Oracle Compute Cloud@Customer Overtakes the Market

Puts Other On-Premises Public Cloud Compute Services on Notice

by, Marc Staimer

October 2023

Research Premise

Public clouds have become ubiquitous and are part of everyday life. But the growth of Amazon Web Services (AWS), Microsoft Azure (Azure), and Google Cloud Platform (GCP) has significantly slowed. Their public cloud growth plans assumed that most or even all new application development would go to their public clouds. They were convinced new application development gravity would then draw many current applications, especially mission-critical ones.

That initially panned out to a lesser degree than expected. In the immortal words of former world heavy weight boxing champ Mike Tyson: “Everyone has a plan until they’re punched in the mouth.” They were punched in the mouth by the market. Although many applications have indeed been developed in their clouds, it turned out the gravity is considerably weaker than was assumed.

They discovered there is a tremendous amount of inertia or friction within many IT organizations when attempting to develop or move their applications off-premises to a public cloud. This has been especially true for mission-critical applications or applications with sensitive data. Laws, regulations, and corporate governance frequently necessitates data residency for some or all data so it never leaves the premises.

There are also latency-sensitive applications that require data and application processing to be co-located in order to deliver the required application response times.

As the lack of expected gravity has slowed growth. AWS, Azure, and GCP grudgingly modified their plans.

The public cloud market buzz morphed into "hybrid" cloud. Hybrid cloud definitions vary by supplier. The common AWS, Azure, and GCP hybrid cloud definition is the migration of some of their cloud services to the customer's premises while running other services on their public clouds. This is strikingly different than the server and storage vendor promotion of moving VMs, containers, and data from the customer's premises to a public cloud.

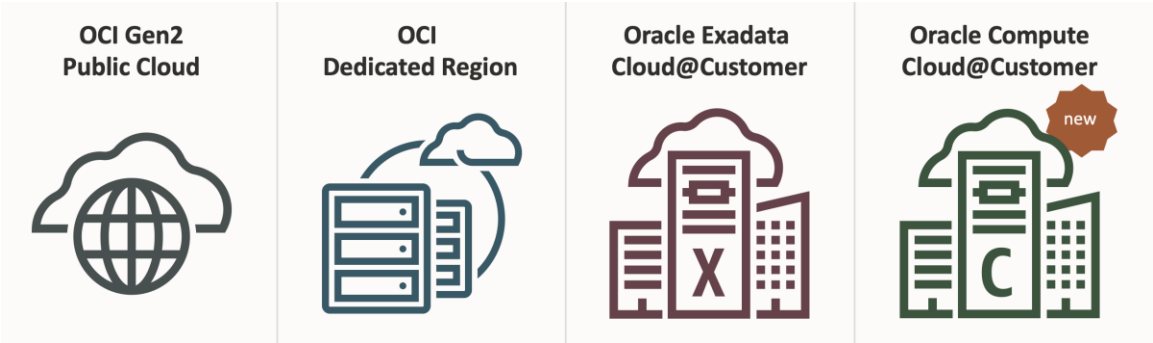
The AWS, Azure, and GCP customer premises programs have many similarities. They're all designed to provide application and data residency on the customer's premises. They provide many of the same tools and APIs of their public cloud that enable software development on the customer premises that can run in their public clouds. They each can deliver cloud services in the customer's data center and at their computing edge.

But the similarities also include common problems:

- Deficient performance
- Restricted scalability
- Inadequate Enterprise database services
- Insufficient elasticity – the ability to scale up when needed and scale down when no longer needed non-disruptively
- Unsatisfactory security across the hybrid or on-premises clouds
- Exclusion of too many essential database cloud services
- Separate vendors for the software and hardware
- Excessive cost

Oracle's Answer to the Cloud On-Premises Market

Oracle just announced their Oracle Cloud Infrastructure (OCI) Compute Cloud@Customer platform. Compute Cloud@Customer is the latest Oracle distributed cloud offering and provides middle compute and edge compute tiers in OCI Cloud@Customer services. It complements the existing Oracle Exadata Cloud@Customer X10M and OCI Dedicated Region offerings by providing OCI compute and storage services in distributed locations. A look at all three shows how.



OCI Dedicated Region

OCI Dedicated Region delivers the complete full range of Oracle Cloud services on-premises in customer data centers. It's a fully managed cloud region with all OCI services and Oracle SaaS applications. As such, Dedicated Region scales from 12 racks to more than 450 racks. The cloud services and low consumption pricing are identical to OCI commercial regions. Dedicated Region helps customers meet regulatory requirements for data residency and privacy and meet customer needs for low latency connectivity to existing resources. It is an ideal way to consolidate legacy applications on high performance managed cloud infrastructure having been adopted globally by major industry leaders such as NRI, Vodafone, and OICT among others.

© 2023 Wikibon

Oracle Exadata Cloud@Customer X10M

Exadata X10M delivers incomparable high performance and availability for all Oracle Database workloads at rack-scale in customer data centers. This helps users of Oracle Database cloud services meet data residency requirements as well as provide lower latency database access than if on-premises applications were accessing database services running in public cloud data centers. It's securely managed through the OCI control plane. No OCI employee ever has access to the data in the Exadata Cloud@Customer X10M. It runs the Oracle Exadata Database Service, the fully managed Oracle Autonomous Database, and can run both concurrently using multiple VM clusters. Exadata Cloud@Customer X10M scales from 380 4th generation AMD EPYC cores in database servers and 240 TB of usable storage to more than 6,000 cores and 5 PB of storage. The 4th Gen AMD EPYC processors continue AMD's x86 architecture innovation record and industry leading performance. Each 4th Gen AMD EPYC processor includes up to 96 cores per socket running at high frequencies, high memory bandwidth, and capacity. Exadata Cloud@Customer has been an enormous success being used in more than 60 countries by large and medium-size organizations including Deutsche Bank, AT&T, HSBC, and TJMG.

NEW: OCI Compute Cloud@Customer

Compute Cloud@Customer delivers OCI compute, storage, and networking infrastructure services using the latest technologies on the customer's premises. It's built, installed, maintained, owned, and fully managed by Oracle. Ideal for modernizing and consolidating traditional, new, and cloud based applications. Compute Cloud@Customer is dynamically elastic with both an infrastructure price component and consumption component.

The low consumption pricing is the same as OCI commercial regions and Dedicated Region. Compute Cloud@Customer helps the customer meet regulatory requirements for data residency, and data privacy. It does so by giving them control over the location of their data and its processing throughout its lifetime. It's an ideal way to consolidate legacy applications on high performance managed cloud infrastructure and meet the needs of applications to have low-latency access to non-cloud or location-specific data.

This Wikibon research examines how Oracle's Compute Cloud@Customer solves the current limitations and high costs faced by customers of AWS, Azure, and GCP offerings in this space. In the Wikibon view, Oracle Compute Cloud@Customer is the perfect complement to Exadata Cloud@Customer X10M. It allows customers to achieve high performance and availability for their entire on-premises cloud infrastructure stack. Compute Cloud@Customer delivers exceptional performance, scalability, security, flexibility, true elasticity, and much lower costs than other systems. It's likely to be highly disruptive to the marketplace. Expect it to win new customers and expand Oracle's cloud market share.

How Oracle Compute Cloud@Customer Overcomes AWS, Azure, & GCP Limitations

Solving those aforementioned problems demanded new hardware and software innovations.

Solving the Cloud On-Premises Performance and Scalability Problems

Compute Cloud@Customer is specifically architected to fully leverage the 4th generation AMD EPYC CPU, allowing it to deliver more cores, faster cores, and greater energy efficiency than its competitors. Compute Cloud@Customer hardware then uses the open-source Oracle Linux distribution to support OCI services, Oracle Applications, Oracle Middleware.

To put Compute Cloud@Customer performance in perspective, AWS Outposts has so many fewer and slower cores that it requires on average 4x the number of racks to equal one Compute Cloud@Customer rack in CPU cores. And that doesn't account for the fact that each Compute Cloud@Customer core is at least 10% faster since AWS Outposts is using Intel-based x86 servers introduced in 2017 or 2018. Compute Cloud@Customer starts out with 552 available compute cores, 6.7 TB of available memory, and 150 TB of storage that can be accessed using block, file, or object protocols. The system can scale compute and memory by 12x, and independently scale storage by more than 20x. This enables the customer to configure up to 6,624 cores, 80.4 TB memory in up to 3 compute racks[1], and 3.4 PB of flexible storage in additional

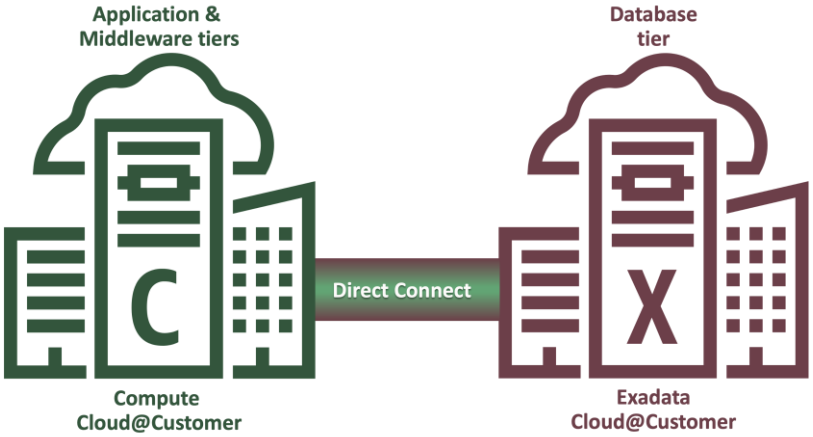
[1] When a customer needs more than 3 racks and less than 12 (the minimum for OCI Dedicated Region), Oracle said they will work with the customer to accommodate.

© 2023 Wikibon

racks. Note that AWS Outposts requires at least 12 racks to match 3 racks of Compute Cloud@Customer in cores alone.

Solving Inadequate Enterprise Database Services On-Premise

Organizations that wish to pair high-performance Oracle Database capabilities with their Compute Cloud@Customer can do so using on-premises Exadata Database Machine or Exadata Cloud@Customer platforms. Compute Cloud@Customer can share a private, very high performance interconnect with Exadata Cloud@Customer. Applications can access databases over up to 800 Gbps of bandwidth without having to contend with other traffic that could increase access latency. This pairing screams performance and will be a welcome addition to Oracle distributed cloud portfolio for existing Exadata Cloud@Customer users.



Solving Insufficient Elasticity

The two parts of Compute Cloud@Customer are "Infrastructure Subscription", and "Metered Consumption". They consists of the following:

Infrastructure Subscription	Metered Consumption
Dedicated infrastructure configured for required services	OCI IaaS Services
Optional upgrades	• *Compute*
Installation	• *Memory*
Fixed monthly costs based on 4 year commitment	• *Storage*
4 different **upgradeable** rack configurations • *Small – 552 cores 6.7TB mem, 150TB usable storage* • *Medium – 1104 cores, 13.3TB mem, 150TB usable storage* • *Large – 1656 cores, 20TB mem, 150TB usable storage* • *XL – 2208 cores, 26.8TB mem, 150TB usable storage*	• *Load Balancer* Takes advantage of universal credits across OCI, Compute Cloud@Customer, and Exadata Cloud@Customer
Expansions and Upgrades Balanced Storage supporting up to **175 TB** usable storage High Performance Storage with up to **60 TB** usable storage E5 Compute with up to **552** OCPUs and **6.7 TB** memory	Pay only for what is used/consumed Metering basis • Compute-Standard-E5: OCPU/hr. • Compute-Standard-E5-Memory: GB mem/hr. • Block Volume Storage-Balanced: GB capacity/mo. • Block Volume Storage-Performance: GB capacity/mo. • File Storage: GB capacity/mo. • Object Storage: GB capacity/mo. • Load Balancer: Load Balancer/hr.

© 2023 Wikibon

Compute Cloud@Customer borrows from the design of Gen2 OCI data centers by implementing three parallel availability domains inside the rack. A major benefit of this design is non-disruptive infrastructure upgrades, software upgrades, and patches. Workloads are auto-moved transparently to other servers while updates take place. Those seamless, non-disruptive software updates, and patches are customer-defined, automated, and delivered from OCI Public Cloud, or Dedicated Region. And consumption elasticity scaling is also non-disruptive in both upwards and downwards directions enabling customers to efficiently manage their costs.

Putting it plainly, Compute Cloud@Customer provides the elasticity customers are demanding.

Solving The Unsatisfactory Security Across Hybrid Or On-Premises Clouds

Security is a major concern for every IT organization today. Malware, ransomware, state run malevolent hacking firms, malicious actors, internal disgruntled employees, seem to be everywhere. Laws, regulations, and corporate governance have made security more important than ever.

Oracle has nearly 5 decades experience in meeting the security needs of enterprise and government IT organizations. They leveraged that experience and expertise with Compute Cloud@Customer. It has built-in unified user identity management that works seamlessly across OCI Public Cloud, OCI Dedicated Region, Exadata Cloud@Customer, and Compute Cloud@Customer. It both strengthens and simplifies user identity management.

Compute Cloud@Customer also comes with highly secure operator controls. Originally developed and introduced to support Exadata Cloud@Customer deployments in the financial services industry, Operator Access Control empower secure, granular, auditable, and high visibility into Oracle Cloud operator access. It enables customers to control when Oracle operators can access their system, how long they can access it, and what operators can monitor or commands they can run. Customer IT can monitor everything that the Oracle Cloud operator does and can immediately terminate their access at any point in time.

Those aforementioned non-disruptive seamless software updates, upgrades, and patching are also a major security factor. This is expressly important for patching. Security vulnerability patches are a fact of life, but having too many patches can be disruptive meaning they have to be scheduled for an “off time.” Off times have become extremely rare in a 7x24x365 world, historically occurring just one weekend a month. The first 24 hours are scheduled to implement the disruptive patches, updates, or upgrades. The next 24 hours are allocated to remove the patches, updates, or upgrades that broke something, which will delay the remediation of security vulnerabilities.

Vulnerability patching has become increasingly problematic as the number of security vulnerability patches has grown exponentially. The Cyentia Institute reports that at the turn of the 21st century the number yearly vulnerabilities requiring patching averaged approximately 1,300. By 2020 that number had risen to more than 18,000 annually. Cyentia’s research revealed the median average time to vulnerability patch remediation was ≥ 100 days with 25% of the vulnerabilities remaining open for more than a year.

When a security vulnerability patch comes out it details the vulnerability or vulnerabilities that are being patched. The malicious actors race to exploit those vulnerabilities before they can be patched. That’s why Compute Cloud@Customer provides customer controlled non-disruptive, seamless, and automated patching.

Compute Cloud@Customer makes security simple and effective.

Solving the Exclusion of Too Many Essential Database Cloud Services

Most of the public clouds extending services to the customer’s premises are primarily infrastructure as a service (IaaS), public cloud APIs and tools, in addition to a smattering of databases. Even the hardware used by Azure Stack and GCP Distributed Cloud is not the same hardware that runs in their public clouds.

Compute Cloud@Customer changes that paradigm. And while it too is an IaaS with the same APIs and tools as OCI, it is specifically architected to be the ideal for the application and middle layer tiers. It’s very high-performance interconnect to Exadata Cloud@Customer X10M provides every essential database model, machine learning (ML), or workload desired, all with a single data copy. Contrast that with AWS Outposts

© 2023 Wikibon

that runs only MySQL, Microsoft SQL Server, and PostgreSQL. Outposts doesn't support Aurora, Oracle Database, Redshift, SageMaker, or any other AWS database model. Azure is similarly hamstrung as is GCP.

Compute Cloud@Customer is not hamstrung and will surely increase service offerings over time, just as they did with Exadata Cloud@Customer.

Eliminating Separate Vendors For Software and Hardware

Few IT pros like troubleshooting software and hardware problems. It can frequently be difficult and time-consuming. The difficulty and time-consumption escalates rapidly when there are different vendors involved, even when those vendors are partnered. Problems are sometimes fronted by one of the partners, but there are still multiple separate organizations involved. Those separate organizations have their own processes, culture, and escalation procedures. That can make solving those difficult problems frustrating and lengthy.

When the hardware for the cloud on customer premises is different from the hardware in the public cloud, developing applications that move seamlessly between the two will frequently have some hiccups. The software code often needs to be adjusted.

Azure Stack Hub, Azure Stack HCI, and Azure Stack Edge all require a separate hardware partner. So do some of the Google Distribute Cloud Edge offerings. The only platform Google provides directly to customers is the small, underpowered Google Distributed Cloud Edge that does not even have shared storage and isn't appropriate for running mission-critical applications or databases. Most of Google's focus is on running their Google Distribute Cloud Edge service called Anthos on hardware platforms that customers must buy from Google partners.

Compute Cloud@Customer does not have a multivendor infrastructure. Not only is the hardware and software both provided by Oracle, but they're tightly co-engineered to work better together. That means problem troubleshooting is simpler and faster. There is only one throat to choke. And any applications that are developed and runs on Compute Cloud@Customer is based on a fully optimized cloud stack that is identical to OCI. That means any application developed for or simply running on Compute Cloud@Customer will also run in the OCI and Dedicate Region Cloud@Customer without changes.

Eliminating Excessive Cost

There are many positive aspects of using public cloud resources on a customer's premises. Aspects such as lower latencies, faster application response times, data residency, improved security, cloud resource consumption and elasticity cost model. But cost has been and continues to be a major obstacle. It turns out customers can be a bit testy about high costs. Compute Cloud@Customer has significant advantages in this area.

It has already been established that customers that have switched from AWS, Azure, and GCP have switched to OCI commonly save more than 50% on their costs. Some save more as much as 83%. The question becomes whether or not those savings translate to the customer's premises, especially with the infrastructure component. It appears it does.

First, a look at how Oracle Compute Cloud@Customer consumption pricing compares with OCI shows they're identical.

	OCI Public Cloud	Compute Cloud@Customer
Cloud Services	Infrastructure, Platform, SaaS apps	Infrastructure
Compute Standard – E5	$0.03 per hour per OCPU	$0.03 per hour per OCPU
Object Storage – Standard	$0.0255 per GB per month	$0.0255 per GB per month
Block Volume Storage – Balanced	$0.0425 per GB per month	$0.0425 per GB per month
File Storage	$0.30 per GB per month	$0.30 per GB per month
Oracle Linux, GraalVM, Java SE	Included at no cost	Included at no cost
Management	OCI engineering	OCI engineering
OCI Consumption	Universal Credits	Universal Credits

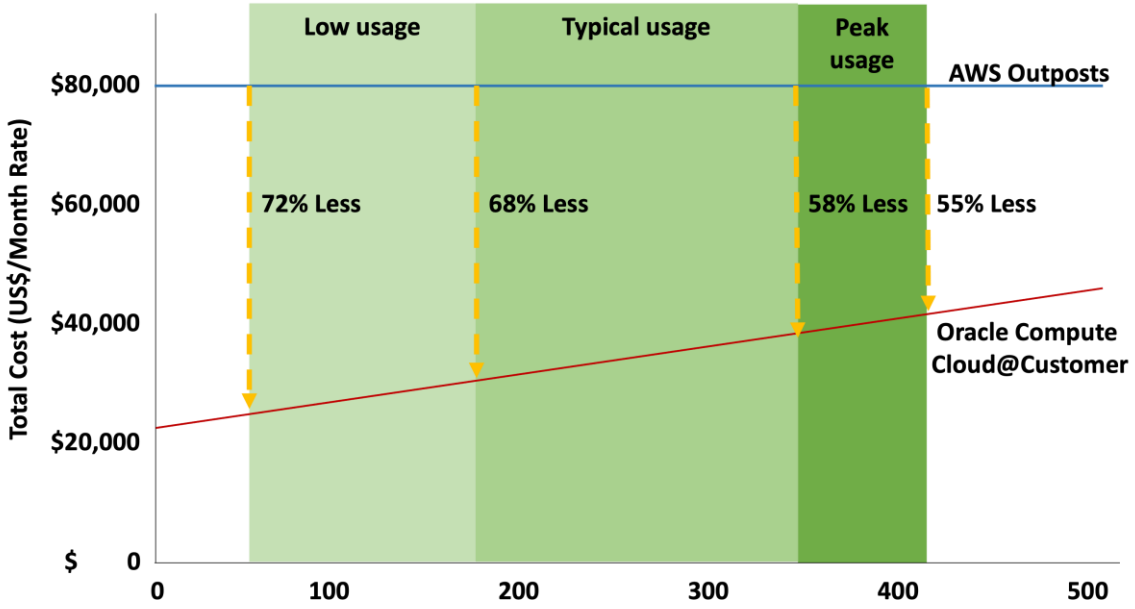
© 2023 Wikibon

More importantly, how does Compute Cloud@Customer overall costs compare to other public cloud providers on-premises services. AWS has been the market leader, and since they provide both the hardware and software similar to Oracle, it's the most straightforward comparison.

It's important to note that AWS changed the way they charge for infrastructure with Outposts. Outposts doesn't provide a separate infrastructure consumption charge. Instead, AWS charges fully loaded consumption pricing for the computer, internal interconnect, and storage in an entire Outposts rack. It does not matter whether the customer uses all of that infrastructure or not. Nobody realistically completely uses all of their Outpost infrastructure. That's exacerbated by the AWS recommendation for customers to "allocate sufficient additional capacity for mission-critical applications to enable recovery and failover if there is an underlying host issue." This results in AWS Outposts customers paying for consumption that they will never use, particularly when they run crucial applications.

Because of architectural and configuration differences, in addition to Oracle not charging for the CPU cores on each compute node that run the hypervisor and various systems, AWS does charge for those cores, it's impossible to do a perfect 1-1 comparison. However, a pretty good estimate of the savings can be calculated by customers based on their utilization level. And the cost differences are startling.

Looking at equivalent configurations in cores, memory, storage, and time frames, it is clearly evident based on MSRP, that Compute Cloud@Customer reduces customer costs by a minimum of 55%. That's when both systems are fully utilized and as much as 72% when they are lightly utilized. Given dynamic scaling, most customers will fall somewhere in the mid 60% savings with Compute Cloud@Customer. It's markedly clear that a combination of Compute Cloud@Customer elasticity, pay for use, and lower pricing all contribute to the much lower costs. See chart below.



Based on the fact that Azure Stack and Google Distributed Cloud Edge have similar costs to AWS, the savings from Compute Cloud@Customer should be similar.

How Oracle Compute Cloud@Customer Stacks Up

Oracle Compute Cloud@Customer provides much more than AWS, Azure, and Google in their cloud on customer premises. More cores per rack, more cores per VM, more storage capacity, more types of storage, more storage flexibility, and a high speed, low latency, no contention interconnect to Exadata Cloud@Customer.

© 2023 Wikibon

	Oracle Compute Cloud@Customer	AWS Outposts	Azure Stack Hub	Google Distributed Cloud Edge
Cores per rack	2,208 cores	576 cores	1,536 cores	288 cores
Cores per VM	96 cores	48 cores	32 cores	48 cores
Storage capacity	3.4 PB	435 TB	1.5 PB	19.8 TB
Storage services (Block, object, and file)	Yes	No File	No File	No Object
Storage flexibility (Change protocol on demand)	Yes	No	No	No
Exadata Cloud@Customer	Yes	No	No	No

Conclusion

The Wikibon conclusion is the Oracle Compute Cloud@Customer unlocks the magic, performance, and affordable pricing of OCI by making it highly available on the customer's premises. That's critically important for those organizations that need data residency, lower application latencies with faster response times, and tighter security. Just as critical is the affordable pricing.

All of this provides a much lower price/core, but also a lower price/performance in a smaller footprint. The lower pricing plus higher performance:

- 4th generation AMD EPYC cores are faster than Intel, previous AMD generations, and ARM. The latest generation of AMD EPYC processors empower faster time-to-value by delivering performance and scalability.
- No cloud on customer premises platform can connect to Oracle Exadata Cloud@Customer X10M with the same low latency high bandwidth interconnect. Nor can any cloud on customer premises come anywhere close to the database performance or automation of Exadata Cloud@Customer X10M.

The combination of Compute Cloud@Customer and Exadata Cloud@Customer X10M delivers the lowest application response times. Nothing else comes close. Those lower application response times translates into:

- Reduced application response times.
- Higher employee productivity, quality of work, morale, and retention.
- Lower employee turnover, training, and cost.
- Lower database and infrastructure cost via consolidation.
- Faster-time-to-market.
- Faster-time-to-actionable-insights
- Faster-time-to-unique-revenues and profits.

When it comes to on-premises cloud services, Oracle's Compute Cloud@Customer is an obvious winner. Being late to the market means it learned from the mistakes of the other public cloud providers.

Wikibon recommends the Oracle Compute Cloud@Customer as it overtakes the market.

More Information

Oracle Compute Cloud@Customer

© 2023 Wikibon