Available online at www.sciencedirect.com

ScienceDirect

Procedia - Social and Behavioral Sciences 195 (2015) 1851 – 1859

Procedia
Social and Behavioral Sciences

World Conference on Technology, Innovation and Entrepreneurship

DBMS as a Cloud service: Advantages and Disadvantages

Ylber Januzaj[a], Jaumin Ajdari[a], Besnik Selimi[a]

[a]*Faculty of Contemporary Sciences, South East European University*
Tetovo, R. of Macedonia

Abstract

The service which enables us to use computing as a service across a product is known as cloud computing. Nowadays the cloud computing paradigm has been receiving significant excitement and attention in technological sphere. Cloud computing shares different resources and information between different devices which are located in different places always based on internet connection. According to this, a cloud DBMS is a database management system which acts through cloud computing. It is worth mentioning that the number of these DBMS which act through cloud computing is expected to increase in the future. Based on related research and results, there is an increment of interest in outsourcing of DBMS tasks to third parties that can afford these tasks with low and cheap cost. In this paper, we discuss about DBMS as a cloud service, advantages and disadvantages, opportunities and limitations, and we focus on the way how to offer a cloud DBMS as one of the best services. We focus on three main characteristics of cloud computing which are considered as the most worried issues of cloud platform. We review cloud database challenges such as: internet speed, multi-tenancy, privacy and security. We also focus on the way how to opposite these challenges in order to provide a successful cloud database. At the end of this paper we explain a specific architecture of cloud DBMS which is known as SCALEDB. We focus on its layer which this architecture contains and the way how these layers works. We thus express the need for a new DBMS, designed specifically for cloud computing environments.

Peer-review under responsibility of Istanbul Univeristy.

Keywords: DBMS; Cloud computing; Database Management System.

1. Introduction

Recently technological advancement, especially those of the cloud services have reached a high point of development and research. However, the technology and principles change rapidly and there is an increasing number of open problems. Nowadays, DBMS outsourcing is one of the most required features of cloud services. Technological advancements around transmission of data through the network, have largely influenced the cost of transmitting the data per terabyte over long distances (Gelogo, Y., E., Lee, 2012). Furthermore, DBMS have achieved progress in two comparisons dimensions: data management and data transfer. Based on the related

Peer-review under responsibility of Istanbul Univeristy.
doi:10.1016/j.sbspro.2015.06.412

research, data management happens to be more costly than data transfer (Gelogo, Y., E., Lee, 2012). In addition, there is an rapidly growing interest in outsourcing DBMS tasks to third parties that can provide these tasks for much lower cost due to economy of scale. Designation of a new outsourcing model has few benefits, but the most significant benefit is reduction of the cost for running DBMS on one's own (Gelogo, Y., E., Lee, 2012), (Buyya, R., Broberg and Goscinski, 2011). In addition, a cloud DBMS is nothing else but a cloud database service which is available and accessible from anywhere. The main way of communication that cloud DBMS uses is over internet, whereby it shares information between multiple devices, and number of these devices is expected to increase. Nowadays, there are a lot of companies that offer DBMS as a cloud service such as: Microsoft Azure, Google, Amazon EC2, GoGrid, Garantia Data, Mongo Lab, etc. These companies offer cloud services with two usual deployment models: we can use database independently from a virtual machine, or we can purchase a cloud service database which are retained by mentioned cloud companies (Abadi, J., D., 2009). These cloud services are much more appropriate to the end-users because they offer services in the so-called "pay – as – you – go " model, and in this way services become cheaper compared to other services. In this paper, we describe the deployment models, architectures and common characteristics of cloud DBMS. We also studying the reasons why one should use DBMS as a cloud service. At the end, we describe an architecture of DBMS in cloud and give some of its advantages.

2. Initial State

As other technology areas that need for changes, DBMS need for improvement when it comes in a specific point. In this section we will talk about the need of using DBMS as a cloud service. We always talk about big companies with large amount of data, with giga's or tera's. Based on related researches year 2013 is considered the year of the CDBMS, thus many companies that offers cloud platforms are enhancing their database service offers and becoming more enterprises in market. There are few reasons why 2013 is considered the year of cloud DBMS but the most important are two: first is the most critical reason because migration to cloud is low cost, and second is about performance because howsoever big is our database we don't need to be worried about our performances because it's in company responsibility (Curino, C., Jones, Popa, Malviya, Wu, Madden, Balakrishnan and Zeldovich, 2011), (Kadam, M., Tamie, Jidge and Tayade, 2014).

2.1 DBMS in cloud platform

The goal of our paper is to clarify to the readers the concept of cloud DBMS, and convince readers to migrate to the cloud service. To do this, first we need to focus on the main characteristics of the cloud computing. There are few, but the most important are: parallelization of computer workload, untrusted host of data storage, replication of data across long distances, in addition these three characteristics are considered the most worried issues of cloud platform. (Gelogo, Y., E., Lee, 2012), (Hogan, M., 2008), (Meng-Ju, H., Chao-Rui, Li-Yung, Jan-Jan and Pangfeng, 2011).

2.1.1 Parallelization of computer workload

In this section we will clarify to the readers the concept: parallelizing of computer workload, how parallel computer works and its advantages. Nowadays many problems that have occurred, the technology of computing have found the solution by applying parallel computing. Parallel computing reaches to solve the problem by scheduling tasks. In addition the main tasks of parallel computing are few, but the most important is separating the problem into a number of parallel tasks, in addition by separating these problems into a number of parallel tasks the outlining results becomes more qualitative. Scheduling of these processes is done by using a specific algorithms, which are different from one to another. We say they are different because there is no single algorithms that is appropriate to all scheduling, in addition each schedule possess its algorithm. As it is mentioned in (Ekanayake, J., and Geoffrey, 2010) many parallel computers use a strategy which is known as space-sharing which is used for more qualitative execution of parallel jobs. Under this DBMS has improved performance through parallelization of different operations, such as: executing queries, building indexes and loading different type of data. Improvement of

cloud DBMS processes is achieved by using multiple CPUs and multiple disks which are allocated in parallel, this is because centralized databases are not able to handle high capacity of data, much less processing them. In addition parallel databases are divided in two groups: multiprocessor architecture and hybrid architecture. Multiprocessor architecture works based on these three architectures: shared memory architecture, shared disk architecture and shared nothing architecture. Regarding to hybrid architecture, it works based on these two architecture: non-uniform memory architecture which is known as NUMA and Cluster which includes shared nothing and shared disks, and it is formed by a group of computers. Next we will give description to the readers about these architectures that we mentioned above. (Gelogo, Y., E., Lee, 2012).

2.1.1.1 Shared memory architecture

In cloud platform shared memory is an architecture which allows multiple CPUs to access on it for the only intent to provide communication between them, or to eliminate redundant copies. Depending on this many programs may run on multiple separate processors or on a single processor. Next we will show a sketch of shared memory architecture.

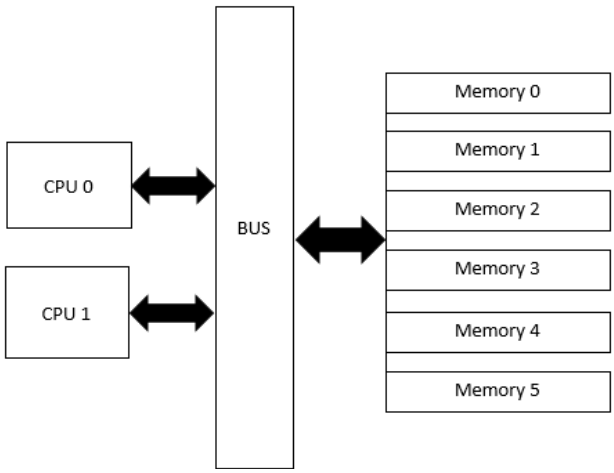

Fig. 1. Shared memory architecture

2.1.1.2 Shared disk architecture

Shared disk architecture is different from shared memory, because here only disks are shared to all processors which communicate in network. In shared disk architecture each processor access to its memory which is controlled by its own CPU. Shared disk architecture have few advantages, but the most important is that is offered with low cost, it provides load balancing and it provides easy migration from centralized form. Next we will show a sketch of shared disk architecture.

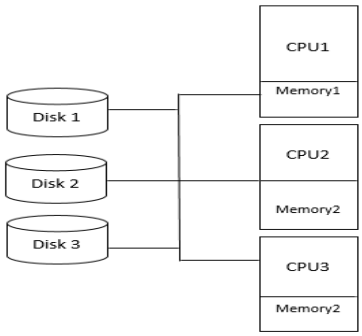

Fig. 2. Shared disk architecture

2.1.1.3 Shared nothing architecture

In this section we will give description to the readers what is shared-nothing architecture. Shared-nothing architecture is known as SN. If we talk about SOA terms, "SN" is each service has a respective database, and it is accessed by only that service and there is no disks or memory shared. Simply if we analyze from a hardware perspective it's a machine which uses local disks and memory.

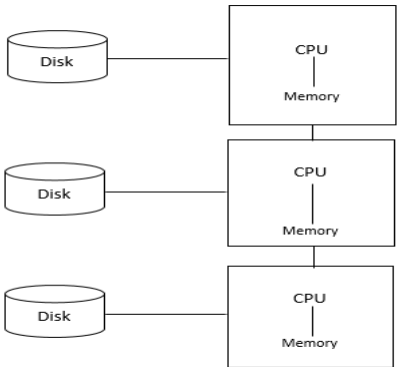

Fig. 3. Shared nothing architecture

2.1.2 Untrusted host

In this section we will clarify to the readers the main security reasons of cloud computing. Cloud computing security it is not a simply security application, it is aggregation of few securities (application security) such as computer security, network security and information security. In addition the main task of cloud security is protection of the data and keeping them more security. Cloud is used by organizations for different issues, such as software as a service (SaaS), product as a service (PaaS), etc., and to provide a secure cloud service there are few issues, but two of them are more important, such as security issues provided by cloud providers, and security issues provided by users. The responsibility goes half in both ways, to the providers and to the users. First providers should provide a secure infrastructure and keep the application protected, while the users need to ensure providing strong usernames and password to fortify their application. Another worrying issue to the users is the privacy of the data, because it may happen that one's user's data can be stored in the same server with another user, but this shouldn't be a worrying issue for us because cloud services provide us a security of data by isolating them and by applying logical storage segregation. Another security issue is the control of data by government. This shouldn't be a worrying issue too because there is no same law in different continents. In addition in United States exists an agreement between cloud providers and government, this agreement says that cloud databases should be accessible by government and they can have full control in data (Gelogo, Y., E., Lee, 2012). In opposite in Europe this

agreement doesn't exist, in addition there is no law that a third party that can control data in cloud (Gelogo, Y., E., Lee, 2012), (Mike, H., 2009). In addition while we have opportunities to decide where we want to purchase cloud database our data can be considered that they are hosted in a trusted host of data storage. Next to give a better clarify to the readers we will give description about security and privacy issues such as identity management, physical and application security and privacy. (Gelogo, Y., E., Lee, 2012), (Mike, H., 2009), (Curino, C., Jones, Zhang, and Madden, 2010).

2.1.2.1 Identity management

Identity management is a security and privacy issue while each enterprise can control information and resources by using its identity management system. To improve the security and privacy, cloud providers integrate the customer's identity management system into their infrastructure using a very powerful and security technology which is known as Single Sign- on technology or SSO.

2.1.2.2 Physical security

Cloud providers also provides physical security which care about physical issue of company such as servers, cables. It worth to be secured because it may happen any unauthorized access, theft or fires.

2.1.2.3 Application Security

Applications which are offered by cloud providers are specified, designed and implemented in secure form. They don't only block any unauthorized access but they are able to identify all the risks which appear to the application. In addition this helps users to grow their faith in cloud platforms.

2.1.2.4 Privacy

Another important issue of cloud computing is the privacy of the data such as credit cards numbers or any other important information. All these important information are held more secure because they are encrypted with special keys, and only authorized users have access on it. Furthermore, cloud computing uses digital identities which helps users to keep their data more protected.

2.1.3 Replication of data across long distances

In this section we will talk for the third characteristic of cloud platform. The main points of cloud database providers are availability and durability of the data, because unavailability of the data can affect the company reputation. The availability and durability of the data can be achieved by under-cover replication. Under-cover replication is an automatic replication of the data without request and permission of the customer. Large companies that offers cloud databases such as: Google, Azure, Amazon etc. have their data centers separated on different parts of the world, and the replication of these data enables users to access on their data not noticing which server is available and which is not (Gelogo, Y., E., Lee, 2012), (Mike, H., 2009), (Curino, C., Jones, Zhang, and Madden, 2010).

3. Dealing With Cloud DBMS

A DBMS is nothing else just a software package which enables us to create and maintain a database. The development of cloud computing affected on DBMS too, and now a cloud provider should offer a stable cloud database in order to convince the customers. Conventional DBMS are not constructed for handling cloud computing demands, and the cloud providers now should be able to offer this migration with cheap and low cost, and help customers planning to use this cloud DBMS for a long time (Gelogo, Y., E., Lee, 2012). First concepts of cloud

DBMS has started early in 1960's, and the point was not changing the function way but only enable users to exchange the data over the network. What makes the difference between conventional DBMS and cloud DBMS is scalability which is the most important characteristics of the cloud DBMS (Gelogo, Y., E., Lee, 2012). With this scalability cloud DBMS are able to handle data and process them without any obstacle, which for a conventional DBMS would be extremely exhaust. Nowadays cloud DBMS have reached a high point of scalability, however cloud DBMS providers are still looking to offer something more stable, and based on the development of technology this is expected to happen. As we mentioned above the ball should be in software side, because they should be able to absorb full advantages of cloud resources. Cloud DBMS now is working on a new model, which is trying to combine different services like data structure and data query language (Gelogo, Y., E., Lee, 2012). Many programming companies are looking forward to suit many programming languages in cloud platform, and this achievement will help the users to save the time on developing cloud DBMS. Despite the benefits that cloud DBMS have, there are still users that have fright migration on it, and the biggest fright is in security section. This is the biggest fright because monitoring of cloud DBMS is more difficult, and as we know the security becomes more serious when we use few virtual machines, which are connected to a cloud DBMS. It becomes more serious because at the same time a database will be accessed by a few number of application, and this might be able to access database without alerting or noticing. So in this situation this will harm our data, and will harm the integral structure of the data, otherwise will put the system in a condition which is known as jeopardy or danger condition (Soror, A., A., Minhas, Aboulnaga, Salem, Kokosielis, and Kamath, 2010), (Watt, A., 2012). When we are in this point, now we should be take care about the security. To take care about the security the only choice is to employ database auditing. Of course we always talk about software which are able to do this auditing of a database such as: analyzing and reporting all activities regarding on database access, notably the doubting database access. All these reports, should be send in another database which is not located on cloud, or should be presented to a responsible person which will maintain the security of our cloud DBMS, despite this we can configure algorithms which can alerts us for all doubting access on our databases, and sure then than measures will be taken to all these threats. Based on researches, this is a new field and all doors are opened for new achievement that happen in this field. This is an important research area because as we mentioned cloud DBMS are designed to run in thousands to millions of nodes, and are able to serve data range starting from petabytes and up. If we try to compare with traditional DBMS, cloud DBMS will offer us less capability of querying and often it might be weak regarding to the consistence, but the scale is bigger offering us availability, elasticity, and load-balancing better than none. A desirable field from all programmers in the world is a technology which will enable us to combine DBMS capability with the cloud scale, and it might be considered too that the research doors are opened for it too (Hashizume, K., Rosado, Fernández-Medina and Fernandez, 2013).

4. Challenges to Cloud Database

To provide a successful cloud database we have to opposite with few challenges, however if we arrive to handle all these challenges we will have a strong and sustainable cloud database (Shehri, W., A., 2013). Next we will mention some of these challenges and we will give description for them.

4.1 Internet speed

As we mentioned above cloud computing is communication between devices over long distances. While our data are in distance, the first thing that we should provide is sustainable internet connection, and a well speed of internet. Internet speed affects directly on the performance of cloud database and it could be considered as a barrier of performance. In cloud databases we send queries and we receive results, while sending queries doesn't require too much internet they will be sent to the database very fast, the problem is with receiving of results while it takes time to retrieve data which directly depends on the internet connection. The solution for this challenge is implementing fast speed cables, always trying to use the latest technology of cables such as fiber technology which is so far the fastest way of communication.

4.2 Query and Transactional Workloads

We have another challenge which is about query workload and transactional workloads. While we are able to control the transactions we are not able to control query workload, because it depends on the number of queries and we don't know the number of the users that are waiting to execute queries. The solution for this challenge is applying parallel processes where a query can be executed in different CPUs giving us the feeling that the query is being executed by a single process.

4.3 Multi-Tenancy

Multi-tenancy is when a single software instance serves hundred users at the same time. In addition cloud providers always think that increasing the number of machines will affect in providing better results but increasing the number of machines it is not always the best solution, because growing the number of machines affects in efficiency of database. As it is mentioned on (Shehri, W., A., 2013) there is a solution which by creating virtual machines for each database but this is not considered as successfully, another idea which is considered as a better solution is to use the same database in different virtual machines which gives us higher performance and speed, and the cloud database can be considered as a well multi-tenancy database (Shehri, W., A., 2013).

4.4 Elastic Scalability

We always try to find better cloud databases which are able to handle any sort of workloads, based on (Shehri, W., A., 2013) one of the best cloud database is considered to be pay per use database and elasticity database. As we mentioned above we always talk about big data which increases day after day, and a critical point for cloud database is when workload becomes increased. The solution for this, is by providing elasticity, in addition elasticity allows a consolidation of the system to consume less performance during high workloads (Shehri, W., A., 2013). Another reason why elasticity of a cloud database is important is about live migration which is the migration from centralized databases into the cloud.

4.5 Privacy

As mentioned above, the privacy is the most important issue of users that plan migrating their databases to the cloud. Due to the fact that cloud databases are accessed and available through the network, it is an interested and important place for hackers trying to break the system even if there is no important information. The current solutions are applying encryption keys to the data stored in cloud databases. The encryption provides additional security, largely reducing the risk of information leak in case of an unauthorized access to the data stored in a cloud database.

5. Cloud DBMS Architecture (SCALEDB)

In this section we give a detailed description of the architecture of a cloud DBMS. The main focus is in SCALEDB, which is a storage engine that is pluggable and compatible with MySQL applications and tools. SCALEDB works on that way that nodes are clustered, and each node in the cluster has access to the entire database. On increased load, additional nodes can be added to increase the scalability and performance. The following figure shows the SCALEDB architecture, which is based on general architecture principles used by cloud DBMS.

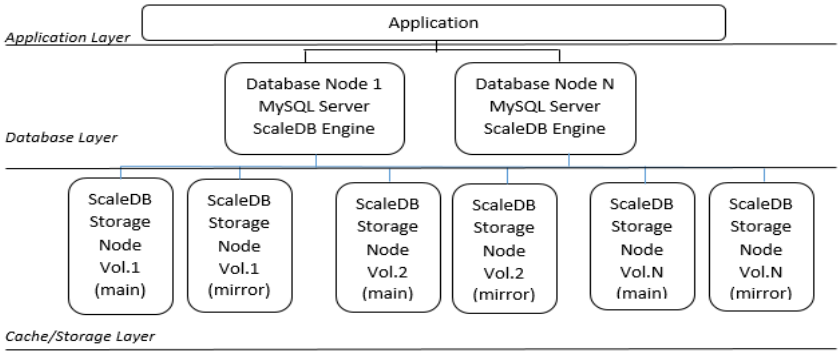

Fig. 4. ScaleDB Architecture (cloud platform)

In this architecture we can see three layers, starting with Storage layer as the first layer, second layer is database layer, and the third layer is Application Layer. In the following we give description how they work. The first layer or storage layer is otherwise known as the cache layer, it is the exact place where the storage is plugged in. It may be local or in cloud, and the most important thing is the recovery of the data which is based on mirroring techniques. If the first drive crashes, the data is available in the second drive. The second layer is the database layer which is separated in multiple nodes, each node having access to the entire database. Each of these nodes works in shared architecture, they share information between them, and if we arrive in a point where the database usage increases then we can simply add new nodes without any problem. We notice that this layer is the splitter between the storage and the application layer. The third, Application layer, is the exact place where the applications are stored. It may be a web server, an application server or a load balancer. The parallelization software is placed in this layer. The three main characteristics of SCALEDB are: it doesn't support distributed transactions, it is not an open source application, and it doesn't support stored procedures (Shehri, W., A., 2013).

6. Conclusion

In this paper, we review database management systems offered as services in cloud platforms. We started with current achievements of cloud databases, continuing with three main characteristics which helps the users to clarify the idea of moving from traditional DBMS to a cloud DBMS, suggesting to readers to migrate to the cloud. At the end we explain a cloud DBMS architecture which is based on SCALEDB and it represents general architecture principles of cloud DBMS. We can conclude that the best choice is moving into the cloud, because it is much cheaper than other services, and it is a long time planning strategy of your work. Additionally, cloud database services can offer more secure environments since it is in the cloud provider responsibility to keep our data secured and recoverable any time that we need for them. Finally, moving to the cloud is the best choice for a company because it decreases investments on storage and network resources that are not used optimally. The pay-as-you-go model used by cloud providers allows more efficient use of these resources by sharing them across customers, and thus it is cheaper because customers pay only for services they are effectively using.

References

Gelogo, E., Y., Lee, S., (2012). Database management system as a cloud service. *International Journal of Future Generation Communication and Networking*, 5.

Curino, C., Jones, C., P., E., Popa, A., R., Malviya, N., M., Wu, E., Madden, S., Balakrishnan, H., Zeldovich, N., (2011). Relational cloud: a database-as-a-service for the cloud. MIT Open Access Articles.

Abadi, J. D. (2009). Data management in the cloud: limitations and opportunities. bulletin of the IEEE computer society technical committee on data engineering.

Kadam, M., Tambe, Sh., Jidge, P., Tayade, E., (2014). Cloud service based on database management system. *International Journal of Engineering Research and Applications*, *4*, 303-306.

Hogan, M., (2008). Cloud Computing & Databases. How databases can meet the demands of cloud computing. ScaleDB Inc.

Meng-Ju, H., Chao-Rui, C., Li-Yung, H., Jan-Jan, W., Pangfeng, L., (2011). SQLMR: A scalable database management system for cloud computing. In Proceedings of ICPP.

Mike, H., (2009). Database Virtualization and the Cloud. ScaleDB Inc.

Buyya. R., Broberg, J., and Goscinski, A. (2011). Cloud computing principles and paradigms. A Jon Wiley & Sons, Inc. Publication.

Chang, F., Dean, J., Ghemawat, S., Hsieh, W., Wallach, D., Burrows, M., Chandra, T., Fikes, A., and Gruber, R., (2006). Bigtable: A distributed storage system for structured data. In OSDI.

Curino, C., Jones, E., Zhang, Y., and Madden, S., (2010) Schism: A Workload-Driven Approach to Database Replication and Partitioning. In VLDB.

Kraska, T., Hentschel, M., Alonso, G., and Kossmann, D., (2009). Consistency rationing in the cloud: Pay only when it matters. *PVLDB*, 2(1).

Soror, A., A., Minhas, U., F., Aboulnaga, A., Salem, K., Kokosielis, P., and Kamath, S., (2010). Automatic virtual machine configuration for database workloads. ACM Trans. *Database Syst*, *35(1)*.

Watt, A. (2012). Characteristics and benefits of database. Database design.

Hashizume, K., Rosado, G., D., Fernández-Medina, E., and Fernandez, B., E., (2013). An analysis of security issues for cloud computing. Hashizume et al. Journal of Internet Services and Applications.

Shehri, W., A., (2013). Cloud Database, Database as a service. Department of computing, Macquarie University, Sydney, NSW 2109, Australia.

Ekanayake, J., Geoffrey, F., (2010). High performance parallel computing with clouds and cloud technologies. Department of Computer Science. Indiana University, Bloomington, IN 47404, USA.