# Cloud Based Distributed Databases: The Future Ahead

Arpita Mathur
Assistant Professor: Dept. of Computer Science
Lachoo Memorial College of Science & Technology
Jodhpur, Rajasthan (India)

Mridul Mathur
Assistant Professor: Dept. of Computer Science
Lachoo Memorial College of Science & Technology
Jodhpur, Rajasthan (India)

Pallavi Upadhyay
Assistant Professor: Dept. of Computer Science
Lachoo Memorial College of Science & Technology
Jodhpur, Rajasthan (India)

***Abstract***

***Fault tolerant systems are necessary to be there for distributed databases for data centers or distributed databases requires having fault tolerant system due to the higher data scales supported by current data centers. In a large distributed database usually data resides on servers which are dedicated servers having backups. Therefore, large amount of servers are used for this purpose.***

***In this paper it is proposed that instead internet can be used as backbone where Infrastructure-as-a-service (Iaas) service of cloud can be used for storage servers. The advantage of this will be that storage location is abstracted and database can be accessed from anywhere. However while the storage allocation is abstracted it also brings in performance concerns in a multi tenant cloud environment where by most of the cloud consumers are geographically dispersed. Recent large Web applications make heavy use of distributed storage solutions in order to be able to scale up. Here we propose that static distributed database is spread over cloud making database dynamic.***

**Keywords-distributed database; cloud; geo-redundancy; API**

## I INTRODUCTION

Data storage is diverse at different remote locations in case of distributed databases. Dedicated servers are used to store these databases; therefore many servers are needed by companies to store their large databases. Those servers were static i.e. their location was fixed and the sites where data was distributed were fixed. Companies needed the infrastructure in the very beginning which costs a lot. Cloud Computing allows users to tap into a virtually unlimited pool of computing and storage resources over the Internet (the Cloud) **[4].** Unlike traditional IT, Cloud users typically have little insight or control over the underlying infrastructure, and they must interact with the computing and storage resources via an Application Programming Interface (API) provided by the Cloud vendors. In exchange for those constraints, Cloud users benefit from utility-like costs, scalability, and reliability, as well as the ability to self-provision resources dynamically and pay only for what they use.

We propose in this paper that the cloud's service IaaS i.e. servers can be used to store these databases at low initial cost. The servers and the sites where the data is distributed can be anywhere in the cloud. Their location and number will not be fixed; it can change dynamically as we are using cloud. There will be no limit to storage space and no fixed number of servers. Their number can increase or decrease as the database grows or shrinks.

Therefore by using internet as backbone the physical data position need not be known and database can be accessed from anywhere as cloud services and storage are accessible from anywhere in the world over an Internet connection.

## II CLOUD

Cloud computing is the convergence and evolution of several concepts from virtualization, distributed application design, grid, and enterprise IT management to enable a more flexible approach for deploying and scaling applications.

Cloud promises real costs savings and flexibility to customers. Through cloud computing, a company can rapidly deploy applications where the underlying technology components can expand and contract with the natural ebb and flow of the business life cycle. Traditionally, once an application was deployed it was bound to a particular infrastructure, until the infrastructure was upgraded. The result was low efficiency, utilization, and flexibility. Cloud enablers, such as virtualization and grid computing, allow applications to be dynamically deployed onto the most suitable infrastructure at run time. This elastic aspect of cloud computing allows applications to scale and grow without needing traditional 'fork-lift' upgrades.

IT departments and infrastructure providers are under increasing pressure to provide computing infrastructure at the lowest possible cost **[2]**. In order to do this, the concepts of resource pooling, virtualization, dynamic provisioning, utility and commodity computing must be leveraged to create a public or private cloud that meets these needs. World-class data centers are now being formed that can provide this IaaS in a very efficient manner **[7]**.

With cloud computing, developers are no longer boxed in by physical constraints. For companies if more processing power is needed, it's always there in the cloud—and accessible on a cost-efficient basis. For end users, applications and documents can be accessed wherever he is, whenever he wants. And, with cloud computing, hardware doesn't have to be physically adjacent to a company's office or data center; cloud infrastructure can be located anywhere.

## III.CLOUD DATABASES

It is evident that storage plays a major part in the data center and for cloud services. The storage virtualization plays a key part in the dynamic infrastructure attribute of Cloud Computing. Which means the storage is provisioned and de-allocated on demand and usage needs. In cloud databases, data is stored on multiple dynamic servers, rather than on the dedicated servers used in traditional networked data storage. When storing database, the user sees a virtual server. In reality, the user's data could be stored on any one or more of the computers used to create the cloud. The actual storage location may differ as the cloud dynamically manages available storage space. But even though the location is virtual, the user sees static location for his data and can manage his storage space as if it were connected to his own PC and this complex stuff is hidden from the cloud consumer.

Currently Cloud platforms have very little support for database design related virtualization enhancements. But in future designing databases specific for Cloud especially for private clouds in large enterprises is a sure possibility. In this context the distributed databases are important when you design database applications which need to be delivered using Cloud platform. Cloud database has both financial and security advantages over traditional storage models.

### A. DISTRIBUTED DATABASES OVER CLOUDS

It is hard to deploy databases to a virtualized, grid or distributed environment. In a distributed database cluster, data must either be replicated across the cluster members, or partitioned between them. In either case, adding a machine to the cluster requires data to be copied or moved to the new node. Since this data shipping is a time-consuming and expensive process, databases are unable to be dynamically and efficiently provisioned on demand. The vendors seeking to create public computing clouds or those trying to establish massively parallel, redundant and economical data driven applications needed a way of managing data that was almost infinitely scalable, inherently reliable and cost-effective.

Let us take example of Google's BigTable solution. It developed a relatively simple storage management system that could provide fast access to petabytes of data redundantly distributed across thousands of machines. As shown in figure 1 physically, BigTable resembles a B-tree index-organized table in which branch

and leaf nodes are distributed across multiple machines. Like a B-tree, nodes "split" as they grow and, since nodes are distributed, it can scale across large numbers of machines.

### B. FEATURES OF CLOUD DATABASES

- Complex attributes: Each “row” often can contain different “columns,” and columns may have multiple values or include a more complex nested structure.
- Automatic geo-redundancy or support for high – availability (HA) configuration: To ensure reliable, fast failover in affordable way. For that elements stored in the database are guaranteed to be replicated across multiple data centers.
- Automatic partitioning across multiple hosts and automatic scale-out: As the size of or demand on the data store exceeds the capability of a single host automatic partitioning is done.
- Support for multiple database management system: To support complex topologies.

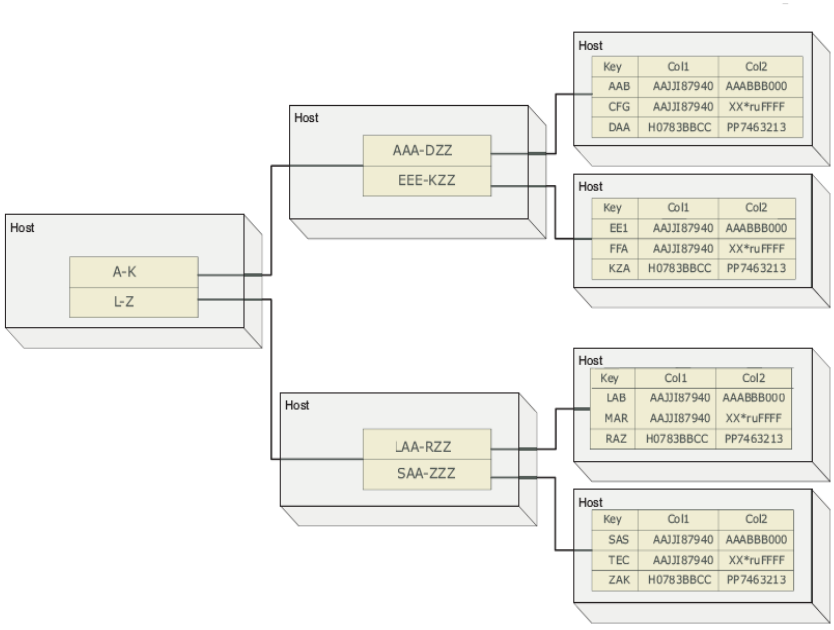


Figure 1: Future database design specific for Cloud

### C. REQUIREMENTS OF CLOUD DATABASES:

1. Fault Tolerance. In the context of distributed databases, one can successfully commit transactions and make progress on a workload even in the face of worker node failure. A fault tolerant distributed DBMS is simply one that does not have to restart a query if one of the nodes involved in query processing fails.
2. Ability to run in a heterogeneous environment. The performance of cloud compute nodes is often not consistent, as all nodes do not attain same performance. A node observing degraded performance would thus have a disproportionate affect on total query latency. A system designed to run in a heterogeneous environment would take appropriate measures to prevent this from occurring.
3. Efficiency. Given that cloud computing pricing is structured in a way so that you pay for only what you use, the price increases linearly with the requisite storage, network bandwidth, and compute power. Efficient software has a direct effect on the bottom line.
4. Ability to operate on encrypted data. Sensitive data may be encrypted before being uploaded to the cloud. In order to prevent unauthorized access to the sensitive data, any application running in the cloud should not have the ability to directly decrypt the data before accessing it.
5. Ability to interface with business intelligence products. As per the other technologies in this case also compatibility is desired. Since variety of data analysis tools like business intelligence tools are already in the market and used by business analysis, the newer technologies and tools must be able to interface with the existing one. So the cloud databases are required to be compatible even to interact with the business analysis tools.

## IV. CASE STUDY

One of the applications of distributed databases over clouds can be in form of network of cellular companies. The cellular companies are having a broad base of clients which used database spread across servers. The servers are dedicated and database is distributed according to cellular companies. If we go for clouds in this case then we may access the customer information through company gateways. The company's database, servers and gateways will be in different distributed clouds and there will be a control cloud on top of all companies' clouds which will control the access of all the distributed clouds.

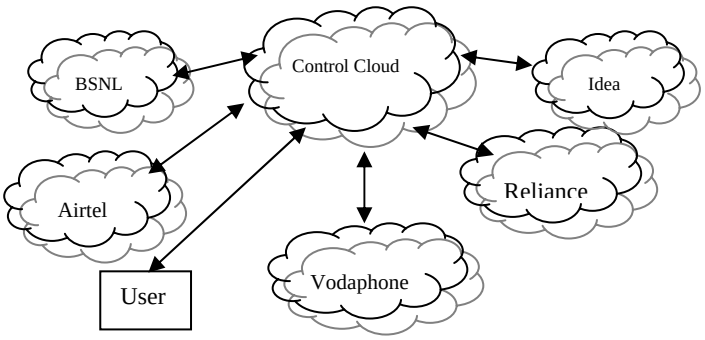


Figure 2: Diagram for cellular distributed cloud database

With reference to number portability in effect to need to know a customer detail is of utmost importance. Thus with the cloud supra structure i.e. having a control domain at top with service provider clouds at first level with distributed databases, the accessibility of required data can be carried out. Henceforth, we can say that distributed database can be configured in a cloud form with having accessibility of data from anywhere over the internet. With recent advances in TRAI (Telephone Regularity authority of India) rules the number portability has taken into force and in this lot of users are switching there providers. According to recent data 1.7 million people have changed there services **[8]**. Here, our proposed scheme plays a significant role because the number and required information can be easily switched or moved through this cluster of clouds which is distributed according to service providers.

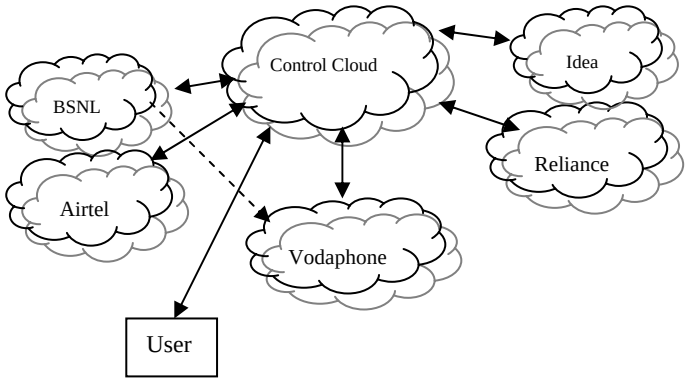


Figure 3: Figure showing user switching there provides.

Figure 3 shows that as the user switches from BSNL to Vodaphone his information is transferred through control cloud from BSNL cloud to Vodaphone cloud as customer data are distributed in clouds according to operator.

## V. CONCLUSION

Today cloud is used mainly for computing purposes. As suggested in this paper clouds can be used with distributed database for handling very large databases maintaining availability, scalability as well as reliability. This will be possible as geographically distributed data is distributed and replicated making data available all the

time. It is needed that the database vendors provide algorithms tailored for virtual storage, especially on the rebalancing based on the disk usage so that skew can be avoided. Cloud databases can be used for data analysis, data warehousing and data mining purposes.

## REFERENCES/BIBLIOGRAPHY


[1] Kephart J.; Chess, David M.; "The Vision of Autonomic Computing", published by the IEEE computer Society, Jan 2003
[2] Armbrust M., Fox A. et al; Above the Clouds: A Berkeley View of Cloud Computing; http://radlab.cs.berkeley.edu/; Feb 10, 2009
[3] R. Agrawal, J. Kiernan, R. Srikant, and Y. Xu, "Order preserving encryption for numeric data". Proc. of SIGMOD, pages 563–574, 2004.
[4] Varia, J.; "Cloud Architectures"; http://jineshvaria.s3.amazonaws .com/ public/cloudarchitectures-varia.pdf ; Jan 2011
[5] Twenty-One Experts Define Cloud Computing; Cloud Computing Journal; http://cloudcomputing.syscon. com/node/612375, July 2010
[6] Geelan J.; A World of Many Clouds; http://cloudcomputing.syscon.com/node/902342; Cloud Computing Journal; Apr 1, 2009
[7] Cloud Database Design, Scale Out Using Shared Nothing Pattern http://www.infoworld.com/d/cloud-computing; Feb 2011
[8] John Ribeiro; Mobile Number Portability http://news.yahoo.com/mobilenumberportabilityinindia; Feb 2011