



GoldenGate Augments Cloud Databases

by, David Floyer

May 21st, 2021

Oracle GoldenGate is now offered as an integral part of Oracle Cloud Infrastructure (OCI), which dramatically lowers the costs of provisioning and maintaining the GoldenGate advanced features for all the Database Platforms supported. Previous Wikibon research established that Convergence & Distribution are important dimensions of any Cloud Database.

Wikibon strongly recommends that 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 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, time-to-value, and functionality when used to implement data-driven strategies.

Premise

Oracle GoldenGate is now offered as an integral part of Oracle Cloud Infrastructure (OCI), which dramatically lowers the costs of provisioning and maintaining the GoldenGate advanced features for all the Database Platforms supported. [Previous Wikibon research established that Convergence & Distribution](#) (see the section below for definitions) are important dimensions of any Cloud Database Platform.

The premise of this research is that GoldenGate improves the convergence-dimension and significantly improves the distribution-dimension for all the Cloud Database Platforms that GoldenGate supports. As a result, GoldenGate improves the integration between different databases and significantly reduces database conversion.

Scope of this Research

This research adds the impact of GoldenGate on the Oracle Cloud Database Platform and adds SAP HANA into the earlier research titled “Cloud Database Platform Positioning,” published February, 19th 2021. The research has also been updated with additional changes and feedback from the community.

GoldenGate 101

The explosion of Databases

Databases were originally for systems of record (transactional databases) and analytic systems (data warehouse databases). In the last decade, enterprises have deployed many new databases, such as NoSQL, log, graph databases, and many others. Wikibon would point out that the reason for these database types is the emergence of new data types. Log files have very different characteristics and processing requirements from systems of record. Splunk is a very effective log database.

The increase in data types is continuing. Blockchain, machine learning, and inference data types are the newest, and Wikibon expects this trend to continue developing.

Changing Enterprise Database Strategy

As a result of the increases in data types, most enterprises have business-critical data distributed across the enterprise in heterogeneous databases. However, enterprise system designers are increasingly focusing on simplifying and automating business processes as part of a drive to increase efficiency and flexibility, i.e., digital transformation.

As a result, the system architects, designers, and programmers need to integrate data from different sources and databases in real-time or near-real-time. This is not easy.

Avoiding Database Conversion

The Irish are famous for their humorous reply when asked for directions, “I wouldn’t start from here.” Many database vendors argue that the solution to heterogeneous databases is to convert every database to their database or set of databases.

Anybody with experience with database conversions knows this solution does not work, whatever gimmicks and magic conversion software are promised. Wikibon knows first-hand how wholesale conversion strategies have caused banks, insurance companies, and many other enterprise types irreparable harm.

The cost of database conversion and the disruption to the lines of business makes a conversion strategy a non-starter. Enterprises should avoid this strategy like the plague and remove any employees or suppliers that advocate such an approach.

Avoiding Database Sprawl

Enterprises need to take advantage of different data types and use databases to manage these new data types. However, introducing a new database for every new data type will lead to database sprawl and make it much more difficult to achieve automation and digital transformation.

Wikibon strongly recommends an alternative strategy of minimizing the number of databases for different data types. Enterprises can achieve this by using converged databases that support multiple data types. For example, Oracle and SAP support almost all data types, whereas each AWS database and Snowflake's database support very few data types.

Wikibon recommends enterprises choose databases carefully, with the main criteria being the ability to support all data types and support distribution of databases across the enterprise in real-time or near real-time.

Augmenting Databases with GoldenGate

Oracle acquired GoldenGate in 2009. It is a comprehensive software package for real-time data integration and replication. It also enables high availability solutions, transactional change data capture, transformations, and verification between operational and analytical enterprise systems.

Oracle GoldenGate has the following features:

- Moves committed transactions, which enables consistency and improves performance.

- Moves data in real-time, which reduces end-to-end latency.

- Supports a wide range of heterogeneous databases running on a variety of operating systems. For example, you can replicate data from an Oracle Database to a different heterogeneous database.

- The GoldenGate microservices architecture is "cloud-native."

- The tight integration allows high performance with minimal overhead on the underlying databases and infrastructure.

Oracle GoldenGate enables the exchange and manipulation of data at the transaction level among multiple, heterogeneous platforms across an enterprise. It moves committed transactions with transaction integrity and low overheads across existing infrastructures. Its modular architecture enables users to extract and replicate selected data records, transactional changes, and changes to DDL (Data Definition Language) across various topologies.

Oracle GoldenGate is used in many environments as a tool to enable database integration. For example, it is certified on the AWS site to support [AWS Aurora](#) and other databases. There are also other tools and software that achieve similar results.

Wikibon believes that the close integration of GoldenGate with OCI significantly changes the skill and time required to set up and maintain these environments. Wikibon believes this integration is a gamechanger and expands its potential use to enterprises of any size.

GoldenGate Business Continuity and High Availability

GoldenGate is an important component of the [Oracle Maximum Availability Architecture \(MAA\)](#). To establish and maintain an MAA environment, data must move between multiple servers and data centers. GoldenGate provides bidirectional active-active replication of Oracle and other databases to support the highest levels of business continuity.

Initial Load and Database Migration

Initial load is extracting data records from a source database and loading those records onto a target database. Initial load is a data migration process that is performed only once. Oracle GoldenGate allows you to perform initial load data migrations without taking your systems offline.

Data Integration

Data integration involves combining data from several disparate sources stored using various technologies and providing a unified view of the data. Oracle GoldenGate provides real-time data integration.

Definitions of Convergence & Distribution

Previous Wikibon research established there are two main dimensions to the evaluation of Cloud Database Platforms.

1. **Convergence** is the ability to use multiple data and database types in real-time or near-real-time with predictable performance. Convergence capabilities include enabling transactional and analytic databases to work together, as the most valuable applications will combine both. Enterprises can use GoldenGate to improve combining different database types.
2. **Distribution** is the ability to:
 - Distribute copies of data to meet availability and consistency requirements of databases in many local domains. GoldenGate can significantly assist in managing this distribution.
 - Distribute domain ownership of databases in the form of a “Data Mesh” while maintaining metadata control to enable enterprise compliance, provenance, and data security. (See the reference to Data Mesh in Footnotes below for more detail.)

GoldenGate & Cloud Database Platforms

GoldenGate Announcement as a Database Service

Oracle has recently announced GoldenGate as an OCI database service. In recent research, Wikibon identified two components that are vital for a successful Cloud Database Platform. These are Database Convergence and Data Distribution. GoldenGate in OCI will significantly enhance the Data Distribution component of most Cloud Database Platforms, including Oracle, AWS RDS, IBM DB2, and Microsoft SQL Server. This capability dramatically extends the range of customers who will take advantage of the GoldenGate functionality. IT does not need to invest resources to install, manage, and maintain the software on-premises.

Data Warehouse & Lake Problems

There are many database analyst’s views available for subsets of the database and database management. Data warehouses and data lakes are often stripped off and reviewed as a stand-alone product. Wikibon and other researchers have shown that the CIO and the lines of business think that the investments in data warehouses and data lakes have not yielded the results that the vendors promised. CIO concerns include late delivery of data, data compliance and provenance, poor documentation, and too many versions of the truth.

Wikibon believes that database convergence and database distribution capabilities are important because they allow enterprises to integrate data sources, ensure consistency between distributed data, and provide automated tools to ensure compliance and provenance. GoldenGate is an important element in achieving these goals.

Technology Impacts on Database Architecture

Wikibon believes that database technologies have evolved to remove the hard break between transactional systems of record and data warehousing. Data in-memory technologies, persistent storage layers, smart flash storage, and mixed columnar and row-based architectures have allowed operational systems of record to use real-time analytics when needed, proven provenance and compliance, and a single version of the truth. With instant scaling of database cloud services, these capabilities are central to allowing full automation of complex business processes.

Technology Impacts on Database Distribution

The other technology impact is the cost of computing. Traditional linear x86 architectures doubled in speed or number of cores every 18 months to two years (Moore's Law). However, new [heterogeneous parallel architectures from Arm](#) have improved processing speed by 118% over the last 5 years, measured in TOPS or Trillions of Operations per Second. The cost of compute is coming down, especially in [Matrix workloads such as ML and AI](#). The comparative cost of compute operations is dropping far faster than the cost of storage and networking.

This means enterprises will push real-time computing to where the data is created, mostly at the Edge. Enterprises will immediately process most of this data at the Edge, keeping only small amounts of valuable resultant data there. Again, enterprises will push requests for analysis to the Edge rather than move data across the network to a traditional data lake.

Databases will need to manage distributed real-time data, manage compliance and provenance, and manage a single version of the truth. GoldenGate will play an important role in this evolution.

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.

1. The first way Cloud Database Platform Providers can tackle the problem of complexity is by providing autonomous capabilities. The Cloud Database Provider delivers automation of upgrades, indexing, performance, recovery, backup, patching, and more. Providers with volume and who utilize machine learning algorithms based on many users will drive continuous improvements.
2. 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.
3. 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 for each one. Examples include Advanced Analytic, AI Inference, AI Learning, Blockchain, Document, Graph, Key-value, Inference, In-memory, Log-file, NoSQL, and 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 reduce 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 automating business processes, making them more valuable in creating data-first business processes. The bottom line is converged databases that scale allow faster automation and simplification of business processes and reduce complex asynchronous business processes.

Oracle, SAP HANA, and to some extent Couchbase, have developed converged databases with a single database engine providing integrated support for 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 faster. The challenge for Arm processors is that vendors have designed almost all database software for x86 processors. However, the manufacture of Arm wafers is over ten times larger than x86 wafers. This volume disparity means that the learning curves of both manufacturing and design have outstripped x86. The latest Apple M1 processors are performing faster than the latest equivalent Intel Tiger Lake i7 chips.

x86 currently 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, and more. Edge devices must often be autonomous, such as autonomous cars, planes, trains, and other machines.

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, and autonomous vehicles. 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 define their data to support their operations and simultaneously 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 to achieve this goal.

Cloud Database Platforms: 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.

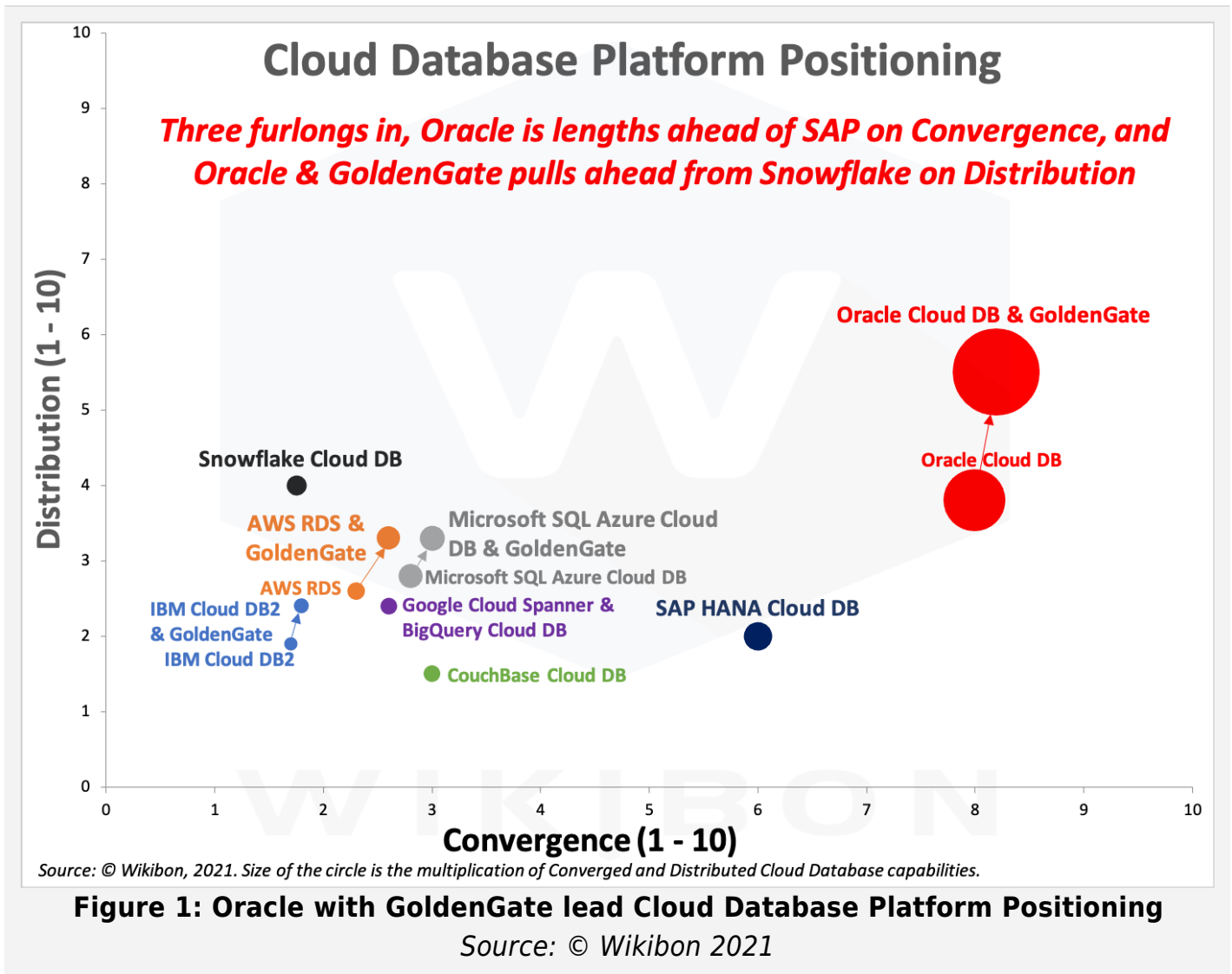


Figure 1 shows the Wikibon assessment of Distribution capability on the y-axis and Convergence capability on the x-axis. The diameter of the circle is the multiplication of the two scores.

GoldenGate support AWS RDS (yellow), IBM DB2 (light blue), Microsoft SQL Server (grey), and Oracle Databases (red). Each of these Cloud Database platforms is assessed with and without GoldenGate, which provides distribution benefit and a small convergence benefit. The arrows connect the original score without GoldenGate to a higher score with GoldenGate.

GoldenGate does not support CouchBase (green) and provides limited support for SAP HANA (dark blue), and Snowflake Cloud Database platforms (black). There is a single point for these platforms. The individual scores on each Database Cloud Platform in Figure 1 are discussed in more detail in the following sections.

Overall, the chart shows that the Oracle Cloud Database platform with GoldenGate is a clear leader on both dimensions. However, Wikibon observes that it is still early in the evolution of Cloud Database Platforms. There is still significant potential for improvement on the distribution dimension, which needs additional features such as distributed data catalogs and additional advanced distribution capabilities to support data-meshing.

AWS

AWS has taken 16 open-source databases, forked most of them with proprietary improvements not given

GoldenGate Augments Cloud Databases

back to the community, and integrated them well but separately into the AWS PaaS. On the Distribution axis, AWS has delivered Outposts, allowing at the moment limited 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 in different countries.

AWS has implemented some enhanced data movement between different databases but has no announced strategy to invest in converged database integration. As a result, enterprises that need to combine data from different databases must perform time-consuming and costly ETL. This architecture makes combining data from different data types (e.g., transactional and analytic data) challenging and time-consuming.

Apart from an [interesting blog on Data Mesh](#), AWS has not announced any data mesh capabilities. GoldenGate provides good support for AWS RDS and is available as a tool on the AWS platform. However, it is supported as an as-is tool and is not available as a fully managed cloud service.

The availability of GoldenGate as a service on OCI is a significant improvement and provides users of AWS RDS with consistent active-active options in real-time analytics on OCI. The GoldenGate score signifies this type of integration, not an as-is implementation.

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 or data mesh capabilities. GoldenGate does not support Couchbase.

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.

GoldenGate gives excellent support for DB2, which is found in many mainframe systems and financial organizations. DB2 databases are often complex, and avoiding conversion and integration will be a sound strategy for many large financial organizations.

IBM is working on a separate distributed architecture platform.

It is unclear if IBM will integrate the two approaches; Wikibon hopes they will, together with Red Hat, provide a Tier-1 Cloud Database Platform. Wikibon also hopes that they will improve the open-source legal foundation to encourage contributions from AWS and others.

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 on Cisco servers that are not used by GCP, meaning a lack of architectural equivalency from cloud to on-premises.

GoldenGate certifies GCP BigQuery and GCP Object Storage as delivery targets.

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. However, GoldenGate certifies SQL Server and Azure Synapse (delivery target only) for Microsoft SQL Server, and enterprises can use it to improve integration between transactional and analytic workloads.

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, lacking equivalency as the Azure public cloud does not deploy any of these hardware systems.

Microsoft has the software and architectural capabilities to become a leader in providing a Cloud Database Platform and building its strong SQL Server customer base. It is giving mixed signals on intent. GoldenGate certifies both SQL Server and Azure Synapse (Synapse is delivery target only).

Oracle

Wikibon believes Oracle has the leading converged database implementation of all other providers in Figure 1. In its latest Oracle Database 21c announcement, Oracle has improved many aspects of its Tier-1 converged Cloud Database Platform, including performance improvements for in-memory, graph, and multi-tenant processing. In addition, JavaScript is 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 ongoing 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 (Real Application Clusters), Active Data Guard (active-passive copy distribution), and Sharding (shared-nothing geo-distribution of horizontally partitioned data). Oracle also announced its distributed sharding performance and flexibility enhancements in Database 21c.

Oracle Database with GoldenGate

Oracle acquired GoldenGate Oracle in 2009 and is an integral part of Oracle Maximum Availability Architecture (MAA). Wikibon has always admired Oracle GoldenGate's unique ability to deploy fault-tolerant database replication combined with real-time operational analytic insights, supporting Oracle and non-Oracle data sources. GoldenGate is installed in the majority of Fortune 2000 enterprises as part of MAA implementations.

The downside of GoldenGate is the complexity and cost of providing and maintaining these capabilities. However, the latest GoldenGate release on OCI is a managed service, which solves the complexity and cost problems and extends these capabilities to every company size.

GoldenGate OCI service in conjunction with the Oracle Database Cloud platform improves the Convergence

dimension slightly and significantly improves the Distribution dimension.

SAP HANA Cloud

SAP HANA is an in-memory SaaS service where the updates and management of SAP are managed by SAP. The fundamental structure of SAP HANA is a data in-memory columnar SQL database that supports a broad range of data and database types, including Analytics, JSON documents, Graph, Spatial, Time-series, Machine Learning, and Blockchain. Its main advantage is to shorten the time required to create SAP analytic reports. SAP has an excellent track record in achieving this benefit.

The SAP HANA architecture also has a Row-based view of the data, enabling SAP to provide transaction processing of SAP systems of record. As discussed in the introduction, this capability is important to enterprises because it allows applications to have consistent and coherent access to real-time analytic data and provide a single source of truth, eliminating the risk of inconsistent data across the organization and improving the ability to automate complex business processes fully.

However, SAP does not currently provide the high availability, recovery, and transaction performance expected for Tier-1 systems-of-record in HANA Cloud S4. SAP has said that HANA will support all environments, including HA transaction services, by December 2025. On December 31, 2027, SAP plans to withdraw support for SAP “Legacy” products running on other databases. This date is the latest of many that SAP has published. Given customers’ reluctance to move, Wikibon believes this date will move out again.

Wikibon would point out there are only three Tier-1 Databases at the moment. They are IBM DB2, Microsoft SQL Server, and Oracle Database. It took the vendors decades to provide and prove the advanced high-availability and recovery features. The software market usually has a long tail, and ERP is no exception. SAP is a leading vendor in ERP, with about 6% of the total ERP market. Wikibon’s research indicates that most of SAP’s largest customers run on Oracle Database. Wikibon believes SAP does not have the volume of SAP HANA database instances to develop full Tier-1 capabilities and will not increase volumes by selling SAP HANA as a general-purpose, stand-alone database.

Wikibon believes the current cloud databases from Google, IBM, Microsoft Azure, Oracle (and AWS, assuming it invests in developing a fully converged database) will radically improve multi-cloud and cloud-on-premises converged database solutions. Wikibon believes it is very likely and SAP customers will demand to continue to run the Tier-1 database of their choice for SAP and other software.

Apart from an [interesting blog on SAP HANA data mesh capabilities](#), SAP has limited capabilities for implementing a distributed data mesh. Some enterprises use GoldenGate with SAP HANA with integration supported via existing certified APIs such as [JDBC](#), [Kafka](#), and [Object Storage](#).

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 cloud-only 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.

Some enterprises use GoldenGate with Snowflake with integration supported via existing certified [APIs such as JDBC](#), [Kafka](#), and [Object Storage](#).

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 three furlongs in, Oracle is lengths ahead on the Convergence dimension. Oracle with GoldenGate has moved ahead of Snowflake on the Distribution dimension.

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. There is still room for innovation in Cloud Database Platforms, and companies like Snowflake have shown the importance of ease-of-use and devolution to the lines of business.

Wikibon believes that large enterprises understand that the choice of Cloud Database Platform 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. AWS is ahead of other vendors moving to Arm technology, reducing costs and increasing performance. However, AWS still has to invest in converged and distributed database software with built-in tier-1 and HA recovery systems to be a long-term enterprise player in cloud databases.

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 an integrated and fully supported Tier-1 level high-availability and recovery capability. To develop an integrated Cloud Database Platform, AWS will need to invest significant resources to change from adding extensions to open-source databases to developing in-house AWS software for an integrated Cloud Database Platform. AWS will need to develop 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. And AWS will need to allow its databases to run on other clouds such as Alibaba, Google, Microsoft Azure, Oracle OCI, and Tencent.

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

GoldenGate Augments Cloud Databases

Database Platform. Wikibon believes IBM and Red Hat should and probably will invest in developing a full-fledged Cloud Database Platform with both convergence and distributed data mesh capabilities.

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. Microsoft has robust distributed services on the distribution axis, 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, lacking architectural equivalency. Overall, Wikibon believes Microsoft has the financial resources and technical capabilities to develop a Cloud Database Platform and upper 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 Exadata Cloud@Customer, delivering architectural equivalency. Oracle Cloud@Customer Cloud Database Platform is moving powerfully down the database convergence dimensions. In addition, the Oracle strategy of making all the cloud database services autonomous is shown by the Oracle Database 21c announcements in early 2021. Oracle Database 21c is initially available on OCI via the Autonomous Database service (in the Always Free Tier) and Oracle Exadata Cloud Service. Oracle Autonomous Database is Oracle's self-managing, self-tuning, and self-securing cloud service built on Oracle's converged database engine.

Oracle has already developed a multi-cloud agreement with Microsoft. Wikibon expects AWS and Oracle to reach an agreement to ensure that Oracle runs well on AWS and can link to Oracle Database hardware.

The announcement of OCI support for GoldenGate has significantly improved the distribution dimension for the Oracle Cloud Database platform. Oracle will need to include data catalogs to help integrate different databases. Oracle will also need to expand its sharding feature further and invest strongly in a Data Mesh architecture and functionality to allow enterprises to devolve and simplify data-led strategies to include the lines of business. In other interesting work, Oracle uses [AI and ML to enrich metadata in an Oracle Cloud Infrastructure Data Catalog](#).

This decentralization 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 simple-to-use 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 the Cloud Database Platform market.

The **SAP HANA Cloud Platform** is difficult to evaluate. SAP started developing SAP HANA when the cloud was still nascent, and other databases did not provide the integration between transactional and analytics. For a time, SAP HANA was the leading platform for ensuring that analytic data could be distributed rapidly to the end-users in the lines of business. However, Wikibon believes that the predicted investments by AWS and Microsoft, and the investments already made by Oracle, reduce any SAP HANA differentiation. The most valuable SAP contribution is its software suites, and Wikibon believes SAP customers would prefer greater flexibility on which cloud and database platform to deploy the SAP software.

SAP is not offering SAP HANA as a general-purpose database, and therefore the potential number of SAP HANA database instances is limited. In software, volume wins. Wikibon believes that SAP will probably

change its current Cloud Database Platform strategy.

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 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, including on-premises deployment, to allow data to remain in place. Also, Snowflake lacks in-database machine-learning algorithms. Although all these requirements are technically challenging, Wikibon considers that Snowflake must take up the challenge or risk being sidelined by Microsoft and Oracle.

Assessment Summary

Wikibon believes that Oracle has the strongest Cloud Database Platform with Autonomous Database and can be integrated with the GoldenGate OCI service. It offers a Tier-1 database foundation, Oracle Exadata Cloud@Customer, which provides identical Exadata X8M hardware and software in on-premises private clouds managed centrally. Oracle also provides Dedicated Region Cloud@Customer, a complete portfolio of public cloud services and Oracle Fusion SaaS applications into an on-premises data center. Wikibon believes that the integration of different databases with the OCI GoldenGate service is a gamechanger. Oracle should shift its marketing to emphasize its ability to provide the best hardware, interconnectivity, and integration for all databases.

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. Microsoft can use the GoldenGate OCI service to improve the integration of multiple heterogeneous databases.

Action Item

Wikibon strongly recommends that senior enterprise executives focus intently on developing 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 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, time-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, which is lengths ahead in the convergence dimension. The ability to use GoldenGate as an OCI service pulls Oracle ahead of Snowflake on the distribution dimension.

Microsoft is a dominant software developer and has a strong Tier-1 SQL Server Cloud Database platform. Its development of Cosmos DB is early but is architecturally sound. Microsoft needs to expand its convergence and distribution capabilities and take advantage of GoldenGate services.

Wikibon recommends that senior executives at large enterprises, especially those with significant Oracle installations, start investing in Oracle Autonomous Database on OCI or on-premises with Cloud@Customer while continuing to track the development of other Cloud Database Platforms. Wikibon also recommends that senior executives invest in GoldenGate as a service to integrate different databases and avoid database conversions. Wikibon believes this can radically improve the cost, functionality, and time-to-value required 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. A Cloud Database Platform is a service delivered from an integrated cloud platform. A Cloud Database Platform enables enterprises to utilize Cloud Database services on demand without an initial investment cost for equipment and licenses. It also allows enterprises to manage distributed databases remotely on private clouds or shared clouds.

A Cloud Database Platform 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