Cloud Databases

Dr.M.Laxmaiah,

Professor & HOD,

Department of CSE-Data Science,

CMR Engineering College,

Hyderabad, Telangana, India.

datasciencehod@cmrec.ac.in.

B. Kumara Swamy,

Assistant Professor,

Department of CSE-Data Science,

CMR Engineering College,

Hyderabad, Telangana, India.

bkumaraswamyse@gmail.com.

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).

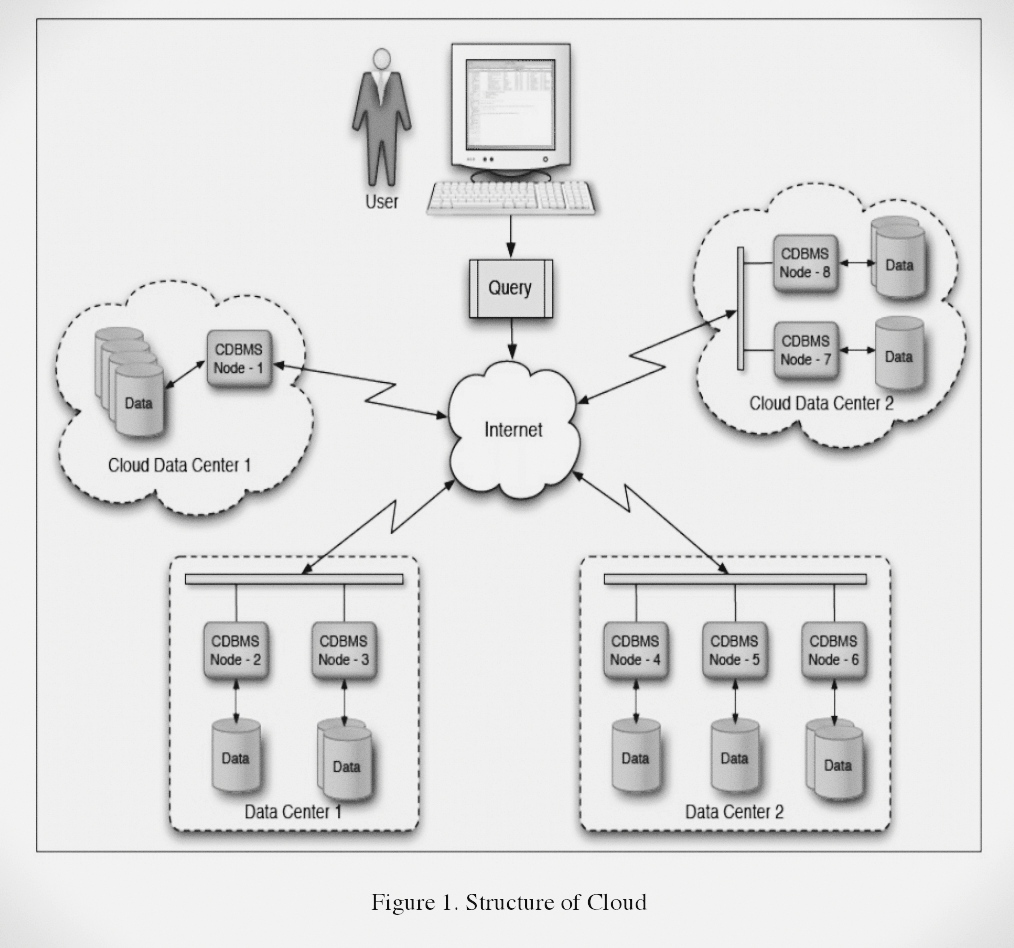
# INTRODUCTION

The cloud database holds the data on different data centers located at different locations. This makes the cloud database structure different from the relational 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.

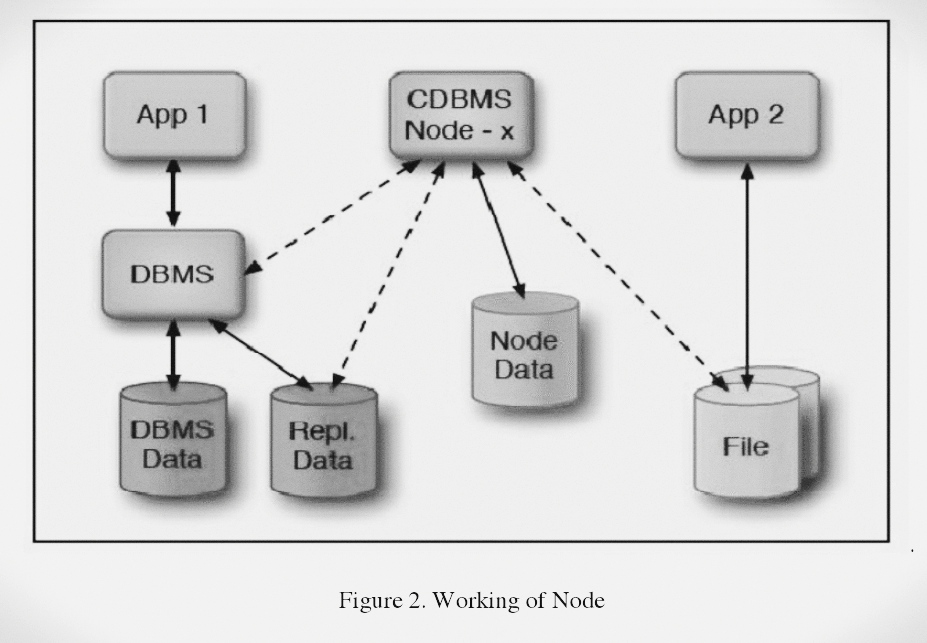
Acloud database is a database service built and accessed through a cloud platform. It serves many of the same functions as a traditional database with the added flexibility of cloud computing. Users install software on a cloud infrastructure to implement the database.Key features:

* A database service built and accessed through a cloud platform
* Enables enterprise users to host databases without buying dedicated hardware
* Can be managed by the user or offered as a service and managed by a provider
* Can support relational databases (including MySQL and PostgreSQL) and NoSQL databases (including MongoDB and Apache CouchDB)
* Accessed through a web interface or vendor-provided API

**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.

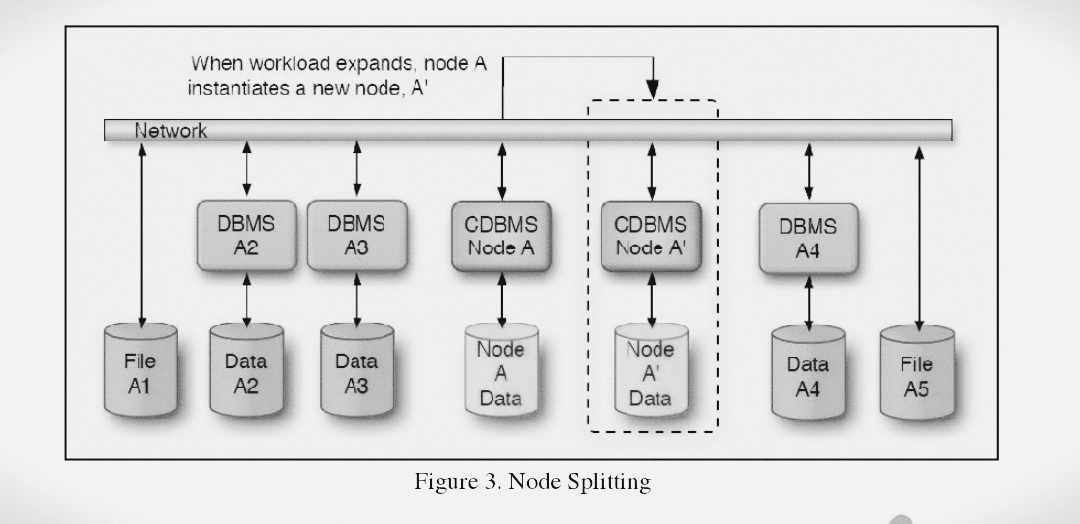
**2.2 Working of Nodes**If there is a node available, either to access the data directly from the database or to get it from another node. The replicated database is not accessed all the time As soon as the database fails to perform, in case of an emergency it is used.As a result, data is fetched more quickly from the database.Application data is stored in CDBMS applications. The CDBMS accesses the data from the files directly.Nodes that directly access the data maintain metadata maps for the files from which they acquired application data.



**2.3 Node Splitting**

BI applications may easily access data without affecting the performance of the database when considering local data within a data center, 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.

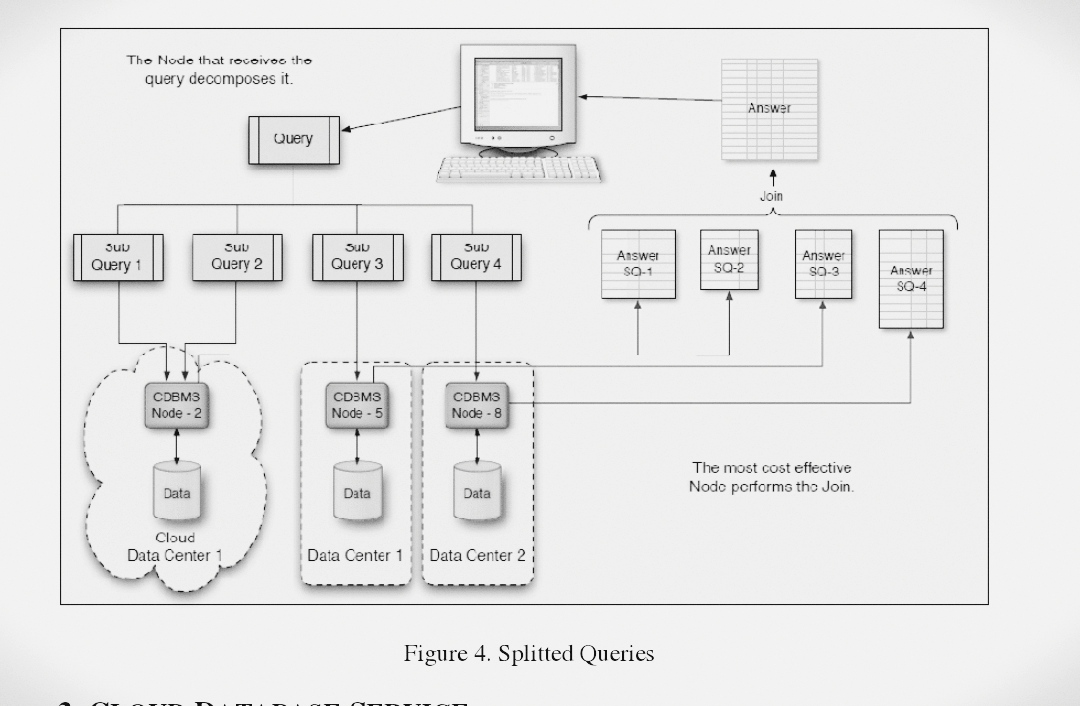


It's critical to handle this query overload as soon as possible. In order to remedy this, CDBMS starts the quick deployment of a new node, which aids in more effectively distributing the query load throughout the database.

The CDBMS remains operating efficiently even as the number of queries rises thanks to this strategy of node splitting. The CDBMS dynamically raises the number of splitting nodes in charge of distributing load as the number of queries rises. The original Node A keeps a record of the workload allocation even after splitting into a new Node A'. This record is essential for effectively allocating inquiries among the nodes, resulting in efficient query handling.

**2.4 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 to 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.

**3. 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 AppEngine Datastore, 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.

**3.1 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.

**3.2 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, whereas SQL Azure only offers 50GB of data storage for one database, Amazon RDS allows users to store up to 1TB of data in a single database.

**3.3 Portability**

Database portability is essential to guarantee uninterrupted user access. There is a chance of data loss or database destruction if a service provider goes out of business. Consequently, it is crucial to have an emergency plan in place. 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.

**3.4 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.

**3.5 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.

**3.6 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 use of interfaces or protocols like Service-Oriented Architecture (SOA), SOAP, or REST constitutes the second method of database accessibility. To provide access to the database services, these interfaces rely on HTTP and more recent API specifications.

**3.7. 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.

**3.8 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 centers, 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 |
| **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 |

**4. 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.

**4.1 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.

**4.2 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.

**4.3 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.

**4.4 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. However, when the workload exceeds the system's capacity, problems could occur in the cloud database. To get around this, the cloud database needs to be able to scale out smoothly 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.

**4.5 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 defense against illegal access.

**5. Advantages of Cloud Database**

1. 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.
2. 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.
3. 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.
4. 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.
5. 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

**6. DISADVANTAGES OF CLOUD DATABASE**

While there are several advantages to using cloud databases, it is essential to be aware of the potential disadvantages, which can be concerning for companies.

I. 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.

II. 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.

III. 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.

IV. 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.

V. 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.

VI. 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

C

LOUD

D

ATABASE

D

ATABASE

AS

A S

ERVICE

C

LOUD

D

ATABASE

D

ATABASE

AS

A S

ERVICE

1. [Huang S](https://dl.acm.org/author/Huang%2C+Shiyue), [Qin Y](https://dl.acm.org/author/Qin%2C+Yanzhao), [Zhang X](https://dl.acm.org/author/Zhang%2C+Xinyi), [Tu Y](https://dl.acm.org/author/Tu%2C+Yaofeng), [Li Z](https://dl.acm.org/author/Li%2C+Zhongliang) and [Cui B](https://dl.acm.org/author/Cui%2C+Bin). (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.WHAT IS 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