

# Oracle Private Cloud Appliance Expansion Options

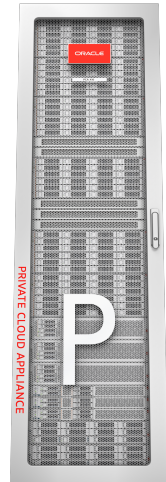
Harness GenAI in your datacenter with Oracle Private Cloud Appliance featuring NVIDIA L40S GPUs, combining performance, data sovereignty, and cloud economics.

## Oracle Private Cloud Appliance

Oracle Private Cloud Appliance enables customers to efficiently consolidate business critical middleware and application workloads. Private Cloud Appliance is a rack-scale engineered system delivering Oracle Cloud Infrastructure (OCI) compatible APIs, SDKs, and management tools on-premises, making workloads, user experiences, tool sets and skills portable between private and public clouds. It enables rapid deployment of applications, middleware, and workloads that are cloud-compatible via automation in an OCI-like environment while being disconnected from the public cloud.

The Private Cloud Appliance can be paired with Oracle Exadata or Oracle Database Appliance to create an ideal infrastructure for scalable, multi-tier applications. It's direct connection to Oracle's database platforms provides extremely low latency and high throughput between applications and databases in a full-stack application environment.

Customers using the Private Cloud Appliance realize “cloud-like” operational benefits, including single-vendor support for their full-stack environment and Trusted Partitions that enables efficient software licensing for Oracle software based on the size of the VMs used.



## GPU Expansion Options for Oracle Private Cloud Appliance

Enterprises around the globe are looking for distributed deployment options for running their AI workloads which can help them cater to data residency, ultra-low latency and regulations. Oracle Private Cloud Appliance featuring [NVIDIA L40S GPUs](#) enables enterprises address these requirements while also delivering cloud-grade AI capabilities directly into enterprise data centers.

This powerful combination empowers businesses to run low-latency AI/ML workloads, high-performance computing (HPC), and graphics-intensive 3D rendering while keeping sensitive data securely on-premises. Whether enterprises focus on generative AI, large language model (LLM) fine-tuning, or real-time digital twin simulations, Oracle Private Cloud Appliance with NVIDIA L40S GPUs offer the scalability and flexibility needed to accelerate innovation.

### Why Oracle Private Cloud Appliance with NVIDIA L40S GPUs?

Oracle Private Cloud Appliance with NVIDIA L40S GPUs provides a scalable, customer managed 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's 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 Private Cloud Appliance can scale up to 48 Nvidia L40S GPUs and is ideal for enterprises looking for a customer-managed Compute and GPU infrastructure.

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>PCA.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 ea.
<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

METRIC PER GPU COMPUTE NODE	GPU COMPUTE
Height	5.1 in, 130.5 mm (3 RU)
Width	19.5 in, 496.5 mm
Depth	32.9 in, 838.1 mm
Weight	<b>117 lbs, 53 kg</b>

## 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**. Outside North America, find your local office at: **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 © 2025, 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.