

# Oracle Compute Cloud@Customer Compute Expansion Options

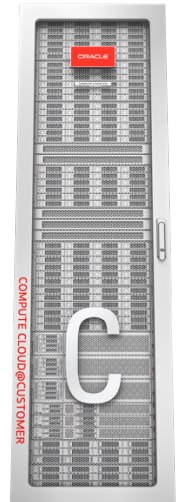
Oracle's fully managed, rack-scale infrastructure that lets customers use OCI compute and harness GenAI on-premises with Oracle Compute Cloud@Customer featuring NVIDIA L40S GPUs, combining performance, data sovereignty, and cloud economics

## ORACLE COMPUTE CLOUD@CUSTOMER

Oracle Compute Cloud@Customer lets you run applications and middleware using OCI Services on high performance infrastructure in your datacenter and is an integral part of Oracle's distributed cloud architecture. This architecture allows customers to overcome some of the challenges of public cloud. Oracle Compute Cloud@Customer is ideal for customers who desire cloud operational and economic benefits, yet require their applications, middleware and data be located on-premises. This could be due to data residency laws, industry regulations, corporate policies, network latency, or the impracticality of moving applications away from other tightly coupled on-premises applications, databases and/or infrastructure.

Oracle Compute Cloud@Customer runs the same compute, storage, and network services as in Oracle Public Cloud, delivering the simplicity, agility and elasticity of a cloud-based deployment. With this service running in your datacenter, customers retain control of data and physical security. The exact same APIs and tools available in Oracle Public Cloud are also available on Oracle Compute Cloud@Customer for a truly consistent application development framework.

With the ability to directly connect to Oracle [Exadata Cloud@Customer](#), customers have the most modern, secure, available and high-performance cloud platform to run applications and databases on-premises, all managed and supported by Oracle, while benefitting from cloud operations, economics and agility.



## GPU Expansion Options for Oracle Compute Cloud@Customer

[Oracle Compute Cloud@Customer](#) featuring [NVIDIA L40S GPUs](#) is a pivotal addition to Oracle's suite of on-premises cloud solutions, which includes Oracle Dedicated Region Cloud@Customer (DRCC) and Oracle Exadata Cloud@Customer. Together, these offerings form Oracle's Distributed Edge, enabling enterprises to bring high-performance AI, machine learning, and database capabilities into their own datacenters, all with the control and security of an on-premises environment. Oracle Compute Cloud@Customer can scale up to 48 Nvidia L40S GPUs and is the ideal choice for organizations seeking high-performance compute without the need for a full cloud region on-premises. With Oracle Cloud Infrastructure, Dedicated Region Cloud@Customer, and Compute Cloud@Customer, Oracle brings AI/ML acceleration and scalable compute power to enterprises in a flexible, cost-effective solution, catering to all deployment scales and requirements of customers.

### Why Oracle Compute Cloud@Customer with NVIDIA L40S?

Oracle Compute Cloud@Customer (C3) GPU provides a scalable platform to build AI and graphics intensive applications at the edge. It is built to power the next generation of datacenter workloads including:

- Generative AI inference: real time inferencing for multi-model generative AI pipelines (text, image, audio, video)
- LLM training and fine-tuning: accelerated performance for fine-tuning medium LLMs and training small LLMs with NVIDIA transformer engine and FP8 support
- Graphics-intensive and VDI applications: 3D graphics and rendering workflows with NVIDIA's RTX and ray tracing capabilities
- Digital twins using NVIDIA Omniverse: Develop and operate complex 3D industrial digitization workflows
- Media streaming: Increased encode/decode density and AV1 support for 4K video streaming
- HPC: Scientific data analysis and simulation workloads with FP32 support

### Configuration

Oracle Compute Cloud@Customer supports up to 48 NVIDIA L40S GPUs across two expansion racks.

HARDWARE	SHAPE	CAPACITY
<b>GPU Node</b> <ul style="list-style-type: none"> <li>• CPU: 2x Intel® Xeon® Platinum 8480+ Processor with 56 cores</li> <li>• GPU: 4x NVIDIA L40S GPU (PCIe) 48GB GDDR6</li> <li>• DRAM: 960GB</li> </ul>	<b>C3.VM.GPU.L40S.1-4</b>  VMs with 1-4 passthrough GPUs <ul style="list-style-type: none"> <li>• 27 OCPUs per GPU</li> <li>• 240 GB memory per GPU (from the same GPU CN)               <ul style="list-style-type: none"> <li>• 200 GB memory per VM</li> <li>• 40 GB memory reserved</li> </ul> </li> </ul>	<b>Minimum per rack</b>  1x GPU-CN, 108 OCPUs, 960GB Memory, 4x NVIDIA L40S with 48GB GDDR6 memory
<b>Storage</b>  Dedicated: ZS9 (Ethernet 100Gb) + DE3-24P (High Performance Storage Disk Tray)		<b>Maximum per rack</b>  6x GPU-CN, 648 OCPUs, 5760GB Memory, 24x NVIDIA L40S with 48GB GDDR6 memory
<b>Networking</b>  Up to 400 Gbps Bandwidth over front end network (no RDMA support)		Non-Clustered GPUs with max. 4 GPUs over PCIe

## Specifications

COMPONENT	HEIGHT	WIDTH	DEPTH	WEIGHT
Oracle Compute Cloud@Customer Infrastructure - GPU Compute with up to 4 GPUs, 48 GB GPU memory, 108 OCPUs and 960 GB CPU memory	5.1 in, 130.5 mm (3 RU)	19.5 in, 496.5 mm	32.9 in, 838.1 mm	117 lbs, 53 kg
Oracle Compute Cloud@Customer Infrastructure – GPU Expansion Rack	2010 mm, 79.1 in	600 mm, 23.6 in	1250 mm, 49.2 in	831 lbs, 377 kg

## Power Consumption

### Maximum Power Consumption Metrics

NODES	WATT	KW	KVA	BTU/HR	KJ/HR	CFM
1	3760	3.76	3.84	12837	13543	594
3	9260	9.26	9.45	31614	33352	1464
6	17510	17.51	17.86	59779	63067	2768

### Typical Power Consumption Metrics at 70% consumption

NODES	WATT	KW	KVA	BTU/HR	KJ/HR	CFM
1	2632	2.6	2.68	8986	9480	416
3	6482	6.5	6.61	22130	23347	1025
6	12257	12.3	12.50	41845	44147	1937

---

## Connect with us

Call +1.800.ORACLE1 or visit [oracle.com](https://www.oracle.com). Outside North America, find your local office at: [oracle.com/contact](https://www.oracle.com/contact).

 [blogs.oracle.com](https://blogs.oracle.com)

 [facebook.com/oracle](https://facebook.com/oracle)

 [twitter.com/oracle](https://twitter.com/oracle)

---

Copyright © 202525, Oracle and/or its affiliates. All rights reserved. This document is provided for information purposes only, and the contents hereof are subject to change without notice. This document is not warranted to be error-free, nor subject to any other warranties or conditions, whether expressed orally or implied in law, including implied warranties and conditions of merchantability or fitness for a particular purpose. We specifically disclaim any liability with respect to this document, and no contractual obligations are formed either directly or indirectly by this document. This document may not be reproduced or transmitted in any form or by any means, electronic or mechanical, for any purpose, without our prior written permission.

Oracle and Java are registered trademarks of Oracle and/or its affiliates. Other names may be trademarks of their respective owners.

Intel and Intel Xeon are trademarks or registered trademarks of Intel Corporation. All SPARC trademarks are used under license and are trademarks or registered trademarks of SPARC International, Inc. AMD, Opteron, the AMD logo, and the AMD Opteron logo are trademarks or registered trademarks of Advanced Micro Devices. UNIX is a registered trademark of The Open Group. NVIDIA is a registered trademark of NVIDIA Corporation.

Disclaimer: This document is for informational purposes. It is not a commitment to deliver any material, code, or functionality, and should not be relied upon in making purchasing decisions. The development, release, timing, and pricing of any features or functionality described in this document may change and remains at the sole discretion of Oracle Corporation.