UNIT-5

Cloud Database
Part-2 of Unit 5

Contents

- Cloud Database- Operational Model for Cloud Database, Types of Cloud Database
- Cloud File System- Distributed File System Basics, Concept of GFS and HDFS, Comparison of Features

Cloud Database

A cloud 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.

Key features:

- A database service built and accessed through a cloud platform
- Enables enterprise users to host databases without buying dedicated hardware.
- There are some SQL-based and some NoSQL- based databases offering.
- Accessed through a web interface or vendor-provided API.

Operation Model for Cloud Database

There are two primary methods to run a database on a cloud platform:

- **Virtual Machine image**: Cloud platforms allow users to purchase virtual-machine instances for a limited time, and one can run a database on such virtual machines.
- Users can either upload their own machine image with a database installed on it, or
- use ready-made machine images that already include an optimized installation of a database.

- **Database-as-a-service (DBaaS):** With a database as a service model, users pay fees to a cloud provider for services and computing resources, reducing the amount of money and effort needed to develop and manage databases.
- Users are given tools to create and manage database instances, and control users. Some cloud providers also offer tools to manage database structures and data. Many cloud providers offer both relational (Amazon RDS, SQL Server) and NoSQL (MongoDB, Amazon DynamoDB) databases. This is a type of software as a service (SaaS).

Architectural and common Characteristics

- **Fast Deployment**: Cloud databases are the perfect choice when you urgently need a database, as they can be up and running in minutes. Cloud databases eliminate the need to purchase and install hardware and set up a network.

- **Accessibility**: Users have quick access to cloud databases remotely through the provider's API or web interface.

- **Scalability**: You can expand cloud database storage capacity without disruptions and meet the requirements. Cloud database scalability is seamless due to DBaaS implementation, which is a major benefit for growing businesses with limited resources.

- **Disaster Recovery**: Data backups are regularly performed on cloud databases and kept on remote servers. These backups enable a business to stay online in cases of natural disasters, equipment failure, etc.

- **Lower Hardware Costs**: Cloud database service providers supply the infrastructure and perform database maintenance. Hence, companies invest less in hardware and have fewer IT engineers for database maintenance.
- **Value for Money**: Many DBaaS solutions are available in multiple configurations, allowing companies only to pay for what they use and turn off services when they don't need them. Cloud databases also save money by not requiring operational costs or expensive upgrades.
- **Latest Tech**: Cloud database providers upgrade infrastructure and keep it updated with new tech. This brings significant savings as companies don't have to allocate funds on new tech or staff training.
- **Security**: Most cloud database providers encrypt data and invest in the best cloud security solutions to keep the databases safe. Although there is no impenetrable security system, it is a safe way to protect data. Since cloud database providers use automation to enforce the best security practices, there is less room for human error compared to using on-premises databases.

Cloud Database Vendors

- Microsoft Azure
- Amazon Web Service (AWS)
- Oracle
- Google Cloud
- Rackspace

Types of Cloud database

- It is also important to differentiate between cloud databases that are relational as opposed to non-relational or NoSQL.

- The details of each type of cloud database are discussed in the following subsections:

- **Cloud Relational Databases**
- **Cloud NoSQL Databases**

Cloud Relational Database

- **Microsoft Azure:**
- Microsoft Azure cloud database is one of the most popular and globally widespread cloud platforms.
- It offers computing, networking, databases, analytics, AI, and IoT services.
- The public cloud computing platform from Microsoft offers various solutions, including Infrastructure as a Service (**IaaS**), Platform as a Serice (**PaaS**), and Software as a Service (**SaaS**).

- **Amazon Web Service (AWS):** AWS is one of the market leaders when it comes to DBaaS. Amazon offers various services for data management and integration. Some of AWS database solutions are:
- **Amazon RDS**. Amazon **R**elational **D**atabase **S**ervice runs on either Oracle, SQL, or MySQL server instances.
- **Amazon SimpleDB**. Designed for smaller workloads, SimpleDB is primarily a schema-less database.
- **Amazon DynamoDB**. DynamoDB is a NoSQL database capable of automatically replicating workloads across three availability zones.
- The downside is that scaling and patching operations require **downtime**.

- **Oracle:** Oracle offers enterprise-scale cloud database technology to its users. The database solution uses machine learning to **automate database management**, ensuring high performance, reliability, and security.
- Oracle cloud database covers hyper-scale Big Data and Streaming workloads, including OLTP- Online Transaction Processing , data warehousing stack, Spark, text search, image analytics, and data catalog. Oltp lile ATM, online booking, credit card payments etc

- The different solutions offered are Infrastructure as a Service (**IaaS**), Platform as a Service (**PaaS**), Software as a Service (**SaaS**), and Data as a Service (**DaaS**).
- The downside is a **lack of integration** with other cloud solutions.

- **Google Cloud:**
- Google Cloud Platform (**GCP**) offers various services that use the same hardware and infrastructure as other Google products. GCP's offer includes a wide range of hosted services for cloud computing, storage, networking, big data, machine learning, IoT, cloud management, etc.
- GCP provides **IaaS**, **PaaS**, and **serverless computing environments**.
- One of the products in Google Cloud Platform is **Cloud Datastore**, a database storage solution for NoSQL non-relational storage.
- Other Google Cloud products are **Cloud SQL** for MySQL fully relational storage and Google's native **Cloud Bigtable** database.

- The downside is a **lack of managed services** and the **high prices**, including a costly support fee.

- **IBM Db2 on Cloud:** IBM Db2 on Cloud is a fully managed SQL database featuring a 99.99% uptime SLA, independent storage and compute scaling through UI and API, several disaster-recovery options, data encryption, and other features.
- IBM's relational database offers advanced data management and analytic capabilities for transactional and warehousing workloads. This database delivers high performance, boasts great insights, data availability, reliability, and broad operating system support.
- The downside of IBM Db2 is that it has fewer regional options, affecting performance in some cases.

- **Rackspace:**

- Rackspace offers scalable, fully managed, or hosted cloud databases, characterized by high performance and a storage area network (SAN) based on the OpenStack platform.
- Rackspace offers easy access to your cloud database via Cloud Control Panel, CLI or API, and features regular backups of all cloud databases.
- Redundant storage and synchronous data replication ensure data protection in case of disaster or hardware failure.
- The downside is a smaller number of data centers compared to the competition.

Cloud NoSQL Databases

- NoSQL database is " not only SQL" database. The evolution of NoSQL database started in early 2009 and has been growing rapidly since because of some limitations with relational databases
- **MongoDB Atlas :** MongoDB Atlas is a cloud database created and managed by the same team that developed MongoDB.
- Mongo's cloud database is a fully managed NoSQL database that features flexibility, scaling, sharding, and **database management automation**. It allows most developers to go through various delivery models without requiring help from a database administrator.
- The downside is that MongoDB Atlas is **NoSQL only**, which means that SQL-reliant applications are not an option with this database.

- Sharding is a type of database partitioning that separates large databases into smaller, faster, more easily managed parts. These smaller parts are called *data shards*. The word *shard* means "a small part of a whole."