European Journal of Academic Essays 1(6): 12-17, 2014
ISSN: 2183-1904
www.euroessays.org

Databases for Cloud Computing: Comparative Study and Review

Pawan Singh[1], Nirayo Hailu [2], Vinod Chandran [3]

[1]Hawassa University, School of Informatics,
5, Awassa, Ethiopia
pawansingh3@yahoo.com, pawansingh51279@gmail.com, dr_pawansingh@hu.edu.et

[2] Hawassa University, School of Informatics,
5, Awassa, Ethiopia
nirayoh@hu.edu.et, nirayo2008@gmail.com

[3] Hawassa University, School of Informatics,
5, Awassa, Ethiopia
vinodchandranvc@gmail.com

Abstract: Unlike last countless years, the present era has no dependency on relational data base due to the rise of technological changes such as web based applications has replaced the stand alone applications , distributed servers has replaced the normal dedicated servers. A giant imagination gets converted into reality in the form of cloud computing next to World Wide Web. Cloud databases are becoming popular and used for data exhaustive applications as data mining, warehousing and intelligences. This paper give a review on an architecture of shared disk database for cloud computing. The database architecture called shared-disk is ideal for cloud databases, which eliminates the need to partition data. It allows the clusters of low-cost servers to use a single assortment of data. It is typically served up by a Network Attached Storage (NAS) or Storage Area Network (SAN). All of the data available to all the servers, there are no partitioning of data is required. For example if we are using two servers and query takes .60 seconds and now we can dynamically add another server, the same query might take .30 seconds. Thus elastic scalability is supported by this type of database.

Keywords: Cloud computing, Cloud database, Shared disk, Share-nothing database, Database.

---

1. Introduction

Cloud computing has emerged as a strong and cost-efficient paradigm for provisioning computing power to users. Within the cloud computing paradigm, users use associate computer network or the net to access a shared computing cloud that consists of an outsized variety (thousands or tens of thousands) of interconnected machines organized joined or a lot of clusters. This provides important advantages each to suppliers of computing power and to users of this computing power [1]. Cloud Computing," to place it merely, suggests that "Internet Computing." the net is often visualized as clouds; thus the term "cloud computing" for computation done through the net. With Cloud Computing users will access information resources via the net from anyplace, for as long as they have, without concerning any maintenance or management of actual resources. Besides, databases in cloud are terribly dynamic and scalable [2]. As simply the general public net spawned personal company intranets, cloud computing is spawning to personal cloud platforms, public and personal cloud platforms are trying to deliver the advantages of cloud computing to their customers. Whether or not this is often a personal or public cloud, the information could be a essential a part of that platform. An information appliance could be a virtual appliance wherever the put in application could be an information system. With the increasing quality of virtualization and cloud computing, we are able to expect that a standard approach of providing information services within the future are through information appliances deployed in SaaS (Software as a Service) clouds. As associate example of this readying mode, Amazon offers MySQL, Oracle, and Microsoft SQL Server virtual appliances for preparation in its EC2(Elastic Computing Cloud) cloud. A vital question to rise is the way to get the simplest information system performance during this setting. Cloud suppliers have an interest in two connected performance objectives: increasing the employment of cloud resources and minimizing the resources needed to satisfy user demand. Users have an interest in minimizing application interval or increasing application turnout. Deploying information appliances within the cloud and standardization the information and virtualization parameters to optimize performance introduces some attention-grabbing analysis challenges.

Corresponding Author: Pawan Singh, School of Informatics, IOT, Hawassa University, Ethiopia, pawansingh3@yahoo.com, +251-938974018

2. Cloud Computing

Vendors offer access to pre-configured computing virtual machines (VMs or cloud servers) on that users will remotely install and run their software system. In the market VM varieties square measure differentiated by the resources they supply like central processing unit power, I/O information measure, and space and memory size [3]. Cloud-based DBMSs area unit generally instantiated through an info appliance i.e., a VM image with a pre-installed pre-configured info engine. Replicated info model on prime of shared networks of multiple VMs, every running as freelance info appliance [4]. Given the target work, QoS goals (expressed as latency bounds), the cloud's resource specification and its financial value model, our framework identifies the foremost cost-efficient VM allocations that might meet the QoS expectations of the work. During run-time, our framework routes incoming user queries to the reserved machines. The goal is to assign queries to machines with enough resources to execute every new and existing queries at intervals their QoS bounds, whereas minimizing any longer money value as a result of per request charges. The restraint convergent thinker is answerable for finding each of the offline resource provisioning and on-line question routing optimization issues. Each issues square measure expressed as constraint programs, which get solved by central organizer victimization associate in nursing ready-to-wear optimization tool.

3. Merits of Cloud Computing

There are different merits of using cloud computing few important of them are mentioned as follows:

3.1. Reduced Cost

There is a shared use of different computing devices as servers, networking equipment etc. This makes more than one applications from midsized to low sized, share their resources to lower their overall cost of computing.

3.2. Elasticity

As the time passes for most of applications the traffic have to face the problems, to overcome it more equipment is required which is easy to add on in case of cloud services.

3.3. Support of different Platforms

Good collection of different platforms support is been provided by most of the cloud computing services ranging from browser to mobiles. This makes wide utility of any application in more number of users.

3.4. Easy Management Approval

Due to the low costing of cloud computing any individual can easily provide the approval of funds, which accelerate the management process of approval.

3.5. Scalable Development

The platforms of cloud computing provides many different core services which can be exploited to accelerate the development process.

4. Cloud Computing Database

There exist some motivating factors to give birth to a concept of Cloud Database as blasting growth in data, strict requirement change to data storage, available high bandwidth etc. There emerge two specific terms along with Cloud database Data as a Service DaaS and Data Base as a Service DBaaS. Both are different as through them the data is stored and managed in different manner. Cloud storage is managed as a virtual storage where user can store his data with the help of internet facility, some of the examples are Dropbox, iCloud, Microsoft SQL Azure, Yahoo's Sherpa, Amazon's SimpleDB and Google Big Table.

With the help of DaaS use can store data to remote disk by using internet. This technology helps merely for backup and basic data management. Unlike DaaS, DBaaShave complete database functionality along with enhanced functions of DaaS. Cloud database may range from traditional database to specific cloud database which may deal with different level of data such as structured, semi structured and non structured data. There exist two type of database architecture Share Nothing and Share Disk in database system.

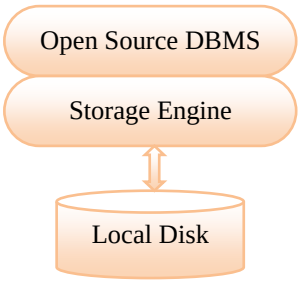

Figure 1: Share Nothing Architecture

In storage architecture of Share Nothing the data is partitioned into independent sets which in turn physically loaded on different database servers. Each server is responsible for making processing and maintenance of their piece of data which makes it scalable as well. Data partitioning required in this architecture has no compatibility with cloud which is working as virtualization and becomes very complex and unmaintanable.

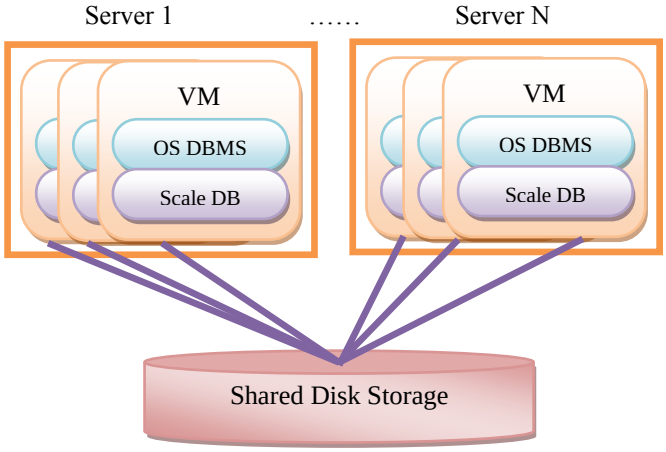

Figure 2: Share Disk Architecture

The whole database is treated by Share Disk database as a single large piece of database stored in Storage Area Network SAN or Network Attached Storage NAS. All node through network can access and share the data storage. It can

be easily be virtualized and have no need to use a middleware to route the request to specific server. Hence it is more suitable for application using online data transactions.

Table 1: Comparative analysis of Share Nothing and Share Disk database architecture

Attributes	*Architecture*	
	Share Nothing	*Share Disk*
Data Partitioning	Yes	No
Distributed	Yes	Yes
Scalability	Yes	Yes
ACID	No	Yes
Adaptability to change workload	Low	High
Maintenance Cost	High	Low
Used for Cloud	Yes	Yes
Analytical	Yes	Yes
Availability	Low	High
Load Balancing	Fixed	Dynamic
Best Performance in	Heavy Read-Write environment	Heavy Read environment

5. Requirement of Cloud Database

Cloud database is using the patterns for evolving business adoption to these technologies that accelerates evolution. Initially, consumer applications are serviced by cloud databases. The early applications give a priority to read access, because the standard ratio of writes to reads was very low. The high-performance read access was the primary acquires criteria for consumer application. Consumer based cloud database applications have been sprouting with the adoption of Web technologies. User generated content, notably within the kind of social networking, have placed somewhat additional prominence on updates. Reads have still outnumbered writes in terms of the ratio, but the gap is going to be tightened. Through support for transactional business applications, this breach between database updates and reads is further recoil. Business applications have additional demand of the cloud database be ACID, providing Atomicity, Consistency, Isolation and Durability [5]. Different cloud database requirement can be understood with the help of following examples.

Consider information powering a consumer-centric cosmetics web site. If the user will a groundwork for an explicit shade of nail paint, it's necessary that the results be delivered outright to stay the user engaged, thus she doesn't click on another cosmetics website. If the positioning same that the chosen lipstick is in inventory and completed the sale, it wouldn't be the tip of the planet to later determine that, as a results of inconsistent knowledge, that lipstick wasn't very in inventory. During this case, the buyer receives associate degree email explaining that it's on backorder and can be shipped soon there is no drawback.

Consider a corporation that sells mechanical devices to makers. An outsized company purchases a load of mechanical devices necessary to stay its mechanical system running. During this example, if the inventory was incorrect, because of inconsistent information, and therefore the cargo is delayed, the corporate agency purchased the widgets is also forced not to run his mechanical system at a value of $250000 per day this creates a great loss. With this understanding of the various stakes concerned, it's simple to grasp however company adoption of cloud databases is dynamical the sport significantly.

6. Database Backup Strategy for Cloud Platform

The best backup strategy for the cloud could be a file-based backup. You lock the information against writes, take a pic, and unlock it. It's elegant, quick, and reliable. The key cloud feature that creates this approach doable is that the cloud's ability to require snapshots of your block storage volumes. While not pic capabilities, this backup strategy would merely take too long. Your backup strategy cannot, however, finish with a file-based backup. Snapshots work superbly inside one cloud; however they can't be leveraged outside your cloud supplier. In different words, associate degree Amazon EC2 elastic block volume pic cannot be leveraged in a very cloud readying, to form certain your application is transportable between clouds, you wish to execute full information exports frequently. However frequently you perform your information exports depend upon what quantity knowledge you'll use. The underlying question you want to boost is, "If my cloud provider suddenly goes down for academic degree extended quantity of some time, what amount data can afford to lose once launching terribly very new environment?" For a content management system, it's reaching to be OK in such academic degree extreme situation to lose each week of knowledge. Associate degree e-commerce application, however, cannot extremely afford to lose any data—even below such extreme circumstances.

7. Advantages of using Share Disk Database in Cloud Computing

Share disc database eliminates the need to partition the data by allowing the clusters of low cost servers to utilize a single collection of data. All servers can access full data as there is no partitioning this reduces the query processing time by adding more servers dynamically. The shared-disk DBMS architecture has other important advantages—in addition to elastic scalability—that make it very appealing for deployment in the cloud. The following are some of these advantages:

7.1. Uses less servers

As shared-nothing databases break the information into distinct items, it's not decent to own one server for every information set, you would like a back-up just in case the primary one fails. This can be known as a master-slave configuration. In alternative words, you want to duplicate your server infrastructure. Shared-disk may be a master-master configuration; therefore every node provides fail-over

for the opposite nodes. This reduces the quantity of servers needed by half once employing shared-disk information [6].

7.2. Utilization of Low Cost Servers

Within a shared-nothing database, every server should run at low processor utilization so as to be able to accommodate spikes in usage for that server's information, this implies that you simply square measure shopping for massive (expensive) servers to handle the peaks [7]. Shared-disk, on the opposite hand, spreads these usage spikes across the complete cluster. As a result, every system may be run at better processor utilization, this implies that with shared-disk info you'll be able to purchase lower-cost artifact servers rather than paying an outsized premium for high-end computers. This additionally extends the time period of existing servers, since they needn't deliver newest performance. If our info backups actually take such an extended time to execute that you simply risk having your slaves falling terribly way behind the master, it is smart to put together multiple slaves and rotate backups among the slaves. This rotation policy can provides a slave sufficient time to catch up with the master once it's dead a backup and before it has to perform its next backup. Once the backup is complete, you ought to move it over to S3 and often copy those backups out of S3 to a different cloud supplier or your own internal digital computer [8].

7.3. Scaling Model

The scale-in model permits cloud suppliers to portion and bill customers on the idea of what percentage instances of an information area unit being run on a multi-core machine. Scale-in permits you to launch one instance of MySQL per CPU core. Consider an instance of an example that a thirty two-core machine might support a cluster-in-a-box of thirty two instances of MySQL [9].

7.4. Easy maintenance & improvement process

It is possible to have an individual up gradation of servers which are part of share disk database, whereas the cluster still remains on-line. You'll be able to select as per requirement a node and make out of service, upgrade them, and place them back in commission whereas the rest of nodes in network nodes still operate. You cannot try this with a shared-nothing information as a result of every individual node owns a particular piece of knowledge, decommissioning a node may result in lack of data required. Cast off one server in exceedingly shared-nothing information may lead to the entire cluster be closed.

7.5. Consistency

The nodes during serving a shared-disk database square measure utterly interchangeable, one may be able to lose nodes and their performance might drop off, however the system keeps in operation. But in a shared-nothing database losing a server leads the system goes down until you manually upgrade a slave to be in the role of master. Additionally, if any time one partition or repartition the database in case of non sharing database [10], one need to take the system to be closed. In alternative words, shared disk database may not require additional scheduled and forced time period unlike the shared-nothing systems.

7.6. Attenuation in efforts of partitioning and tuning

It is required in shared-nothing cloud database to partition the information. The simplest way is to partition the data among some server, this partitioning information to attenuate the traffic between nodes within the cluster is also called data shipping which requires a good deal of progress analysis and tuning. Making an attempt to accomplish this during a static shared-nothing cluster could be an important challenge, however making an attempt to try to do thus with a dynamically scaling information cluster could be a herculean task [11].This all necessary exercise has no significance in shared disk database.

7.7 Attenuation in cost for support

Exiting advantages of cloud databases is that they transfer abundant of the low-level DBA functions to experts who are dealing the databases in an exceedingly centralized manner for each and every user. Though, standardization a shared-nothing information needs the coordinated involvement of each the DBA and also the application programmer [4,5, 11]. This considerably will amplify support prices. The ideal condition for Shared-disk databases is to cleanly separate the working of the DBA and also the application developer. In cloud environment the support cost is further reduced by shared-disk databases as they conjointly give seamless load-balancing.

8. Issues and Challenges of Cloud Database

Creating an information appliance that may simply be deployed during a cloud, associate degreed getting an accessible, and usable information instance from this appliance need addressing several problems associated with preparation. These straightforward and mundane tasks are often terribly tough and time intense.

8.1. Localization

When we begin a VM from a duplicate of an info appliance, we want to present this new VM and therefore the info system running on that a definite "identity." we tend to discuss with this method as localization, as an example, we want to present the VM a Macintosh address, AN IP address, and a bunch name. We tend to additionally got to adapt (or localize) the info instance running on this VM to the VM's new identity. as an example, some info systems need each info instance to possess a singular name, that is usually supported the host name or IP address. The VMM [3] and therefore the underlying OS and networking infrastructure might facilitate with problems like distribution IP addresses, however there's generally very little support for localizing the info instance. The particular localization needed varies from info system to info system, which will increase the trouble needed for making info appliances.

8.2. Routing

In addition to giving each VM and info instance a definite identity, we have a tendency to should be able to route application requests to the VM and info instance. This includes the IP-level routing of packets to the VM, however it additionally includes ensuring that info requests square measure routed to the proper port and not blocked by any firewall, that the show is routed back to the consumer console if required, that I/O requests square measure routed to the proper computer storage device if the "compute" machines of the IaaS cloud square measure totally different from the storage machines, and so on.

8.3. Authentication

The VM should bear in mind of the credentials of all shoppers that require to attach thereto, freelance of wherever it's run within the cloud.

9. Pit Falls of Cloud Database

One of the most worrisome principal of the cloud computing is Dynamic scalability for databases. The cause is simple; the majority databases use a shared-nothing design. In the share nothing databases the data are partitioned in such a way that one silos comes in share of one server because the share nothing database architecture is depending on the strategy of splitting data. One may assume that dynamically accumulating another information server is as straight forward as cacophonic the info across an added server. For instance, if one have got three servers, every with thirty three percent of the entire information, and add a fourth server, one will only take a fourth of the info from every server and currently one have got total four servers every owning twenty five percent of the info. Sadly, it is the most difficult task. Several user requests require connected information, for instance, you may wish to seek out all customers who placed an order within the present month. You would like to travel to the invoices table and realize the invoices dated for last month then you follow an information key to the client table to gather their contact information. If this is often unfolded across multiple servers, you end-up process info on one machine then passing that information to the second d machine for process. Data base performance may be degraded if this transferring of information takes places repeatedly for this reason the partitioning of the information should be done terribly rigorously to attenuate it. Partitioning information, a long method, is noted as dark art due to the amount of talent needed. The automation of this method is vague goal. Positively one could be able to use a middleware to mechanically repartition the information to accommodate an ever-changing variety of information servers, however your performance will quickly go down to ground.

10. Conclusion and Future Scope

In either of case of management development or construction of a plate form for cloud computing, one have a urgent requirement of cloud compatible database. Dynamic scalability is the main foundation of cloud computing which cannot be achieved in share nothing as its compatibility does not match because share nothing database have a need of data partitioning. Unlike the share nothing, share disk database architecture comes in accordance to elastic scalability. It provides high accessibility in support of service level agreements (SLAs). Due to their compatibility, cloud computing can introduce associate ascendance of the shared-disk information. As cloud computing becomes additional widespread as a resource provisioning paradigm, we are going to progressively see shared disk info systems being deployed as virtual appliances on Infrastructure as a Service (IaaS) clouds like Amazon's EC2. In this paper, we tend to illustrate a number of the challenges related to deploying these appliances and conjointly illustrate the subject area read of the shared disk information for cloud computing. We tend to mention the techniques needed to deal with these challenges. In future we tend to work on their problems and challenges illustrated in it. And conjointly style a reliable information system for cloud computing.

References

[1] Curino, C., Jones, E., Popa, R., Malviya, N., Wu, E., Madden, S., Balakrishnan, H., Zeldovich, N, "Relational Cloud: A Database Service for the Cloud", CIDR. pp. 235–240 (2011).

[2] Curino, C., Zhang, Y., Jones, E.P.C., Madden, S, "Schism: a workload-driven approach to database replication and partitioning", PVLDB 3(1), 48–57 (2010).

[3] R. H. Katz. "Tech titans building boom", IEEE Spectrum, Feb (2009).

[4] Das, S., Agrawal, D., El Abbadi, A, " G-Store: A Scalable Data Store for Transactional Multi key Access in the Cloud", ACM SoCC. pp. 163–174 (2010).

[5] Das, S., Agarwal, S., Agrawal, D., El Abbadi, A, " ElasTraS: An Elastic, Scalable, and Self Managing Transactional Database for the Cloud", Tech. Rep. 2010-04, CS, UCSB (2010).

[6] Das, S., Nishimura, S., Agrawal, D., El Abbadi, A, " Live Database Migration for Elasticity in a Multitenant Database for Cloud Platforms", Tech Rep 2010-09, CS, UCSB (2010).

[7] Dean, J, "Talk at the Google Faculty Summit", (2010).

[8] M. Ahmad, A. Aboulnaga, S. Babu, and K. Munagala, "Modeling and exploiting query interactions in database systems", ACM Int. Conf. on Information and Knowledge Management CIKM, (2008).

[9] M. Armbrust, A. Fox, R. Griffith, A. D. Joseph, R. H. Katz, A. Konwinski, G. Lee, D. A. Patterson, A. Rabkin, I. Stoica, and M. Zaharia.,"Above the clouds: A Berkeley view of cloud computing", Technical report, EECS Department, University of California, Berkeley, (2009).

[10] A. Ganapathi, H. Kuno, U. Dayal, J. Wiener, A. Fox, M. Jordan, and D. Patterson, "Predicting multiple metrics for queries: Better decisions enabled by machine learning", IEEE Int. Conf. on Data Engineering ICDE, (2009).

[11] Das, S., Agrawal, D., El Abbadi, A, "ElasTraS: An Elastic Transactional Data Store in the Cloud", USENIX HotCloud (2009).

Authors Profiles

Dr. Pawan Singh received the B.E. (Computer Science) from CCS University, Meerut, M.Tech. (Information Technology) from GGSIPU, New Delhi and PhD (Computer Science) from Magadh University, Bodh Gaya in 2013. His research areas are Software

engineering, web mining, Wireless and Cloud Computing. He has published number of research papers in various reputed journals. Presently he is serving as a Assistant Professor in School of Informatics, Institute of Technology, Hawassa University, Awassa, Ethiopia.

Mr. Nirayo Hailu received his B. Sc. degree in Computer Science from Hawassa University Department of Computer Science and M. Sc. degree in Computer Science from Addis Ababa University in 2008 and 2011, respectively. Since 2008 he is working in Hawassa University school of Informatics, during 2009-2011 he stayed in Addis Ababa Department of Computer and Mathematical Sciences to study computer science. He is now school head and lecturer at Hawassa University Institute of Technology (HIoT), School of Informatics.

Mr. Vinod Chandran received MCA [with specialization in Software Engineering] degree from Mahatma Gandhi University, Kottayam, Kerala, India in 1999. He has about 15 years of teaching experience in graduate and post graduate studies of Universities in India [1999 to 2008] and abroad [Ethiopia from 2008 – till date]. He has served in many bodies of universities associated with quality teaching and assessment including the present Chairmanship of Academics Standards and Quality Assurance Committee of School of Informatics in Hawassa University as he has obtained special training for Teaching & Assessment of Computing programmes by Edexcel Foundation & London Examinations(University of London) at University of Mysore campus, Manasagangotri,India (in 2000).His area of interest includes Computer Graphics, Systems Modeling and Simulation, Operating Systems, Database Management Systems, Wireless and Cloud Computing, Artificial Intelligence and Software Engineering. He is presently continuing with School of Informatics of Hawassa University, Ethiopia.