Cloud Database Platform Positioning   by, David Floyer

February 18th, 2021

Enterprise Cloud Database Platforms will need two main characteristics for a future multi-cloud distributed environment.

Convergence is the ability to use multiple database and data types in real-time or near-real-time with predictable performance. Convergence support includes transactional and analytic databases types working together, as the most valuable applications will combine both.

Distribution is the ability to:

Distribute copies of data to meet availability and consistency requirements of databases in many local domains. Distribute domain ownership of databases in the form of a "Data-mesh" while maintaining control of metadata to enable enterprise compliance, provenance, and data security.

Wikibon believes Oracle has the strongest Cloud Database Platform with Autonomous Database. It offers a Tier-1 database foundation, Oracle Cloud@Customer, which provides identical X8M hardware and software in on-premises private clouds managed centrally. Oracle also provides a Dedicated Region Cloud@Customer, which is a complete portfolio of public cloud services and Oracle Fusion SaaS applications into an on-premises data center.

Wikibon believes that Snowflake has impressive ease-of-use for end-users and a potentially impressive data-mesh vision. Snowflake will need to execute well and quickly on its overall vision.

Wikibon is impressed with the vision for Microsoft Cosmos DB. Microsoft also has a Tier-1 database foundation in SQL Server. Wikibon believes that Microsoft will need to integrate its Tier-1 transactional SQL Server with Cosmos DB in the future.

Cloud Database Platform Positioning

Premise

Enterprise Cloud Database Platforms will need two main characteristics for a future multi-cloud distributed environment.

1. **Convergence** is the ability to use multiple database and data types in real-time or near-real-time with predictable performance. Convergence support includes transactional and analytic databases types working together, as the most valuable applications will combine both.
2. **Distribution** is the ability to:
   1. Distribute copies of data to meet availability and consistency requirements of databases in many local domains.
   2. Distribute domain ownership of databases in the form of a "Data-mesh" while maintaining control of metadata to enable enterprise compliance, provenance, and data security. (See the reference in Footnotes below on Data-mesh for more detail.)

Challenges of Traditional Databases; How Cloud Databases Help

Complexity

Traditional databases need expensive people specialized in software and hardware to run large database systems. DBAs, storage specialists, system administrators, and network specialists must keep databases and infrastructure up to date, optimize performance, isolate bottlenecks, and keep fragile systems running. Data workflows are complex and tortuous for transactional, analytic, and other single-purpose database systems, requiring data managers, data architects, data scientists, data engineers, statisticians, storage specialists, and many more specialized and expensive people to manage the data stored in many silos. Data is moved and transformed without context. The resulting centralized data lakes and data warehouses are the triumph of hope over experience.

The first way Cloud Database Providers can tackle the problem of complexity is by providing autonomous capabilities. The Cloud Database Provider delivers automation of upgrades, indexing, performance, recovery, backup, etc. Providers with volume and who utilize machine learning algorithms based on many users will drive continuous improvements.

The second way Cloud Databases can provide radical simplicity is to support the physical distribution and management of data close to where the enterprise first creates and uses it. The people responsible for data creation are usually best positioned to define what the data means in their domain and keep the data in context with other data and processes to support its business needs.

The third requirement for an enterprise Cloud Database is to automate the collection and distribution of metadata between all the domains and provide the services to optimize and predict workloads' performance using data over multiple locations.

Wikibon believes that autonomous databases are essential to reduce the cost and complexity of using data. This simplicity and automation will allow domain business staff to directly define and consume their data and broadcast the metadata to other domains.

Database and Data Types

Database and data types have exploded in the last two decades, with many providers introducing different database types. Examples include Advanced Analytic, AI Inference, AI Learning, Blockchain, Document, Graph, Key-value, Inference, In-memory, Log-file, NoSQL, Relational-operational Time-series databases. Each of these databases deals with different data types and provides specialized structures to improve performance and reduce complexity.

Some of these database types have robust independent implementations, such as MongoDB (Document), SAP HANA (In-memory), and Splunk (Log-file). Cloud providers support them all; for example, AWS has sixteen different database types. However, the user workflows between the different database types have become extraordinarily complex and insecure. Data transformations and the time to move data result in higher costs, loss of data context, and significantly reduces the data's business value.

The longer elapsed times mean that synchronizing applications using transaction and analytic databases is much more challenging (and usually impossible) to achieve. Synchronous applications enable a more significant potential for automation of business processes, making them more valuable in creating data-first business processes. The bottom line is converged databases that scale allows faster automation and simplification of business processes and reduce the number of complex asynchronous business processes.

Oracle, and to some extent Couchbase, have developed converged databases with a single database engine providing integrated support of different databases and data types. The converged database supports transactional and analytic database types working together. Equally important is the performance and automation of each database type within the converged database. Converged databases provide an essential game-changing reduction in complexity with equal or better performance than specialized databases.

Database Performance

Cloud Database Platforms need to support synchronous business processes in real-time or near-real-time. These database platforms are complex and require specialized hardware and platform services to optimize performance. Wikibon is predicting that flash drives will be the same cost as HDDs by 2026. Flash and other non-volatile memory technologies will allow simple single-tier storage solutions that help enable distributed database solutions. Improved protocols such as NVMe and RoCE (RDMA over Converged Ethernet) radically reduce protocol overhead and improve latency. NVMe also provides faster and lower costs for any-to-any connectivity of processors and storage. Non-volatile memory technology also simplifies recovery and restart and provides lower-cost shared caches.

Wikibon believes that elastic scalability is a crucial attribute for successful Cloud Databases. These databases must use cloud services to scale resources instantaneously for short periods while the database keeps running to minimize elapsed time-to-value.

Arm processors and systems are now faster and cheaper than x86 processors, and the rate of improvement is significantly quicker. The challenge for Arm processors is that vendors have designed almost all database software for x86 processors. However, the manufacturer of Arm wafers is over ten times larger than x86 wafers. This volume means that the learning curves of both manufacturing and design have outstripped x86.

x86 has an advantage because vendors will need time and resources to migrate software to Arm. However, Apple and Microsoft are converting their PC platforms to run primarily on Arm and extend the platform to support Arm mobile applications. AWS has invested in Arm-based Graviton and is migrating its platform to ARM at a rapid rate. Wikibon expects cloud platforms and Cloud Database Platforms to migrate to work with Arm, mainly because Apple and Arm have developed heterogeneous architectures that can accelerate specific workloads by a hundred times compared with the general-purpose x86 architectures. Wikibon

believes that Cloud Database Platforms will adopt heterogeneous Arm technology early because of databases' particular performance requirements.

Data Distribution

Enterprises are increasingly distributing the location of data creation. The continued reduction in the cost of Micro-Electrical Mechanical Systems (MEMS) capabilities is pushing enormous amounts of data creation to the Edge in warehouses, retail outlets, energy distribution, etc. Edge devices must often be autonomous, such as autonomous cars, planes, trains, and more.

Moving large amounts of data is very expensive. It takes a significant amount of time, and (as previously discussed) data loses context if not processed at the point of creation. It follows that databases must support on-premises equipment and processing, in addition to the cloud. This distributed data processing's primary use is almost always to support remote business operations, such as plants, warehouses, autonomous vehicles, etc. The secondary benefit is to provide data in context to other parts of the business.

Cloud Database Platforms must therefore be available in both public and private clouds. The business people who support the business operations at each location must be able to define their data to support their operations and at the same time provide access and context to other sites.

At the same time, organizations must ensure compliance with legal requirements and ensure all data's provenance and safety at the local and national level. Databases will need to support a "Data-mesh" approach and the metadata requirements.

Cloud Database Platform: Horses on the Track

Wikibon believes there are two fundamental requirements for a Cloud Database Platform: the ability to support converged databases and support a devolved distributed business environment. Figure 1 provides Wikibon's assessment of the current position of Cloud Database Platform providers.
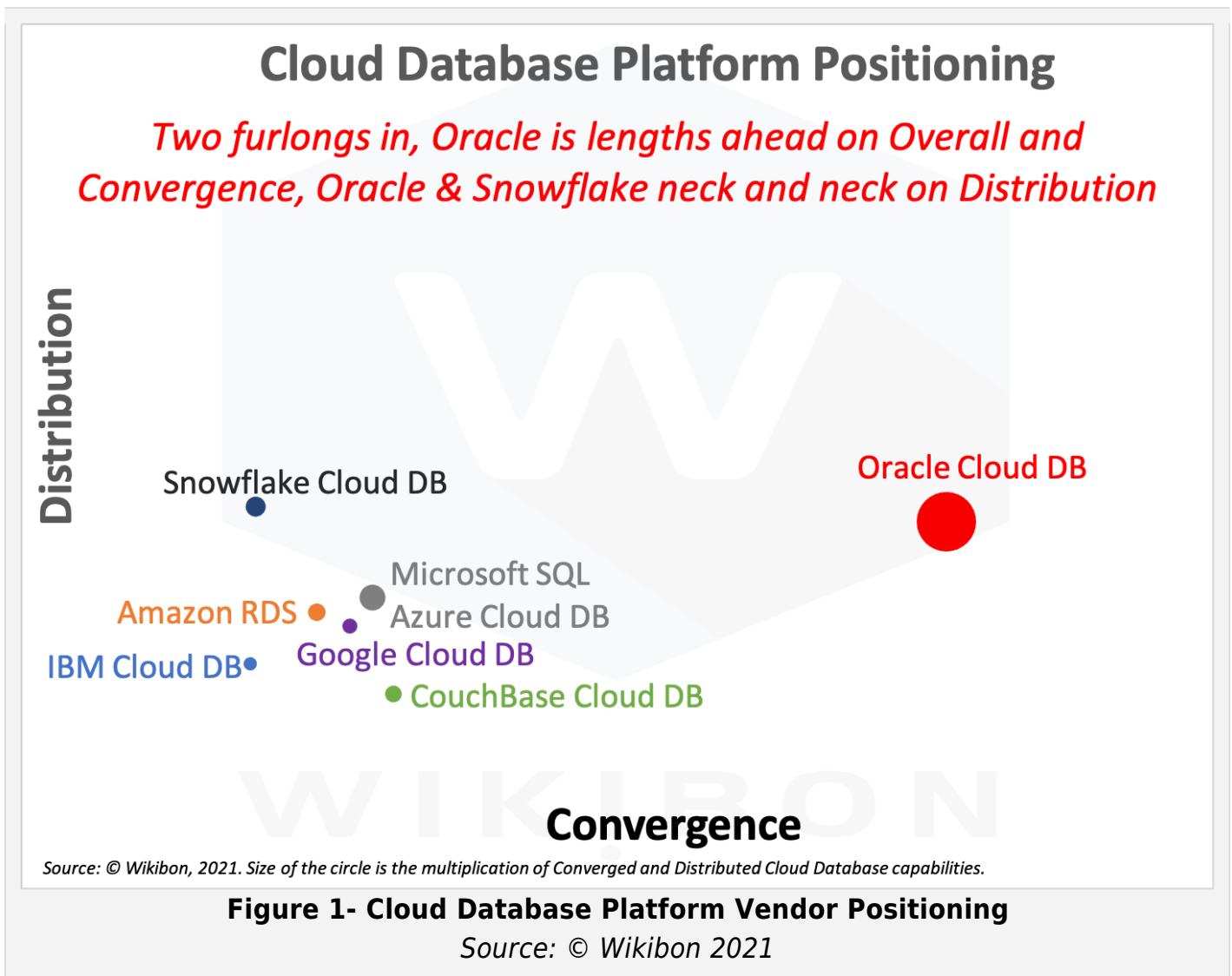
Cloud Database Platform Positioning

Two furlongs in, Oracle is lengths ahead on Overall and Convergence, Oracle & Snowflake neck and neck on Distribution

Snowflake Cloud DB

Oracle Cloud DB

Microsoft SQL
Azure Cloud DB
Amazon RDS
Google Cloud DB
IBM Cloud DB
CouchBase Cloud DB

Convergence

Source: © Wikibon, 2021. Size of the circle is the multiplication of Converged and Distributed Cloud Database capabilities.

Figure 1- Cloud Database Platform Vendor Positioning
Source: © Wikibon 2021

Figure 1 shows the Wikibon assessment of Distribution capability on the y-axis and Convergence capability on the x-axis. The size of the circle is the multiplication of the two scores.

AWS

AWS has taken 16 open-source databases and integrated them well but separately into the AWS PaaS. On the Distribution axis, AWS has delivered Outposts, allowing on-premises RDS implementations, which is currently a small subset of the 16 databases. AWS also has good support for data protection across different regions.

AWS has implemented some enhanced data movement between different databases but has no announced strategy to invest in these databases' converged integration. As a result, enterprises that need to combine data from different databases must perform time-consuming and costly ETL. This architecture makes combining transactional and analytic data extremely difficult.

Couchbase

Couchbase focuses explicitly on providing support for transactional and analytic databases. It claims a NoSQL heritage but has implemented a SQL-compliant combination of analytical and transactional databases. Couchbase has limited cloud database distribution capabilities.

IBM

IBM offers a Tier-1 (See Footnotes for a definition of Tier-1) DB2 Cloud Database, which provides integrated transactional and analytic capabilities. IBM provides good support for data protection but little logical distribution support. IBM is working on a separate distributed architecture platform. It is unclear if IBM will integrate the two approaches.

Google

Google offers BigTable and Spanner databases for OLTP, and BigQuery is a data warehouse database. There is little converged capability between the OLTP and data warehouse databases. BigQuery is at heart a columnar database written to take advantage of cloud scalability. BigQuery is serverless, has hybrid NoSQL capabilities such as record type, and can hold and address raw JSON documents. BigQuery is a lightly converged analytic database that excels when database sizes are massive.

Google offers Google Anthos for on-premises Cloud Database Platform requirements. Google Anthos typically runs Cisco servers that are not used by GCP, meaning a lack of architectural equivalency from cloud to on-premises.

Microsoft

Microsoft has a Tier-1 SQL Server Database for on-premises deployment. Currently, it offers Azure SQL Cloud Database for transactional cloud services and Azure Synapse as a data warehouse Cloud Database. Wikibon assesses that users will find little converged capability between the SQL Cloud and Synapse Database offerings.

Microsoft is developing Azure Cosmos DB as a globally distributed, scalable, multi-model database cloud service extension to Azure Synapse. It is building this service from the ground up. Cosmos DB provides native support for NoSQL and OSS APIs, including MongoDB, Cassandra, Gremlin, etcd, Spark, and SQL. It offers multiple consistency models from strong to eventual and says it supports low read and write latencies. Wikibon understands that the emphasis of Cosmos DB is to integrate analytic requirements. The Microsoft Azure SQL services separately support transactional needs.

Microsoft offers Azure Stack for on-premises Cloud Database requirements. As noted in prior Wikibon research, Azure Stack requires an Azure Stack operator on-site. Microsoft offers a selection of qualified systems from five different vendors. Wikibon does not rate this approach as satisfactory, as the Azure public cloud does not deploy any of these hardware systems.

Oracle

Wikibon believes Oracle has the leading converged database implementation of all providers. In its latest Oracle Database 21c announcement, Oracle has improved all aspects of its Tier-1 converged Cloud Database Platform, including performance improvements for in-memory, graph, multitenant processing, as well as the addition of JavaScript in-database. AutoML for in-database machine learning (ML) is an additional automation capability. Blockchain Tables provide immutable insert-only tables in Oracle Database. A native JSON binary data type was introduced, which increases Document Database performance and function. Oracle has a clear on-going strategy of reducing complexity for DBAs and data users through automation, performance, and integration of all database and data types and is executing this strategy well.

On the distribution side, Oracle has excellent cluster and distributed processing with RAC, Active Data Guard (active-passive copy distribution), Sharding (shared-nothing geo-distribution of horizontally partitioned data), and GoldenGate (active-active copy distribution). Oracle also announced its distributed sharding performance and flexibility enhancements in Database 21c.

Snowflake

Snowflake has created significant momentum, focusing on reducing complexity with improved ease-of-use and time-to-value for data warehouses. Snowflake has also introduced enhanced capabilities for sharing data warehouses, with some Data-mesh capabilities. However, Snowflake has an isolated analytic SQL database with limited advanced functions and is still in the gate regarding convergence. Snowflake currently has an immature ability to integrate machine learning.

Conclusions and Recommendations

General Assessment

An early study of the automobile market concluded that the number of chauffeurs available would constrain the market's size. Ford showed that making cars simple and in volume made chauffeurs redundant. Similarly, the demand for household telephones was thought to be constrained by the number of telephone operators before AT&T automated phone calls.

In the same way, Wikibon believes that Cloud Database Platforms will remove the dependence on expensive IT staff by automating many mundane tasks and devolving data management and exploitation to the lines of business.

Figure 1 shows the horses on the Cloud Database Platform track. Wikibon's overall assessment is that two furlongs in, Oracle is lengths ahead on Overall and Convergence, Oracle and Snowflake are neck and neck on Distribution.

Both of these dimensions are critical to providing a Cloud Database Platform with the automation, ease of use, robustness, and flexibility to support data-led enterprises without crippling IT staff overheads. Figure 1 also shows that across providers, convergence is further ahead than distribution.

Wikibon believes that large enterprises understand that the choice of Cloud Database Platforms is a more critical and strategic decision than the choice between alternative Cloud IaaS or PaaS platforms in the data-led journey.

Vendor Future Assessments

Databases are complex technologies where, in addition to convergence and distribution, automation, performance, and reliability are critical.

Wikibon assesses that **AWS** is doing well within its limitations. It currently supports mainly smaller organizations with smaller-scale database requirements. As Wikibon noted in prior research, the more databases and data types exist, the more specialized transfer systems are required. Sixteen databases would require 120 different transformation transport systems. Fifty database types would require 1,225.

AWS will undoubtedly provide a stable and highly performant IaaS and PaaS platform for itself and other providers. In particular, AWS is ahead of other vendors moving to ARM technology, which can reduce costs and increase performance. However, AWS still has to invest in converged and distributed database software and hardware integration to utilize this infrastructure fully.

Suppose AWS wants to move upmarket and meet the requirements of enterprise-level, mission-critical databases and provide for the aggressive scope of future automation applications. In that case, Wikibon believes AWS must radically change its strategy. Instead of leaving developers to integrate the databases with its platform's high-availability capabilities, AWS must provide a fully supported high-availability and recovery capability. To develop an integrated Cloud Database Platform, AWS will need to invest significant resources to move from upgrading open-source databases to developing in-house AWS software for an integrated Cloud Database Platform. AWS will need to supply both convergence and distribution capabilities

in such a platform. Also, AWS will need to bring more of its databases to Outposts to meet the distribution capabilities fully.

Google has developed its databases primarily for its own use in their particular and unique business but has struggled to make its Cloud Database offerings relevant to enterprises. Wikibon believes that Google will probably partner with other vendors.

IBM has significant experience in Tier-1 databases, has strong enterprise service and sales capabilities, and has developed impressive potential Data-mesh capabilities. Although IBM is nascent in cloud services, IBM and Red Hat have the financial, technical, and research capabilities to invest in developing a Cloud Database Platform. Wikibon believes IBM and Red Hat should and probably will invest in developing a full-fledged Cloud Database Platform.

Microsoft has significant experience in developing a Tier-1 database and has strong enterprise marketing and software distribution capabilities. It has built a strong SaaS presence around its Office and Teams software. Also, Microsoft is a primary IaaS/PaaS cloud infrastructure provider and is supporting a multi-cloud strategy.

Wikibon is impressed with the vision for Microsoft Cosmos DB. Wikibon believes that Microsoft will need to integrate its Tier-1 transactional SQL Server with Cosmos DB in the future. On the distribution axis, Microsoft has robust distributed services, and Cosmos DB allows data to be placed close to the users. However, Microsoft will need to enhance its Azure Stack offering based on third-party hardware, which Wikibon assesses as inadequate. Overall, Wikibon believes Microsoft has the financial resources and technical capabilities to develop a Cloud Database Platform and the management flexibility to partner with others. Wikibon expects Microsoft to invest heavily in developing a Cloud Database Platform and maintain its lead over AWS and Google.

Figure 1 shows **Oracle** has by far the best Tier-1 Cloud Database Platform and has developed a robust infrastructure platform with Oracle Exadata X8M that is the basis of database services on Oracle Cloud Infrastructure and Cloud@Customer, delivering architectural equivalency. Oracle Cloud@Customer Cloud Database Platform is moving powerfully down the database convergence and autonomous dimensions, as was shown by the Oracle Database 21c announcement in early 2021. Oracle has already developed a multi-cloud agreement with Microsoft. Wikibon expects that AWS and Oracle will reach an agreement to ensure that Oracle runs well on AWS.

Oracle will need to build on its sharding feature and invest strongly in a Data-mesh architecture to allow enterprises to devolve and simplify data-led strategies to include the lines-of-business. This strategy may bring them into potential conflict with centralized IT development executives who value and evaluate Oracle's Database products. However, Wikibon believes the Cloud Database Platform is ripe for disruption driven by simpler database and data management software that will empower less technical users in the lines of business. Wikibon believes that Oracle has the vision and management drive to extend its lead in Cloud Database Platform.

Snowflake has made good initial progress on ease-of-use and devolved data warehouses. Snowflake needs to expand its TAM to justify its market cap and has experienced technical architects. Wikibon expects that Snowflake will invest in convergence with a combination of acquisition and integration. Wikibon understands that the current Snowflake data-mesh implementation is currently envisioned as an ability to connect with any other domain, internal and external, and move any required data. Wikibon believes that Snowflake will need to develop a distributed database capability to allow data to remain in place. Although this is technically challenging, Wikibon considers that Snowflake must take up the challenge or risk being sidelined by Microsoft and Oracle.

Assessment Summary

Wikibon believes Oracle has the strongest Cloud Database Platform with Autonomous Database. It offers a

Tier-1 database foundation, Oracle Cloud@Customer, which provides identical X8M hardware and software in on-premises private clouds managed centrally. Oracle also provides a Dedicated Region Cloud@Customer, which is a complete portfolio of public cloud services and Oracle Fusion SaaS applications into an on-premises data center.

Wikibon believes that Snowflake has impressive ease-of-use for end-users and a potentially impressive data-mesh vision. Snowflake will need to execute well and quickly on its overall vision.

Wikibon is impressed with the vision for Microsoft Cosmos DB. Microsoft also has a Tier-1 database foundation in SQL Server. Wikibon believes that Microsoft will need to integrate its Tier-1 transactional SQL Server with Cosmos DB in the future.

Action Item

Wikibon strongly recommends enterprise senior enterprise executives focus intently on the development of Cloud Database Platforms, which are fully converged and excel across both transactional and analytic data types. The converged database must also support data-mesh features and allow both distributions of data where the data is created and the ability of lines of business to work directly and own the data that is vital to them. They will radically improve the cost, speed to value, and functionality when used to implement data-driven strategies.

At the moment, Oracle Database is a Tier-1, mission-critical converged Cloud Database Platform, is lengths ahead in the convergence dimension, and is neck and neck on the distribution dimension with Snowflake. Microsoft is early in the development of Cosmos DB but is architecturally sound. Microsoft needs to expand its convergence and distribution capabilities.

Wikibon recommends large enterprise senior enterprise executives with significant Oracle installation start investing in Oracle Autonomous Database on OCI or Cloud@Customer while continuing to track the development of Cloud Database Platforms. Wikibon believes this can radically improve the cost, speed to value, and functionality used to implement data-led strategies.

Footnotes

Data-mesh Reference

Zhanak Dehghani of ThoughtWorks is an authority on the simplification and devolution of Enterprise Data Management. This discussion between Dave Vellante and Zhanak Dehghani is an excellent introduction to the subject.

Cloud Database Platform Definition

Cloud Databases are an emerging category of databases. Cloud Databases are a service delivered from an integrated cloud platform. This Cloud Database Platform (CDP) enables users to utilize Cloud Databases Services on demand without the initial investment cost for equipment and licenses. The CDP also allows enterprises to manage distributed databases remotely on private clouds or shared clouds.

A CDP can reside in a private cloud, public cloud, hybrid cloud, and multi-cloud environments. From an application perspective, the database services are identical. The only difference lies in where the database resides.

Tier-1 Database Definition

Tier-1 Databases have a strong track record of performance and reliability for large-scale mission-critical applications. Such a track record takes many years to achieve. At the moment (2021), Wikibon considers

only three vendors offer Tier-1 databases. The three are IBM DB2, Microsoft SQL Server, and Oracle Database.

David Floyer spent more than 20 years at IBM, holding positions in research, sales, marketing, systems analysis and running IT operations for IBM France. He worked directly with IBM's largest European customers, including BMW, Credit Suisse, Deutsche Bank and Lloyd's Bank. Floyer was a Research Vice President at International Data Corporation (IDC) and is a recognized expert in IT strategy, economic value justification, systems architecture, performance, clustering and systems software.

David Floyer

@dfloyer
david.floyer@wikibon.org