CLOUD DATABASE
DATABASE AS A SERVICE

Waleed Al Shehri

Department of Computing, Macquarie University
Sydney, NSW 2109, Australia
waleed.alshehri@students.mq.edu.au

ABSTRACT

Cloud computing has been the most adoptable technology in the recent times, and the database has also moved to cloud computing now, so we will look into the details of database as a service and its functioning. This paper includes all the basic information about the database as a service. The working of database as a service and the challenges it is facing are discussed with an appropriate. The structure of database in cloud computing and its working in collaboration with nodes is observed under database as a service. This paper also will highlight the important things to note down before adopting a database as a service provides that is best amongst the other. The advantages and disadvantages of database as a service will let you to decide either to use database as a service or not. Database as a service has already been adopted by many e-commerce companies and those companies are getting benefits from this service.

KEYWORDS

Database, cloud computing, Virtualization, Database as a Service (DBaaS).

1. INTRODUCTION

A database can be accessed by the clients via the internet from the cloud database service provider and is deliverable to the users when they demand it. In other words, cloud database is designed for virtualized computer environment. The cloud database is implemented using cloud computing that means utilizing the software and hardware resources of the cloud computing service provider. Cloud computing is growing at a very high pace in the IT industry around the world. Many companies have started moving towards cloud computing and accessing their data from cloud database. A survey has shown that almost 36 percent of the companies are running applications through cloud services (Mimecast Survey, 2011). Cloud computing can be referred as a new dimension in IT world in terms of cost saving and faster application performance. This trend of the companies shows that in the near future, companies will start relying on the cloud applications. Cloud database is mostly used as a service. It is also called Database as a Service (DBaaS).

The cloud database will become the most adopted technology for storing huge data by many companies of the world. It is not as simple as taking the relational database and deploying it over a cloud server. It is more than that. It means that adding of additional nodes when required online, and increasing the performance of the database. There is need to distribute the data over different

data centers distributed over different locations. The database must be accessible all the time so that the user can get the data whenever he needs. The cloud database must be easy to manage and it should reduce the costs as well (Curino, Madden, and et.al.). Cloud computing is very efficient in recovering the information after a disaster in the database.

The usage patterns over the cloud database are invented as the requirements and the advancement in the technology is increasing. At the beginning of the cloud database, there was only read facility available to the customer accessing the cloud database. However, on the demands of the customer requests, write query was also involved. This all was possible by the introduction of Web 2.0. It is observed that the number of read requests in the database is still greater than the write request. But in the near times, the number of read will also increase in the cloud database as the business applications are also depending on cloud computing (Hogan, 2008). This trend has started shrinking the gaps between the read and write requests to cloud database.

2. STRUCTURE OF CLOUD DATABASE

The cloud database holds the data on different data centers located at different locations. This makes the cloud database structure different from the rational database management system. This makes the structure of the cloud database a complex one. There are multiple nodes across a cloud database, designed for query services, for data centers that are located in different geological locations and the corporate data centers as well. This is linking is mandatory for the easy and complete access of the database over the cloud services. There are different methods for accessing the database over the cloud services, the user can access it via computer through the internet, or a user using a mobile phone can access the cloud database via 3G or 4G services (Pizzete and Cabot 2012). To better understand the structure of the cloud database we will demonstrate the example of a Business Intelligence application. The BI applications are used for storing huge data as the corporations use it for storing data for their customers.

Here we assume that the user is accessing the cloud database from a computer through the internet. The internet is the joining point; that act as a bridge among the data centers, cloud data centers and the user who is accessing the data. It is important to note here that only a single node is not used in cloud database; however there are different nodes, that are used for the cloud database (Curino, Madden, and et.al.). For this purpose, peer-to-peer communications are preferred. The purpose to adopt peer-to-peer communication is that, a single node can handle any sort of the query implemented by the user. This seems complex, but an easy solution for this sort of node system that each node in the cloud database has the map to the data stored in each node. This map to the data stored helps in the easy access of data for the specific query.

2.1 Overview

Once the query is generated from the user via computer, the node first decides the sort of query, and which node will be best for the query. After the query is identified by the node, then it is transferred to that specific node. Then the specific node takes care of the query and responds to the user. For example, when the query is received then maybe it is first sent to Node 1, then Node 1 identifies that which Node will solve the query will be suitable. May be Node 7 holds the data, Node 1 will send the query to Node 7 after checking the data map. Once the query is sent to the specific query, then data is directly sent to the user without any further delay. The figure below shows the basic architecture of the cloud database; or it can be considered as an overview.
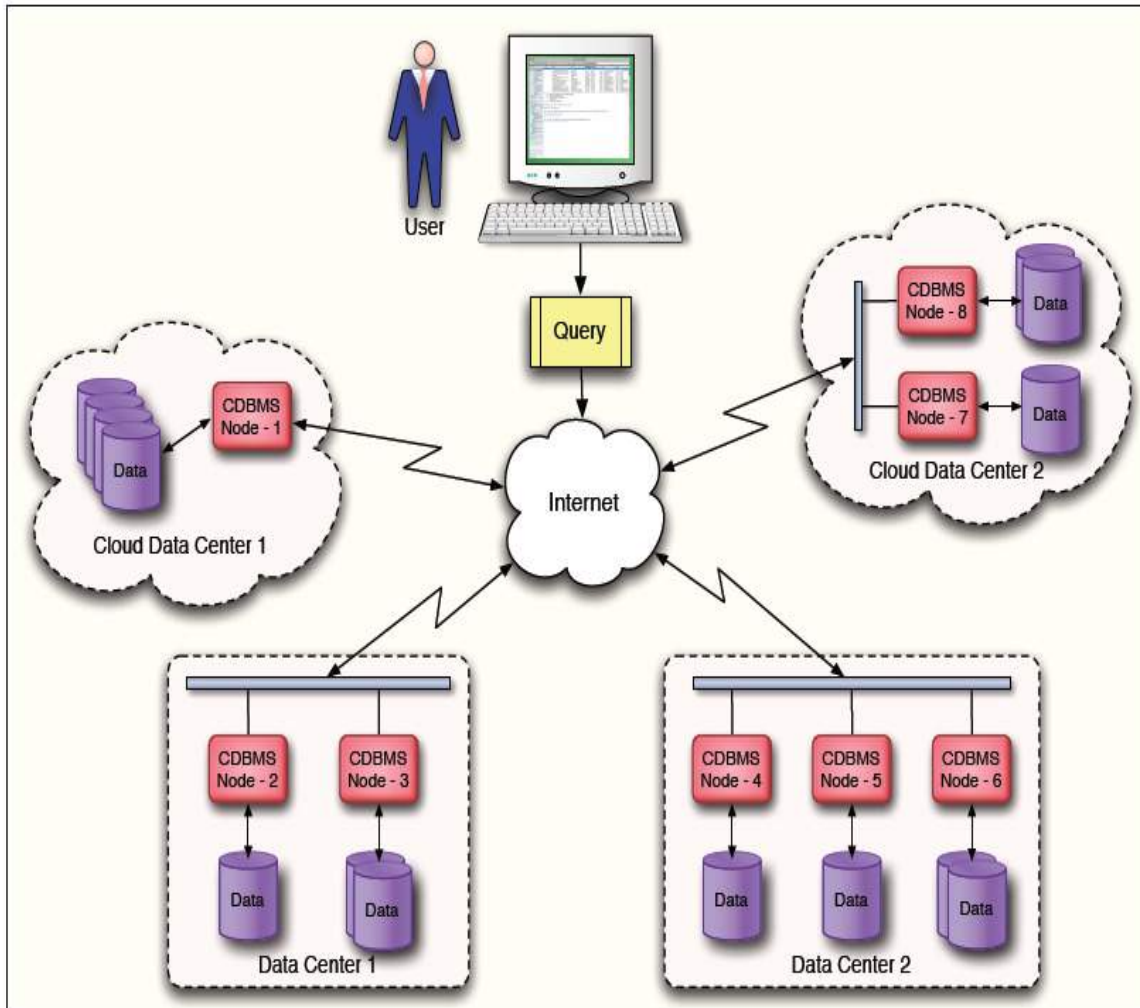
Figure 1. Structure of Cloud

2.2 Working of Nodes

In this section we will be discussing about the working of a node in the cloud database management system. The working process of node in the CDBMS for accessing data that is stored in the database is something like when a query is given to the node, there are two options to a node available, either to access the data directly from the database or by getting it from replicating database. The replicated database is not accessed all the time because it is meant for emergency purposes, when the database fails to perform (Bloor, 2011). Mostly the node gets data from the database as it increases the performance of fetching data. In CDBMS; the application data is stored in the applications. The CDBMS accesses the data from the files directly. If the node accesses the data directly, then the node keeps a metadata map of the file from which the application data was acquired. The following figure shows the working of a node in the cloud database management system
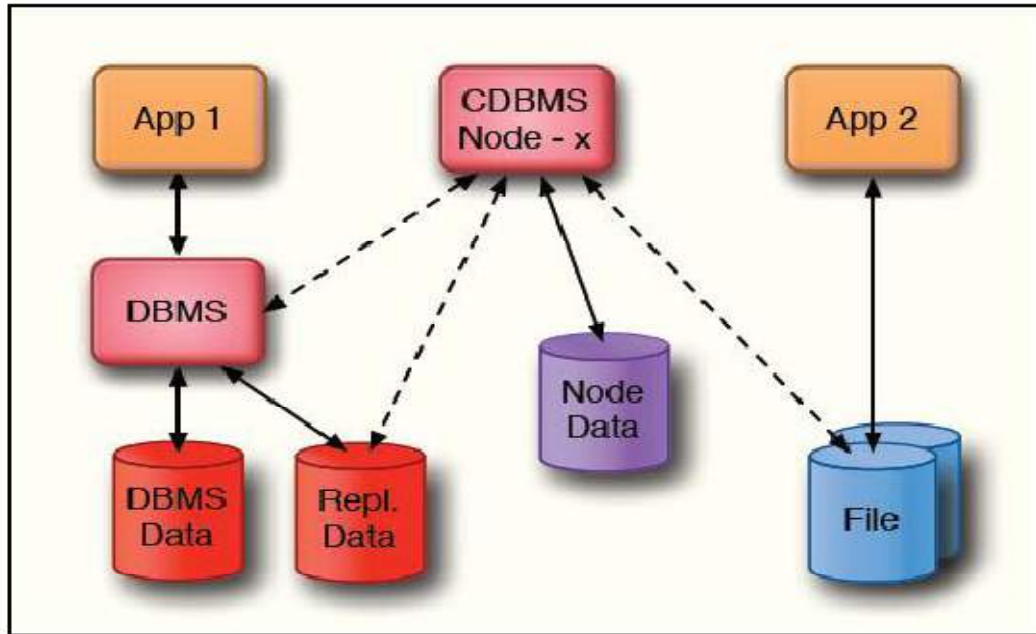
Figure 2. Working of Node

The above figure shows the working of a node for fetching data from DBMS data and files. Moreover, the CDBMS will also maintain its database for storing the data that is being frequently used by the nodes. This improves the performance of CDBMS.

2.3 Node Splitting

If we talk about the local data that is available in the data center at one place, and there is capacity for tera bytes data, then it is very easy for the BI applications to access data without any decay in the performance of the database. On the other hand, if this much huge that is to be handles by the cloud system, then it is very hard for the cloud to manage the increasing number of queries. This can be a complex mechanism for the cloud DBMS. However, there are multiple nodes in a data center, and every data center that is located on the other geological location may have many nodes (Bloor, 2011).

Most of the larger organizations like marts prefer to use cloud services for database. It is a good practice but there are scalability issues with it. As the cloud database may have a huge number of queries as expected, then handling more queries, the CDBMS may face performance issues. It is known that there are many nodes in a cloud DBMS, but these nodes are not enough all the times. Mostly the number of queries keeps on increasing. This overload of queries needs to be handled immediately. For this purpose, CDBMS instantly initiate a new node that shares the load of queries to the database.

This concept needs to be elaborated with the use of an example that is given below. In the figure we can see that the Node A is handling the data for the database, A2, A3, A4 and on the other hand files for A1 and A5. When the node gets more number of queries creating a burden, then the same node splits up into a new node who will distribute the queries with the original node. After the splitting process, the original Node A will be handling data for files A1 and databases A2 and

A3. On the other hands, the new Node A' will handle A4 and A5. This is a very good practice and this scalability of cloud databases makes it possible to handle huge amount of data.
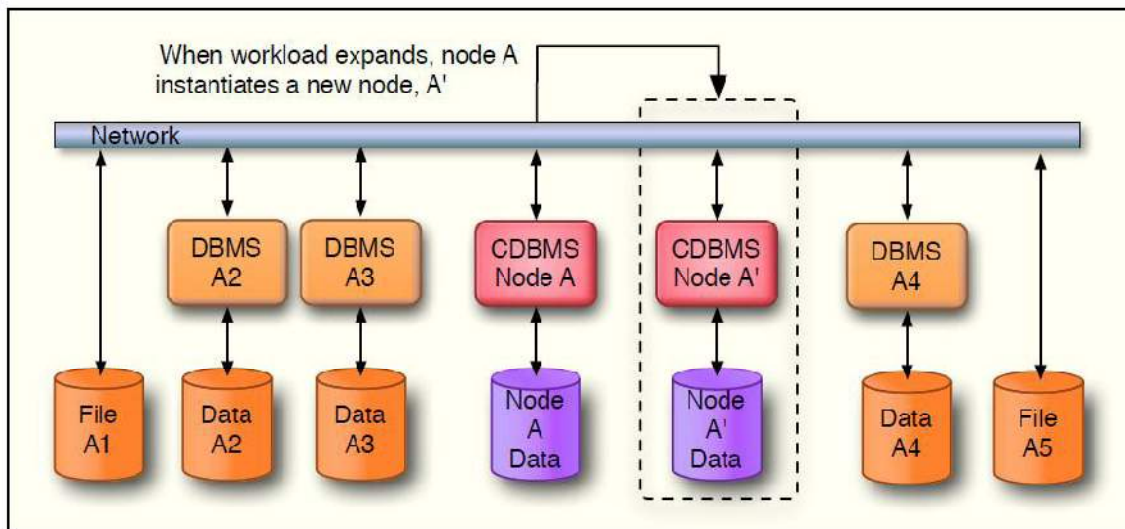

Figure 3. Node Splitting

This Node splitting practice keeps the CDBMS perform well even the number of queries keep on increasing. As the number of queries will increase, the number of splitting nodes will increase as well who will work to distribute the load of queries in the node. When Node A splits up into a same Node A' then Node A will keep a record of all the workload distribution. This record helps in the distribution of the queries.

The splitting nature of the cloud database helps in handling a number of queries for keeping up the performance of the database. Once the queries have been resolved, those split queries are joint back to the Main Node A.

2.4 Distributed Queries

If we consider a database of a major company who is holding a large amount of data distributed like, products, customers, staff and company policies, then in this case different sorts of queries can be involved to get data (Curino, Madden, and et.al.). In CDBMS these different entities may end up in different applications. Resolve to each query; different nodes may be involved. In CDBMS; there are different methods for storing data in DBMS like in query oriented database or column store database. However the most effective way to handle the database is by having distributed queries.

The distributed query can be understood as the combination of many queries, and each query will make contact to each distributed node for the retrieval of the information. As there are different queries; so the number of results can be multiplied as well (Bloor, 2011). As the answer that are distributed; they are joined at the end.

Let us consider an example of the following distributed query that is further divided into sub queries. The query is generated from the computer via the internet, the query is further divided into sub queries; each sub query is forwarded to the specific node. In the following example, the Sub Query 1 and Sub Query 2 are directed to Node 2 of CDBMS. The sub query 3 is moved to

Node 5 and sub query is moved to Node8. Once the nodes are resolved, the answers are also returned in distributed form. Those answers of the nodes are combined and then sent back to the user.
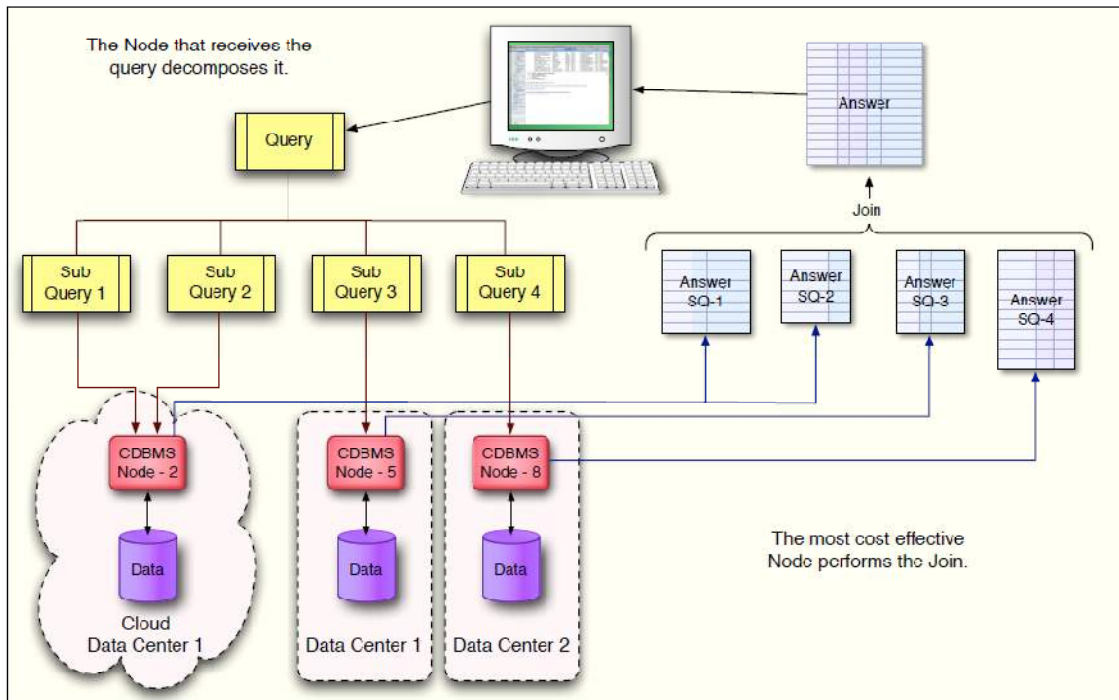

Figure 4. Splitted Queries

3. CLOUD DATABASE SERVICE

Many different cloud database service providers are working who provide database as a service that is further divided into major three categories. There are rational database, non-rational database and operating virtual machine loaded with local database software like SQL.

There are different companies offering database as a service, DBaaS like Amazon RDS, Microsoft SQL Azure, Google AppEngine Datastore and Amazon SimpleDB (Pizzete and Cabot 2012). Each service provider is different from the other depending upon the quality and sort of services being provided. There are certain parameters that can be used to select the best service that will suit for your company. This is not limited to a certain company; these parameters can help in deciding the best service provider depending upon the requirements of any company.

3.1 Choosing best DBaaS

The selecting of DBaaS depends not only on the services being provided by the company, but it also depends on the requirements of the company as well. There are certain parameters that can be taken as a guide to choose the best DBaas.

3.2 Data Sizing

Every DBaaS provider has a different capacity of storing data on the database. The data sizing is very important as the company will need to be sure about the size of data that it will be stored in its database. For example, the Amazon RDS allows the user to store up to 1TB of data in one database on the other hand SQL Azure offers only 50GB of data for one database.

3.3 Portability

The database should be portable as the database should never be out of the access of the user. The service provider may go out of business, so the database and the data stored can be destroyed. There should be an emergency plan if such things happen. This can be resolved by taking cloud services from other companies as well so that the database is accessible even in the case of emergency.

3.4 Transaction Capabilities

The transaction capabilities are the major feature of the cloud database as the completion of the transaction is very important for the user. The user must be aware if the transaction has been successful or not. There are companies who mostly do transact money, in this situation the complete read and write operations must be accomplished. The user needs a guarantee of the transaction he made, and this sort of transaction is called an ACID transaction (Pizzete and Cabot 2012). If there is no need of the guarantee then the transactions can be made by non ACID transactions. This will be faster as well.

3.5 Configurability

There are many databases that can easily configurable by the user as most of the configuration are done by the service provider. In this way there are very less options available left to the administrator of the database and he can easily manage the database without more efforts.

3.6 Database Accessibility

As there are different number of databases, the mechanism for accessing the database are different as well. The first method is the one that is RDBMS being offered through the standards of the industry drivers such as Java Database Connectivity. The motive of this driver is that allows the external connection to access the services through the standard connection. The second accessibility of the database is that by the usage of interfaces or protocols like, Service-Oriented Architecture (SOA) and SOAP or rest (Pizzete and Cabot 2012). These interfaces use HTTP and some new API definition.

3.7. Certification and Accreditation

It is better to get the services of the cloud database provider, who have got certification and accreditation. It helps in mitigating the risks of services for the company to avoid any inconvenience. The companies who have certifications like FISMA can be considered reliable as compared to other DBaaS provider.

3.8 Data Integrity, Security and Storage Location

Security has been the major threat to the data stored in the cloud storage. The security also depends on the encryption methods used and the storage locations of the data (Hacigumus, Iyer and Mehrotra 2004). The data is stored in the different locations in data centers.

	Amazon RDS (MySQL)	Microsoft SQL Azure	Google Datastore	Amazon SimpleDB
Type	RDBMS	RDBMS	NoSQL	NoSQL
Maximum amount of data that can be stored	1 terabyte per database[2]	50 gigabytes per database[3]	Not published for entire database, but 1 MB limit on a subset of data (called an entity). Limit to the number of indexes.	10 gigabytes per database domain (roughly equivalent to an RDBMS table)[4]
Ease of software portability with similar, locally hosted capability	High. MySQL instantiation in cloud is very similar to the local instantiated version.	High. Most SQL Server features are available in SQL Azure.	Medium/Low. Requires Java Data Objects or Datastore-specific interface and use of App Engine.	Medium. Requires SimpleDB-specific interface.
Transaction capabilities	Yes	Yes	Yes	Yes
Configurability and ability to tune database	High. MySQL instantiation in cloud.	Medium. Can create indexes and stored procedures, but no control over memory allocation or similar resources.	Low	Low
Database accessible as "stand-alone" offering	Yes	Yes	No. Requires Google App Engine application layer.	Yes
FISMA Certified	No	No	No	No
Can designate where data is stored (e.g., region or data center)	Yes	Yes	No	Yes
Replication	Yes	Yes	Yes	Yes

Figure 5. Common Consideration

4. CHALLENGES TO CLOUD DATABASE

The implementation of cloud database has certain challenges in its implementation and its successful working. However, after all these challenges, the cloud database is becoming the best option for the companies. Below are some of the challenges to the cloud computing.

4.1 Internet Speed

The speed of data transfer in the data center is comparatively very high as compare to the speed of the internet that is used to access the data center. This is a barrier to the performance of the cloud database. This affects the performance of the cloud database (Bloor, 2011). The queries sent to the database are very fast, but the time taken to retrieve data from data center depends on the speed of the internet. The solution to this challenge is that to have faster speed cables, but that will cost very high and the motive of having a cloud database will waste.

4.2 Query and Transactional Workloads

There is a major difference between the query workload and the transaction work load. When we talk about the transactional workload, we can get an estimate about the time that will be required while on the other hand, we cannot estimate about the time of query workload. In query workload, it depends on the number of queries, and it is not known how many users will be there who will be making queries to the database.

4.3 Multi-Tenancy

There may be a database and a workload that needs to be handled, but the main thing to ponder over is that what is the best way to get the maximum perform from the given machine. In this regard, it is important that the number of machines should be lesser and the efficiency should not decrease. The system should be able to understand the number of hardware resources that are required for each of the workload. The workloads may be located on the same machines and which mechanism is used when they need to be joined. The best solution for this is to make virtual machines for each database and many virtual machines for a number of databases built on the same machine (Curino, Madden, and et.al.). There are more machines are required let's say 2 to 3 machines that will be required to share the same workload. This eventually reduces the performance and the speed 6 to 10 times. The reason behind this lower performance is that each of the virtual machine has its own operating system and its own database. When these two major components are separate for each virtual machine, each virtual machine has its own buffer loop. The better idea is to use the same database server on different machines that will increase the performance as well.

4.4 Elastic Scalability

As we are talking about the cloud database, then a good cloud database as a service is the one that can handle any sort of the work load. However in the cloud database, the problem arises when the workload increases the capacity of the system. The cloud database must be able to scale out itself when the workload increases. The scaling out of the database helps in the best performance and efficiency of the cloud database.

4.5 Privacy

Privacy has been the most important issue when it comes to cloud computing. The cloud computing is a more advanced in terms of the accessibility to the users and hackers who like to break into the system. The privacy in the cloud database is the very important thing that keeps the record of the customers of the companies (Curino, Madden, and et.al.). The companies cannot afford to leak out the information that is stored in their database. If there is encryption of data in database, then it is quite easy to store in a secure way.

5. Advantages of Cloud Database

The cloud computing has given a new dimension to IT industry and the companies are looking to adopt cloud services rather than investing a huge money in getting the infrastructure for own database system. This advent in computing and cloud computing, the cloud database is also picking up its pace in making its permanent place in IT world. There are a number of advantages that make it preferable and adoptable by a huge number of companies for its matchless services in

a very cost saving manner. If the companies do not get the services of a cloud database, then they will have to invest huge money for setting up their own data centers and then hiring separate staff to manage and take care of all the data center processes. Here are few advantages of adopting cloud database.

I. The technology has changed the way of business, and now the people use to shop over the internet and they rely on shopping for saving their time. This change in the business has let the companies think about the fastest way they can do business over the internet. There was a time when software needed to be installed to access the database of the company but now a day the employees even don't have time to install software on their computer rather they prefer to use a ready to available resources. They prefer to use the cloud database so that they can access the information stored in their database without wasting any time.

II. The other advantage of using a cloud database is that it saves a lot of money. The company does not need to invest money in setting up their own data centers and then managing it by hiring extra staff for this purpose. Moreover, after setting up a data center, the company will need to buy the softwares as well and their maintenance is also required.

III. The cloud database service providers of DBaaS providers also make the customer free from the tensions of making any immediate changes in the database. On the other hand, the cloud database providers also offer scalability on the peak times that does not let the performance of the company go down.

IV. Cloud computing has given the freedom to access the information from anywhere without any boundaries of getting to your personal computer at home. This makes it a very powerful technology and the companies prefer it as the customers, employees or the authorities of the companies can get the formation they want from anywhere at any time.

V. There are many other benefits of cloud database as well, that makes it the best option available to the larger organizations and companies who need to hold terabytes of data. The cloud database makes the availability of data possible anytime from anywhere.

6. DISADVANTAGES OF CLOUD DATABASE

As there are advantages of using a cloud database, there are disadvantages as well. The disadvantages can be alarming sometimes for the companies.

I. The companies have to pay for the usage of the cloud database as per decided. Every time the data is transferred from the database, the company will have to pay each time. If the traffic of the company for transferring data with the database is high then the company may be paying than its expectations.

II. The other disadvantage of using a cloud database is that, we do not have a full control over the server where our database is being held. We do not have the control over the softwares installed on those computers. You cannot do anything to make the security of cloud database strong. The client will have to rely on the provider only. The security issues can be a big problem for the companies.

III.   The data you have hosted on the cloud database is totally dependent on the service provider. The data and information about a company are the most important asset for the organization. The organizations cannot afford to lose its information about its customers and company policies. If the information is given in the wrong hands then the company or the organization may face heavy losses.

IV.   As there are masses of data hosted on the cloud database so it is very difficult to transfer that data to your computer. For this purpose, internet speed must be high. On the other hand, the traditional database can transfer data at a very high speed.

V.   If the client wants to switch database from one service provider to new one, then he may face problems. The reason is that each service provides use their own methods and techniques for storing data. The organization must be very careful about the selection of DBaaS provider.

VI.   In case of cloud database, the data is to be fetched via internet, so if the server is down, then it may cause inability to access the data from the server. This causes huge losses when the information is not available when needed.

7. CONCLUSION

In conclusion, this report has outlined the concept of cloud database, and presented some of its main aspects.The companies have started relying on the cloud computing for several reasons and a trend has started by adopting cloud computing services for better and faster availability of the information rather than setting up an individual data center for each organization or the company. The organizations always look for the ways that is effective and is cost saving. The same is the process with the database. Earlier the organizations set up their own data center and have their traditional database. Now the cloud database has evolved a new dimension Database as a Service (DBaaS). This allows the companies and organizations to use the resources of the DBaaS providers and without any hassle to invest and maintain the hardware and software for their data centers that hold all the information in the database. They get services from DBaaS provider and enjoy the freedom of 24/7 available database. There are advantages and disadvantages as well; however the adoption the cloud database has proven that the advantages are more than the disadvantages. The cloud database services have offered many benefits and different companies are in the race. The organization chooses the one that suits its requirements.

REFERENCES

[1]   Bloor, R. 2011. WHAT IS A CLOUD DATABASE? Retrieved 25th November 2012 from http://www.algebraixdata.com/wordpress/wp-content/uploads/2010/01/AlgebraixWP2011v06.pdf
[2]   Curino, C., Madden, S. and et.al. Relational Cloud: A DatabaseasaService for the Cloud. Retrieved 24th November 2012 from http://www.cidrdb.org/cidr2011/Papers/CIDR11_Paper33.pdf
[3]   Finley, K. 2011. 7 Cloud-Based Database Services. Retrieved 23rd November 2012 from http://readwrite.com/2011/01/12/7-cloud-based-database-service
[4]   Hacigumus, H., Iyer, B. and Mehrotra, S. 2004. Ensuring the Integrity of Encrypted Databases in the Database-as-a-Service Model. Retrieved 24th November 2012 from http://link.springer.com/chapter/10.1007%2F1-4020-8070-0_5?LI=true
[5]   Hacıgumus, H., Iyer, B. and Mehrotra, S. Providing Database as a Service. Retrieved 25th November 2012 from http://archive.systems.ethz.ch/www.systems.ethz.ch/education/past-courses/fs09/HotDMS/pdf/daas.pdf

[6]   Harris, D. 2012. Cloud Databases 101: Who builds 'em and what they do. Retrieved 25th November 2012 from http://gigaom.com/cloud/cloud-databases-101-who-builds-em-and-what-they-do/

[7]   Hogan, M. 2008. Cloud Computing & Databases:How databases can meet the demands of cloud computing. Retrieved 23rd November 2012 from http://www.scaledb.com/pdfs/CloudComputingDaaS.pdf

[8]   Mykletun, E. and Tsudik, G. 2006. Aggregation Queries in the Database-As-a-Service Model. Retrieved 24th November 2012 from http://link.springer.com/chapter/10.1007%2F11805588_7?LI=true

[9]   Oracle. 2011. Retrieved 23rd November 2012 from http://www.oracle.com/technetwork/topics/entarch/oes-refarch-dbaas-508111.pdf

[10]   Pizzete, L. and Cabot, T.2012. Database as a Service: A Marketplace Assessment. Retrieved 23rd November 2012 from http://www.mitre.org/work/tech_papers/2012/11_4727/cloud_database_service_dbaas.pdf

[11]   Postgres Plus. 2012. Cloud Database: Getting started Guide. Retrieved 23rd November 2012 from http://get.enterprisedb.com/docs/Postgres_Plus_Cloud_Database_Getting_Started_Guide.pdf

[12]   Rouse, M. 2012. Cloud Database. Retrieved 25th November 2012 from http://searchcloudapplications.techtarget.com/definition/cloud-database-database-as-a-service

[13]   Saini, G.P. 2011. Cloud Computing: Database as a Service. Retrieved 24th November 2012 from http://cloudcomputing.sys-con.com/node/1985543

[14]   VMware. 2012. Getting Started with Database-as-a-Service. Retrieved 23rd Novermber 2012 from http://www.vmware.com/pdf/vfabric-data-director-20-database-as-a-service-guide.pdf

[15]   Zhang, J. 2011. Database in the Cloud Retrieved 25th November 2012 from http://www.ibm.com/developerworks/data/library/dmmag/DMMag_2011_Issue2/cloudDBaaS/

Image Source

[16]   Bloor, R. (Author). 2011. WHAT IS A CLOUD DATABASE ? Retrieved 25th November 2012 from http://www.algebraixdata.com/wordpress/wp-content/uploads/2010/01/AlgebraixWP2011v06.pdf

[17]   Pizzete, L. and Cabot, T. (Authors). 2012. Database as a Service: A Marketplace Assessment. Retrieved 23rd November 2012 from http://www.mitre.org/work/tech_papers/2012/11_4727/cloud_database_service_dbaas.pdf

AUTHOR

Waleed Al Shehri received his bachelor degree in computer science from King Abdulaziz University, Jeddah, Saudi Arabia (2005), MSc degree in information technology form Macquarie university, Sydney, Australia (2011). His current research interests in databases and cloud computing specially Database as a Service (DBaaS) . Currently working in the Department of IT in Royal Saudi Air Force (RSAF).