The world of booming data

PLUS: THE NEW CLOUD DATABASE TOP 20

CLOUD DATABASE REPORT 2022

ANALYSIS BY JOHN FOLEY

Contents

NEW WEBSITE, PODCAST & NEWSLETTER

Welcome to Cloud Database Report 2022 and Cloud Database Top 20. This Report was first published in February 2021 on the Cloud Wars website. A lot has changed since then:

•The Report has a new website: View it here

•The Report is comprised of weekly news & analysis, podcasts, and newsletter

•The Cloud Database Top 20 has been updated with three new vendors and others have dropped

I launched the Cloud Database Report after working for years as a tech journalist covering data management, then as a communications strategist with three leading database providers—Oracle, IBM, and MongoDB. I focus on all aspects of the database market: platforms, providers, data strategy and architecture, and business deployment. I'm briefed regularly by market leaders including AWS, Google Cloud, Microsoft, Oracle, SAP, and Teradata, and well as emerging players such as Cockroach Labs, Neo4j, SingleStore, Yugabyte, and many others.

My analysis also appears regularly on the Acceleration Economy website, where I host the Data Revolution channel.

The cloud database market is transforming as a new generation of cloud-native database providers challenge long-established vendors with 30-plus years of experience. I invite everyone to subscribe to the Cloud Database Report to stay up to date on this rapidly changing market.

John Foley, Founder & Editor, Cloud Database Report

Subscribe to the Cloud Database Report

DATABASE RENAISSANCE

In early 2021, I began this Report with the following observation: "The cloud database market is growing and changing, as incumbents led by Oracle compete with the Big 3 cloud service providers—AWS, Google, and Microsoft—and dozens of specialists and startups."

As it turns out, that was an understatement. The innovations in the cloud database market are coming so fast that it's hard to keep up with it all. I counted no fewer than 200 announcements made by the Big 3 public cloud providers at their annual conferences in Q4 of 2021—AWS re.Invent, Google Cloud Next, and Microsoft Ignite. Those announcements included databases, AL/ML frameworks, analytics, edge servers, development and management tools, and other cloud services and capabilities that can be stitched together to create "data clouds." (More on data clouds later in this report.)

And that's just the Big 3. Oracle, IBM, SAP, Teradata, and other industry incumbents are in high gear developing their own new capabilities. In 2021, Oracle introduced Oracle Database 21c with loads of new features, upgrades to its Apex low-code development environment, and a new and improved Autonomous Data Warehouse with Auto ML and support for "citizen data scientists," as just a few examples. Even Teradata, sometimes viewed as the classic "legacy" data warehouse vendor, struck up a product development and integration deal with AWS. No one is resting on their laurels in this market—they would be left behind if they did.

And there are the emerging "cloud-native" database providers, some of which—like Couchbase and SingleStore—have been growing into serious contenders over 10 years, while others—such as Firebolt, Pinecone Systems, and TileDB are in their early stages. Venture funding is pouring into these upstarts. In December, I reported that 10 database startups—including Cockroach Labs, Redis, and Yellowbrick Data—received a total of more than $2.9 billion in funding in 2021.

- See my article, "Cockroach Labs, with New $5 Billion Valuation, Caps Big Year in Database Startup Funding"

One of the biggest investment rounds of 2021 went to Databricks—$1.6 billion in Series H. Bloomberg referred to Databricks as "a $38 billion startup."

This near-frantic pace of funding and development causes me to say that the 50-year-old database market is going through a Renaissance. It's not your father's (or

grandfather's) relational DBMS managing data in tables and rows on servers in data centers. Cloud databases are the beating heart of the data-driven digital enterprise.

5 FACTORS DETERMINE CLOUD DATABASE LEADERS

With so many options, how does an organization choose the best one for its particular needs? The Cloud Database Report has identified 20 cloud database providers that, in our assessment, are the leaders. They represent a cross-section of the market—incumbents, public cloud providers, and challengers.

As you would expect in this fast-changing market, a few newcomers have joined the Cloud Database Top 20 this year, while others have dropped off. New on the list are PlanetScale, SingleStore, and Yugabyte. Gone are Cloudera, MarkLogic, and Vertica. You will find the complete Cloud Database Top 20 below, with my take on each vendor.

Why the changes? In short, challengers are displacing incumbents. It's also a reflection of a single but significant change to the criteria for being included on the Top 20. This year, I have added "cloud native" as a determining factor. Five considerations now go into the evaluation process.

Cloud Database Top 20 criteria:

1. **Cloud-native architecture.** The cloud database services are designed to work across hybrid and multi-clouds and integrate with other cloud-based infrastructure.

2. **Enterprise capabilities.** Vendors with a complete range of services and support that enterprises may want or need. Fully managed services are a plus.

3. **Platform adaptability.** Tools, services, and APIs for data integration/migration and application compatibility are must-haves.

4. **Innovation.** A steady pipeline of new, modern, differentiating capabilities.

5. **Demonstrated business value.** Customer success is the #1 proof point.

Subscribe to the Cloud Database Report

IRRESISTIBLE FORCES MEET IMMOVABLE OBJECTS

I sometimes use the paradox of irresistible forces (cloud-native startups) meeting immovable objects (incumbent vendors) in characterizing the competitive landscape. In some parts of the tech market—processors, mobile devices, servers—it's not unusual to see turnover as users replace yesterday's products with new and better ones.

But that's less likely in the database market because database management systems are deeply rooted in the business, with applications, custom code, queries, and team skills all woven into the data environment.

It can be complex and risky to swap databases. What often happens is that businesses will leave their existing database platforms unchanged (though perhaps migrate them to the cloud), while deploying new cloud databases for modernization and new development. In this way, irresistible databases and immovable databases coexist.

For this reason, Oracle and other incumbent database vendors have shown themselves to be resilient and have staying power. And in general, their cloud businesses are growing, albeit not as fast as the cloud-native providers.

Meanwhile, the Big 3 public cloud providers—AWS, Google Cloud, and Microsoft Azure—have a few advantages in the cloud database market that put them in a unique position of strength. They share a few common characteristics:

- **Continuous innovation.** Each of the Big 3 offers a steady stream of improvements. For example, at re:Invent, AWS introduced Private 5G Clouds and serverless analytics services.
- **Fast & easy integration.** Surrounding services such as storage, data migration, IoT, ML frameworks are part of an integrated, overarching infrastructure and architecture.
- **Purpose-built databases.** Each vendor offers a range of database options, both home-grown and third-party, for different types of workloads, such as data warehousing, transactions, application modernization, graphs, etc.
- **Flexible deployment options.** The cloud providers offer both fully managed and self-managed database services, and solutions for hybrid and multi-clouds.

It should be noted that each of the Big 3 cloud providers also partners with many of the same companies they compete with in databases. So, for example, DataStax's Cassandra-

compatible NoSQL database is available on AWS, Google Cloud, and/or Microsoft Azure. At the same time, those companies also offer their own NoSQL databases.

BE LIKE SNOWFLAKE

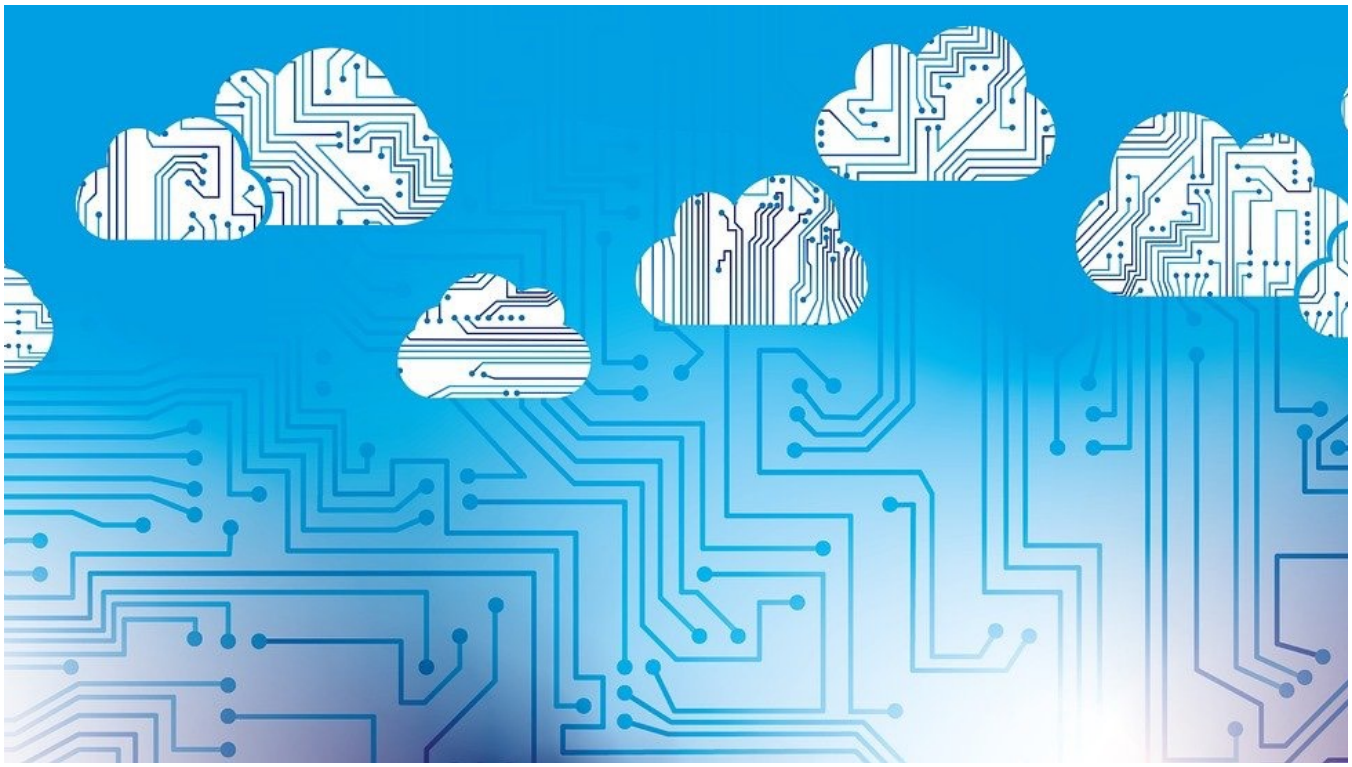
That leaves one other major player that must be considered in this database market overview—Snowflake. Snowflake grabbed the attention of investors with its IPO in September 2020, as its share price doubled on the first day of trading.

What explains that kind of dramatic growth? Stock market exuberance aside, Snowflake has revolutionized the data warehouse and analytics market by talking not about the guts of database technology, as some vendors do, but about data sharing. I'll bet most people are not even aware of, nor do they care about, the underlying database system of the Snowflake platform. Snowflake emphasizes its data architecture, which it describes as "a single, integrated platform delivered as a service."

In essence, Snowflake has simplified a process that business leaders have considered a priority for more than 20 years, but which has been complex, expensive, and only scratched the surface of its potential—and that is putting actionable business intelligence into the hands of more people. More than 800 data sets were available from third-party providers in Snowflake's data marketplace, as of October 2021.

The Snowflake platform can be used as a data warehouse or data lake by data scientists and data engineers. Snowflake is all about data. One of my predictions for 2022 is that more database providers will try to be like Snowflake, by developing business models and messaging around business value rather than speeds and feeds.

I should add that there are potential issues with the Snowflake model that bear watching. Snowflake CEO Frank Slootman admitted last year that database migrations can be lengthy, expensive, and risky. Another potential gotcha is Snowflake's consumption-based pricing model; as usage increases, so do costs.



DATA PLATFORMS & DATA CLOUDS

As mentioned earlier, the cloud database market increasingly is less about the core database management system itself and more about the data platform and overall cloud environment inclusive of AI/ML, analytics, data distribution, migration tools, etc.

The platform model is one of the great advantages of the cloud. It's much easier for vendors and customers alike to assemble comprehensive data environments that make it possible to import data from disparate sources, run algorithms, distribute data across regions, and more. It's not uncommon for projects to weave together a half-dozen or more cloud services across hybrid and multi-clouds.

One related concept that seems to be on the rise is the "data cloud." Google Cloud, Oracle, and Snowflake offer data clouds, and Cloudera and Pinecone Systems made reference to data clouds last year.

What exactly is a data cloud? Sudhir Hasbe, product management leader for Google Cloud Data Analytics, told me in a briefing that customers view data clouds as a way to bring together AI, ML, business intelligence, and analytical and operational data.

"The way customers are thinking about data clouds is pretty extensive—everything that they're doing," he said. "They are looking at it like, there's a cloud for all of their data and how they can get value out of it. It's a holistic view of data, not just a platform or infrastructure."

That's how data clouds differ from long-established platforms such as data warehouses and data lakes. Data clouds are more comprehensive—the sum total of databases, data warehouses, lakes, and fabrics across an enterprise. Essential components include data integration and replication, data engineering, sharing, and analytics, all overlaid with end-to-end security and governance.

Snowflake sums up the data cloud simply as: "For all of your data and all of your users."

CIOs and CTOs care about data platforms and data clouds because it's often the surrounding tools and services that account for the greatest business value. If data integration is fast and easy, if third-party data sets and tools are only a click away, if resiliency can be improved, all of that results in less complexity and faster pay off.

PURPOSE-BUILT VS. ALL-PURPOSE

As noted, innovation is thriving in the database market. One point of reference is Carnegie Mellon University's Database of Databases, which at last count provided an archive of 795 different database management systems. And that number is growing.

There are many different kinds of databases: relational, key value, document, time series, graph, search, object-oriented, wide column, spatial, vector, and more. Which brings me to our next topic—the ongoing debate between purpose-build databases and multi-purpose databases.

Let's start with purpose-built. The advantage with purpose-built databases is that they excel at certain kinds of workloads. For example, a time-series database might be ideal for IoT data, a graph database for social networks, and so on. The big benefit of this approach is optimized application performance.

AWS, with more than a dozen databases in its portfolio, has thrown its weight behind the purpose-built model.

But there are potential drawbacks to the purpose-build approach, as well. Imagine a data environment that includes relational, document, graph, search, and OLAP databases, all from different vendors. That could stretch the administrative and management capabilities of your in-house team.

Alternatively, all-purpose databases—also called multi-model or universal databases—may be able to do all of the above from one platform. Oracle describes 21c as a "converged database" capable of managing multiple data types. Likewise, SingleStore serves as a "single store" for various workloads.

The good news is that IT and dev teams do not need to come down on one side or the other of the purpose-build vs. all-purpose decision. They can do both—use a multi-model database for some mixed workloads and purpose-built databases where application performance is a priority. Many IT organizations will not limit themselves to a single cloud database or cloud database provider but will use the best platforms for the job. The decision often comes down to the nature of the application supported and performance requirements.

In fact, some leading database providers offer both purpose-built and all-purpose databases. For example, Google's Cloud's BigQuery is a data warehouse, and its Firestore is a NoSQL document database for new applications. On the other hand, Google's Cloud SQL and Cloud Spanner are capable of handling different data types and workloads.

Another decision point is SQL or NoSQL, although the distinction here is not as clear as it may seem. Despite the name, many NoSQL databases do in fact support SQL queries. And several of the cloud database providers in our Top 20, including AWS, Google Cloud, IBM, Microsoft, and Oracle, offer both SQL and NoSQL options.

With so many databases to choose from, it underscores the importance of a data architecture that fits the pieces together in support of the business strategy.

Key architecture components

An adaptable data architecture—featuring cloud-native databases, data distribution, resilience, and security—is essential to modern data management.

A few of the key capabilities to look for in cloud databases include:

- Multiple deployment options - hybrid, multi-cloud, edge, on premises, mobile.

- Data distribution - capabilities that support data distribution across regions and availability zones include replication, geo-partitioning, change data capture, data streaming, APIs, and prebuilt connectors.

- Managed services and automation - some cloud databases are self-managed by the user, while others are fully managed by the service provider. In addition, cloud databases are increasingly offered with some level of automation, such as serverless provisioning, auto-sharding, and, in the case of Oracle's Autonomous Database, self-patching and self-healing.

DATABASE MIGRATION

As Gartner has rightly pointed out, there's a major trend underway to deploy new databases in the cloud, and to migrate existing databases from on-premises to the cloud. I wrote quite a lot about these projects in 2021, and there's no doubt they will continue to be a heavy workstream—and in some cases a headache—for many organizations.

So-called "same-to-same" database migrations, which involve moving one type of database to a similar database in the cloud (for example, SQL Server to SQL Server) are generally the easiest. Case in point: Google Cloud's DMS can be used to migrate on-prem PostgreSQL databases to Google Cloud's PostgreSQL (which is an option with Google Cloud's Cloud SQL service).

Database migrations get more complicated when they involve different source and target databases—say from an Oracle database to PostgreSQL—or when there are hundreds of terabytes to move. That's because these projects often involve more than just the data; there are custom code and applications tied to those data sets. And full-scale data warehouses represent an even bigger lift.

In fact, it can take weeks, months, or even years to move a company's "data estate" from one database environment to another. In one case study, AWS said it took two years to move 7,500 Oracle databases with a total of 75 petabytes of data, used by Amazon.com, to AWS's own Aurora and special-purpose databases.

Migration scenarios include application modernization, data governance (such as locating data in a specific country), and offloading mainframe or other legacy workloads to the cloud. But the starting point for many IT teams is new development.

All cloud database providers offer data-migration tools and services to facilitate the shift of database workloads to their platforms. AWS's Database Migration Service, can be used to move databases to Amazon Aurora, Redshift, DynamoDB, or DocumentDB. And Google Cloud's Database Migration Service supports the migration of MySQL, PostgreSQL, and SQL Server workloads to Google's Cloud SQL. Other vendors have similar tools.

Just remember that database migration is seldom as easy as it sounds, so plan carefully.

WHAT'S NEXT

There's no reason to expect the pace of change in the cloud database market will slow down. If anything, the opposite is true—database innovation is coming like a firehose as businesses look to accelerate product development, customer relationships, and growth through smart, well-orchestrated data strategies.

There's tremendous opportunity. Businesses are generating terabytes, petabytes, and in extreme cases even exabytes of data, but they are in desperate need of new ways to wring value from all of that data. Cloud databases are often a better way to aggregate, store, analyze, distribute, and drive insights and actions.

I invite you to subscribe the Cloud Database Report to keep up with it all. You will receive weekly articles, podcasts, exclusive interviews, and analysis on database industry leaders and emerging players. Thank you and have a great data-filled year!

John Foley, Founder & Editor, Cloud Database Report

Subscribe to the Cloud Database Report

THE CLOUD DATABASE REPORT TOP 20

The following vendors (in alphabetical order) are the Cloud Database Report's Top 20 as of January 2022. These companies were selected based on five criteria: Cloud-Native Architecture, Enterprise Capabilities, Platform Adaptability, Innovation, and Demonstrated Business Value.

Vendor	Database Offerings	Differentiators	My Take
AWS	Aurora, DocumentDB, DynamoDB, Redshift, RDS, others	Wide choice of purpose-built databases; integration with AWS services	Extensive choices, easy integration with cloud infrastructure; but potential competitor to its customers
Cockroach Labs	CockroachDB, CockroachCloud	Distributed database with PostgreSQL compatibility	Fast-emerging enterprise player with $5 billion market valuation (Dec. 2021)
Couchbase	Couchbase Cloud	NoSQL document database with cloud, server, mobile versions	Keeping pressure on MongoDB in NoSQL document databases
Databricks	Databricks Lakehouse Platform	Software environment based on Apache Spark for data engineering and analytics	Powerhouse in data engineering; investors include Alphabet, Amazon, Microsoft, Salesforce
DataStax	DataStax Enterprise, Astra Cloud Service	Apache Cassandra NoSQL database as a service	Riding popularity of Apache Cassandra among developers
Google Cloud	Cloud Spanner, BigTable, BigQuery, Firestore, Firebase, others	Google-developed, cloud-native databases plus popular SQL and NoSQL databases	Continues to solidify its transformation into an enterprise-class cloud provider with industry expertise

Vendor	Database Offerings	Differentiators	My Take
IBM	Cloudant, DB2, Informix, Netezza, others	More than a dozen IBM and non-IBM databases offered on IBM Cloud	Not much innovation; of interest mainly to existing customers
InterSystems	InterSystems IRIS	Cloud database supports SQL and NoSQL, transactions, analytics	Established enterprise-class database provider with expertise in healthcare
MariaDB	MariaDB, SkySQL	Open source relational SQL database from the developers of MySQL; SkySQL cloud service	MySQL alternative with an impressive lineage
Microsoft	Azure SQL, Cosmos DB, others	SQL Server in deployment options; Cosmos DB managed NoSQL database; Azure Synapse Analytics	#1 ranked cloud provider in the Cloud Wars Top 10; SQL Server popular with developers; many database options
MongoDB	MongoDB Atlas, Realm	NoSQL document DB	Popular with developers for new apps, but niche enterprise player
Neo4j	Neo4j, Aura Cloud	Open source NoSQL graph DB; Aura Cloud available o AWS, GC, Azure	A leader in the fast-growing market for graph databases; $325M funding in 2021
Oracle	Oracle Autonomous Database, MySQL, Oracle NoSQL, others	Oracle DB supports converged workloads; Autonomous Data Warehouse for analytics	Still the leader in enterprise-class database capabilities
PlanetScale	PlanetScale	MySQL database + open source Vitess for scalability; available in 3 deployment options	Newcomer to the Cloud Database Top 20; grew from work at Google/YouTube
Redis	Redis Enterprise Cloud	NoSQL in-memory data store supports abstract data, such as strings, lists, maps	Commercial version of open source database with subsecond responsiveness

Vendor	Database Offerings	Differentiators	My Take
SAP	SAP HANA, SAP IQ, Adaptive Server, SQL Anywhere	SAP HANA & HANA Cloud are in-memory relational databases; others support OLTP, analytics, mobile	Comprehensive family of data management software for SAP environments
SingleStore	SingleStore	Multi-model SQL database supports transactions, analytics, various data types	Emphasis on "data intensity" and converged workloads in one system is compelling
Snowflake	Snowflake Data Cloud	SQL-compatible data warehouse and data lakes	Fresh approach to analytics and data sharing puts data into hands of more users
Teradata	Teradata Vantage	SQL database with scalable MPP architecture for analytics workloads	Proven, enterprise-class data warehouse used across industries
Yugabyte	YugabyteDB	Open source distributed SQL database, available on-prem, self-hosted, or as managed service	Cloud-native, Postgres-compatible database provides modernization path for relational databases

Subscribe to the Cloud Database Report