IJITKMSpecial Issue (ICFTEM-2014) May 2014 pp. 27-32 (ISSN 0973-4414)

A Review on Cloud Computing and Data Management in Cloud

Dr. Ajay Jangra[1], Vandana Bhatia[2]
[1]Assistant Professor, U.I.E.T, Kurukshetra University, Kurukshetra
[2]Research Scholar, U.I.E.T, Kurukshetra University, Kurukshetra
[1]Er_jangra@yahoo.co.in, [2]Vbhatia91@gmail.com

Abstract:-Cloud Computing is a latest effort in delivering computing resources either hardware or software as a service over a network. Cloud computing gives a recent trending in IT that moves computing and data away from desktop and portable PCs to large data centers storing peta-bytes of information. Cloud Computing deals with various aspects like availability, scalability, virtualization, interoperability, quality of service and the delivery models of the cloud, namely private, public and hybrid. Cloud databases are mainly used for data intensive applications such as data warehousing, data mining and business intelligence. Cloud database is required to apace effectively to reduce the burdens during routing configuration. The goal of this paper is to review the concept of Cloud Computing, Cloud database and providing database as a service over a network.
Keywords: Cloud Computing (CC), IaaS, PaaS, SaaS, DaaS, Cloud Database, Cloud Storage

1. Introduction

Cloud computing is an emerging technology which provides on demand network access to shared pool of resources like networks, servers, storage, applications or services resulting in minimal management effort and service provider interaction[1]. Grid computing, virtualization, automatic computing, Service Oriented Architecture (SOA), Peer-to-Peer computing and many others all are coalesced in cloud computing in many ways as it uses existing functionality of all and add other perks too. It is a paradigm shift from the parallel and distributed computing, where an organization uses resources as service to "utility computing" where you pay-as-you-go like electricity bill. Clients have nothing to do with management and maintenance problems of the resources machines in Cloud Computing [2]. It is the way of future as it makes sharing of system or information much easier but some security issues may arise due to easy sharing of data inside cloud. For avoiding such issues every cloud computing service has an interface and employs a different access control protocol. Sometimes to ensure security, strong password policies are enforced.

For different people, CC has different meanings so there is no univocally accepted definition of the term. The National Institute of Standards and Technology (NIST) defines cloud computing as "A model for convenient, on demand network access to computing resources such as networks, servers, storage, applications and services that can be quickly deployed and released with very little management by the cloud provider"[3]. Pinase et al. [4] define CC as a distributed logical entity with managed computing resources deployed in big data centers around the globe and connected using public networks, like the Internet. As for the purpose of such entity, Maia et al. [5] define CC as a service with remote access to hardware and software in a highly reliable and transparent way like the electrical network.

Cloud Computing is gaining a high popularity in the I.T industry as it offers significant benefit by freeing them from the low level tasks of setting up basic hardware and software infrastructures and thus enabling them to focus on innovation and strengthen the business value for their services by providing benefaction of reduced cost and easy and fast deployment of resources. It is a paradigm shift into next generation data centres hosted by large infrastructure companies such as Amazon, Google, Yahoo, Microsoft, or Sun.

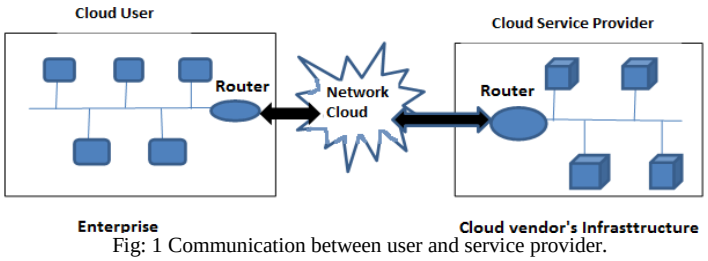

Fig: 1 Communication between user and service provider.

As shown in fig 1, Cloud providers are the companies which manage large datacenters and have expertise of operating these datacenters. Cloud users are the organizations which use services from providers. There is no need to

deploy computing resources at Cloud user's site. These resources are available from the cloud providers on utility basis and charged on usage basis. Thus an organization can improve efficiency while minimizing the expenditure and the operation overhead. [3] Since user requirements for cloud services are varied, service providers have to ensure that they can be flexible in their service delivery while keeping the users isolated from the underlying infrastructure.

2. SERVICES AND CHARACTERISTICS

The current widely accepted Cloud computing definition is based on the NIST definition that identifies five essential Cloud characteristics [3]

- On-demand self-service.
- Broad network access and diversity of client devices.
- Resource pooling that allows providers to serve multi-tenant customers by managing resource utilization more efficiently using virtualization, resource partitioning and workload balancing
- Rapid elasticity that allows scaling resources dynamically.
- Measured service with the pay-per-use business model.

Other additional feature is the heterogeneity on both provider and customer sides, and multi-provider services. The increase in popularity of concepts like Infrastructure as a Service (IaaS), Platform as a Service (PaaS), and Software as a Service (SaaS), presented by Luis et al. [6] was the next step in the evolution of CC. This challenged the Web 2.0 companies which were already facing huge data and infrastructures growths. Companies like Google and Amazon provide users with hardware resources, computational resources, and software resources which have properties like elasticity, availability, and cost effectiveness in mind. At times, the concept of Virtualization is used by maximizing the output of Individual server in which client thinks there are multiple Servers having multiple operating systems performing their specific operations.

The concept can be extended to Database as a Service (DaaS) or Storage as a Service. The idea behind DaaS is to avoid the complexity and cost of running your own database. According to Elena Ferrari [7], there is an addition to this list of services, called Database-as-a-Service (DaaS). In this service, organizations host their own databases in Cloud Computing. This service provides the access for DML (Data Manipulation Language) statement features (strore, retrieve, update and delete the data) via the internet following [7]. Multitanency is one of the key features of DaaS offering, where a large volume of databases with different Service Level Agreement (SLA) requirements are co-located in one environment and sharing resources.

3. ISSUES IN CLOUD COMPUTING

Behl [8] explores cloud computing security issues and highlights the key research challenges that include:

- Availability and Performance — this issue can be resolved through well-formed SLA (service level agreement) coined with real-time monitoring.
- Malicious Insiders — the cloud service providers cannot restrict their employees, contractors and other trusted people who have access to the secure data of customers through supply chain management.
- Outside Attacks — for example, the hackers can get access to the data; to resolve this issue, the network perimeter should be protected through firewalls.
- Service Disruptions — it can occur when no more resources are available for other customers and this may cause customer dissatisfaction. This issue can be resolved by ensuring that connections are coming from known IP pool and DNS

4. CLOUD STORAGE DEPLOYMENT MODELS

Data is stored online in cloud so that it can be accessed by multiple clients over a network. For efficiency, the cloud storage must be agile, scalable, multitenant and secure. Cloud storage is generally classified according to the following configuration:

- **Public Cloud:** It providers offer service publically. It is cheap and standardized. The Data must be scalable and flexible.
- **Private Cloud:** Its service is limited to one enterprise so it requires customized, dedicated functions, local control and is generally expensive. It consists of one or few tenants that trust each other. Security Criteria must be Strong in this configuration.

- **Hybrid Cloud:** In this Private infrastructure is integrated with public cloud. Therefore trust in private cloud is combined with scalability of public cloud.

5. CLOUD DATABASE AS STORAGE

Cloud database is a networked online stored database that typically runs on a cloud computing platform, such as Amazon EC2, GoGrid and Rackspace. From music files to pictures to sensitive documents, the cloud can invisibly backs up files and folders and alleviates the potentially endless and costly search for extra storage space as long as user have internet capability. An alternative to buying an external hard drive or deleting old files to make room for new ones, cloud storage is convenient and cost-effective as user can store files on a server out in the internet somewhere rather than on your local hard drive [9].The Cloud Database is constructed by collecting a number of sites. The sites are also called as nodes which are interlinked by a communication network. Every single node is a database class. Each database class has its own database, terminals, the central processor and their individual local database management system.

Amazon was one of the first companies to offer cloud services to the public. Amazon SimpleDB is a web service for running queries on structured data in real time. It works in close conjunction with Amazon simple storage Service (Amazon S3 which allows to store items up to 5GB in size in Amazon's virtual storage device) and Amazon Elastic Compute Cloud (Amazon EC2 which offers virtual machines and extra CPU cycles), collectively providing the ability to store, process and query data sets in the cloud.

Few examples of cloud storage are as follows [10]

- Google Docs allows users to upload there documents on its server.
- Web e-mail providers like gmail, yahoo etc.
- Picasa hosts millions of digital photos.
- YouTube allows users to host video files.
- Web site hosting companies like GoDaddy, hostmonster allows its clients to store files and data of their web site.

Cloud Storage, Data as a service (DaaS) and Database as a service (DBaaS) are the different terms used for data management in the Cloud on the basis of how data is stored and managed. Cloud storage is virtual storage that enables users to store documents and objects. Dropbox, iCloud etc. are popular cloud storage services. DaaS allows user to store data at a remote disk available through Internet. It is used mainly for backup purposes and basic data management. Cloud storage cannot work without basic data management services. So, these two terms are used interchangeably. DBaaS is one step ahead. It offers complete database functionality and allows users to access and store their database at remote disks anytime from any place through Internet. Amazon's SimpleDB, Amazon RDS, Google's BigTable, Yahoo's Sherpa and Microsoft's SQL Azure Database are the commonly used databases in the Cloud [11].

Cloud database Development issues

Daniel J. Abadi [12] in his work discussed the limitations and opportunities of deploying data management issues on these emerging cloud computing platforms (e.g., Amazon Web Services). He speculated that large scale data analysis tasks, decision support systems, and application specific data marts are more likely to take advantage of cloud computing platforms than operational, transactional database systems[12].

To provide Scalability, developers have to develop databases in such a way that they can support and handle unlimited number of concurrent users and data growth. Enterprises deal with huge volumes of data. Availability in Cloud database implies that database is up and running 365 * 24 * 7. So it becomes necessary to replicate data across large geographic distances to provide high data availability, durability and high levels of fault tolerance. Service level agreement (SLA) should address performance, availability, latency, and QoS (quality-of service) issues.Amazon's S3 cloud storage service replicates data across "regions" and "availability zones" to provide better availability. Cloud database must be portable such that it can be accessed from different locations and devices such as mobiles, tablets, notepads and computers.Cloud databases must support eventual consistency as data is replicated at multiple distributed locations like in No-SQL databases. It becomes difficult to maintain the consistency of a transaction in a database which changes too quickly especially in the case of transactional data. To make the cloud database secure, sensitive data is encrypted before being uploaded to the cloud to prevent unauthorized access. Any application running in the cloud should not have the ability to directly decrypt the data before accessing it. Providing security and privacy to different databases on the same hardware is also a big challenge.

Das et al. [13] proposes Elastrans where transactions are allowed but only at partition level. A second solution proposed by Francisco M. et al. [2] builds on top of the former, expanding the consistency to a group of partitions by introduction of a new layer of replication which also ensures higher availability.
There are two common deployment models: users can run databases on the cloud independently, using a virtual machine image, or they can purchase access to a database service, maintained by a cloud database provider. Of the databases available on the cloud, some are SQL-based and some use a NoSQL data model. Ordinary Database like RDBMS is not suitable for handling massive sparse data sets with loosely defined schemas. The need to store and process such big data defined the role of NoSQL databases in the database technology as Cloud databases.

6. CLOUD DATABASE STORAGE MANAGEMENT

Because of the ubiquitous nature of cloud computing, there is a rapid growth in the number of applications which leverage various cloud platforms, resulting in a tremendous increase in the scale of the data generated as well as consumed by such applications. Scalable database management systems (DBMS)—both for update intensive application workloads, as well as decision support systems—are thus a critical part of the cloud infrastructure. Conventional DBMS deals with structured data which is held in databases along with its metadata. While Cloud databases can be used for unstructured, semi-structured data or structured data. Data stored in files of various types where the metadata was either unavailable or incomplete is called unstructured data. Cloud databases are able to support changing storage requirements of Internet-savvy users who deal more with unstructured data, user created content such as documents and photos.
As shown in fig 2, a data management life cycle is depicted to utilize large scale data. The first step is to collect digital data in the field, followed by data organization and annotation, which requires sophisticated computer infrastructure to streamline the workflow and continuous stream of collected data. A system which fascinates Cloud Computing has a large amount of data, which can be collected by using database and data warehouse for analysis. The data has to be transferred between two spatially separated, tightly coupled locations that is front end and back end. To bridge this gap, we need a mobile data management system equipped with abstracted program and data hosting infrastructure and means of data access.

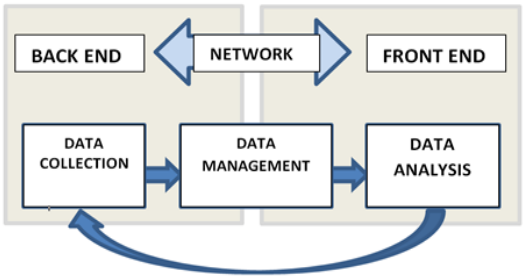

Fig 2: Data management cycle

Amazon Identity and Access Management (IAM) is an implicit service and the authentication infrastructure used to authenticate access to the various services. AWS Management Console (AWS Console), a web-based point and click interface to manage and monitor the Amazon infrastructure suite including (but not limited to) EC2, EBS, S3, SQS, Amazon Elastic MapReduce, and Amazon CloudFront. Amazon also makes available a mobile application for the Android which has support for some of the management features from the console.

Previous and on-going work

Scalable and distributed data management has been the vision of the database research community for more than three decades. In [15], authors conducted various experiments on On-premises traditional database in terms of IBM'S DB2, Oracle database and Microsoft SQL Server. The performance of the Cloud Database is evaluated in this research and a comparison is made with that of an on premises traditional database.
Much research has focussed on designing scalable systems for both update intensive workloads as well as ad-hoc analysis workloads. Initial designs include distributed databases for update intensive workloads, and parallel database systems for analytical workloads. Parallel databases grew beyond prototype systems to large commercial systems, but distributed database systems were not very successful and were never commercialized – rather various ad-hoc approaches to scaling were used. Changes in the data access patterns of applications and the need to scale out to thousands of commodity machines led to the birth of a new class of systems referred to as Key-Value stores

which are now being widely adopted by various enterprises. In the domain of data analysis, the MapReduce paradigm and its open-source implementation Hadoop has also seen widespread adoption in industry and academics [16]. Solutions have also been proposed to improve Hadoop based systems in terms of usability and performance [17]. Jeffrey et al. [18] introduces MapReduce as a master-slave model. The failure of a slave is managed by re-assigning its task to another slave, while master failures are not managed as considered unlikely to happen. Users specify a map and a reduce function. The map function processes key/value pairs and generates a set of intermediate key/value pairs. The reduce function merges all intermediate values associated with the same intermediate key and produces a result as a list of values

7. CONCLUSION

Cloud Computing makes a better use of distributed resources, combine them to achieve higher throughput and be able to solve large scale computation problems. Cloud computing provides a simultaneous way to lower costs, increase responsiveness and flexibility, and improve quality of service. Google, Amazon, Yahoo and other Internet service providers, as well as IBM, Microsoft and other IT companies have proposed their own cloud computing strategy, while various telecom operators paid great attention to cloud computing. In this paper the basic concepts like different cloud models, deployment models and cloud storage techniques are discussed. Various risk factors involved in storing data in cloud storage and its solution are also briefly illustrated.

References

[1] YashpalsinhJadeja and KiritModi, "Cloud Computing - Concepts, Architecture and Challenges", 2012 International Conference on Computing, Electronics and Electrical Technologies [ICCEET] co-published by IEEE in 2012, pp.1 to 4

[2] M. Armbrust, A. Fox, R. Griffith, A. Joseph, R. Katz, A. Konwinski, G. Lee, D. Patterson, A. Rabkin, I. Stoica, and M. Zaharia,

[3] P. Mell and T. Grance, "The NIST Definition of Cloud Computing," National Institute of Standards and Technology Special Publication 800-145, 2011

[4] F. Pianese, P. Bosch, A. Duminuco, N. Janssens, T. Stathopoulos and M. Steiner, "Toward a cloud operating system," in Network Operations and Management Symposium Workshops (NOMS Wksps), 2010 IEEE/IFIP, 2010, pp. 335-342.

[5] F. Maia, J. Armendáriz-Iñigo, M. Ruiz-Fuertes and R. Oliveira, "Scalable transactions in the cloud: Partitioning revisited," in Proceedings of the 2010 International Conference on on the Move to Meaningful Internet Systems: Part II, Hersonissos, Crete, Greece, 2010, pp. 785-797.

[6] L. M. Vaquero, L. Rodero-Merino, J. Caceres and M. Lindner, "A break in the clouds: towards a cloud definition," SIGCOMM Comput.Commun.Rev., vol. 39, pp. 50-55, December, 2008.

[7] Elena Ferrari, Database as a Service: Challenges and Solutions for Privacy and Security, 2009 IEEE APSCC pp. 2 to 4

[8] A. Behl,"Emerging Security Challenges in Cloud Computing", IEEE international Conference Information and Communication Technologies (WICT), 2011. Pp.1 to 4

[9] Yuma Matsui, "Portable Data Management Cloud for Field Science" in 2012 IEEE Fifth International Conference on Cloud Computing

[10] Cloud Security Alliance ,: http://www.cloudsecurityalliance.org , accessed on 5 march 2014

[11] How Cloud Storage Works, [Accessed:feb 2014] , available at: HowStuffWorks "How Cloud Storage Works" computer.howstuffworks.com

[12] Daniel J. Abadi. "Data Management in the Cloud: Limitations and Opportunities", IEEE 2009. pp. 1 to 10

[13] S. Das, S. Agarwal, D. Agrawal and A. E. Abbadi, "ElasTraS: An elastic, scalable, and self-managing transactional database for the cloud," CS, UCSB, 03/2010. 2010.

[14] Database as a Service: Reference Architecture – An Overview, An Oracle White Paper on Enterprise Architecture September 2011

[15] "Comparing Real World Database Performance" [Online]. [Accessed: Dec 2013] Available at : http://public.dhe.ibm.com/common/ssi/ecm/en/iml14276usen/IML14276USEN.PDF

[16] Lin Weiwei, Liang Chen and Liu Bo, "A hadoop-based efficient economic cloud storage system" in Inter. Conf on Circuits, Communication and System, pp. 1-4, 2011

[17] E. Yoon. Hadoop Map/Reduce Data Processing Benchmarks. Hadoop Wiki. [Accessed: Feb 2014] Availaibleat:http://wiki.apache.org/hadoop/DataProcessingBenchmarks.

[18] J. Dean and S. Ghemawat, "MapReduce: simplified data processing on large clusters," Commun ACM, vol. 51, pp. 107-113, January, 2008