Oracle Exadata X8M Powers Multi-Cloud Strategy

By David Floyer

December 10, 2019

PREMISE AND EXCERPT

This Wikibon research focuses on the Oracle Exadata X8M Multi-Cloud Strategy and examines the following premises:

1. Oracle's new Exadata X8M technology together with the Oracle Autonomous Database and multi-cloud innovations is the most profound update in Oracle's history. These technologies will allow mission-critical systems of record to be connected to an autonomous DbaaS (Database-as-a-Service) on a private on-premises cloud, with the future option of migrating some or all of the databases to an Oracle Cloud Database service.
2. The time for deploying "Do-it-Yourself" (DiY) systems for any Tier-1 database (i.e. IBM DB2, Microsoft SQLServer, and Oracle Database Enterprise Edition) supporting larger-scale mission-critical systems of record is well and truly over.
3. AWS and some consultants are promoting the idea that converting tier-1 databases to a different cloud-native database is "the only" long-term strategy. The premise of this research is that conversion of larger mission-critical databases is a very expensive, protracted and risky strategy, and a sentence to 5-years hard labor.
4. The Exadata X8M and Oracle's DbaaS strategy offer a lower cost and more flexible strategy for retaining cloud options. These options include on-premise private cloud, Cloud-adjacent architectures, Oracle Cloud, and fully integrated multi-cloud, cloud-native options with Microsoft and VMware.

Wikibon advises senior enterprise executives in the strongest possible terms to avoid conversions of complex databases supporting larger-scale mission-critical systems of record. In addition, avoid like the plague any systems integrators and consultants who are suggesting such conversions.

Wikibon also strongly recommends that enterprise IT migrate away the traditional ways of running Oracle Databases, either on-premises or in the cloud, and use Oracle DBaaS offerings instead, starting with Exadata X8M. Enterprise IT should evaluate different ways of running applications using Oracle Database, such as using Microsoft Azure or VMware to run the application and Oracle DBaaS for the database.

EXECUTIVE SUMMARY

Wikibon concludes that the Exadata X8M update and the Oracle DBaaS cloud services based on Exadata are the broadest and strongest improvement in the Oracle Database platform ever announced by the company. We assert this based on the following key points, split into technology and business impacts.

TECHNOLOGY IMPACTS

- The Exadata X8M infrastructure is a combination of scale-out and scale-up technologies, where compute, networking and storage have been optimized to run the Oracle Database.
- Oracle has by far the most advanced deployment of Intel Optane persistent memory, using RoCE in App Direct Mode for critical Oracle Database environments. This enables Oracle Database to use RDMA to read remote persistent memory, bypassing network and IO software, interrupts and context switches. This delivers an end-to-end latency of less than 19 microseconds. Oracle claims that log writes are 8X faster, and specialized algorithms have improved performance for OLTP, analytic workloads, and workload combinations.
- Oracle's objective is to take full responsibility for providing integrated updating of hardware, operating systems and middleware, and meeting enterprise SLAs for compliance and performance.
- In addition, Oracle claims that almost all of the DBA and operational functions will be fully automated, with dramatic reductions in enterprise staffing. An early example is the Automatic Indexing feature, which instead of years of expert DBA indexing can accomplish this task in less than 24 hours.
- There are sound technical and business reasons why larger-scale mission-critical systems of record have not moved to traditional centralized clouds. The Oracle Exadata X8M is architecturally identical with Oracle's multiple database public cloud services today, including Autonomous Database, the Exadata Cloud Service and Gen 2 Exadata Cloud at Customer. This provides enterprise IT with a path to the cloud model of their choice with the same infrastructure, software, APIs and management tools.

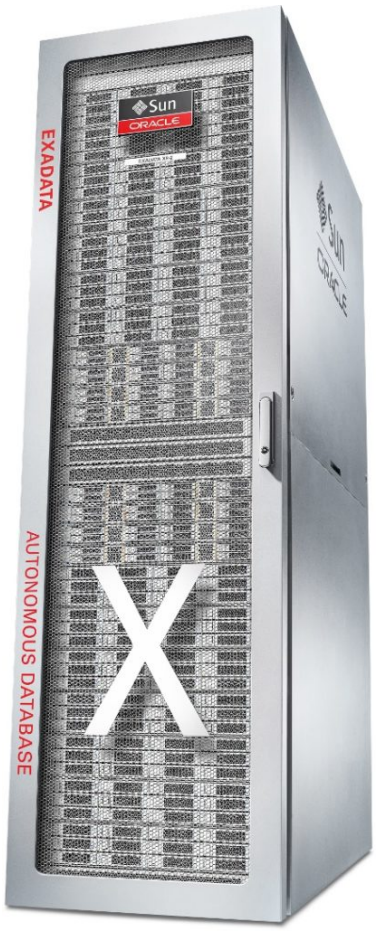


Figure 1: Oracle Exadata X8M is the Foundation for Oracle Autonomous Multi-cloud Database.

Credit: Oracle 2019

- Exadata X8M can be the foundation for the next generation systems of record, which will integrate vast amounts of data to streamline and automate core business processes.

BUSINESS IMPACTS

- This Wikibon financial study compares the 4-year cost of a Full Stack Oracle Database System supporting larger mission-critical applications to a traditional best-of-breed DiY system. The reference system is Exadata X8M running Oracle Database Enterprise Edition with additional options. The study confirms that the very latest traditional best-of-breed DiY systems are currently 48% more expensive than the Exadata X8M . Moreover, the Autonomous functionality on Exadata X8M will improve the cost differential substantially.
- The time for building bespoke systems and processes from best-of-breed components for any Tier-1 database supporting larger-scale mission-critical systems of record is over.
- Converting Tier-1 databases to a different cloud "native" database is being promoted as a long-term strategy. Wikibon strongly reminds CIOs that conversion of larger mission-critical databases is a very expensive, protracted and risky strategy. Wikibon strongly recommends the alternative strategy of moving to the current autonomous Tier-1 DbaaS platform for these workloads.
- Notwithstanding some cautions we outline, the technical evaluations and business benefit analysis support the conclusion that Oracle Exadata is the safest bet for existing and next generation Oracle Database mission critical systems of record.
- This research concludes that larger-scale mission-critical database deployments need highly specialized hardware and software to meet service level objectives. For Oracle Database workloads, Exadata provides that specialized hardware.
- Wikibon also concludes that Oracle's multi-cloud strategy allows the greatest flexibility and retains the broadest set of alternative options for the future. The Exadata X8M is a foundational technology that retains future cloud-native options of integrated Oracle Cloud, and multi-cloud, cloud-native options to integrate the Oracle Autonomous Database with Microsoft Azure and VMware Cloud Foundation.

EXADATA X8M TECHNOLOGY ANALYSIS

The Exadata platform has just had the most extensive technology infusion ever. The traditional x86 processor has become the bottleneck to computer processing, with the slow-down and eventual ending of Moore's Law. This section addresses what technology is needed to enable high performance computing and how Oracle is providing it for larger mission-critical Oracle Database workloads.

High performance systems (HPS) used to be mainly in the realm of universities and government agencies, and require highly specialized skills to design, manage and utilize. HPS technologies have now moved into the

enterprise, from sandals to suits. HPS techniques are at the heart of AI, machine learning, and inference workloads.

The philosophy of next generation ultra-high-performance systems is to offload as much processing as possible to other distributed processors, especially in the system components of storage and networking. These processors can be specialized x86 processors as in the case of the Exadata storage servers. Increasingly, they are Arm processors, as in Mellanox network adapters (See Figure 4 below) and inside SSD drives.

This distributed end-to-end architecture is supported by the full stack of Oracle Linux OS, middleware and Database to provide integrated performance, redundancy, and recoverability.

The details of the advanced technologies that are incorporated in the Exadata X8M, and how they combine to improve database functionality and performance, are included in the Footnotes below. These major technology advances include:

- *New Intel Processors for Exadata X8M*
- *Exadata X8M RoCE Networking*
- *New Intel Processors & Software for Storage Servers*
- *New Persistent Memory/Storage Layer (PMSL)*
- *In-Flash Columnar Cache*
- *Migration to KVM*
- *ZDLRA X8M*

Wikibon believes that the Exadata X8M and the ZDLRA X8M will become available as full Oracle Cloud offerings in the near future, in line with Oracle's overall cloud strategy. Autonomous backup and recovery as part of an autonomous database is no longer an impossible dream.

EXADATA PLATFORM BUSINESS ANALYSIS

NEXT GENERATION MODERN WORKLOADS

Figure 2 below shows different workload types (Traditional, Virtualized, and Next Generation) placed on two dimensions, latency (y-axis) and complexity (x-axis). The traditional workloads in Figure 2 have a white background. Virtualized workloads, such as those running on VMware, cover a large space in the middle, shown with a grey background. The next generation application types are shown with a yellow background.

The most challenging workloads require low-latency and are very complex. They are positioned in the top-right hand corner of Figure 2. Low latency is very important to traditional real-time database workloads, such as

systems of record (e.g., order entry, supply chain management and optimization), and mission-critical analytics against data warehouses. The speed of functions such as locking/unlocking parts of the database and log-writes are critical to overall database performance and throughput.

The availability of ultra-low latencies for IO in Exadata is a key enabler for these new workload types, with much higher complexity and/or performance requirements. Examples of modern applications that can run well on Exadata X8M include arbitrage, high frequency trading, AI inference applications, real-time fraud detection, and AI machine learning.

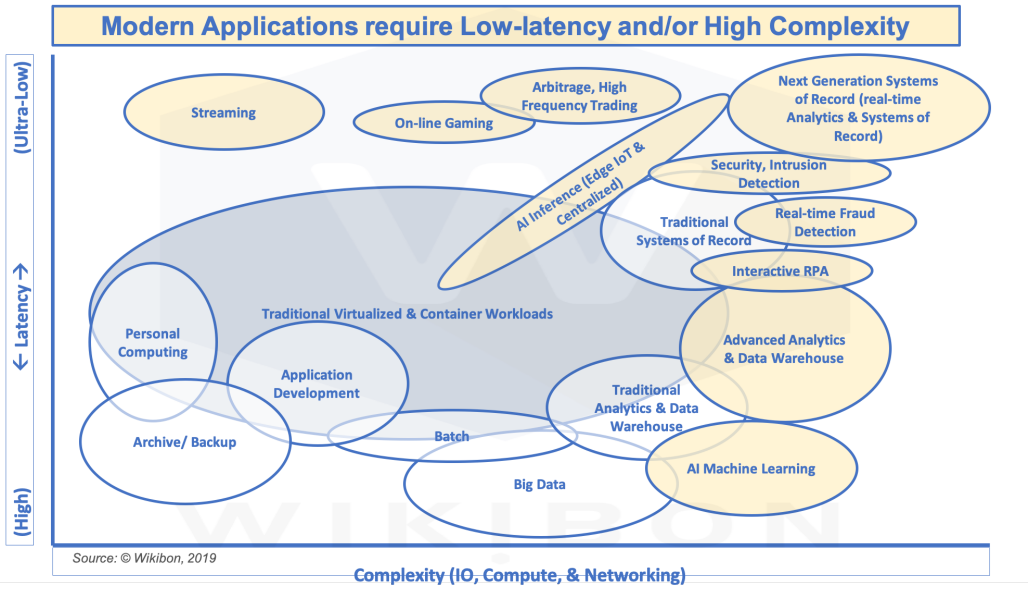


Figure 2: Emerging Ultra-low Latency & High Complexity Workloads

Source: Wikibon © 2019. Larger-scale enterprise workloads plotted against a "y" axis of latency and a "x" axis of application complexity (taking into account complexities of IO, compute, and Networking) Traditional applications are shown with a white background, virtualized applications with a grey background, and modern applications with a yellow background.

Of particular interest and value to the lines of business are next generation systems of record. The vast majority of traditional larger mission-critical systems of record for enterprises have remained in on-premises private clouds, because of performance requirements, compliance concerns, and the high cost and business risks of conversion to a public cloud.

An example of a next generation system of record is an upgraded traditional system enhanced with real-time fraud detection driven by AI inference code. To give the same SLA as the original system of record, the original system must operate much faster, and leave room for the inference fraud-detection system to complete within (say) a 200-millisecond envelope.

The Exadata X8M technologies enable existing systems of record to be upgraded to next generation applications by allowing the parallel execution of real-time analytics or AI inference applications to be driven by the system. As a result, much greater automation of business processes can be supported by the database system. Alternately, many more divisions of a large organization can be supported on a single system of record, instead of having to

operate separate systems for each division. For most applications, the Exadata X8M is the fastest available Oracle Database infrastructure platform.

THE ECONOMICS OF EXADATA X8M COMPARED TO A "DO-IT-YOURSELF" INFRASTRUCTURE STACK

Wikibon researched the financial case for Gen 2 Exadata Cloud at Customer in a recent research report titled "Oracle Ups its Game with Gen 2 Exadata Cloud at Customer" where we also looked at using the Exadata in an on-premises Cloud at Customer mode. In this research report the focus is on Exadata X8M in stand-alone mode. This could be as a replacement or expansion of an existing Exadata installation, or as a replacement for traditional do-it-yourself (DiY) best-of-breed data center technologies.

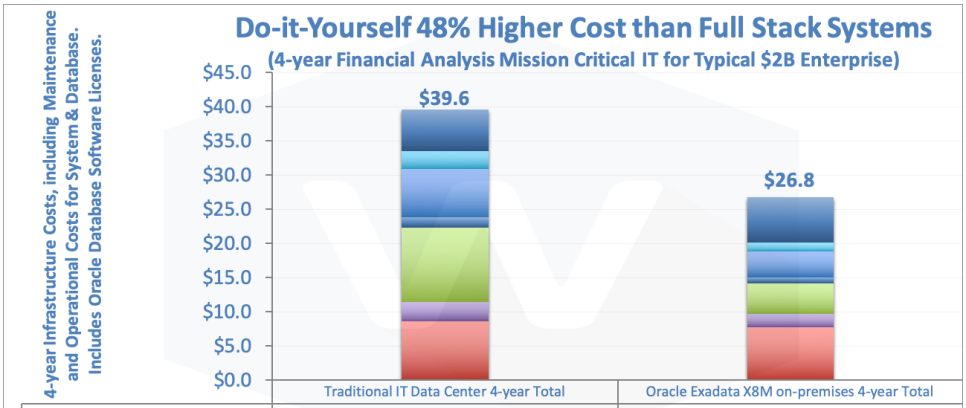
The financial data in this research report uses the same assumptions as in the earlier 2019 research. The previous research takes both an expanded view of the IT budget (including development) and the business revenue impact on the enterprise users of the IT applications. For this report, however, the focus is narrower, specifically, "What are the changes to the IT operational Oracle Database costs using Exadata X8M on-premises, compared with the traditional "DiY best-of-breed" approach?"

The following are the differences between the line items in the two complementary but separate analyses:

- The IT Infrastructure line item has been separated out into the following components
  - Server Costs
  - Storage Costs
  - Networking Costs
  - System Software Costs
- costs for power and space have been extracted from other costs and included as a separate line item.
- remainder of the other costs are not included, as they do not impact the database costs significantly.
- Oracle Database costs are included.
- same workload, IT and business assumptions (e.g., $2B in revenue) are used as in the earlier 2019 earch.
- operational costs now include a small portion of the developer costs. These include developers who st in operational work as DBAs and application specialists.
- remaining developer costs are now not included, as it is not part of the business questions being asked.
- The remaining "Other" costs are not included - for the most part, all relevant database costs are already included.

The results of the analysis are shown in Figure 3 which compares:

- The 4-year TCO of a traditional "DiY" IT datacenter for Oracle-based mission critical systems of record applications in a typical enterprise with $2B in revenue.
- Deploying an Exadata X8M for the same workloads in the same enterprise at the same location.



	Traditional IT Data Center 4-year Total	Oracle Exadata X8M on-premises 4-year Total
Servers	$6.1	$6.6
Infrastructure Software	$2.6	$1.4
Tier 1 & Tier-2 Storage	$7.0	$3.8
Network	$1.6	$0.8
Operational Support (including DBA)	$10.8	$4.4
Environmentals (Power/Space)	$2.8	$2.0
Oracle Database	$8.6	$7.8
Total Database Budget	$39.6	$26.8

Source: © Wikibon 2019 Note: Full Stack Systems assumes Oracle Exadata X8M

Figure 3: Comparison of the 4-year costs of traditional IT DiY Infrastructure running larger-scale Mission-critical Oracle Database workloads vs. Oracle Exadata X8M on premises over 4 years.

Sources: © Wikibon Analysis 2019, Table 2 Oracle Ups its Game with Gen 2 Exadata Cloud at Customer, Wikibon 100+ in-depth-interviews with Exadata and other hyper-converged environments.

The main comparative analysis shown in Figure 3 show that the Do-it-Yourself traditional IT datacenter approach costs 48% more than the Exadata X8M-based integrated full stack solution designed for Oracle database specifically. The line items in Figure 3 are as follows.

1. Exadata X8M servers are 8% more expensive than currently available DiY traditional servers.
2. Exadata infrastructure software is 47% lower cost than infrastructure software to support functions such as replication, backup, recovery, utilities, et al.
3. Exadata storage is 46% lower cost than the highest performance DiY traditional storage arrays, because of built-in storage enhancement features in Exadata X8M storage.
4. Exadata built-in network hardware & software are higher performing with higher functionality and 49% lower cost than traditional high-performance networks.
5. The operational support and DBA costs for Exadata X8M are 59% lower than the operational support required by the DiY traditional datacenter solution.
6. Exadata power, cooling & space are 30% lower cost than the larger number of x86 controllers and equipment within the DiY traditional datacenter solution.

7. The Oracle Database license and maintenance costs are about 10% lower cost than the DiY solution. Wikibon assumed that additional Enterprise Edition In-Memory Database features are required for the new In-Flash Columnar Cache X8M features. Most of the savings will come from Oracle Database maintenance, and redeploying licenses for new projects.

So, what is the answer to the question "What are the changes to the IT operational Oracle Database costs from using Exadata X8M on-premises, compared with the traditional "DiY best-of-breed" approach?" In summary, it is likely that deploying your own DiY system will be significantly more costly (Wikibon concludes 48% more costly) than deploying an integrated Exadata X8M solution. Wikibon believes this percentage will increase rapidly over the next few years with the further development and deployment models of the Oracle Autonomous Database.

The simple conclusion is that the time for DiY infrastructure is well and truly over. The benefits of Exadata either on-premises, in a cloud datacenter, or in an Oracle Cloud are just too great. An Exadata-based strategy in general allows the best choice of options for larger enterprises to reduce running costs, take advantage of cloud economics and create next generation hybrid applications. Oracle is already successfully using Exadata to run its own SaaS and PaaS portfolio.

Oracle is the most important database for the majority of larger mission-critical systems of record. Oracle now offers a clear lower-cost path forward to a fully autonomous DBaaS.

Wikibon recommends that Enterprises with larger-scale mission-critical Oracle Databases running bespoke DiY hardware should evaluate Exadata X8M as a stepping stone on a journey to a fully autonomous DBaaS environment. Wikibon also recommends Enterprises with installed Exadatas to migrate over time to Exadata X8M and future Exadata releases as the start of the journey to an autonomous DBaaS environment on-premises.

ALTERNATIVE STRATEGIES

AWS CLOUD-NATIVE

As mentioned in the premise, AWS proposes an alternative strategy of converting an Oracle Database to a "cloud-native" AWS database with everything running in the AWS cloud. The AWS transactional database is Aurora (based on open-source MySQL), and the data warehouse database is Redshift (also open-source). The AWS "proof-point" is that Amazon retail has converted everything from Oracle. This strategy has a sales line that suggests that Aurora/Redshift is the only way digital transformation will happen, and moving to an Oracle Cloud is a transition.

AWS is and will continue to be a very successful cloud provider, and has made enormous strides in IaaS functionality. AWS offers a wide selection (eleven) of open source databases and can demonstrate outstanding value for many workload types. However, AWS databases are not suitable for all workloads. AWS does not have a Tier-1 relational database suitable for workloads such as complex larger-scale mission-critical systems of record.

Wikibon believes the AWS Tier-1 database conversion strategy is flawed for six main reasons:

1. The AWS PaaS layer and the AWS databases are not offered as an integrated product. AWS does show ways that the PaaS layer can be used to address some of the recovery functions of Oracle. However, enterprise IT has the responsibility for developing, testing and maintaining the integration of AWS PaaS and AWS databases. AWS does not offer any support for function or performance after conversion. AWS does not offer a database offering anywhere near the integrated functionality of Oracle, IBM, or Microsoft Tier-1 databases, or the development and production support services they all offer.
2. Complex database conversion is extremely risky, expensive, and takes a very long time to complete. Amazon took over five years to migrate the data warehouse and transactional applications from Oracle to Aurora and Redshift, at a reported cost of over 1,000 engineer-years of effort. Wikibon has analyzed the AWS references on the Amazon site, and there are no references of conversions of larger-scale mission-critical systems of record to AWS databases (or even to Oracle running on AWS).
3. Conversion means freezing code and functionality of existing systems, which would seriously delay the introduction of application & process functionality to the systems of record vital for digital transformation initiatives.
4. The Oracle Database development, operational tools, support, and ecosystems are more highly developed for large-scale mission-critical application development and operations. Conversion from Oracle to AWS Aurora leads to higher IT costs for development and operations.
5. Wikibon strongly believes that complex large-scale databases require their own highly integrated hardware and software stacks to achieve functionality, automation and autonomy. Next-generation systems of record will need to exploit these capabilities. Wikibon believes Oracle has a path to achieve this disruptive outcome without disruptive conversion.
6. Our wide experience in conversion leads Wikibon to believe the expected outcome for adoption of this AWS strategy is five-years hard labor with little or no return. Conversion would slow down digital transformation.

The Oracle strategy is to develop an automated and autonomous Tier-1 DbaaS, based on Exadata X8M functionality. Oracle has announced plans to deeply integrate this DbaaS with its own cloud-native services and other cloud-native platforms and applications. Wikibon expects most of this to be completed within about 2 to 3 years.

The bottom line is that any transformation plan has to ensure that the core systems of today continue to work. Tier-1 databases are and will still be necessary. No platform is perfect or complete. One of the major benefits of a Multi-Cloud strategy is the ability to make things work when a piece is missing.

DELL HYPER-CONVERGED POWERONE

Dell PowerONE is an automated infrastructure solution for VMware with Dell PowerEdge servers, PowerMax and PowerProtect storage, and PowerSwitch and SmartFabric networking. The most important aspect is that the complete stack in integrated, is maintained by Dell as a single component, with automation leading to what Dell claims is a 95% reduction in support effort. This solution, together with other similar solutions such as VxRack, are great infrastructure solutions for general-purpose VMware workloads.

However, this approach does not allow the much greater benefits of Oracle Database autonomous functionality, and most of the critical database performance and low-latency features on the Exadata X8M listed in the technical sections in Footnotes. In addition, this approach makes it much harder to take advantage of the strategic autonomous cloud-native offerings from Oracle and Oracle partners.

IBM POWER

IBM Power9 processor cores are more powerful than Intel processors, both in cycle time, and the very fast IO speed. Power9 supports Gen 4 PCIe, NVLink v.2, CAPI 2.0, and OpenCAPI.

However, the business case outlined in a previous section shows much higher operational support costs for Power9. In addition, a major disadvantage of the Power9 systems is that the Core Processor Licensing Factor is 1 for Power9 processors, and only a half (0.5) for Intel Processors. This means that the Oracle licenses on Power servers are twice the cost of Oracle licenses on Intel servers. The net of this is that Power9 is rarely the lower cost option. Power9 is not part of a converged solution. Again, this approach makes it much harder to take advantage of the longer-term strategic autonomous cloud-native offerings from Oracle and Oracle partners.

MICROSOFT AZURE-ORACLE & VMWARE-ORACLE MULTI-CLOUD PARTNERSHIPS

The win-win agreements between Microsoft and Oracle, and Oracle and VMware are of profound importance, and increase future options for enterprises. They offer a way of offloading the Oracle Database component running on Microsoft Azure or VMware to Oracle DBaaS running in the same Equinix location. These tightly-coupled integrated multi-cloud solutions help to reduce the cost and increase the functionality of Oracle workloads running on Microsoft Azure or VMware cloud platforms. For example, an enterprise with broad use of Microsoft infrastructure software (e.g., Active Directory, Hyper-V) can easily migrate the application layers to Microsoft Azure, and offload the Oracle Database to the Oracle Cloud.

Wikibon, along with enterprises that use other Oracle and AWS products and services would applaud the announcement of a similar multi-cloud integrated platform initiative from AWS IaaS/PaaS and Oracle Autonomous Database. This is what AWS and Oracle customers really want.

ONE LAST ISSUE: ORACLE AS A SOLE SUPPLIER

Wikibon analysts are very rarely as enthusiastic about a set of technologies as this research on Exadata X8M has been. Oracle is the best, most functional database system on the market for large-scale mission-critical systems of record. Wikibon believes that running Oracle DbaaS on Exadata X8M has the potential to become the platform of choice for the next generation of large-scale cloud-native mission-critical applications.

However, there is one area where enterprise business executives should ask Oracle enterprise executives a tough question: is it safe to accept Oracle as a sole supplier? This is a key issue which needs to be clearly addressed by Oracle. Is Oracle serious in aiming to become a true volume cloud provider, and provide true on-demand cloud services? Oracle currently has a $8B global cloud business, largely driven by its database and SaaS applications, which suggests it is serious. Enterprise executives should seek a compelling answer to this question.

Wikibon believes that if Oracle can convert the most functional database into the most popular DBaaS at a competitive price, they have a bright future. If Oracle takes a cash-cow strategy and optimizes on profits, other DBaaS cloud vendors will overtake Oracle in about five years, and the opportunity will disappear.

CONCLUSIONS

Wikibon believes the Oracle DBaaS based on the Exadata X8M will become the default system for larger-scale Oracle mission-critical workloads. The integrated technology is more advanced than any other standalone or cloud alternative running Oracle. The Exadata X8M and Oracle Autonomous Database software will become more functional and tightly integrated over time.

The Oracle strategy is to make identical software and hardware available for on-premises Exadata, Exadata Cloud at Customer, and Exadata Cloud Service and Autonomous Database. The Oracle multi-cloud strategy is to deeply integrate the Oracle Cloud with other cloud-native platforms such as Microsoft Azure and VMware. This range of solutions offer better database performance and price performance than any other integrated hardware solution from any vendor, and better than DiY from any cloud provider.

However, the most important reason for adoption is not price or performance alone. It is because of the potential innovative applications that this technology enables. The fastest time to value is if the developers of the next generation of enterprise applications can build on the existing systems of record. This can be achieved by tightly integrating systems of record with real-time analytic systems, and increasingly using advanced analytics and artificial intelligence in both. The amount and range of data accessed will increase dramatically, as the functionality and latency of hybrid and multi-cloud improves.

These systems will offer enterprises significantly greater business orchestration and automation of their current business models. Even more importantly, they will enable the development of completely new and innovative business models for both new and existing enterprises. Wikibon believes that both approaches have the capability to dramatically reduce enterprise costs and significantly increase competitiveness, and achieve it faster than any conversion strategy.

The next generation of integrated database systems also offer a unique capability for new SaaS systems to innovate in advanced solutions for specific verticals and cross-industry markets. Oracle may be able to demonstrate the methodologies with its own SaaS offerings, but history is not kind to the feet of cobbler's children. Therefore, it is a business imperative that both the integrated transactional and analytic components are available at a price that will incent ISVs to invest in the Oracle Autonomous Database technology vision.

Oracle has created solid partnerships with Microsoft and VMware, with more to follow. These allow great flexibility in running both Oracle and Microsoft applications, the two biggest providers of enterprise software packages. Wikibon believes the DBaaS model will be very popular with larger enterprise IT. Wikibon believes that AWS and Oracle will eventually listen to what their customers want, and agree to provide a similar service.

Lastly, multi-cloud coupling of Oracle (and other Tier-1) DBaaS with other platforms will also increase the pressure to provide low and ultra-low latency connections between these clouds. Equinix offers 2 ms connections between adjacent hybrid clouds, but will need to offer much lower latency solutions in the future.

Lastly, the financial analysis in this research shows that Exadata X8M together with Oracle Database is a much lower cost (48%) than any DiY alternative, and the difference in cost will grow larger as additional autonomous functionality is added. In addition, the business case for avoiding complex database conversion is overwhelming. For very large enterprises it could save 5-years hard labor, and billions in engineering time. And it could save the enterprise.

ACTION ITEM

Wikibon advises senior enterprise executives in the strongest possible terms to avoid conversions of complex databases supporting larger-scale mission-critical systems of record. In addition, avoid like the plague any systems integrators and consultants who are suggesting such conversions.

Wikibon also strongly recommends that enterprise IT migrate away the traditional ways of running Oracle Databases, either on-premises or in the cloud, and use Oracle DBaaS offerings instead, starting with Exadata X8M. Enterprise IT should evaluate different ways of running applications using Oracle Database, such as using Microsoft Azure or VMware to run the application and Oracle DBaaS on Exadata X8M for the database.

Wikibon believes that other IaaS and SaaS cloud providers will enable Oracle DBaaS in the same way as Microsoft Azure.

Wikibon strongly recommends Exadata X8M as a platform for next generation enterprise applications that tightly integrate systems of record with real-time analytic systems, and increasingly use advanced analytics and artificial intelligence in both. These systems will offer enterprises significantly greater business orchestration and business automation of current and future business models.

Wikibon advises CXOs to include Oracle Exadata X8M in RFPs and evaluate Oracle as a strategic partner for digital transformation of Oracle-based systems of record and advanced data analytics.

FOOTNOTES: EXADATA X8M TECHNOLOGY ANALYSIS

NEW TECHNOLOGY

The Exadata platform has just had the most extensive technology infusion ever. The traditional x86 processor has become the bottleneck to computer processing, with the slow-down and eventual ending of Moore's Law. This section addresses what technology is needed to enable high performance computing and how Oracle is providing it for larger mission-critical Oracle Database workloads.

High performance systems (HPS) used to be mainly in the realm of universities and government agencies, and require highly specialized skills to design, manage and utilize. HPS technologies have now moved into the enterprise, from sandals to suits.

The philosophy of next generation ultra-high-performance systems is to offload as much processing as possible to other distributed processors, especially in the system components of storage and networking. These processors can be specialized x86 processors as in the case of the Exadata storage servers. Increasingly they are Arm processors, as in Mellanox network adapters (See Figure 4 below) and inside SSD drives.

This distributed end-to-end architecture is supported by the full stack of Oracle Linux OS, Middleware and Database to provide integrated performance, redundancy, and recoverability.

The next sections review the advanced technologies that are incorporated in the Exadata X8M, and how they combine to improve database functionality and performance. The major headings for these discussions are:

- New Intel Processors for Exadata X8M
- Exadata X8M RoCE Networking
- New Intel Processors & Software for Storage Servers
- New Persistent Memory/Storage Layer (PMSL)
- In-Flash Columnar Cache
- Migration to KVM
- ZDLRA X8M

NEW INTEL PROCESSORS FOR EXADATA X8M

The Exadata X8M uses the same latest 24 core Intel Cascade Lake found in the Exadata X8, with a 15% faster clock compared to X7. The Spectre & Meltdown security problems that were managed by software in the X7 are now resolved in silicon. These Intel x86 chips are now generally available to other server vendors.

The performance of processors has stalled at no more than 4 Ghz, and there is a limit to the number of processor cores that can be used efficiently. This means that the central CPU is a choke point for computing. Systems

architects are finding ways to process more information in parallel, and to push processing tasks to specialized processors. Examples of these are heterogeneous architectures in which multiple processor types, GPUs, neural networks, controllers, and ASICs address new challenges and opportunities.

EXADATA X8M RoCE NETWORKING

The Importance of Low-latency IO

Low-latency reads of non-cached data are critical to minimizing database and application-level latency in transactional systems. Latency for log writes is another critical area. This section dives into how Exadata X8M's combination of RoCE networking and persistent memory working in application-direct mode is designed to minimize these latencies. Even more important, ISVs and enterprise developers are not required to make any application modifications.

Introducing RoCE

The Exadata X8M has made radical improvements in system networking with RoCE (RDMA over Converged Ethernet). Converged Ethernet has overtaken InfiniBand and Fibre Channel in investment and speed to market. The same Remote Direct Memory Access (RDMA) protocol that was previously used over the Exadata 40Gb InfiniBand link is now supported over a much faster 100Gb converged Ethernet network. This will allow further expansion to 200Gb and 400Gb bandwidths in the not too distant future.

Earlier we discussed the importance of multiple distributed heterogeneous processors that manage different parts of the total system. Mellanox is the leading supplier of converged Ethernet. Mellanox also deploys ASICs and Arm processors on its network adapters, and at the same time has introduced the protocols, APIs and software support to complete the solutions. Oracle has been able to take the Mellanox RoCE technology and apply it to improve IO latency.

The Mellanox ConnectX-5 EN 100Gbit type Network Adapter in Figure 4 directly reads and writes to memory with no extra copying or buffering. It is dual-ported, and the latency is ultra-low (about 760ns).

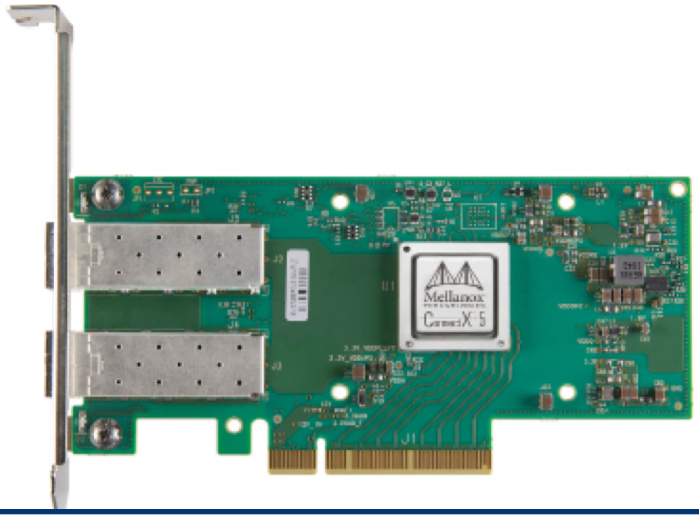


Figure 4: Mellanox ConnectX-5 EN Network Adaptor

Source: Mellanox ConnectX-5 EN 100Gbit type Network Adapter downloaded September 2019.

RoCE and Storage Networks

Initially, flash drives used the same slow and chatty SCSI Protocols as HDDs, to minimize adoption friction. Now these protocols are a major inhibitor to flash performance. NVMe is a new lower-overhead protocol for SSD IO. NVMe drive volume uptake has been dramatic,

and Wikibon expects PCIe SSDs will be 50% of the SSD market by the end of 2019. Exadata was a very early adopter of NVMe drives.

NVMe over Fabrics (NVM-oF) is also an important development for storage in general. The fabrics can be Fibre Channel, InfiniBand, iWarp, RoCE, and TCP. The previous generations of Exadata used RDMA over InfiniBand as the fabric. As we discussed earlier, converged Ethernet has emerged as the "winner" in high-bandwidth/low-latency networking. NVMe over RoCE and TCP will be important options for general purpose storage. However, there is a very big difference in latency between NVMe over RoCE using the NVMe IO protocol, and RDMA (Remote Direct Memory Access) over RoCE issuing reads and writes to the Intel Optane Persistent Memory/Storage. RDMA, as its name implies, can issue commands which access memory directly. The protocol latency difference between the two methods is about a factor of 10. Only the Exadata X8M supports the advanced RDMA over Converged Ethernet.

Introducing RDMA

RDMA has been and will continue to be an integral part of the Exadata high-performance architecture. Oracle has used its low-latency and high-bandwidth capabilities to radically improve clustered database performance. Some of the recent RDMA improvements on the X8M for Oracle Database include the following:

- Large data transfers benefit from very high throughput with minimal involvement from the CPU, improving elapsed time and consistency.
- Inter-node OLTP clustering requires the lowest possible latency to enable effective scaling. A unique Direct-to-Wire Protocol means that inter-node OLTP cluster messaging is three times (3X) faster than before.
- Read and log-write latencies are absolutely key to complex data workloads and high-performance multi-node scaling. RDMA addresses these problems by means of:
    - A Unique RDMA protocol to coordinate transactions between nodes.
    - An Ultra-low-latency of read IOs to the new persistent memory/storage layer in the storage servers (see "Exadata X8M Read Performance with PMSL NVDIMM" section below).
    - A unique "Smart Fusion Block Transfer" (SFBT) that eliminates log-writes on inter-node block moves.
    - Ultra-low-latency writes of database logs to an ultra-fast persistent memory/storage layer in the storage servers (see "Persistent Memory/Storage Layer Commit Accelerator" section below).

RoCE for Network Prioritization

There are many latency-sensitive database algorithms, especially in the areas of locking, cluster heartbeats, transaction commits, cache fusion, and many others. The RoCE Class of Service (CoS) allows network prioritization for these algorithms. CoS allows packets to be sent with multiple classes of service, each with separate and

messages requiring low latency are not slowed by high throughput messages from workloads such as backup, reporting, and batch. The blue in Figure 5 represents the longer queue for lower priority messages.

The Exadata X8M uniquely chooses the most optimal class of service for each Oracle Database message. Again, this is an example of the X8M pushing distributed processing down to the place where it is best suited to do the job, and providing the architecture and software layers to make this part of a coherent whole.

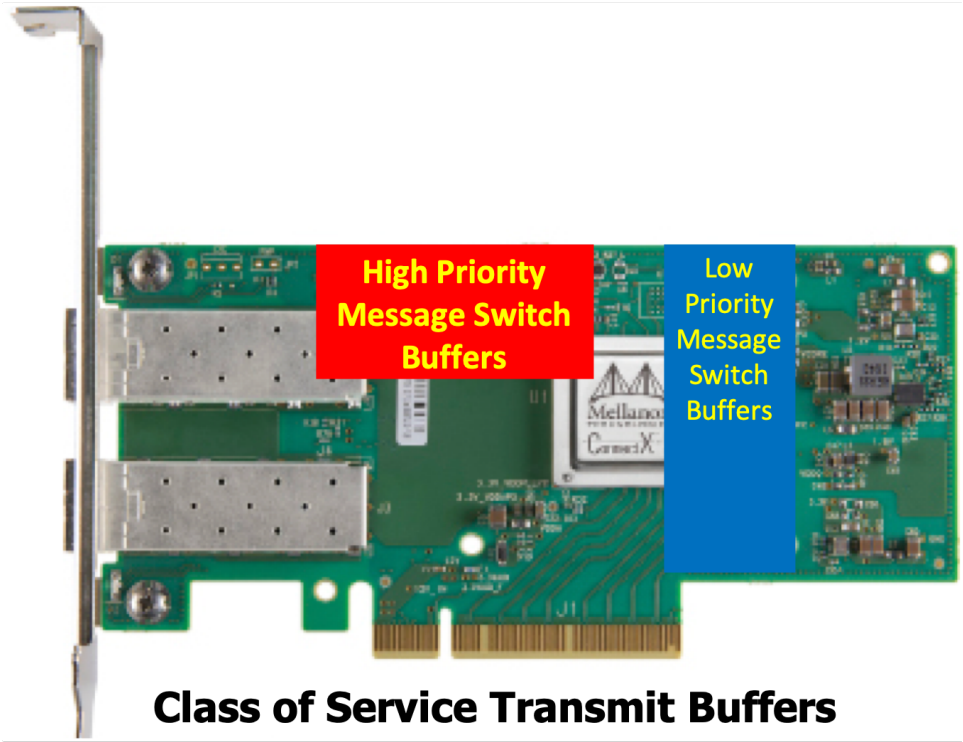


Figure 5: Class of Service Transmit Buffers on Mellanox ConnectX-5

Source: Mellonex Card from Figure 4 & Class of Service .Implementation Wikibon 2019

Figure 5

There are many latency-sensitive database algorithms, especially in the areas of locking, cluster heartbeats, transaction commits, cache fusion, and many others. The RoCE Class of Service (CoS) allows network prioritization for these algorithms. CoS allows packets to be sent with multiple classes of service, each with separate and independent network buffers. Figure 5 shows the CoS transmit buffers in the Mellanox network adapters. The red represents the shorter high-priority queue for the high priority message switch buffers. This ensures that messages requiring low latency are not slowed by high throughput messages from workloads such as backup, reporting, and batch. The blue in Figure 5 represents the longer queue for lower priority messages.
The Exadata X8M uniquely chooses the most optimal class of service for each Oracle Database message. Again, this is an example of the X8M pushing distributed processing down to the place where it is best suited to do the

RoCE for Network Prioritization

Ethernet has traditionally dealt with network congestion by silently dropping packets, and expecting the sender to detect lost packets and retransmit them. A less common but equally important source of packet losses are switch and link failures.

The results of these packet "drops" on conventional Ethernet is a drastic hit to latency and throughput, and the reason why native Ethernet is rarely if ever deployed in large latency sensitive database workloads.
Exadata X8M's RoCE implementation avoids packet drops by using RoCE Priority-based Flow Control (PFC). For example, if the low-priority buffer in Figure 5 above is full or nearly full, the RoCE adapter tells the sender to pause sending until the buffer is less full.

If the problem is endemic, another mechanism, the RoCE Explicit Congestion Notification (ECN), enables the RoCE switch to mark the packet flow as "too fast." It then instructs the source to slow down packet sends.

RoCE for Instant Failure Detection

All database systems use frequent heartbeat messages between all components and nodes to detect possible failures. Some failures, in particular server failures, normally require a long timeout to avoid evicting servers from a cluster prematurely. In these situations, it is difficult to distinguish quickly between a slow response to a heartbeat because of a high CPU load, and a true server failure.

The X8M uses RDMA to quickly confirm server failure. RDMA communicates using hardware, so that remote ports can respond even if the software is running slowly. The process used is to send groups of four (4) RDMA reads to the suspect server. These are sent across all combinations of source and target ports. If all four RDMAs fail, the server is safely evicted from the cluster with a very low probability of a false positive.

RoCE as an Industry Standard

RoCE is now an open standard, defined by an open consortium. The code and protocols are developed as open-source and maintained in upstream Linux. A key contributor to this consortium is the InfiniBand Trade Association (IBTA). It is actively supported by the major network adapter vendors such as Broadcom, Intel and Mellanox as well as by the major top-of-rack switch vendors, such as Arista, Cisco, Juniper and Mellanox. The X8M uses Mellanox network adapters and Cisco Ethernet switches.

NEW INTEL PROCESSORS & SOFTWARE FOR STORAGE SERVERS

A key offloading feature of the CPU in the Exadata X8M is a scale-out network of storage servers. The value of the new Intel chips and RoCE for the storage network is discussed in the next sections below.

New Intel Chips for Storage Servers

This offloading of additional work from database processors to storage cores has required some improvements to the storage server technology. In the X8M, the improvements are 16 core Intel Cascade Lake CPUs, a very fast Persistent Memory/Storage Layer (discussed later in detail), flash NVMe drives, and (in high capacity Exadata storage servers) high-density 14TB HDD Helium drives.

Increased core count in the storage servers assists both transaction processing and analytic systems. The storage servers in the X8M are offloading an increased amount of processing from the database servers, both for transaction processing and analytic workloads.

NEW PERSISTENT MEMORY/STORAGE LAYER (PMSL)

NVDIMMs in General

Wikibon has discussed in previous research that non-volatile storage added to DRAM will produce a very cost-effective high-performance additional storage layer, which can be addressed as memory. This is generally called a NVDIMM, non-volatile DIMM. There are a number of technologies that have been developed in this space that use different forms of NVDIMMs. They all add additional tiers of capacity, performance, and price between DRAM and slower forms of non-volatile storage. One example is the Kioxia XL-Flash, which uses single-cell flash to achieve a claimed read-time of 5 microseconds.

Oracle has selected Intel Optane DC persitent memory based on 3D CrossPoint technology to add to the DIMM form factor in the storage servers. More importantly, Oracle has developed RDMA-based software solutions to take advantage of this new low-latency storage technology.

Intel Optane can be configured in two ways, either in Memory Mode, or in App Direct Mode. The Memory Mode is essentially a cache, and is not persistent. App Direct allows persistence. The Oracle storage servers only address Optane storage using the App Direct Mode.

In App Direct Mode there are two types of direct memory load/store available. One is DRAM, and the other is the Intel Optane NVDIMM. The operating system in the storage controller and the code running under this OS need to be explicitly aware of the two types of direct load/store memory available, and must decide which data reads or writes are suitable for DRAM and which are suitable for Intel Optane. Operations that require the lowest latency and don't need a permanent data storage can be executed on DRAM, such as database scratch pads. Data that needs to be made persistent or very large data structures can be routed to the persistent Optane, assuming the higher and more variable latency can be tolerated.

The Optane persistent memory NVDIMM has a DIMM form factor. Inside this NVDIMM, there is a front-end DRAM capacitor-protected buffer, which normalizes transfer speed in and out of the slower Optane persistent storage. The capacitor protection allows sufficient time to complete operations to the Optane persistent storage in the case of power failure. This internal architecture is transparent to the RDMA code.

The reads which end up going to PMSL will be significantly slower and have greater variance than reads direct from DRAM. However, the use of the RDMA memory protocol and the faster speed of the Optane storage results in much faster reads within the storage servers.

Persistent writes are more complex and require a much more innovative software design. The recovery mechanisms call special Intel CPU instructions to flush data from the CPU cache to the PMSL. These either complete a sequence of writes to the NVDIMM, or backs them out. The "Persistent Memory/Storage Layer Commit Accelerator" section below shows how this is used to enhance log writes.

Exadata X8M Read Performance with PMSL NVDIMM

The Exadata X8M storage servers have a persistent memory storage layer in front of the flash storage layer. This delivers a practical 2.5 times higher IOPS than the previous model, up to about 16 Million 8K SQL IOPS for each storage server.

Figure 6 shows the Persistent Memory/Storage Layer (PMSL) NVDIMM in the storage servers of the X8M. This technology has a DIMM form factor. Inside the PMSL is a capacitance protected DRAM front-end, in front of a very fast storage layer of Intel Optane technology.

A great advantage of the X8M is that the Oracle Database uses RDMA reads to directly read the PMSL, instead of traditional IOs. This bypasses the network and NVMe IO software, all the interrupts, and all the context switches. The result is an impressive 8 times (8X) better latency for reads. The Oracle measured figure is 19µs (microseconds) or less for an 8K database read. Wikibon believes this is the lowest latency in the industry.

The PMSLs are automatically tiered and shared across databases, and used as a cache for the most active data. Wikibon recommends enterprise IT test this feature on their own workloads to determine actual performance improvements. As should be expected, the persistent memory/storage layer is also mirrored automatically across storage servers for fault-tolerance.

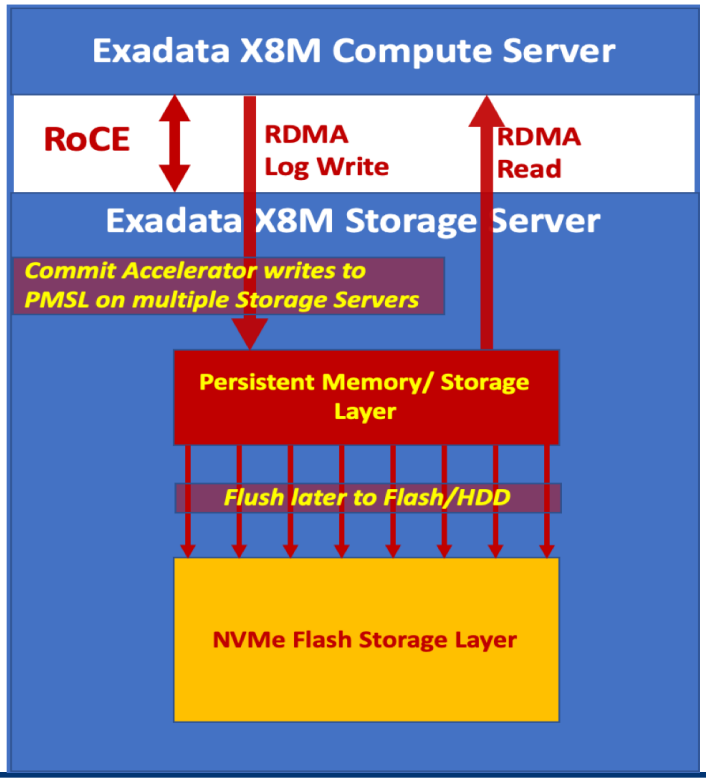


Figure 6: Persistent Memory/Storage Layer in the Exadata X8M storage server

Graphic Source: © Wikibon 2019

Persistent Memory/Storage Layer Commit Accelerator

Earlier in this report, Wikibon pointed out the benefits of reducing log-write latency, which is critical in OLTP performance. Any log-write slowdown can cause a commit backlog, and severely impact performance throughput. Figure 6 above also shows that the Commit Accelerator uses RDMA writes to write logs directly to the PMSL. This can be up to eight times faster than without the PMSL.

Figure 6 also shows the highly parallel interface between the PMSL and the NVMe flash. After log-writes are written, they are batched up by the Commit Accelerator Write-back function and moved to the NVMe flash. The parallel bandwidth between the PMSL and the NVMe flash is more than sufficient to free up PMSL capacity in real-time, and ensure effective use of an expensive resource.

The Commit Accelerator allows Oracle Databases to issue one-way RDMA writes to the PMSL on multiple storage servers. This again bypasses the network and storage software, interrupts, context switches, and more, and improves latency.

The overall benefit is much faster, more consistent and higher overall log-write rates. Wikibon again recommends testing this feature with in-house workloads that would benefit from higher log write rates with improved latency. Faster log-writes usually mean faster commit times, improved locking, and improved database throughput and performance consistency. Wikibon does not know any DBA who will not be delighted with this outcome.

IN-FLASH COLUMNAR CACHE

In-flash columnar scans can be offloaded to the storage server. The creation time for this has been significantly reduced in the Exadata Software version released with X8M, Exadata System Software 19.3. A runtime analyzer finds the best compression algorithm, and the dictionary created during this analysis is reusable. This leads to reducing the time to create an in-flash columnar cache by up to a third.

Sum and group aggregations can be enabled with in-memory columnar format. This reduces the traffic to the database server and improves CPU utilization of the database server. More importantly, it reduces query time by up to 50%.

Smart scans have been extended to wide tables, which can be up to 3 times faster. By using in-memory format for DMLs, smart scans can be up to 5 times faster.

These columnar cache features all require Exadata X8M hardware to function. These features require the Enterprise in-memory optional feature within Oracle Enterprise Database19.3.

ZDLRA X8M

Wikibon has written in depth on ZDLRA (Halving Downtime Costs for Mission Critical Apps) and believes that it is the most complete Oracle Database backup and recovery platform available. The ZDLRA is fed directly from the database memory (not from the storage layer), and provides end-to-end validation of database consistency without having to recover the database.

The ZDLRA is based on an efficient incremental-forever architecture, which allows fine levels of recovery. The new ZDLRA X8M uses the same 100 Gb RoCE technology internally, with the benefits of high bandwidth.

Wikibon believes that the Exadata X8M and the ZDLRA X8M will become available as full Oracle Cloud offerings in the near future, in line with Oracle's overall cloud strategy. Autonomous backup and recovery as part of an autonomous database is no longer an impossible dream.

MIGRATION TO KVM

Exadata X8M now uses Linux Kernel Virtual Machine (KVM) which performs better than the Xen-based virtualization used in InfiniBand versions of Exadata. KVM is only supported with X8M. RoCE and persistent storage are only supported on the X8M in either KVM or bare-metal deployments.

The advantages of KVM include much better performance. Oracle claims KVM supports:

- 2X more guest VM Memory, up to 1.5 TB/server
- Faster client network latency with RoCE and RDMA
- Significantly more guest VMs per server (about 50% more)

Virtual Machines are a critical capability for providing isolation within any database consolidation environment. KVM enables full use of the configurable memory of Exadata, and this allows 50% more VM guests per database server. KVM also provides improved performance. These are all valuable enhancements for database consolidation on Exadata.

WIKIBON TEAM

David Floyer

Chief Technology Officer

@dfloyer

david.floyer@wikibon.org

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.