ISSN: 2277-9655
Scientific Journal Impact Factor: 3.449
(ISRA), Impact Factor: 1.852

IJESRT

INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY

Cloud Databases: Future of Distributed Databases

Vipin Nagar [*1], Kavita Tiwari [2]

[*1, 2] Assistant Professor, Department of Computer Science, Babaria Institute of Technology,Vadodara

vipinnagar810@gmail.com

Abstract

Hardware failures in current data centers are very frequently because of high volume data scales supported. Data replication is the better option for this condition. Distributed database is a concept of distribution data storage at different remote locations. Many of servers are used to store these databases; therefore many of servers are used for this purpose. We can use internet as backbone where Iaas service of cloud can be used for storage servers. The advantage of this will be that exact data position need not be known and database can be accessed from anywhere.

The geographical distribution of data locations is considered in this paper, it is proposed that distributed database at cloud which is a cost-efficient replication scheme across data centers that dynamically take the number of replicas employed per partition to the query load. Now a day's large Web applications make heavy use of distributed storage solutions in order to be able to scale up. As data scales up its availability becomes more crucial and more important, as even reliable hardware may fail. Static database is converted to dynamic database when using cloud..

Keywords: Cloud database.

Introduction

Distributed database is a concept of distributing data storage at different remote locations. Dedicated servers are used to store these databases; therefore many servers are required by organization to store their large databases. Those servers were static i.e. their location was fixed and the sites where data was distributed were fixed. Organization must have appropriate infrastructure to place these fixed server at fixed position.

What we propose in this paper that the cloud's service Infrastructure-as-a-service (IaaS) 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 physical location and number will not be fixed; location can be 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 exact 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.

Why Cloud Computing?

Cloud computing represents a dramatic change in how we store information and how we run computer applications. Instead of running programs and storing data on an individual desktop computer, everything is hosted in the "cloud". A collective set of computers and servers accessed via the Internet. Cloud computing dynamically scalable and virtualized resources are available as a service over the Internet.

Cloud computing provides increased amounts of storage and processing power to run the applications. Cloud computing also shows new ways to access information, process and analyze data, and connect people and resources from any location anywhere in the world.

Cloud computing offers more flexibility in computing power, often at lower costs. With cloud computing, companies don't have to engineer for peak-load capacity, because the peak load can be spread out among the external assets in the cloud. And, because additional cloud resources are always at the ready, companies no longer have to purchase assets. If more processing power is needed, it's

ISSN: 2277-9655
Scientific Journal Impact Factor: 3.449
(ISRA), Impact Factor: 1.852

always there in the cloud—and accessible on a cost-efficient basis.

Applications and documents can be accessed wherever by the end client. The fear of losing data if a computer crashes is no more there. Documents hosted in the cloud always exist, no matter what happens to the user's machine. The advantage of group collaboration is also there with cloud computing for both individuals and organizations.

Cloud computing provide its services in low costs, because the cloud enables more efficient sharing of resources than does traditional network computing. With cloud, you can do individuals and small businesses. And, with cloud computing, hardware doesn't have to be physically placed to a company's office .Cloud infrastructure can be located anywhere.

Cloud databases

With cloud databases, data is stored on multiple third-party servers, rather than on the dedicated servers used in traditional networked data storage. When storing database, the user interface with 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.

Cloud database has both financial and security advantages over traditional storage models. Financially, the cloud's virtual resources are typically cheaper than dedicated physical resources connected to a personal computer or network. Data stored in the cloud is secure from accidental erasure or hardware crashes, because it is duplicated across multiple physical machines.

Why clouds is the future scope of distributed database?

It is difficult to implement databases to a virtualized, distributed environment. In a distributed database cluster, data must either be replicated across the multiple servers, 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 HugeTable solution. It developed a relatively simple storage management system that could provide fast access to petabytes of data that is redundantly distributed across thousands of servers or machines. As shown in figure 1 physically, HugeTable 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.

Features of cloud databases

a. A hierarchical structure resembling a B-tree index: So that rapid lookup via a single key value is provided.
b. Complex attributes: Each "row" often can contain different "columns," and columns may have multiple values or include a more complex nested structure.
c. 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.
d. 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.
e. Support for multiple database management system: To support complex topologies.

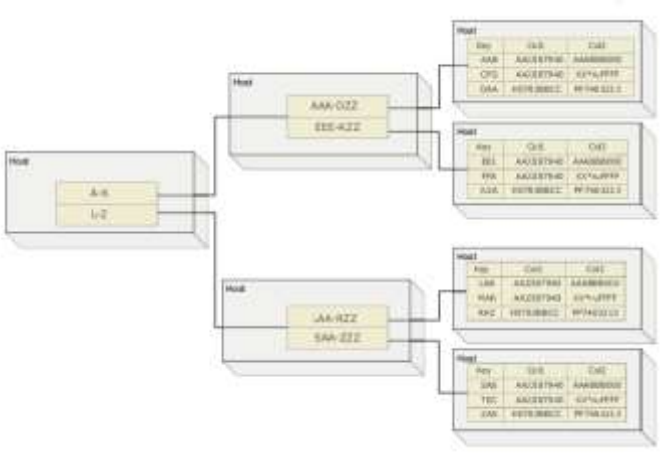

Figure 1: Clouds Future Data Centers

ISSN: 2277-9655
Scientific Journal Impact Factor: 3.449
(ISRA), Impact Factor: 1.852

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. 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.

Drawbacks of Cloud Database

1. Transactional support and referential integrity: Applications using cloud databases are largely responsible for maintaining the integrity of transactions and relationships between "tables."

2. Complex data accesses: Cloud databases; excel at single-row transactions. Most applications use joins and other operations.

3. Business intelligence: Application data has value not only to power applications, but as information for business intelligence. Businesses will not return willingly to the pre-relational database days when business data was locked in impenetrable application data stores.

4. The cloud is not a predictable and stable environment. Performance is also not guaranteed unless running on dedicated nodes which would defeat the purpose of using the cloud.

5. Scaling databases: Scaling a database elastically can be very complex and tedious to manage.

6. Distributed databases are not the same as distributed applications: Maintaining multiple active/master copies of a database in multiple locations and/or clouds requires building logic to handle conflicts, network problems, or latency while attempting to maintain single source of truth at all times.

7. Infrastructure-as-a-Service (IaaS) scenario; the database is expected to support multi-tenancy to enable a cost-effective and operationally efficient framework. It may cause more headaches and management overhead to keep it running.

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 as well as reliability. This will be possible as geographically distributed data is distributed and replicated making data available all the time. Cloud databases can be used for data analysis, data warehousing and data mining purposes.

ISSN: 2277-9655
Scientific Journal Impact Factor: 3.449
(ISRA), Impact Factor: 1.852

Reference

1. *Armbrust M., Fox A. et al; Above the Clouds: A Berkeley View of Cloud Computing;*
2. *Kephart J.; Chess, David M.; "The Vision of Autonomic computing" published by the IEEE computer Society, Jan 2003.*
3. *R. Agrawal, J. Kiernan, R. Srikant, and Y. Xu, "Order preserving encryption for numeric data". Proc. of SIGMOD, pages 563–574, 2004.*
4. *Amazon Web Services. SimpleDB. Web Page.*
5. *Erenben C.; "Cloud Computing: The Economic Imperative"; IBM eSchool News;*
6. *Varia, J.; "Cloud Architectures"; July 2008*
7. *Twenty-One Experts Define Cloud Computing; Cloud Computing Journal, July 2008;*
8. *Geelan J.; A World of Many Clouds; Cloud Computing Journal; Apr 1, 2009*