CLOUD DATABASES

Abstract

In recent times, cloud computing has emerged as one of the most versatile technologies, and as a result, databases have also transitioned to the cloud. In this context, we will delve into the intricacies of Database as a Service (DBaaS) and explore its functioning. This article encompasses all the essential details regarding Database as a Service (DBaaS). The operation of the database as a service and its issues are explored using an appropriate approach. Database as a Service (DBaaS) delves into the architecture of cloud-based databases and its collaborative functionality with nodes. This paper will highlight key considerations when choosing the optimal Database as a Service (DBaaS) solution. It will assess the advantages and disadvantages of DBaaS to assist in making an informed decision about its adoption. Numerous e-commerce enterprises have already embraced DBaaS, experiencing its advantages firsthand.

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

Authors

Dr. M. Laxmaiah
Professor & HOD
Department of CSE-Data Science
CMR Engineering College
Hyderabad, Telangana, India.
laxmanmettu.cse@gmail.com

B. Kumara Swamy
Assistant Professor
Department of CSE-Data Science
CMR Engineering College
Hyderabad, Telangana, India.
bkumaraswamyse@gmail.com

Avudurthi Saikiran
Assistant Professor
Department of CSE-Data Science
CMR Engineering College
Hyderabad, Telangana, India.
saikiran231091@gmail.com

I. INTRODUCTION

The cloud database stores data across multiple data centers situated in different regions, setting it apart from the structure of a traditional relational database management system. A cloud database comprises numerous nodes specifically designed for query services, data centers distributed across various geographical regions, and corporate data centers. There are multiple methods for accessing the database through cloud services, including internet-based access from a computer or mobile access via 3G or 4G services on a mobile phone. To gain a deeper understanding of the architecture of a cloud database, let's consider the scenario of a Business Intelligence application. These BI apps are employed by organizations to manage substantial datasets for their clients. For our example, we'll assume that the user is accessing the cloud database from a computer via the internet. The internet acts as the conduit, bridging connections between data centers, cloud data centers, and the end user who is accessing and viewing the data. For this purpose, peer-to-peer communication is the preferred approach. The objective of employing peer-to-peer communication is to enable a single node to handle any user-implemented query. While it may seem complex, a straightforward solution within this node system is to equip each node in the cloud database with a map detailing the data stored across all nodes. This data map simplifies the retrieval of data for any given query.

A cloud database is a database service created and accessible through a cloud platform. It shares many similarities with traditional databases but enjoys the added advantage of flexibility offered by cloud computing. Users deploy the database by installing software on cloud infrastructure.

Key Features

- A cloud-based database service created and accessible via a cloud platform.
- Empowers enterprise users to host databases without the need for dedicated hardware purchases.
- Can be either self-managed by the user or provided as a managed service by a service provider.
- Offers support for both relational databases (such as MySQL and PostgreSQL) and NoSQL databases (including MongoDB and Apache CouchDB).
- Accessible through a web interface or vendor-provided API.

II. OVERVIEW

When a computer generates a query, the node initially identifies the query type and the most suitable node to process it. Once identified, the node forwards the query to the designated node, which subsequently processes the query and provides a response to the user. For instance, upon receiving a query, it might be initially routed to Node 1, which then assesses the appropriate Node best suited to resolve the query. In this scenario, if Node 7 holds the required data, Node 1, after consulting the data map, transfers the query to Node 7. Once the inquiry is directed to the pertinent Node, the data is promptly delivered to the user without additional delay. The diagram below provides an outline of the fundamental architecture of the cloud database.

1. **Working of Nodes:** When a node is accessible, you can either directly access the data from the database or retrieve it through an alternate node. The replicated database isn't in continuous use; it serves as a backup resource to be called upon when the primary database encounters performance issues. CDBMS applications store application data, and the CDBMS retrieves data directly from the files. Nodes that directly access the data maintain metadata maps for the files from which they acquired application data.

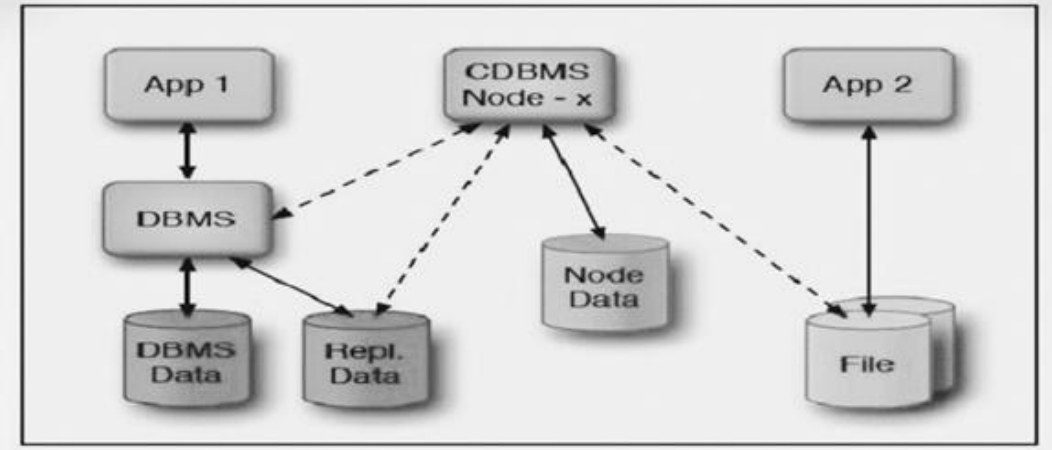

Figure 1: Working of Node

2. **Node Splitting**: BI applications may easily access data without affecting the performance of the database when considering local data within a data canter, where terabytes of data can be handled. On the other hand, managing such a huge amount of data on a cloud system is difficult as it struggles to handle the increasing number of queries, thereby complicating the cloud DBMS. However, data centers have numerous nodes, and those situated in various locations may house a large number of nodes.

   The use of cloud services for databases is frequently preferred by bigger businesses, especially marts, and is usually thought to be advantageous. Scalability problems, however, might appear in this situation. The Cloud Database Management System (CDBMS) may face performance issues as a result of the cloud database's anticipated high volume of queries. Although a cloud DBMS usually has several nodes, they might not always be enough to manage the steadily rising volume of queries.

   It's critical to handle this query overload as soon as possible. To remedy this, CDBMS starts the quick deployment of a new node, which aids in more effectively distributing the query load throughout the database.

   The CDBMS remains operating efficiently even as the number of queries rises thanks to this strategy of node splitting. As the volume of queries grows, the CDBMS dynamically scales up the number of splitting nodes responsible for load distribution. Even after the creation of a new Node A', the original Node A maintains a workload allocation history. This record is essential for the efficient distribution of queries among nodes, enhancing query processing efficiency.

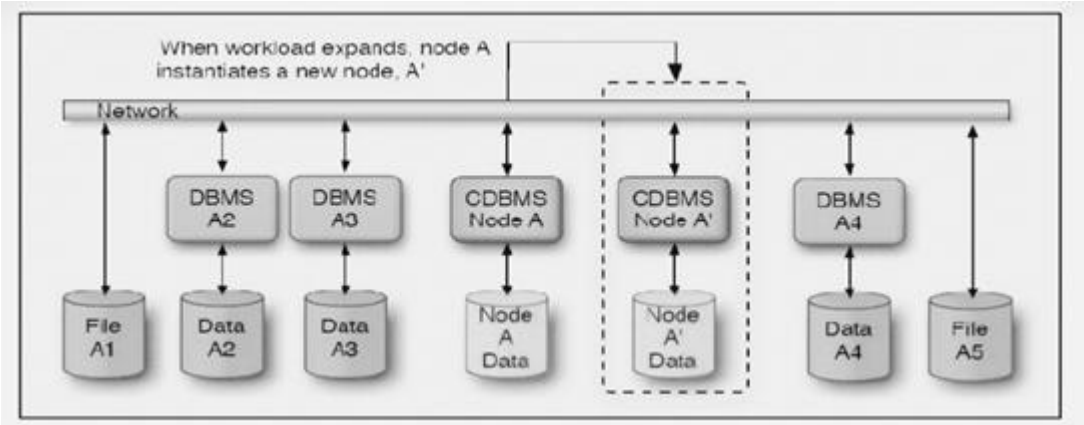

Figure 2: Node Splitting

3. **Distributed Queries:** Think about a large company's huge database, which includes a variety of data categories like products, clients, employees, and corporate policies. The process of retrieving data from this database involves responding to different kinds of inquiries. To efficiently answer each query in a Cloud Database Management System (CDBMS), separate apps may be used to manage these various entities.

Data can be stored in a CDBMS in a variety of ways, including query-oriented databases and column-store databases. However, using distributed queries to manage the database is the most effective strategy.

One way to think of a distributed query is as a composite of several queries, each of which reaches out to a different distributed node to retrieve data. There may be a large rise in the number of results while several queries are active.

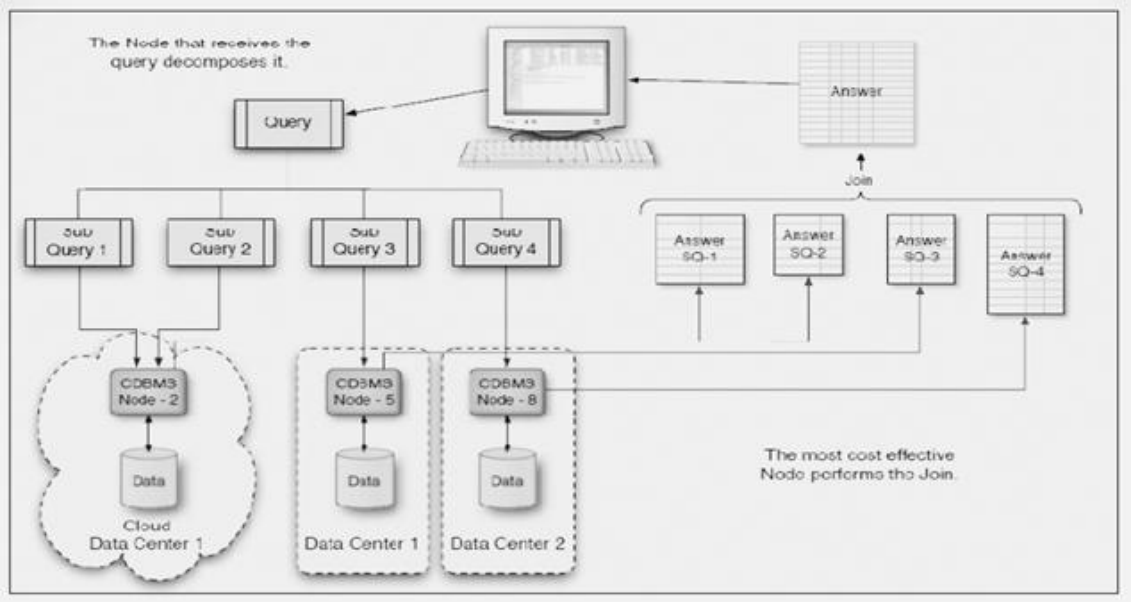

Figure 3: Splitted Queries

III. CLOUD DATABASE SERVICE

Database as a Service (DBaaS), which is provided by several cloud database service providers, can be divided into three primary categories: rational databases, non-rational databases, and virtual computers running regional database software like SQL.

Several businesses offer DBaaS, including Google AppEngineDatastore, Microsoft SQL Azure, Amazon RDS, and Amazon SimpleDB. Every service provider sets themselves apart based on the caliber and kind of services they deliver. Specific factors can be taken into consideration while choosing the best service provider for a business. These standards are not exclusive to any one business and can help identify the finest service provider depending on the special needs of every enterprise.

1. **Choosing Best Dbaas:** The services that the company provides as well as the particular needs of the organization have an impact on the decision to use DBaaS. Choosing the best DBaaS can be based on a number of factors, some of which can act as helpful suggestions for making an informed choice.

2. **Data Sizing:** The amount of storage space each DBaaS provider makes available for its databases varies. Companies must precisely estimate the volume of data they anticipate storing in their databases, therefore data sizing is an important factor to take into account. For instance, while SQL Azure offers a maximum of 50GB of data storage for a single database, Amazon RDS permits customers to store up to 1TB of data within a single database.

3. **Portability:** The ability to move databases is crucial for ensuring uninterrupted user access. In the event of a service provider ceasing operations, the risk of data loss or database termination exists. Therefore, having a contingency plan in place is of paramount importance.A solution to this issue is to take into account cloud services from various providers. By utilizing a variety of cloud service providers, the database is always available, protecting user data and assuring continuous availability.

4. **Transaction Capabilities:** The effective execution of transactions, which is of the utmost significance to users, is ensured by transaction capabilities in cloud databases. Complete read and write operations must be performed in order to maintain transaction integrity for businesses that deal primarily with money or financial transactions. These transactions that demand a success assurance are known as ACID transactions.On the other hand, non-ACID transactions can be used in situations when a guarantee is not required. Since they don't require the same level of assurance and may be more appropriate for some applications or use cases, these non-ACID transactions allow speedier Processing.

5. **Configurability :** Many databases are made to be easily changeable by the user, with the service provider handling the majority of configurations. This strategy drastically limits the options the database administrator must consider, streamlining database management with the least amount of effort needed.

6. **Database Accessibility:** Database accessibility is made possible by two independent processes. Utilizing RDBMS (Relational Database Management System) provided by industry-standard drivers like Java Database Connectivity (JDBC) is the first approach. This driver enables external connections so that users can access services over a common connection. The alternative method for accessing databases is through interfaces or protocols, including Service-Oriented Architecture (SOA), SOAP, or REST. To provide access to the database services, these interfaces rely on HTTP and more recent API specifications.

7. **Certification and Accreditation:** It is strongly advised that a firm use the services of a recognized and accredited cloud database provider because it helps to reduce any potential dangers and makes the transition easier for the business. When compared to other DBaaS providers, businesses who hold certifications like FISMA can be seen as more dependable, adding an extra degree of confidence and trust.

8. **Key Considerations: Data Integrity, Security, and Storage Location:** The security of data kept in cloud storage is still a major worry. The data storage locations and the encryption techniques used both have an impact on the level of security. Cloud data is dispersed over different areas of data canters, which is essential to guaranteeing its security.The security of data kept in cloud storage is still a major worry. The data storage locations and the encryption techniques used both have an impact on the level of security. Cloud data is dispersed over different areas of data canters, which is essential to guaranteeing its security.

Feature	Amazon RDS	Microsoft SQL Azure	Google Cloud Datastore	Amazon SimpleDB
Database Engine	MySQL, PostgreSQL, SQL Server, and others	Microsoft SQL Server	NoSQL (schemaless)	NoSQL (schemaless)
Scalability	Vertical and Horizontal	Vertical and Horizontal	Horizontal	Horizontal
Data Consistency	ACID compliant	ACID compliant	Eventual Consistency	Eventual Consistency
Data Model	Relational	Relational	Document	Key-Value
Query Language	Sql	Sql	Gql (Google Query Language)	Simpledb Query Language
Secondary Indexes	Yes	Yes	Yes	Yes
Data Storage	Ebs (Elastic Block Store)	Azure Storage	Google Cloud Storage	Proprietary Storage

Feature	Amazon RDS	Microsoft SQL Azure	Google Cloud Datastore	Amazon SimpleDB
Data Security	Vpc, Encryption At Rest/Transit	Vnet, Encryption At Rest/Transit	Encryption At Rest/Transit	Encryption At Rest/Transit
Managed Service	Yes	Yes	Yes	Yes
Backup & Restore	Automated Backups, Snapshots	Automated Backups, Geo-Replication	Backup and Export	Data Export
Performance	Instance Types, Read Replicas	Performance Levels	Automatic Sharding	Proprietary Scaling
Use Cases	General Purpose Relational DB	Relational Databases	Web & Mobile Apps	Small-scale Applications
Cost Model	Pay-as-you-go, Instance-based	Pay-as-you-go, DTUs	Pay-as-you-go	Pay-as-you-go

IV. CHALLENGES TO CLOUD DATABASE

Cloud database setup and successful operation provide unique difficulties. However, despite these challenges, cloud databases are gradually replacing on-premises databases for many businesses. These are a few of the difficulties that cloud computing faces.

1. **Internet Speed:** When compared to the speed of internet access utilized to get there, the data center's internal data transmission rate is far faster. This discrepancy hinders the cloud database's ability to operate at peak efficiency, which lowers overall effectiveness. While database queries are performed quickly, internet speed has an impact on how quickly data is retrieved from the data center.

   Using faster speed connections is one solution to this problem, but doing so can be quite expensive and could reduce the cost-effectiveness of having a cloud database in the first place. In order to overcome this problem, finding the ideal balance between cost and performance continues to be important.

2. **Query And Transactional Workloads:** The workload associated with queries and the workload associated with transactions differ significantly. It is possible to predict how long a transactional task will take to complete. Predicting the time required becomes difficult in the query workload scenario, though. The duration of the query burden relies on how many queries are made, and it is unknown how many users will query the database. The specific period for query processing is therefore yet unknown.

3. **Multi-Tenancy:** The main goal is to handle a database and its workload while maximizing the performance of a specific system. The need to maintain efficiency while reducing the number of machines cannot be overstated. To accomplish this, even when

several workloads are operating on the same machines, the system must efficiently distribute hardware resources to each workload.

One approach is to build separate virtual machines for each database; they can be constructed on the same real machine. However, when multiple machines must share the same task, performance and speed can be significantly decreased, often becoming up to six to ten times slower. Because each virtual machine runs on its own operating system, there is a performance loss.

4. **Elastic Scalability:** A top-notch Database as a Service (DBaaS) should be able to effectively manage a variety of workloads when selecting a cloud database. Nevertheless, when the workload exceeds the system's capacity, challenges within the cloud database can emerge. To address this, the cloud database must possess the capability for seamless scalability as the workload grows. The database is scaled out to provide maximum performance and effectiveness, enabling the cloud database to effectively react to changing workloads.

5. **Privacy:** As cloud computing provides increased accessibility to both authorized users and possible hackers, privacy is a major concern. It is crucial to maintain privacy in cloud databases, especially when dealing with sensitive consumer data. The saved data cannot be exposed by the companies. Using encryption for the data kept in the database is one efficient technique to improve security. Data storage security is ensured through encryption, which also adds another line of defence against illegal access.

V. ADVANTAGES OF CLOUD DATABASE

Efficient Business Operations: The business environment has been completely transformed by cloud computing, which makes transactions quicker and more effective. Employees can now easily access information utilizing freely available resources, such as cloud databases, whereas in the past accessing company databases required program installs. This increases productivity and saves crucial time.

Cost Savings: By using a cloud database, businesses can avoid spending money on building and maintaining their own data centers, saving them a lot of money. Additionally, there is a decrease in the cost of purchasing and maintaining software.

Flexibility and Scalability: The benefit of using a cloud database service provider is that they can handle immediate database changes, relieving consumers of these worries. Additionally, these providers offer scalability during peak hours, guaranteeing constant performance standards even during times of high demand.

Access Anywhere, Anytime: By enabling users to access information from anywhere, cloud computing frees them from the constraints of being confined to a personal computer at home. Customers, staff members, and corporate executives all benefit from the convenience of having access to the necessary information whenever they need it.

Superior Data Availability: For larger organizations and businesses that handle enormous amounts of data, cloud databases are a great answer. By guaranteeing data

availability at all times and from any location, cloud databases enable easy management and access. Top of Form

VI. DRAWBACKS OF CLOUD DATABASES

While there are numerous advantages associated with adopting cloud databases, it is essential to be aware of the potential disadvantages, which can be concerning for companies.

1. **Cost Considerations:** Payment for using a cloud database is depending on consumption. The business spends money every time data is transmitted from the database. If data transfer traffic is heavy, costs could end up being more than anticipated.

2. **Lack of Full Control:** Users' access to the server that hosts their database is limited. Additionally, they have no control over the software that is installed on the server and must rely on the provider to take all necessary security precautions. The risks associated with this reliance on the supplier for security can be high.

3. **Data Dependency:** The service provider is the only source of authority for the data stored in the cloud database. Important firm data and information lost might have serious repercussions, including financial losses and privacy violations.

4. **Data Transfer Speed:** Due to internet speed restrictions, retrieving huge amounts of data from the cloud database may take some time. In contrast, conventional databases might provide faster data transfer rates.

5. **Switching Providers:** Since many service providers use different methods and approaches for data storage, moving the database from one to another might be challenging. Selecting a new DBaaS provider must be done with care in order to prevent issues with the move.

6. **Server Downtime:** Access to cloud databases requires a live internet connection. Users might have trouble obtaining crucial information if a server goes down, which could cause significant losses if the information isn't available when it's needed.

REFERENCES

[1] Huang S, Qin Y, Zhang X, Tu Y, Li Z and Cui B. (2023). Survey on performance optimization for database systems. Science China Information Sciences. 10.1007/s11432-021-3578-6. 66:2. Online publication date: 1-Feb-2023.

[2] D. -P. Tan, L. Li, Y. -L. Zhu, S. Zheng, H. -J. Ruan and X. -Y. Jiang, "An Embedded Cloud Database Service Method for Distributed Industry Monitoring," in IEEE Transactions on Industrial Informatics, vol. 14, no. 7, pp. 2881-2893, July 2018, doi: 10.1109/TII.2017.2773644.

[3] P. T. Anh Mai, J. K. Nurminen and M. Di Francesco, "Cloud Databases for Internet-of-Things Data," 2014 IEEE International Conference on Internet of Things (iThings), and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom), Taipei, Taiwan, 2014, pp. 117-124, doi: 10.1109/iThings.2014.26.

[4] Al Shehri, Waleed. (2013). Cloud Database Database as a Service. International Journal of Database Management Systems. 5. 1-12. 10.5121/ijdms.2013.5201.

[5] 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

[6] Bloor,R.2011.WHATIS A CLOUD DATABASE? Retrieved 25th November 2012 fromhttp://www.algebraixdata.com/wordpress/wpcontent/uploads/2010/01/AlgebraixWP2011v06.pdf
[7] Finley, K. 2011. 7 Cloud-Based Database Services. Retrieved 23rd November 2012 from http://readwrite.com/2011/01/12/7-cloud-based-database-service
[8] 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/pastcourses/fs09/HotDMS/pdf/daas.pdf
[9] 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
[10] Rouse, M. 2012. Cloud Database. Retrieved 25th November 2012 from http://searchcloudapplications.techtarget.com/definition/cloud-database-database-as-a-service
[11] Saini, G.P. 2011. Cloud Computing: Database as a Service. Retrieved 24th November 2012 from http://cloudcomputing.sys-con.com/node/1985543
[12] 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

