist cloud customers know, analyze and further cut down significant cost over the consumption of resource.

5. State-of-the-art

This section is about the state-of-the-art related to cloud computing. Initially, the basic technologies for cloud computing get

Volume 1 Issue 1, September (2014)

ELK Asia Pacific Journals

www.elkjournals.com

noted. We further surveyed popular products of cloud computing.

5.1 Cloud computing technologies

In this section reviews on implied technologies about cloud computing environment get presented.

5.1.1 Data centres' *Architectural Design*: Data centre is a home for storage and power of computation and remains at the centre of cloud computing with 1000s of devices, such as switches, servers and routers. Right kind of planning for respective network architecture turns up critical, as the same heavily lay impact on performance application like environment of distributed computing. Moreover, resiliency and scalability must get well. In the current scenario, layered approach remains as the base for network architecture. This has been well analysed and tested in selected largest deployed centres of data. Core data layers comprise core, aggregation, followed by access layers (see Fig. 3). The access layer servers in racks respond to physical connectivity of the network. There are basically 20-40 servers/rack, where each gets connected to switch for accessibility with 1 Gbps link. Access switches in general are connected to two aggregation switches in order to get redundancy with 10 Gbps links.

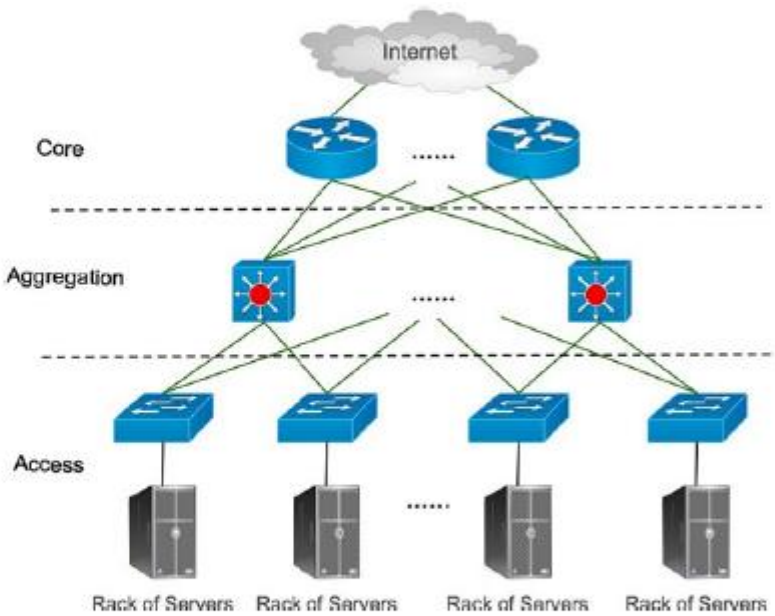


Fig. 3 Basic-layered design of data-centre network infrastructure10 Gbps links

Aggregation layer, in general offers basic functions, like services for domain, location, server-load balancing, etc. Core layer offers connectivity towards switches of multiple aggregations and adds resilient routed fabric without any single-point failure. Core routers are liable to manage traffic in relation with data centre. Common practice is that of leverage commodity noted for the activity of Ethernet switches as well as routers towards the creation of network infrastructure. Indifferent modes of business solutions, along with infrastructure layered network can get illustrated in reference to meet determined challenges of the business.

Network architecture for a data centre must comprise of [1, 21–23, 35]:

Uniform high capacity: Maximum server to-server traffic flow rate needs to get limited through the available networking interface cards' capacity for the purpose of sending as well as receiving servers, added by assigning servers towards respective service that needs to remain independent over network topology. There is the possibility for arbitrary host in communication relation with data centre with other networking host at full bandwidth in the local network interface.

Free VM migration: the process of virtualization permits entire state of VM in terms of being transmitted over migration of networking towards VM migration from a physical machine to other. Hosting of cloud computing service can migrate VMs in order to get statistical multiplexing or otherwise attain changing patterns of communication for the attainment of high bandwidth towards tightly synchronised hosts or get variable distribution of heat and availability of power at data centre. Communication topology needs to get designed for supporting rapid migration of virtual machine.

Resiliency: Instances of failures remains at common at scale, whereby the network infrastructure need to remain tolerant to fault against different kinds of failures in servers, outages of links, or failures in server-rack. Current unicast as well as multicast communications need to get affected towards the points of underlying physical connectivity.

Scalability: infrastructure of network needs to assess many servers and further permit incremental expansion.

Backward compatibility: Infrastructure of network must remain with backward compatibility, along with switches as well as routers that are running Ethernet as well as IP. As the current data centres get leveraged by commodity Ethernet as well as IP-based tools, there is the need to get used by newer architecture without any big modification. Another relevant area for the attainment of rapid innovation is design and deployment approach based on shipping-container, and noted as modular data centre (MDC). In general, MDC to fewer thousands relate to servers, get interconnected through switches towards the networking form of infrastructure. Applications of highly interactive aspects that remain sensitive to the time of response and are apt for geo-diverse MDC close to population areas. Moreover, MDC supports the redundancy approach in every area that can lose power, get earthquake, or can suffer from riots. Instead of 3-layered approach as mentioned above, where Guo et al. [22, 23] proposed the approach of server-centric, MDC under recursively defined network-structures.

5.1.2 Clouds' Distributed File System: GFS or the Google File System [19] stands as a proprietary mode of distributed file system initiated by Google and designed particularly for reliable and efficient accessibility of the data towards the usage of huge clusters of determined commodity servers. Various files gets categorised in different chunks that comprises of 64 megabytes, and further append to or otherwise read and remains extremely rare to the state of overwritten/shrunk status. As against the traditional mode of file systems, the approach of GFS is subject to run over the data centres in order to offer extremely higher inputs of data, along with lower latency as well as survival of individual failures of the server.

GFS has inspired open source for the Hadoop Distributed File System (or the HDFS) [24] storages related to larger files in the multiple machinery system. The approach attempts get reliability through the process of replicating data from multiple servers. In the same way, GFS stores data over multiple categories of geo-diverse nodes. File system has been created from cluster of respective data nodes that can serve blocks related to data over determined network through block protocol meant for HDFS. Data get collected over HTTP, and further permits accessibility to all the relevant content from respective web browser or some of the other clients. Data nodes can communicate to each other in order to rebalance distribution process and can even shift copies to necessary destination. It can keep replication of various high data.

5.1.3 Clouds' Distributed Application Framework:

Applications based TTP in general conform structure of web application like, Java EE. In response to the modern data centre, respective environments, servers' clusters, utilised for computation as well as jobs of data-intensive modes like, analysis of financial trend or animation of film. The software, MapReduce [16] initiated by Google is meant to support aspects of distributed computing over larger sets of data clusters over computers and comprises of one Master, where applications of the client get submitted to jobs of MapReduce. Master is liable to push work out towards available nodes of task within data centre, with the aim to keep tasks close to data. Master understands the content of node with data, that are related to other hosts being nearby. In case the task is not hosted over node, the stored data attains priority over nodes under same rack. Through this approach, network traffic over respective backbone gets reduced, and further assist in terms of developing all through, since the backbone remains as the bottleneck. In case of any failure of the task or times out, the same gets rescheduled. In case of failure of the Master, all tasks are subject to get lost. Records of Master as in the file system, as the same starts, it appears for any kind of data, for restarting respective work from the position the same has been left. Open source for Hadoop MapReduce project [25] is from by Google. In the current scenario, there are various organizations implementing Hadoop MapReduce towards the process of running large data-intensive in terms of computations.

5.2 Commercial products

This section is about the survey related to dominant products from cloud computing.

5.2.1 Amazon EC: The Amazon Web Services (or the AWS) [3] has been recognised as the set of cloud services that are liable to offer cloud-based computation, necessary amount of storage and all the other features that organizations need and are preferred by individuals, for the purpose of deploying services and applications over an on-demand aspect and at the rate of commodity prices. Offerings of AWS are accessible on HTTP, by REST and SOAP protocols. Amazon Elastic Compute Cloud (or the Amazon EC2) is liable to enable users of the cloud get launch and further manage instances of the server in data centres by the implication of APIs or necessary available tools along with utilities. EC2 remains as virtual machines that are been running over top of Xen virtualization engine [55]. As the creation, users are free to upload software and thereby initiate necessary changes to the same. As the changes get accomplished, the same can be bundled in the new image of the machine. There is an identical copy that is liable to get launched in due course of time. Users are almost with full control over the entire software stack related to the instances of EC2, resembling hardware. In contrast, this is an inherent feature that turns hard for Amazon in terms of offering automatic resources for scaling. EC2 offers the scope to place instances towards multiple locations that comprise Regions as well as Availability Zones. Regions with 1 or 2 Availability Zones get noted for geographically dispersed conditions. Availability Zones remain distinct to respective locations that get engineered to remain as insulated from failures over other Availability Zones. This further offers inexpensive, low-latency network connectivity towards other Availability Zones under the respective Region. Images of EC2 get stored and further retrieved from Amazon Simple Storage Service (or the Amazon S3), that stores necessary data as "*objects*" grouped in "*buckets*." Every object comprises 1 byte-5 gigabytes data. Objects are URI [6] pathnames. Buckets need to remain explicitly created before implication. This can also be stored in either 1 or many Regions. Users are open to make selection about Region towards the mode of optimizing latency, cost minimization, or address the demands of regulatory. Amazon's Virtual Private Cloud (or the VPC) is about securing and bridging the gap between current IT infrastructure as well as AWS cloud. Amazon's VPC is very effective in enabling enterprises towards the process of connecting current infrastructure towards isolated AWS resources of the computation through Virtual Private Network (or the VPN) connection, added by the extension of management scopes like firewalls, security services and detection systems with the inclusion of resources of AWS. In respect of the cloud users, respective Amazon Cloud Watch remains effective to the management tool that can collect raw data towards partnered AWS services like Amazon EC2 added by data processing approaches in the readable form, close to real-time metrics. Metrics are related to EC2 with aspects like CPU use, networking or out bytes, reading disk read or write operations, etc.

5.2.2 Microsoft Windows Azure platform: Windows Azure platform from Microsoft [53] comprises of 3 components, offering determined service sets of cloud users, offering Windows-related environment in order to run storing and application of data over servers within data centres; SQL Azure offers data services under SQL Server based on cloud and .NET Services under the note of distributed services of infrastructure towards cloud-based as well as local applications. Platform for Windows Azure are subject to get used through both running of applications within cloud as well as local systems. Windows Azure supports in building .NET Framework added by other ordinary languages under Windows systems, such as C#, Visual C++, Basic and others. It supports general programs, instead of single computing class. Developers are liable to create web-applications through the use of technologies like ASP.NET as well as Windows Communication Foundation (or the WCF), respective applications running over independent background or applications combining both. It allows the process of storing data within blobs, queue and tables, all having accessibility under RESTful style through HTTP/s. SQL Azure comprises of SQL Azure Database along with "Huron" Data Sync. It is created on Microsoft SQL Server, offering database management system (or the DBMS) within cloud. Data can get accessed through the usability of ADO.NET added by other Windows data with access of interfaces. Respective users can imply on-premises software towards work along with cloud based data. "*Huron*" Data Sync is liable to synchronize data over diversified on-premises DBMSs.

Table 1 A comparison of representative commercial products

Cloud Provider	Amazon EC2	Windows Azure	Google App Engine
Classes of Utility Computing	Infrastructure service	Platform service	Platform service
Target Applications	General-purpose applications	General-purpose Windows applications	Traditional web applications with supported framework
Computation	OS Level on a Xen Virtual Machine	Microsoft Common Language Runtime (CLR) VM; Predefined roles of app. instances	Predefined web application frameworks
Storage	Elastic Block Store; Amazon Simple Storage Service (S3); Amazon SimpleDB	Azure storage service and SQL Data Services	BigTable and MegaStore
Auto Scaling	Automatically changing the number of instances based on parameters that users specify	Automatic scaling based on application roles and a configuration file specified by users	Automatic Scaling which is transparent to users

Facility of .NET services for the creation related to the distributed applications. Component of Access Control offers a cloud-based approach related to verification of single identity over the applications as well as companies. Service Bus assists exposure to application of web services at the endpoints with access towards other kinds of applications, even if on-premises/cloud. Every exposed endpoint gets assigned to a URI that clients can imply to locate as well as access service. Every physical resources, application and VMs under the data centre get monitored by software, fabric-controller. By means of every application, users upload configuration file offering XML-based illustration of what is needed for the application. As per this file, fabric controller is subject to make decisions about the newer applications that must run, selecting physical servers towards optimized mode of hardware utilization.

5.2.3 Google App Engine: It GAE [20] has been noted as a platform in respect to traditional mode of web applications within Google-managed data centres. In the current scenario, languages for supported programming get noted as Python as well as Java.

Web frameworks based on Google App Engine remain inclusive of Django, Pylons, CherryPy, added by web2py, and customised Google-written web structure same as JSP or ASP.NET. Google is liable to handle deployment code towards cluster, process of monitoring, failure over as well as application of launching aspects as needed. Current mode of APIs are subject to support aspects like storing as well as retrieving data from respective Big Table [10] that remains non-relational database, creating HTTP requests along with caching. Determined developers are having read only accessibility towards file system over App Engine. The Table 1 sums up 3 examples related to popular offerings of the cloud for the utility of the classes for computing, types of target application, added by computation models, storage as well as auto-scaling. Such offerings of the cloud are on the basis of different abstract levels as well as resource management. Users are liable to select single or multiple type different types of cloud in order to offer satisfaction to the demands of the business.

6. Research challenges

Though cloud computing gets widely noted by industry, yet research over cloud computing is still in the initial stage. There are various issues that are not addressed fully, whereas newer challenges are still emerging from the applications of the industry. This section sums up some challenges related to the issues of research within cloud computing.

6.1 Automated service provisioning: This stands as a chief feature for cloud computing that is about the acquisition capability as well as releasing of resources on-demand. The core aim of service provider is about the allocation of de-allocate resources from cloud in terms of satisfying service level objectives (or SLOs), whereas minimization of coast of operation. Still, it is hard to detect the way service provider is attaining determined objectives. On a specialised note, it gets hard to determine the way map SLOs like QoS demands low-level resource as the CPU and demands of the memory. Moreover, in order to attain higher agility as well as respond towards rapid fluctuations of the demands like flash crowd effect, offering resource based decisions that are made online. Offering of automated service is never a novel issue. Dynamic resource for Internet applications gets researched extensively from past [47, 57]. These include:

(1) Constructing model for application performance with scopes to predict various applications that is needed to handle demand determined level for the satisfaction of QoS features;

(2) on a periodical basis prediction of future demand as well as the determination of demands for resource through performance model; and

(3) automatic process of allocating resources by the application of predicted demands of the resources.

Model for the application performance is liable to get structured through different approaches like Queuing theory [47], Statistical Machine Learning [7] and Control theory [28]. Moreover, distinct difference gets noted among the control of proactive and reactive resources. Proactive approach implies predicted demand towards periodically noted allocation of resources prior to the need. The reactive approach is liable to react towards the immediate fluctuations of demand prior to periodic demand-prediction. Both these approaches are effective resource for controlling dynamic operations.

6.2 Migration of Virtual Machine: it can offer benefits in cloud computing through migration of virtual machine to balance load in data centre. Moreover, it can enable robust as well as highly responsive data centres. It is from migration techniques [37] to which Xen [55] and VMW [52] get implemented for "live" migration towards VMs with extremely short down-times from tens of milliseconds to a second. As per Clark et al. [13] migrating is OS and its applications over one unit permits ignorance of many issues for process level migration and assess benefits related to VMs' live migration. Basic benefits of VM migration aim in avoiding hotspots.

Recently, detection of workload hotspots as well as initiation of migration lacks response for sudden workload changes. Moreover, memory state must get transferred efficiently and consistently with integrated physical servers and resource applications.

6.3 Consolidations of Server: It is effective for maximizing resource use and minimizing consumption of energy in cloud computing. Live migration of VM technology gets implied to consolidate VMs over multiple under-used servers over single server, for remaining servers for energy-saving. Optimal issue of consolidating servers get formulated as varied vector in issue of bin-packing [11], as NP-hard optimization issue. Many heuristics get proposed to respective problem [33, 46]. Moreover, VMs relies on communication demands [34]. Still, such activities should never hurt performance of the application, as resource usage (footprint [45]) of individual VMs in due course of time [54]. In case of server resources with shared VMs, like bandwidth, cache memory and disk I/O, max consolidated server leads to congestion as VM changes footprint over server [38]. Thus, fluctuations over VM footprints should be investigated and used for effectiveness. Lastly, system should be quick to react towards resource congestions [54].

6.4 Energy management: development of energy efficiency in cloud computing is for cost of powering with cooling accounts to 53% of total cost of operation in data centres [26]. In the year 2006, the US data centres consumed as much as 1.5% of total energy with growth of 18% per year [33]. Thus, infrastructure providers face hurdles to reduce consumption of energy. Aim is to cut cost of energy and meet regulations of government and standards of environment. Energy-efficient data centres are receiving enough attention and analysed from instances like, hardware for energy efficiency, enabling slowing of CPU speeds as well as turning off some of hardware components [8]. Energy-aware scheduling of job [50] and consolidation of server [46] are ways to reduce consumption of by turning-off unused machines. Latest research is also researching energy-efficient protocols for the network and infrastructures [27]. Basic challenge is to gain good trade-off among savings of performance and application performance, whereby many researchers are coordinating solutions for managing performance and power under dynamic environment of cloud [32].

6.5 Management of Traffic and Analysis: Data analysis for traffic has been noted as significant in reference to current data centres. As for instance, various web applications depend on traffic analysis to optimize experiences of customer. Network operators should know traffic flows for managing and planning decisions. Still, many challenges are for traffic analysis and measurement in Internet Service Providers (or ISPs) to extend towards data centres.

Firstly, links' density remains higher than ISPs/enterprise networks, making worst methods.

Secondly, most popular methods can compute matrices of traffic among few hundreds of send hosts, yet modular data centre consist thousand servers.

Lastly, current methods assume flow patterns with reasonable aspect in Internet as well as enterprising networks, yet applications get deployed over data centres, like jobs of MapReduce, changing traffic pattern.

Moreover, tighter application coupling of network, storage and computing is noted in other settings. Various measurement as well as analysis is needed for traffic of data centre. As per Greenberg et al. [21] features of data centre is related to the flow sizes as well as concurrent flows, added by the application of guiding network design.

As per Benson et al. [16] traffic at data centre undergoes relevant research for examining traces of SNMP from routers.

6.6 Data security: It is another significant topic of research in cloud computing. As the service providers usually do not have access towards physical security system related to data centres, they need to depend on provider of infrastructure in order to attain full mode of data security. Even in case of virtual mode of private cloud, service provider can just determine security setting on a remote basis, without understanding if the same is fully implemented. Infrastructure provider needs to follow:

(1) *confidentiality*, in order to secure accessibility towards data as well as transfer of the same, and

(2) *auditability*, in order to attest if setting of security for the applications as tampered or not.

Confidentiality is attained through the cryptographic protocols, and auditability is attained through the application of remote attestation approaches. Remote attestation demands for trusted platform module (or the TPM) in order to generate summation of non-forgeable system (which is system state encrypted through the application of TPM's private key) since the proof of system security. Still, a virtualized environment such as clouds, determined VMs are noted to remain dynamically migrating from a location to other, thus, directly implementing remote attestation remain insufficient. This is a complex way for the creation of trust mechanisms in each of the cloud's architectural layer.

Firstly, layer of hardware needs to remain trusted through the implication of hardware TPM.

Secondly, platform for virtualization needs to remain trusted through the application of secure mode of monitoring virtual machine [43].

Migration of VM needs to get allowed in case both the sources, with destination servers get trusted. In the recent work designing is noted for its efficient protocols towards the process of establishment of trust and management [31, 43].

6.7 Software frameworks: Cloud computing offers platform for compelling in terms of hosting larger applications of data-intensive that leverage MapReduce like Hadoop for scalable as well as fault-tolerant kind of data processing. Current work notes performance as well as consumption of resource related to the job of MapReduce with highly dependent application type [29, 42, 56]. As for example, Hadooptasks like I/O intensive, in place of requiring significant resources of CPU. Moreover, allocation of VM towards Hadoop node can have heterogeneous features. As for instance, availability of bandwidth towards VM remains dependent over other kinds of collocated VMs in relation with same server. Thus, it turns up to remain optimizing in context of cost and performance of application of MapReduce for a careful selection of right configuration values for the parameter [29] and further designing algorithms with more efficient scheduling [42, 56]. Through mitigation bottleneck resources, applications can develop. Basic challenges are –performance modelling, Hadoop jobs (online/ offline), and dynamically conditioned adaptive schedule. More argument is about structure of MapReduce energy-aware venture [50] that is to turn Hadoop node in sleep mode as job gets accomplished in the process of waiting newer assignments. For this, Hadoop and HDFS need to remain energy-aware. Moreover, trade-off gets noted between energy-awareness and performance. As per objective, desirable trade-off remains unexplored to the respective research

6.8 Storage and data management: structured software like MapReduce, by Hadoop and Dryad are to distribute data-intensive jobs. As mentioned above, this structure operates through Internet-scale file systems like HDFS and GFS. These files are different from traditional mode of distributed file systems within storage structure, pattern of access and application of determined programming interface. Further, they are not for the purpose of implementing standard POSIX interface, thus got compatibility with legacy file systems. Many researches dealt with this problem [4, 40], like [4] with proposed MapReduce by cluster file system, as GPFS of IBM. According to Patil et al. [40] newer API primitives remain constant for scalable as well as concurrent data access.

6.9 Architectures for novel cloud: in the current scenario, various commercial clouds get implemented for huge data centres and further get operated for centrally. Though design meets economy-of-scale added by high manageability, it has limitations like high-initial investment and high-energy expense for construction of data centres. In recent researches [12, 48] small sized data centres get advantageous over big data centres, whereby small centre never consume much power, thus does not need powerful mode, yet remains expensive in cooling system.

Small data centres remain cheaper towards the creation of and better mode of geographically troubled state than large data centres. Geo-diversity is desired for the response time in relation with critical services like delivery of content and interactive gaming. As for instance, Valancius et al. [48] researched feasibility related to hosting services for video-streaming with the application gateways (nano-data centres). Further research is for the application of voluntary resources (such as resources donated through end-users) for applications of hosting cloud [9]. Clouds create through the voluntary resources, or amalgamating voluntary resources and dedicated resources being cheaper towards the process of operating and remaining suitable towards non-profit applications, like scientific computing. Still, architecture imposes challenges like managing heterogeneous resources and recurrent churn events. Moreover, schemes for devising incentive related to architectures are open issues.

7. Conclusion

Cloud computing has emerged in the recent times and compels management and delivery of services through Internet. Development of cloud computing is changing on a rapid pace, whereby IT and utility of computing are turned into reality. Still, benefits from cloud computing as per current technologies are not noted as matured approach in terms of realising potentiality. Various challenges are like automatic provision of resource, management of power and security that are attaining importance in current research community. Thus, we considered scopes of ground breaking aspects of this field by noting impact of development towards respective industry. This paper surveys state-of-the-art related to cloud computing, and further covers necessary concepts, designs of architecture, prominent features, and basic technologies like research directions. Since the development in the field of cloud computing is still new, we consider this research can clarify many points related to challenges and lead to more research in future.

References

[1] Al-Fares M et al (2008) A scalable, commodity data center network architecture. In: Proc SIGCOMM

[2] Amazon Elastic Computing Cloud, aws.amazon.com/ec2

[3] Amazon Web Services, aws.amazon.com

[4] Ananthanarayanan R, Gupta K et al (2009) Cloud analytics: do we really need to reinvent the storage stack? In: Proc of HotCloud

[5] Armbrust M et al (2009) Above the clouds: a Berkeley view of cloud computing. UC Berkeley Technical Report

[6] Berners-Lee T, Fielding R, Masinter L (2005) RFC 3986: uniform resource identifier (URI): generic syntax, January 2005

[7] Bodik P et al (2009) Statistical machine learning makes automatic control practical for Internet datacenters. In: Proc HotCloud

[8] Brooks D et al (2000) Power-aware microarchitecture: design and modeling challenges for the next-generation microprocessors, IEEE Micro

[9] Chandra A et al (2009) Nebulas: using distributed voluntary resources to build clouds. In: Proc of HotCloud

[10] Chang F, Dean J et al (2006) Bigtable: a distributed storage system for structured data. In: Proc of OSDI

[11] Chekuri C, Khanna S (2004) On multi-dimensional packing problems. SIAM J Comput 33(4):837–851

[12] Church K et al (2008) On delivering embarrassingly distributed cloud services. In: Proc of HotNets

[13] Clark C, Fraser K, Hand S, Hansen JG, Jul E, Limpach C, Pratt I, Warfield A (2005) Live migration of virtual machines. In: Proc of NSDI

[14] Cloud Computing on Wikipedia, en.wikipedia.org/wiki/Cloudcomputing, 20 Dec 2009

[15] Cloud Hosting, CLoud Computing and Hybrid Infrastructure from GoGrid, http://www.gogrid.com

[16] Dean J, Ghemawat S (2004) MapReduce: simplified data processing on large clusters. In: Proc of OSDI

[17] Dedicated Server, Managed Hosting, Web Hosting by Rackspace Hosting, http://www.rackspace.com

[18] FlexiScale Cloud Comp and Hosting, www.flexiscale.com

[19] Ghemawat S, Gobioff H, Leung S-T (2003) The Google file system. In: Proc of SOSP, October 2003

[20] Google App Engine, URL http://code.google.com/appengine

[21] Greenberg A, Jain N et al (2009) VL2: a scalable and flexible data center network. In: Proc SIGCOMM

[22] Guo C et al (2008) DCell: a scalable and fault-tolerant network structure for data centers. In: Proc SIGCOMM

[23] Guo C, Lu G, Li D et al (2009) BCube: a high performance, server-centric network architecture for modular data centers. In: Proc SIGCOMM

[24] Hadoop Distributed File System, hadoop.apache.org/hdfs

[25] Hadoop MapReduce, hadoop.apache.org/mapreduce

[26] Hamilton J (2009) Cooperative expendable micro-slice servers (CEMS): low cost, low power servers for Internet-scale services In: Proc of CIDR

[27] IEEE P802.3az Energy Efficient Ethernet Task Force, www.ieee802.org/3/az

[28] Kalyvianaki E et al (2009) Self-adaptive and self-configured CPU resource provisioning for virtualized servers using Kalman filters. In: Proc of international conference on autonomic computing

[29] Kambatla K et al (2009) Towards optimizing Hadoop provisioning in the cloud. In: Proc of HotCloud

[30] Kernal Based Virtual Machine, www.linux-kvm.org/page/ MainPage

[31] Krautheim FJ (2009) Private virtual infrastructure for cloud computing. In: Proc of HotCloud

[32] Kumar S et al (2009) vManage: loosely coupled platform and virtualization management in data centers. In: Proc of international conference on cloud computing

[33] Li B et al (2009) EnaCloud: an energy-saving application live placement approach for cloud computing environments. In: Proc of international conf on cloud computing

[34] Meng X et al (2010) Improving the scalability of data center networks with traffic-aware virtual machine placement. In: Proc INFOCOM

[35] Mysore R et al (2009) PortLand: a scalable fault-tolerant layer 2 data center network fabric. In: Proc SIGCOMM

[36] NIST Definition of Cloud Computing v15, csrc.nist.gov/groups/SNS/cloud-computing/cloud-def-v15.doc

[37] Osman S, Subhraveti D et al (2002) The design and implementation of zap: a system for migrating computing environments. In: Proc of OSDI

[38] Padala P, Hou K-Y et al (2009) Automated control of multiple virtualized resources. In: Proc of EuroSys

[39] Parkhill D (1966) The challenge of the computer utility. Addison- Wesley, Reading

[40] Patil S et al (2009) In search of an API for scalable file systems: under the table or above it? HotCloud

[41] Salesforce CRM, http://www.salesforce.com/platform

[42] Sandholm T, Lai K (2009) MapReduce optimization using regulated dynamic prioritization. In: Proc of SIGMETRICS/Performance

[43] Santos N, Gummadi K, Rodrigues R (2009) Towards trusted cloud computing. In: Proc of HotCloud

[44] SAP Business ByDesign, www.sap.com/sme/solutions/business

management/businessbydesign/index.epx

[45] Sonnek J et al (2009) Virtual putty: reshaping the physical footprint of virtual machines. In: Proc of HotCloud

[46] Srikantaiah S et al (2008) Energy aware consolidation for cloud computing. In: Proc of HotPower

[47] Urgaonkar B et al (2005) Dynamic provisioning of multi-tier Internet applications. In: Proc of ICAC

[48] Valancius V, Laoutaris N et al (2009) Greening the Internet with nano data centers. In: Proc of CoNext

[49] Vaquero L, Rodero-Merino L, Caceres J, Lindner M (2009) A break in the clouds: towards a cloud definition. ACM SIGCOMM computer communications review

[50] Vasic N et al (2009) Making cluster applications energy-aware. In: Proc of automated ctrl for datacenters and clouds

[51] Virtualization Resource Chargeback, www.vkernel.com/products/EnterpriseChargebackVirtualAppliance

[52] VMWare ESX Server, www.vmware.com/products/esx

[53] Windows Azure, www.microsoft.com/azure

[54] Wood T et al (2007) Black-box and gray-box strategies for virtual machine migration. In: Proc of NSDI

[55] XenSource Inc, Xen, www.xensource.com

[56] ZahariaMet al (2009) ImprovingMapReduce performance in heterogeneous environments. In: Proc of HotCloud

[57] Zhang Q et al (2007) A regression-based analytic model for dynamic resource provisioning of multi-tier applications. In: Proc ICAC