A Review Paper for better Understanding of Cloud Computing Layered and Network Architecture with its Effect in Research Challenges

Vipin Kumar
Department of Computer Applications
KIET Group of Institutions, Ghaziabad
vipin.kumar@kiet.edu

Ankit Verma
Department of Computer Applications
KIET Group of Institutions, Ghaziabad
ankit.verma@kiet.edu

***Abstract*- Cloud computing has recently came into existence as a new technology for providing services like software as a service, platform as a service and infrastructure as a service with low cost over the Internet. Cloud computing is attractive to business owners as it eliminates the requirement for users to plan ahead for provisioning, and allows enterprises to start from the small and increase resources only when there is a rise in service demand. However, despite the fact that cloud computing offers huge opportunities to the IT industry, the development of cloud computing technology is currently at its infancy, with many issues still to be addressed. In this paper, I am giving a survey of cloud computing layered and network architecture with highlighting its key concepts, architectural principles, state-of-the-art implementation and it's in effect research challenges. The aim of this paper is to provide a better understanding of the design challenges of cloud computing and identify important research directions in this increasingly important area.**

***Keywords*- Cloud computing, Data centers, Virtualization.**

1. INTRODUCTION

Due to rapid growth in the field of processing and storage technologies and the huge success of the Internet, computing resources have become cheaper, more powerful and more ubiquitously available than ever before. This technological trend has enabled the recognition of a new computing model called cloud computing, in which resources (e.g., Mail, Application, CPU and Storage) are provided as general utilities that can be leased and released by users through the Internet in an on-demand fashion.
There are so many definition available for cloud computing, Here I am giving some of the main definitions offered by important organizations.

According to encyclopedia whatis.com defines cloud computing as:
"A general term for anything that involves delivering hosted services over the Internet."… "A cloud service has three distinct characteristics that differentiate it from traditional hosting. It is sold on demand, typically by the minute or the hour; it is elastic -- a user can have as much or as little of a service as they want at any given time; and the service is fully managed by the provider (the consumer needs nothing but a personal computer and Internet access)."
According to the IEEE (Institute of Electrical and Electronics Engineers) cloud computing is
"A paradigm in which information is permanently stored in servers on the internet and cached temporarily on clients"
According to the most famous encyclopedia in Internet, Wikipedia, cloud computing is: "Computation, software, data access, and storage services that do not require end-user knowledge of the physical location and configuration of the system that delivers the services".
According to the University of Berkeley, one of the most important technical universities over the world defines the term cloud computing as:
"Cloud Computing refers to both the applications delivered as services over the Internet and the hardware and systems software in the datacenters that provide those services"
So from all these definition we can summarize that cloud computing is a computational paradigm based on resource consumption, applications sharing, hardware utilization or computation, offered by internet and consumed under demand. The rapid development of cloud computing has made a tremendous impact on the Information Technology (IT) industry over the past few years, where large companies such as Google, Amazon and Microsoft strive to provide more powerful, reliable and cost-efficient cloud platforms, and business enterprises seek to reshape their business models to gain benefit from this new paradigm. Indeed, cloud computing provides several compelling

features that make it attractive to business owners, as shown below.

1.1 Less Initial Investment: Cloud computing uses a pay-as you-use pricing model. A new service provider does not need to invest in the networking infrastructure to start gaining benefit from it. It simply rents resources as per your requirement from the cloud and pay for the usage that is very less as compare to direct investment.

1.2 Very Low Operating Cost: Services or Resources in a cloud computing environment can be rapidly allocated and de-allocated on demand. Hence, a service provider no longer needs to pay as per the peak load. Less payment is given when resources demand or service load is low.

1.3 Highly Expendable or Scalable: Cloud computing provider's pool large amounts of resources that can be easily accessible or expand as per service provider needed. A service provider can easily expand its service to large scales in order to handle rapid increase in service demands.

1.4 Easy Access: Cloud computing services generally having web-based interface those are easily accessible through a variety of devices like desktop, laptop, cell phone and PDAs over Internet connections.

1.5 Low Business Risks and Maintenance Cost: In Cloud computing a service provider outsource services or resources due to it a service provider shifts its business risks (such as hardware or application failures) to infrastructure providers, who often have better expertise and are better equipped for managing these risks as compare to service provider. Due to this, a service provider can cut down the hardware maintenance and the staff training costs.

2. Types of Clouds

There are so many type of cloud available in commercial market, so it is not easy to select could services at once glace, it need to consider many issues when moving an enterprise application to the cloud environment. For example, some service providers prefer lowering operation cost, while others may prefer high reliability and security. Accordingly, there are different types of clouds, each with its own benefits and drawbacks:

2.1 Public clouds: It is type of cloud in which service providers offer resources as a services to the general public. Public clouds offer several key benefits to service providers, like no initial investment on infrastructure and putting all risks on infrastructure providers. However, public clouds provide lack of fine-grained control over data storage, network and security settings, which hampers their effectiveness in many business scenarios

2.2 Private clouds: Private cloud is also known as internal clouds, this internal cloud specially designed for exclusive use by a single organization within it premises. A private cloud may be development and managed by the organization itself or by external infrastructure providers. A private cloud offers the highest degree of control over performance, reliability and security. However, they are often criticized because of do not provide benefits such as no initial capital cost and for being similar to traditional proprietary server farms.

2.3 Hybrid clouds: It is a combination of private and public cloud models that overcome the limitations of both the approaches. In a hybrid cloud, services infrastructure runs in private clouds while the remaining part runs in public clouds. A hybrid cloud is more secure or flexibility than public and private clouds. Specifically, its provide tighter control and security over application data compared to public clouds, while still facilitating on-demand service expansion and contraction. On the down side, designing a hybrid cloud requires carefully determining the best split between public and private cloud components.

2.4 Virtual Private Cloud: An alternative solution for service provider to overcome the limitations of both public and private clouds is called Virtual Private Cloud (VPC). A VPC runs on top of public clouds. The main difference between VPC and public cloud is that a VPC leverages virtual private network (VPN) technology that allows service providers to design their own topology and security settings over public cloud such as firewall rules. VPC is essentially a more holistic design since it not only virtualizes servers and applications, but also the underlying communication network as well. Additionally, for most companies, VPC provides seamless transition from a proprietary service infrastructure to a cloud-based infrastructure, owing to the virtualized network layer.

3. Cloud Computing Architecture

Cloud Computing Architecture can be divided into 4 layers: the hardware layer, the infrastructure layer, the platform layer and the application layer, as shown in Fig. 1. We describe each of them in detail:

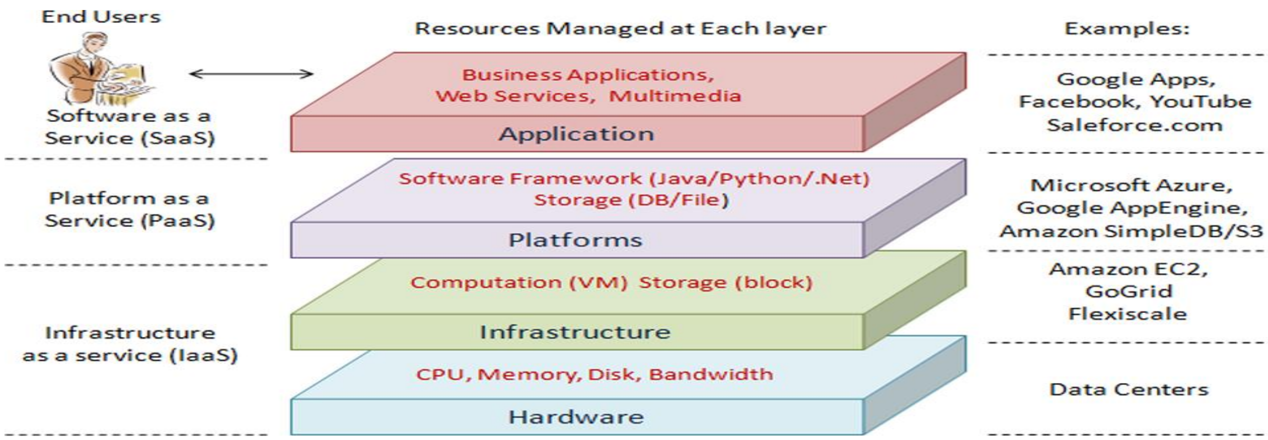

Fig. 1: Cloud Architecture

3.1 The hardware layer: This is the lowest layer of the cloud used to manage the physical resources of the cloud, including physical servers, routers, switches, power and cooling systems. In general, the hardware layer is implemented in data centers. A data center contains thousands of servers that are organized in racks and interconnected through switches, routers, cable or other fabrics. Typical issues rise at hardware layer are hardware configuration, Ault tolerance, traffic management, power and cooling resource management.

3.2 The infrastructure layer: This is the virtualization layer of the cloud; the infrastructure layer creates a pool of storage and computing resources by partitioning the physical resources using virtualization technologies available in the market like Xen, KVM and VMware. The infrastructure layer is an core component of cloud computing with many key features, such as dynamic resource assignment, are only made available through virtualization technologies.

3.3 The platform layer: On top of the infrastructure layer platform layer exists, this layer consists of operating systems and application frameworks. The purpose of the platform layer is to minimize the burden of deploying applications directly into VM container. For example: At the platform layer Google App provide API support for implementing storage, database and business logic for web applications.

3.4 The application layer: This is the highest layer of the cloud hierarchy, which consists of the actual cloud applications. Different from traditional applications, cloud applications can leverage the automatic-scaling feature to achieve better performance, availability and lower operating cost.
Compared to traditional service hosting environments such as dedicated server farms, the architecture of cloud computing is more modular. Each layer is loosely coupled with the layers above and below, allowing each layer to evolve separately. This is similar to the design of the OSI model for network protocols. The architectural modularity allows cloud computing to support a wide range of application requirements while reducing management and maintenance overhead.

4. CLOUD COMMERCIAL MODEL

Cloud commercial model is a service-driven business model. Cloud computing architecture provides three types of services software as a service, platform as a service and infrastructure as a service on-demand basis. Conceptually, every layer of the cloud computing architecture described in the previous section can be implemented as a service to the layer above. However, in practice, clouds offer services that can be grouped into three categories as mention below:

4.1 Infrastructure as a Service: IaaS provisions resources such as servers in term of virtual machines (VM)), network bandwidth, storage, and related tools necessary to build an application environment from scratch. The user has high level of usability and developers can still deal with low level details such as starting VMs or mapping static IP-Addresses to VMs. As a VM behaves almost similar to a physical server, virtually any web-application can be mapped to this type of service. As the management of these customizable environments is in most cases highly application-dependent, it is difficult for the providers to provide automatic scalability and fail over. The cloud owner who offers IaaS is called an IaaS provider. Examples of IaaS providers include Amazon EC2 [2], GoGrid [15] and lexiscale [18].

4.2 Platform as a Service: PaaS provides a high-level environment, a domain-specific platform, on which developers write customized applications (e.g, Google's App Engine is targeted exclusively at traditional web applications). The developer can focus on the main functionality of his application and program to a more or less open specification. After deployment, the service provider automatically takes care of maintenance, scale out and load balancing. Unfortunately, SaaS applications are often restricted to the platforms abilities and cannot be easily migrated to other providers. It would be difficult to start with a toy application on a platform provider and then switch to an infrastructure cloud when it matures. Examples of PaaS providers include Google App Engine [20], Microsoft Windows Azure and Force.com.

4.3 Software as a Service: SaaS refers to special-purpose software made available via the internet. These services are not suitable for building individual applications and are restricted to what the application is and can do. There is only little information published about the underlying technology. Examples of SaaS application are public email providers (Gmail, Hotmail, etc.), the Google apps, various search engines, etc.

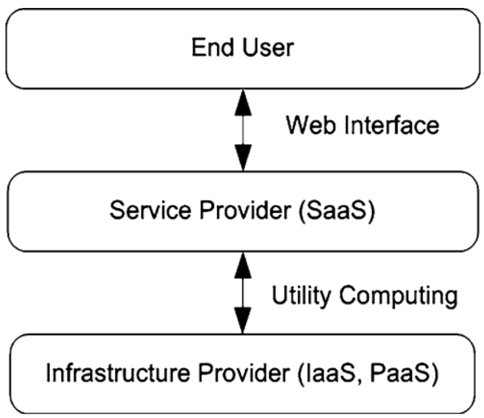

Fig. 2: Cloud Commercial Model

5. DATA CENTERS ARCHITECTURAL MODEL

A data center, which is home to the computation power and storage, is central to cloud computing and contains thousands of devices like servers, switches and routers. Proper planning of this network architecture is critical, as it will heavily influence applications performance and throughput in such a distributed computing environment. Further, scalability and resiliency features need to be carefully considered. Currently, a layered approach is the basic foundation of the network architecture design, which has been tested in some of the largest deployed data centers. The basic layers of a data center consist of the core, aggregation, and access layers, as shown in Fig. 3. The access layer is where the servers in racks physically connect to the network. There are typically 20 to 40 servers per rack, each connected to an access switch with a 1 Gbps link. Access switches usually connect to two aggregation switches for redundancy with 10 Gbps links. The aggregation layer usually provides important functions, such as domain service, location service, server load balancing, and more. The core layer provides connectivity to multiple aggregation switches and provides a resilient routed fabric with no single point of failure. The core routers manage traffic into and out of the data center. A popular practice is to leverage commodity Ethernet switches and routers to build the network infrastructure. In different business solutions, the layered network infrastructure can be elaborated to meet specific business challenges. Basically, the design of a data center network architecture should meet the following objectives [1, 21–23]:

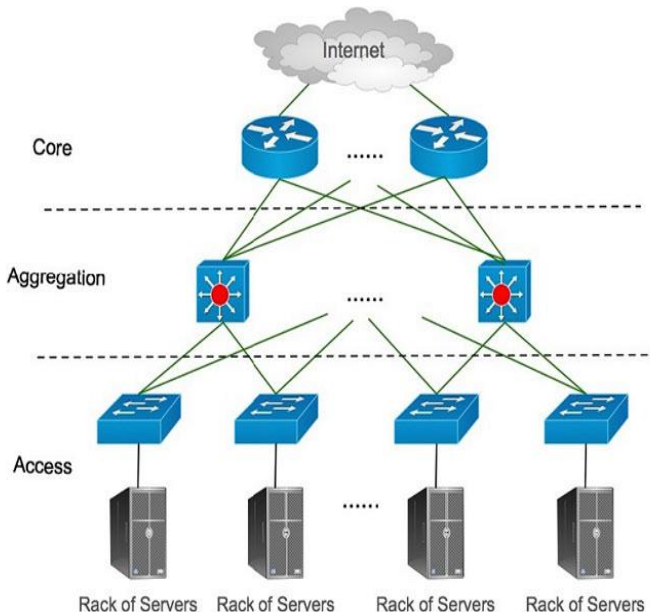

Fig 3: Cloud Data Center Architecture

5.1 Uniform high capacity: The maximum rate of a serverto-server traffic flow should be limited only by the available
capacity on the network-interface cards of the sending and receiving servers, and assigning servers to a service should be independent of the network topology. It should be possible for an arbitrary host in the data center to communicate with any other host in the network at the full bandwidth of its local network interface.

5.2 Free VM migration: Virtualization allows the entire VM state to be transmitted across the network to migrate a VM from one physical machine to another. A cloud computing hosting service may migrate VMs for statistical multiplexing or dynamically changing communication patterns to achieve high bandwidth for tightly coupled hosts or to achieve variable heat distribution and power availability in the data center. The communication topology should be designed so as to support rapid virtual machine migration.

5.3 Resiliency: Failures will be common at scale. The network infrastructure must be fault-tolerant against various types of server failures, link outages, or server-rack failures. Existing unicast and multicast communications should not be affected to the extent allowed by the underlying physical

5.4 Scalability: The network infrastructure must be able to scale to a large number of servers and allow for incremental expansion.

5.5 Backward compatibility: The network infrastructure should be backward compatible with switches and routers running Ethernet and IP. Because existing data centers have commonly leveraged commodity Ethernet and IP based devices, they should also be used in the new architecture without major modifications. Another area of rapid innovation in the industry is the design and deployment of shipping-container based, modular data center (MDC). In an MDC, normally up to a few thousands of servers, are interconnected via switches to form the network infrastructure. Highly interactive applications, which are sensitive to response time, are suitable for geodiverse MDC placed close to major population areas. The MDC also helps with redundancy because not all areas are likely to lose power, experience an earthquake, or suffer riots at the same time. Rather than the three-layered approach discussed above, Guo et al. [22, 23] proposed server-centric, recursively defined network structures of MDC.

6. Research Challenges

Although cloud computing has been widely adopted by the industry, the research on cloud computing is still at an early stage. Many existing issues have not been fully addressed, while new challenges keep emerging from industry applications. In this section, we summarize some of the challenging research issues in cloud computing.

6.1 Automated Service Provisioning

One of the key features of cloud computing is the capability of acquiring and releasing resources on-demand. The objective of a service provider in this case is to allocate and de-allocate resources from the cloud to satisfy its Service Level Objectives (SLOs), while minimizing its operational cost. However, it is not obvious how a service provider can achieve this objective. In particular, it is not easy to determine how to map SLOs such as QoS requirements to low-level resource requirement such as CPU and memory requirements. Furthermore, to achieve high agility and respond to rapid demand fluctuations such as in flash crowd effect, the resource provisioning decisions must be made online. Automated service provisioning is not a new problem. Dynamic resource provisioning for Internet applications has been studied extensively in the past. These approaches typically involve: (1) Constructing an application performance model that predicts the number of application instances required to handle demand at each particular level in order to satisfy QoS requirements; (2) Periodically predicting future demand and determining resource requirements using the performance model; and (3) Automatically allocating resources using the predicted resource requirements. Application performance model can be constructed using various techniques, including Queuing theory, Control theory [28] and Statistical Machine Learning [7]. Additionally, there is a distinction between proactive and reactive resource control. The proactive approach uses predicted demand to periodically allocate resources before they are needed. The reactive approach reacts to immediate demand fluctuations before periodic demand prediction is available. Both approaches are important and necessary for effective resource control in dynamic operating environments.

6.2 Virtual Machine Migration

Virtualization can provide significant benefits in cloud computing by enabling virtual machine migration to balance load across the data center. In addition, virtual machine migration enables robust and highly responsive provisioning in data centers. Virtual machine migration has evolved from process migration techniques. More recently, Xen and VMWare have implemented "live" migration of VMs that involves extremely short downtimes ranging from tens of milliseconds to a second. Clark et al. [13] pointed out that migrating an entire OS and all of its applications as one unit allows to avoid many of the difficulties faced by processlevel migration approaches, and analyzed the benefits of live migration of VMs. The major benefits of VM migration is to avoid hotspots; however, this is not straightforward. Currently, detecting workload hotspots and initiating a migration lacks the agility to respond to sudden workload changes. Moreover, the inmemory state should be transferred consistently and efficiently, with integrated consideration of resources for applications and physical servers.

6.3 Server Consolidation

Server consolidation is an effective approach to maximize resource utilization while minimizing energy consumption in a cloud computing environment. Live VMmigration technology is often used to consolidate VMs residing on multiple under-utilized servers onto a single server, so that the remaining servers can be set to an energy-saving state. The problem of optimally consolidating servers in a data center is often formulated as a variant of the vector bin-packing problem [11], which is an NP-hard optimization problem. Various heuristics have been proposed for this problem [33]. Additionally, dependencies among VMs, such as communication requirements, have also been considered recently [34]. However, server consolidation activities should not hurt application performance. It is known that the resource usage (also known as the footprint) of individual VMs may vary over time. For server resources that are shared among VMs, such as bandwidth, memory cache and disk I/O, maximally consolidating a server may result in resource congestion when a VM changes its footprint on the server. Hence, it is sometimes important to observe the fluctuations of VM footprints and use this information for effective server consolidation. Finally, the system must quickly react to resource congestions when they occur.

6.4 Energy Management

Improving energy efficiency is another major issue in cloud computing. It has been estimated that the cost of powering and cooling accounts for 53% of the total operational expenditure of data centers [26]. In 2006, data centers in the US consumed more than 1.5% of the total energy generated in that year, and the percentage is projected to grow 18% annually [33]. Hence infrastructure providers are under enormous pressure to reduce energy consumption. The goal is not only to cut down energy cost in data centers, but also to meet government regulations and environmental standards. Designing energy-efficient data centers has recently received considerable attention. This problem can be approached from several directions. For example, energyefficient hardware architecture that enables slowing down CPU speeds and turning off partial hardware components [8] has become commonplace. Energy-aware job scheduling and server consolidation are two other ways to reduce power consumption by turning off unused machines. Recent research has also begun to study energy-efficient network protocols and infrastructures [27]. A key challenge in all the above methods is to achieve a good trade-off between energy savings and application performance. In this respect, few researchers have recently started to investigate coordinated solutions for performance and power management in a dynamic cloud environment [32].

6.5 Traffic Management and Analysis

Analysis of data traffic is important for today's data centers. For example, many web applications rely on analysis of traffic data to optimize customer experiences. Network operators also need to know how traffic flows through the network in order to make many of themanagement and planning decisions. However, there are several challenges for existing traffic measurement and analysis methods in Internet Service Providers (ISPs) networks and enterprise to extend to data centers. Firstly, the density of links is much higher than that in ISPs or enterprise networks, which makes the worstcase scenario for existing methods. Secondly, most existing methods can compute traffic matrices between a few hundreds end hosts, but even a modular data center can have several thousand servers. Finally, existing methods usually assume some flow patterns that are reasonable in Internet and enterprises networks, but the applications deployed on data centers, such as MapReduce jobs, significantly change the traffic pattern. Further, there is tighter coupling in application's use of network, computing, and storage resources, than what is seen in other settings. Currently, there is not much work on measurement and analysis of data center traffic. Greenberg et al. [21] report data center traffic characteristics on flow sizes and concurrent flows, and use these to guide network infrastructure design. Benson et al. [16] perform a complementary study of traffic at the edges of a data center by examining SNMP traces from routers.

6.6 Data Security

Data security is another important research topic in cloud computing. Since service providers typically do not have access to the physical security system of data centers, they must rely on the infrastructure provider to achieve full data security. Even for a virtual private cloud, the service provider can only specify the security setting remotely, without knowing whether it is fully implemented. The infrastructure provider, in this context, must achieve the following objectives: (1) *confidentiality*, for secure data access and transfer, and (2) *auditability*, for attesting whether security setting of applications has been tampered or not. Confidentiality is usually achieved using cryptographic protocols, whereas auditability can be achieved using remote attestation techniques. Remote attestation typically requires a trusted platform module (TPM) to generate non-forgeable system summary (i.e. system state encrypted using TPM's private key) as the proof of system security. However, in a virtualized environment like the clouds, VMs can dynamically migrate from one location to another, hence directly using remote attestation is not sufficient. In this case, it is critical to build trust mechanisms at every architectural layer of the cloud. Firstly, the hardware layer must be trusted using hardware TPM. Secondly, the virtualization platform must be trusted using secure virtual machine monitors. VM migration should only be allowed if both source and destination servers are trusted. Recent work has been devoted to designing efficient protocols for trust establishment and management [31].

6.7 Software Frameworks

Cloud computing provides a compelling platform for hosting large-scale data-intensive applications. Typically, these applications leverage MapReduce frameworks such as Hadoop for scalable and fault-tolerant data processing. Recent work has shown that the performance and resource consumption of a MapReduce job is highly dependent on the type of the application [29]. For instance, Hadoop tasks such as sort is I/O intensive, whereas grep requires significant CPU resources. Furthermore, the VM allocated to each Hadoop node may have heterogeneous characteristics. For example, the bandwidth available to a

VM is dependent on other VMs collocated on the same server. Hence, it is possible to optimize the performance and cost of a MapReduce application by carefully selecting its configuration parameter values [29] and designing more efficient scheduling algorithms. By mitigating the bottleneck resources, execution time of applications can be significantly improved. The key challenges include performance modeling of Hadoop jobs (either online or offline), and adaptive scheduling in dynamic conditions.
Another related approach argues for making MapReduce frameworks energy-aware. The essential idea of this approach is to turn Hadoop node into sleep mode when it has finished its job while waiting for new assignments. To do so, both Hadoop and HDFS must be made energy-aware. Furthermore, there is often a trade-off between performance and energy-awareness. Depending on the objective, finding a desirable trade-off point is still an unexplored research topic.

6.8 Storage Technologies and Data Management

Software frameworks such as MapReduce and its various implementations such as Hadoop and Dryad are designed for distributed processing of data-intensive tasks. As mentioned previously, these frameworks typically operate on
Internet-scale file systems such as GFS and HDFS. These file systems are different from traditional distributed file systems in their storage structure, access pattern and application programming interface. In particular, they do not implement the standard POSIX interface, and therefore introduce compatibility issues with legacy file systems and applications. Several research efforts have studied this problem [4]. For instance, the work in [4] proposed a method for supporting the MapReduce framework using cluster file systems such as IBM's GPFS. Patil et al. proposed new API primitives for scalable and concurrent data access.

6.9 Novel cloud architectures

Currently, most of the commercial clouds are implemented in large data centers and operated in a centralized fashion. Although this design achieves economy-of-scale and high manageability, it also comes with its limitations such high energy expense and high initial investment for constructing data centers. Recent work [12] suggests that smallsize data centers can be more advantageous than big data centers in many cases: a small data center does not consume so much power, hence it does not require a powerful and yet expensive cooling system; small data centers are cheaper to build and better geographically distributed than large data centers. Geo-diversity is often desirable for response time-critical services such as content delivery and interactive gaming. For example, Valancius et al. studied the feasibility of hosting video-streaming services using application gateways (a.k.a. nano-data centers). Another related research trend is on using voluntary resources (i.e. resources donated by end-users) for hosting cloud applications [9]. Clouds built using voluntary resources, or a mixture of voluntary and dedicated resources are much cheaper to operate and more suitable for non-profit applications such as scientific computing. However, this architecture also imposes challenges such managing heterogeneous resources and frequent churn events. Also, devising incentive schemes for such architectures is an open research problem.

7. CONCLUSION

Cloud computing has recently emerged as a compelling paradigm for managing and delivering services over the Internet. The rise of cloud computing is rapidly changing the landscape of information technology, and ultimately turning the long-held promise of utility computing into a reality. However, despite the significant benefits offered by cloud computing, the current technologies are not matured enough to realize its full potential. Many key challenges in this domain, including automatic resource provisioning, power management and security management, are only starting to receive attention from the research community. Therefore, we believe there is still tremendous opportunity for researchers to make groundbreaking contributions in this field, and bring significant impact to their development in the industry. In this paper, we have surveyed the state-of-the-art of cloud computing, covering its essential concepts, architectural designs, prominent characteristics, key technologies as well as research directions. As the development of cloud computing technology is still at an early stage, we hope our work will provide a better understanding of the design challenges of cloud computing, and pave the way for further research in this area.

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:Procof 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 dataprocessing 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 filesystem. 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
[21] Guo C et al (2008) DCell: a scalable and fault-tolerant network structure for data centers. In: Proc SIGCOMM
[22] Guo C, Lu G, Li D et al (2009) BCube: a high performance, server-centric network architecture for modular data centers. In:
[23] 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 servicesIn: 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-configuredCPU 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: Procofinternational conference on cloud computing
[33] Li B et al (2009) EnaCloud: an energy-saving application live placement approach for cloud computing environments.In:Procof 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] Qi Zhang, Lu Cheng, Raouf Boutaba for Spinger 2010 "Cloud computing: state-of-the-art and research challenges"