

the
GORILLA
GUIDE[®] to...



Modernizing IT Infrastructures with HPE ProLiant Servers

Security, Generative AI, High
Performance and Compute Ops
Management

ED TITTEL

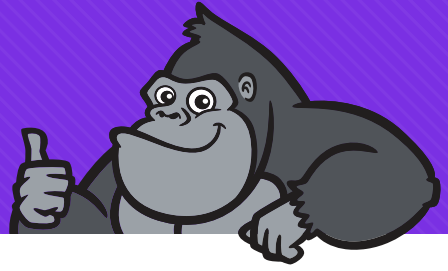


Hewlett Packard
Enterprise

intel[®]

POWERED BY  **ActualTech**
MEDIA

the
GORILLA
GUIDE[®] to...



Modernizing IT Infrastructures with HPE ProLiant Servers

Security, Generative AI, High
Performance and Compute
Ops Management

By Ed Tittel

POWERED BY  **ActualTech**
MEDIA

Copyright © 2025 by Future US LLC
Full 7th Floor
130 West 42nd Street
New York, NY 10036

All rights reserved. This book or any portion thereof may not be reproduced or used in any manner whatsoever without the express written permission of the publisher except for the use of brief quotations in a book review. Printed in the United States of America.

www.actualtechmedia.com

PUBLISHER'S ACKNOWLEDGEMENTS

DIRECTOR OF CONTENT DELIVERY

Wendy Hernandez

CREATIVE DIRECTOR

Olivia Thomson

SENIOR DIRECTOR OF CONTENT

Katie Mohr

ABOUT THE AUTHOR

Ed Tittel is a 30-plus year veteran of the IT industry who writes regularly about cloud computing, networking, security, and Windows topics. Perhaps best known as the creator of the *Exam Cram* series of certification prep books in the late 1990s, Ed writes and blogs regularly for GoCertify.com, TechTarget, ComputerWorld, TekkiGurus, and other sites. For more information about Ed, including a resume and list of publications, please visit EdTittel.com.

ENTERING THE JUNGLE

Introduction: Enhance and Transform Your Organization with IT Modernization	8
Chapter 1: Setting the Stage for Server Security	10
Cybercrime Is on the Rise, with Costly Consequences	10
Case in Point: Change Healthcare	11
Surviving the Onslaught	13
Chapter 2: HPE and Intel Deliver Trusted Security by Design	15
A Closer Look at HPE Security Protection	17
Chapter 3: How Intel-Powered HPE ProLiant Gen11 Delivers Business Security	19
Chapter 4: How HPE ProLiant Gen11 and Intel Technology Deliver Computer Operations Security	22
Chapter 5: How Intel-Based HPE ProLiant Gen11 Enables Secure Apps, Services, and Solutions	25
Chapter 6: Initial Considerations for Business Use of GenAI	29
Making GenAI Work for Your Business	32
Common Use Cases for GenAI	33
Why HPE and Intel Excel at GenAI Workloads	35
Chapter 7: Getting Value from GenAI	37
Harness GenAI and Make It Your Own	39
Are You Ready to Benefit from GenAI?	40
Other Key HPE ProLiant Gen11 Factors: Business-Critical for Today's Enterprise Customers	42

Why HPE and Intel Can Help Your Advance Your Initial GenAI Moves	44
How HPE ProLiant GenAI Solutions Meet Challenges	45
Chapter 8: Build Your First GenAI App with HPE ProLiant Gen11 Intel Xeon Servers	47
Choosing a First GenAI Application	48
Corresponding HPE ProLiant Gen11/12 Servers and GPU Options	50
HPE Services Does AI for Customers	52
Chapter 9: Delivering Data Analytics	54
The Private Cloud Base	59
Chapter 10: Supporting Monster Databases	61
Reduce Costs and Improve Performance	61
Chapter 11: Use Virtualized VMware Storage	65
Benefits of VMware vSAN ReadyNodes	66
Maximizing VMware vSAN Original Storage Architecture	67
Advancing to Express Storage Architecture (ESA)	70
Modern Management Mindset	72
Chapter 12: Introducing Modern Compute Operations Management	73
The COM Breakdown, Part 1: What Is Compute?	75
The COM Breakdown, Part 2: What About Operations?	75
The COM Breakdown, Part 3: Why management Is Key	76
Compute Operations Management à la HPE and Intel	77
Chapter 13: Starting from Silicon: HPE Integrated Lights Out (iLO) Management	79
Basic Characteristics of iLO	80
Why Do Organizations Need iLO?	81
Where Does iLO Get Its Value?	81
Introducing HPE Intelligent Provisioning	82
iLO Provides Critical Connection Capabilities	82

Chapter 14: HPE OneView	84
OneView Remains Alive and Well	85
What HPE OneView Brings to the Party	85
OneView Also Leads into GreenLake Compute Ops Management	86
Chapter 15: HPE Compute Ops Management	90
HPE Compute Ops Management Features	90
Key Elements in HPE Compute Ops Management:	92
Benefits of HPE Compute Ops Management	94
Working Toward Modernization with HPE	94

CALLOUTS USED IN THIS BOOK



SCHOOL HOUSE

In this callout, you'll gain insight into topics that may be outside the main subject but are still important.



FOOD FOR THOUGHT

This is a special place where you can learn a bit more about ancillary topics presented in the book.



BRIGHT IDEA

When we have a great thought, we express them through a series of grunts in the Bright Idea section.



DEEP DIVE

Takes you into the deep, dark depths of a particular topic.



EXECUTIVE CORNER

Discusses items of strategic interest to business leaders.



DEFINITION

Defines a word, phrase, or concept.



GPS

We'll help you navigate your knowledge to the right place.



KNOWLEDGE CHECK

Tests your knowledge of what you've read.



WATCH OUT!

Make sure you read this so you don't make a critical error!



PAY ATTENTION

We want to make sure you see this!



TIP

A helpful piece of advice based on what you've read.

INTRODUCTION

Enhance and Transform Your Organization with IT Modernization

Welcome to The Gorilla Guide To...[®] Modernizing IT Infrastructures with HPE ProLiant Servers.

In this guide, we'll explain what's involved in modernizing your IT infrastructure, and explore key factors that HPE ProLiant Servers powered by Intel Xeon Scalable processors bring to that exercise. To begin our journey, please understand that the fundamental value of IT modernization comes from its ability to enhance and transform an organization's operations, to make them more agile, efficient, and competitive. This generally means closely and carefully examining existing systems, software, and infrastructures, then deciding on how to upgrade or replace them for best effect. Overall, key benefits from modernization include:

- **Improved efficiency:** Modernized IT systems streamline processes, reduce redundancies, and automate tasks, for faster and more capable operations
- **Enhanced agility:** A modernized technology environment lets organizations adapt quickly and robustly to changing markets, customer demand, and new business opportunities

- **Cost savings:** A modernized IT environment helps optimize infrastructure, lowers support costs for aging legacy system, to reduce TCO and improve resource usage
- **Better security:** Modernized systems and services generally bring more advanced security features, improved data protection and access controls, and reduce security exposures
- **Room to innovate:** Up-to-date technology enables and anticipates leading-edge/bleeding-edge tools and platforms, at the edge, in the cloud or on-premises so that organization can focus on meeting strategic goals and growing the business

To those ends, HPE ProLiant Servers powered by Intel Xeon Scalable processors have been designed specifically for end-to-end security, to support generative Artificial Intelligence (GenAI) use cases and workloads, to deliver high performance and reliability, and to promote a fully modern approach to compute operations management. In the upcoming sections of this eBook we'll tackle those four subjects—namely, security, GenAI, performance, and compute ops management—in that order, and explore them in sufficient detail to make HPE ProLiant Server capabilities in those areas intelligible and compelling for your organization.

In the first section of this guide (Chapters 1 through 5), we'll explore how HPE designs security into its ProLiant Server family from the silicon all the way to the application layer, including support for business security, computer operations security, and secure applications, services, and solutions.

CHAPTER 1

Setting the Stage for Server Security

From the perspective of implementing IT systems and infrastructures, security and data protection must be among any organization's primary requirements. That's because any gap or breach can wreak untold havoc on a business or organization, including taking it offline for days or weeks. Sometimes outages can drag out for months! This is simply intolerable for businesses to achieve their goals and meet their financial responsibilities. Ditto for governments, non-profits, organizations, and agencies so they can provide services and information they're supposed to deliver.

Cybercrime Is on the Rise, with Costly Consequences

Today's cyber security news just keeps getting more alarming, expensive, and pervasive. Here are some recent statistics to demonstrate these dramatic downsides:

- **The cost of cybercrime is expected to increase by nearly 70%** between 2023 and 2028 (\$5.7 trillion), says [Statista](#).

- **The average cost of a data breach in the United States was \$9.48 million in 2023** (\$5.13 million in Canada, \$4.45 million globally), also from [Statista](#).
- **With a CAGR over 15%, Cybersecurity Ventures says, “cybercrime represents the greatest transfer of economic wealth in history”** ([Embroker](#), January 2024). That’s *not* the kind of wealth transfer to which anyone aspires, except crooks.
- **Ransomware figures into about one-quarter of all security attacks, but affected 63% of organizations in 2023** ([TechTarget](#), January 2024). The average cost of such an attack in 2023 was \$5.23 million ([Morgan Lewis](#), March 2024).
- **And, finally, Terranova Security reports that, “98% of web applications are vulnerable to attack ...”** (January 2024) with consequences such as malware infection, redirection to malicious websites, or other forms of unwanted or unauthorized access.

Given that IT’s mission is to serve customers, stakeholders, and employees, maintaining security means maintaining normal operations. Simply put, IT can’t afford to use systems or services that might fall prey to cyberattacks. It certainly can’t afford to put entire supply chains on hold until things get fixed, as with the pharma companies, pharmacies, medical professionals, and patients all stymied by a recent meltdown at a national clearinghouse for medical prescriptions, as covered in the next section.

Case in Point: Change Healthcare

On Feb. 21, 2024, the Blackhat hacking gang (closely tied to the same group that brought down Colonial Pipeline in 2021) foisted a successful ransomware attack against a United Healthcare subsidiary

named Change Healthcare (CHC). This might not sound serious, but it is deadly so, with increasingly costly or health-challenging consequences unfolding every day.

Why? Because CHC acts as a clearinghouse for medical prescriptions for most of the United States (millions to tens of millions of scrips daily). CHC also handles payments at pharmacies, and then works in the background with doctors and hospitals, insurance companies and Medicare/Medicaid to obtain and disburse funds for their shares in such transactions (both coming and going). That means millions of people trying to get prescriptions filled can't access copay or insurance coverage amounts, use coupons, or pay by credit card. This leaves many facing a stark choice: pay full price, or do without.

The consequences are already staggering. The U.S. Department of Health is talking about emergency loans to hospitals and hospital systems. Private physicians in solo or group practices need emergency loans or lines of credit to make payroll and keep going. Pharmacies are stuck between having no way to get reimbursed to keep their stocks up, and difficulty collecting from patients who need prescriptions where it can literally be a matter of life or death if they go unfilled. It's a huge mess, with CHC systems slowly returning to service on March 18 (26 days after the initial attack).

Multiple sources report that an as-yet-identified party is reported to have paid the hackers \$22 million by tracking the balance in one of their known Bitcoin wallets. Even so, there's still no estimate as to when CHC will resume normal operations. Indeed, CHC's VP for Integrated Services and Managed Care, Nicole Fallon, is [quoted as saying](#) "...the full extent of the CHC cyberattack's impact ... is simply not yet clear..." and speaks to "...a significant disruption in operations" ([March 2024](#)).

Businesses and organization have no choice but to consider the shoes that CHC currently occupies, and do everything possible to avoid ever having to fill them under any circumstances. HPE ProLiant Gen11

servers with the latest Intel technology can help them with a big part of the due diligence involved in avoiding such a fate. That's because they're designed to be secure from the silicon up to the distributed (and cloud-based) services and applications on which today's IT infrastructures rely.

Surviving the Onslaught

Zero trust and secure operations are vital to providing timely, efficient IT services, especially given increasing dependence on hybrid cloud IT architectures. Powerful personnel problems can present, too, in finding qualified security professionals: with ISC2 [reporting](#) over 4 million open, unfilled security jobs expected in 2025. Increasingly IT regards cyber security as its No. 1 concern, with at least 12% of IT expenditures going toward security coverage.



According to the ISC-squared [Cybersecurity Workforce Study](#) (November 2023), the gap between cybersecurity jobs available and people able to fill them will hit 4 million in 2025. That's over 70% of the total number employed in the field in late 2023 (5.5 million). ISC-squared CEO Clar Rosso says "... we must double this workforce to adequately protect organizations and their critical assets." Businesses and organizations must be equally kind to find and retain new cybersecurity professionals, while doing everything reasonable to retain those already working in this area. HPE's trusted security helps keep attack surfaces small, but it's vital to see qualified cybersecurity pros as the treasured human resources they represent.

In the next chapter, we switch gears from surveying the scary and sometimes sorry state of threats and security, to exploring how HPE in collaboration with Intel can help organizations survive the threats (and thrive despite them) thanks to its long-standing approach known as trusted security by design.

CHAPTER 2

HPE and Intel Deliver Trusted Security by Design

HPE and Intel understand the risks and threats that modern organizations must counter. By careful and deliberate design, the HPE ProLiant Gen11 server family and Intel 5th Gen Xeon processors help organizations deploy technology with confidence and speed. Security is designed and built in from the ground up, starting with a silicon-based root of trust that secures millions of lines of firmware. Currently, there are over 4 million HPE servers, most using Intel technology, in IT operations around the globe that provide fundamental, innovative security, reliability, and performance. Hardware-based security is especially important to enterprise or government agency deployments that might be sensitive or even formally classified (secret, top-secret, and so on). In such scenarios, security is not just good practice, it's mandated by policy, rules, and even law.

The HPE integrated Lights Out (iLO) management system verifies all devices during system boot to make sure they're intact, secure, and tamper-free. Indeed, HPE iLO is a do-it-all server management facility. In fact, it simplifies server setup, supports large-scale deployment and management, provides access to power and thermal controls, and offers basic remote access and administration (see **FIGURE 1**, which shows the HPE iLO security dashboard). Thus, HPE iLO is tailor-made to help IT manage servers across their entire

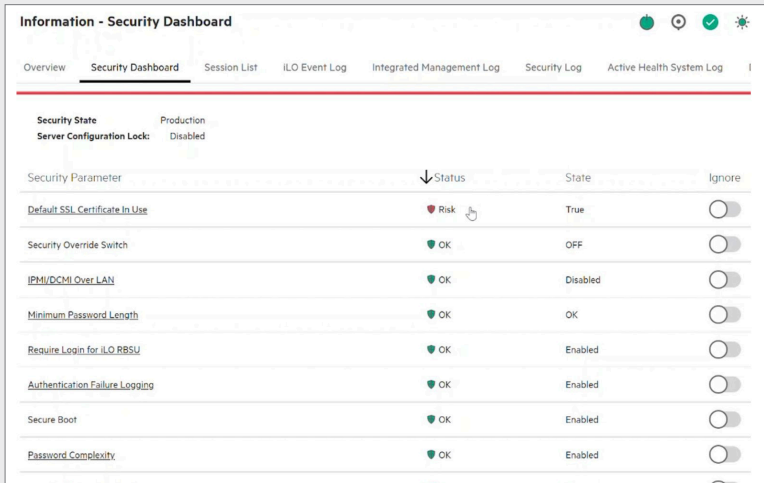


FIGURE 1: The integrated Lights Out (HPE iLO) security dashboard shows key security indicators

lifecycles. The [HPE iLO6 1.56 User Guide](#) explains all of its features and functions, including security, in profuse detail (see the [HPE iLO 6 online demo](#) for a more complete walk-through).

To extend its perimeter for hardware-level security, HPE works with Intel and other partners who provide system components and assemblies under its security umbrella. This includes a security protocol and data model (the industry SPDM standard) to verify component integrity and authenticate add-in interfaces. This adds the same level of protection to third-party elements during boot-up that HPE iLO performs on internal components.

HPE also extends global security coverage to all parties involved in its trusted supply chain. This includes protection and confidence for HPE ProLiant Gen11 servers obtained from trusted suppliers and manufacturing facilities, and extends to protection and tamper-proofing while servers are in transit for delivery. This includes such well-known companies as Intel, Broadcom (which now owns VMware), and others.

A Closer Look at HPE Security Protection



In creating its Intel-based HPE ProLiant server products, including the current Gen11 iteration, HPE and Intel have assiduously focused on security from the hardware level all the way up to the software level (including layers for the OS, platforms or frameworks in use, and the applications or services that use them). This is depicted in **FIGURE 2**, which shows the entire HPE protection path from the bottom up to the top.

From the very first moments of start-up, power-on, and self-test, security and IT protection is established and maintained via the HPE iLO management system. This verifies all devices during system boot to make sure they're intact, secure, and tamper-free. Furthermore, HPE embraces the partners within its security ecosystem to extend protection to them. This includes a security protocol and data model to verify the integrity of components and to authenticate supported option cards. Nothing runs that's not designed, vetted, and checked in real time for security.

HPE also extends security coverage to all parties involved in its trusted supply chain now available around the globe. This includes protection and confidence for HPE ProLiant servers obtained from trusted suppliers such as Intel and manufacturing facilities, and

extends protection against tamper-proofing while servers are in transit for delivery. See the September 2022 blog post, “[HPE trusted supply chain expands to worldwide customers](#),” for more information, or check out the overall description at [HPE Server Security Optimized Service for HPE ProLiant](#) (it explains the complete offering in full detail).

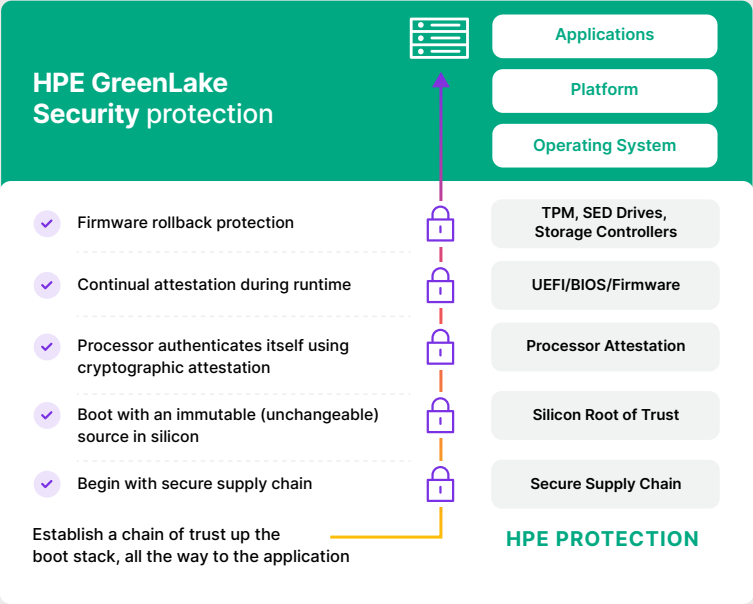


FIGURE 2: HPE Protection spans the whole path from the supply chain, to the hardware, up through the stack into the OS, platform, and application layers

CHAPTER 3

How Intel-Powered HPE ProLiant Gen11 Delivers Business Security

HPE's silicon root of trust defines a continuous, non-stop zero trust architecture. By checking firmware, UEFI, the Intel Xeon processors, and components during start-up, and at every subsequent access, HPE can protect against advanced and persistent threats. Among its other protection techniques, HPE maintains a known, good, and validated copy of UEFI, firmware, and so forth in read-only form. If any change is detected during start-up, or any time thereafter, HPE can replace a suspect in-memory or rewritable entity (e.g., firmware, UEFI, controller code, and more) with the locked down copy to make sure what runs is nothing more nor less than what's supposed to. Should such repairs fail, the runtime environment will shut down rather than risk compromise or breach. This applies to all Intel Xeon-based HPE ProLiant Gen11 servers, regardless of location (edge, cloud, or data center).

These various protections are calculated to protect businesses and organizations from risk. They make sure that what's running is known to be known, good, working and secure before it's turned loosed to do its thing. If what's running cannot be validated, verified, or attested as such—known, good, working and secure—the server

will refuse to run rather than risk potential compromise, exposure, or loss. That's essential for businesses' peace of mind when operating on a rude and dangerous security landscape.

Hardware security gets overall coverage only so far, as you can see in the previous **FIGURE 2**. That's why the company also builds security protection into its HPE GreenLake environment, designed for hybrid cloud use at the edge, in the cloud, or in data center racks. HPE embeds security technologies into its software offerings, starting with attestation services to validate and verify data structure integrity. It also includes built-in firewalls, intrusion detection and prevention systems (IDS/IPS), and identity and access management (IAM) systems.

The overarching idea is to secure software at all key levels—including the operating system, the runtime platforms and frameworks they support, and the applications and services that run inside them—to prevent unauthorized access or disclosure and to fend off

Closer Look: SPDM Defined and Explained

The Security Protocol and Data Model, aka SPDM, is a standard ([DSP0274](#), version 1.01) from the Desktop Management Task Force (DMTF). This specification describes how platform components can register and communicate using SPDM to send messages, create, access, and manage data objects. Key capabilities include authentication of hardware identities and measurement and verification of firmware identities. The guiding principle is to enable "efficient access to low-level security capabilities and operations." That's why HPE uses SPDM to help ensure hardware and firmware security in its servers.



any potential cyberattacks. This lets IT concentrate on operating its applications and services, and shields them from the time and effort required to manage security in reactive mode. This approach is also calculated to deliver safe, secure, and positive user experiences by checking hardware and software before it's used, and then continuously validating and checking it while it's being used (which provides opportunities for monitoring performance, latency, response time and user experience, with automated optimization and/or fixes at the ready should measurements wander outside carefully circumscribed boundaries).

In summary, HPE ProLiant Gen11 servers using the Intel 5th Gen Xeon processor provide the following business benefits:

- Safe, trusted components that stay safe across the lifecycle (servers' silicon root of trust, trusted supply chain, device validation and attestation, secure delivery, set-up, and maintenance)
- Limit business risk and exposure to loss to maximize business opportunities, and leverage growth and innovation
- Ensure and deliver safe, secure, and positive user experiences

CHAPTER 4

How HPE ProLiant Gen11 and Intel Technology Deliver Computer Operations Security

On the hybrid cloud front, HPE’s Intel-powered GreenLake—a potent portfolio of cloud and as-a-service options designed to simplify and accelerate business innovation—includes its own suite of security and data protection capabilities. These are designed to enhance data value through attestation (proof of identity, and showing that OS and application software remain intact, secure, and trustworthy, based on certification authority registered TPM sign-offs) and verification (automated and ongoing checks on data integrity, coherence, validity, and relevance). In fact, HPE GreenLake covers the gamut of possible deployments from platforms (individual servers) to entire fleets (multitudes of servers in multiple data centers).

By design, HPE GreenLake delivers secure, authenticated, and encrypted connections between compute devices and the HPE GreenLake platform itself, as well as the [HPE GreenLake Compute Ops Management](#) service (see “Understanding HPE GreenLake Compute Ops Management Subscription Model”). This latter element is key, because such management lets organizations remotely access any and all parts of their runtime environment (on-premises, at the edge, and in the cloud) and maintain constant, consistent, and policy-driven control over everything.

HPE GreenLake security, integrity, and verification mechanisms help to speed innovation and new technology adoption by creating and building upon a zero trust foundation. This means all parties must continuously prove their identities, and demonstrate verifiable access rights or permissions with each and every interaction with applications, services, and so forth. Zero trust makes it much easier to identify actual or potential attacks. It also helps organizations protect their investments by minimizing attack surfaces, thereby limiting their risks of exposure or compromise

Understanding HPE GreenLake Compute Ops Management Subscription Model

The [HPE GreenLake for Compute Ops Management](#) offering is available in an “as-a-Service” (aaS) subscription model. Organizations can use this environment to unify compute management, quickly update server firmware, and obtain real-time access to servers. This means terrific visibility into your distributed environment, so you can identify issues and update servers in just a few mouse clicks. Better yet, this subscription supports agile server lifecycle management. That means less manual effort, more efficient server deployments and updates, and quick access to server health (with alerts for critical hardware failures so you can respond immediately and appropriately). Based on a zero trust approach that uses multifactor authentication, security certificates, and the industry’s most secure server architecture (HPE ProLiant and Intel) to establish governance and compliance controls for your entire enterprise (edge, data center, and cloud).



As depicted in **FIGURE 2** (Chapter 2), HPE GreenLake handles the upper software-focused layers in your enterprise infrastructure. That includes the following layers in that model: the OS (6, usually Windows or some form of Linux/Unix), platform (7, e.g., Ansible, Puppet, Chef, and HPE GreenLake's own Edge-to-Cloud Platform Services), as well as applications and services. Indeed, HPE GreenLake also includes a comprehensive collection of Software-as-a-Service capabilities, available as a monthly, pay-as-you-go subscription to HPE ProLiant customers.

In summary, HPE's Intel-powered ProLiant Gen11 servers provide the following benefits to computer operations:

- Fast, easy access to server setup, deployment and action, thanks to HPE iLO
- Zero trust approach ensures hardware, infrastructure, applications, and data start out protected, and stay protected in use, in motion, and at rest
- HPE GreenLake delivers comprehensive protection at the OS, platform, and application layers at the edge, in the data center, and in the cloud—everything's integrated

CHAPTER 5

How Intel-Based HPE ProLiant Gen11 Enables Secure Apps, Services, and Solutions

In building the tools, platforms, and environments that support the Intel-powered HPE ProLiant Gen11 server family, the company has been extraordinarily careful to craft extensions, APIs, and tools to open its environment to third parties and technology partners. This means extending security to all entrants willing to abide by and adhere to the HPE GreenLake security model.

That model includes the following characteristics:

- Core device management functionality with full visibility, dashboards, and health indicators
- Common console for server status available to any designated administrator on any Internet-attached device
- Lifecycle operations one device at a time, including uptake, setup, deployment, updates, and retirement/safe disposal or reassignment
- Utilize all features and functionality that HPE GreenLake Compute Ops offers, including visibility, remote access and operation, status and deployment, policy enforcement, and more

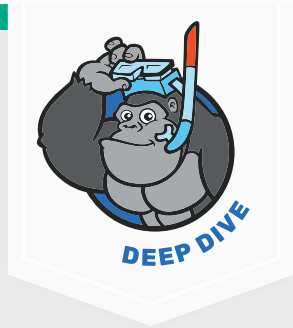
- Policy-driven to support security, governance, and compliance mandates
- Automated and group actions to simplify administrative overhead, reduce (or eliminate) routine manual tasks, speed incident or outage response, and so forth
- REST APIs: an open, fully documented [application programming interface](#) for MSPs or customers who wish to manage devices via software or the command line

In addition, HPE GreenLake offers more of the same kind of benefits it brings to individual servers for fleet-scale deployments (many servers in one or more data centers, on-premises or in the cloud). Most importantly, HPE GreenLake security protection provides attestation, verification, and validation for hybrid clouds at the top of the security stack—that is, for the operating system, runtime platform and environment, and applications or services in use. Here, attestation and validation mean that integrity, intactness, and proper use of access controls also apply to all of these key elements in constant use:

- Operating systems, such as Windows, Linux ,and more
- Platforms, such as containers, virtualization tools, runtime libraries, and other general software supports (e.g., databases, AI models, data analytics, and so forth)
- Applications and services, for everything from simple apps to complex systems for email, learning management, collaboration, content management, and most anything else you might think of

Thus, HPE in collaboration with Intel provides solid security from end-to-end all the way from silicon into applications and services. And it does so at scale all the way into and across the hybrid cloud combinations typical in most modern enterprises, organizations, and agencies.

How HPE GreenLake for Compute Ops Management Creates MSP Opportunities



As explained earlier, the [HPE GreenLake for Compute Ops Management](#) offering is available in an “as-a-Service” (aaS) subscription model. It’s also useful to understand that HPE GreenLake is inherently distributed for use in hybrid and multi-cloud environments. That means it extends security from individual servers acting as service or application hosts to the fleet level, where hundreds or even thousands of servers may be deployed, configured, monitored, and managed through its dashboard and other interfaces.

Some of the biggest beneficiaries of this capability include managed service providers (MSPs). If they operate their own Intel Xeon-based HPE ProLiant Gen11 servers (or remotely manage such servers at customers’ requests) they can use all the capabilities of HPE GreenLake Computer Ops management as their customers’ needs, situations, and budgets allow. As explained in this [November 2022 CRN](#) story, HPE partners can resell Compute Ops Management subscriptions (supports both Gen10 and Gen11 servers), either standalone or tied to specific servers. Indeed, HPE urges those partners to bundle such a subscription with every server sale. Why? Because it brings benefits to MSPs and their customers. MSPs can bundle their own services using open APIs. Customers gain through simpler, more understandable management, automation, and security/policy controls.

In summary, HPE ProLiant Gen 11 servers and the Intel Xeon 5th Gen processors that power them can support secure solutions, services, applications, and data because:

- Security is fundamental, from edge to cloud, so that all systems, OSES, platforms, software, and services run in a protected, zero trust environment
- HPE GreenLake extends security into the OS, platform, and applications/services layers thanks to careful use of multifactor authentication, digital certifications, and constant attestation and validation of identity, privileges, access rights, and so forth
- Third parties can (and do) operate under the HPE umbrella, thanks to partner programs that extend the same checks and controls to their offerings as to those directly from HPE

With a solid foundation in security, HPE ProLiant servers are well-suited to provide the kinds of data protection, access controls, robust remote access, and secure transport and communications that modernized infrastructures need. That's how they can survive the cyber threat landscape, and thrive in the face of increasing numbers and severity of attacks.

In the next section of this guide, we explore the exploding, top-of-mind area known as GenAI which provides new insights, information, predictions, and innovation for organizations to put to work. In addition, GenAI supports ready automation of repetitive and routine tasks—for example, monitoring, code migration, or regulatory updates—along with data analysis and organization so that organizations can make more and better use of their data and knowledge. Chapters 6 through 8 cover business considerations, how to use GenAI to create value, and a primer on building a first GenAI app with HPE ProLiant servers powered by Intel Xeon Scalable processors.

CHAPTER 6

Initial Considerations for Business Use of GenAI

There's no doubt that artificial intelligence (AI) is taking the technology world by storm. Indeed, AI [software spending](#) is expected to double from \$33 billion in 2022 to \$64 billion by 2025, with AI-related [overall investment](#) expected to jump from \$156 billion in 2022 to as much as \$275 billion by 2025. Of that amount, spending on AI-related infrastructure is jumping from over half of that total, to an amazing \$300-plus billion by 2031 (a CAGR of nearly 30%). What the fuss is all about, of course, is that AI is already delivering on its promised impacts on business, which include faster and better innovation, increased flexibility and insights, and a shorter time-to-value or time-to-market when turning ideas into cash flow. This has spawned a market efflorescence that [The Economist](#) says has not been seen since the dot-com era.

But AI comes with a unique and formidable set of challenges. On the one hand, organizations need to identify applications or services that can benefit from the insights and content that AI can generate on their behalf. This requires careful observation of what's already available and what's most urgently needed, a profound understanding of the underlying data and an appreciation of where value can be added, productivity gained, and time-to-market accelerated. On the other hand, incorporating and integrating Generative AI (GenAI)

means adopting one or more technology stacks, identifying and preparing data for use in training and refining AI models, and standing up applications to take advantage of the power and potency of GenAI in useful and effective ways. Those are big tasks for organizations to tackle and solve so they, too, can put GenAI to work—preferably sooner, rather than later.

In this Chapter, you should find the information you need to better understand the values and costs of your own adoption, creation, and piloting of GenAI technologies. This begins with an assessment of impacts and common use cases for GenAI, along with an explanation of how HPE and Intel are uniquely suited to handle such workloads.

Leaving aside the breathless excitement and universal enthusiasm surrounding AI, it's important to understand that running GenAI applications represent a different kind and class of computing workload than many other typical IT tasks. The characteristics of GenAI workloads are tough and demanding, with a special emphasis on compute. Indeed, GenAI workloads require significant processing power to train models and generate content. Large models such as Llama3 or GPT-4o run best on specialized hardware designed to offer them the best possible speed and efficiency. Other aspects of GenAI workloads include the following:

- **Memory consumptive:** AI workloads demand substantial memory where local, fast stores for model parameters and intermediate values eat large amounts of RAM. Certain transformation-based models (GPT-4 or GPT-4o) may use hundreds of millions to billions of values: this requires substantial memory, storage, and bandwidth at great scale to accommodate heavy access to huge datasets.
- **Data dependencies:** AI models, especially those for generative AI (aka GenAI: see “Generative AI (GenAI) Defined” on p. 4) rely on reams of high-resolution data, with many processing steps devoted to data preparation to drive model training.

- **Latency limits:** Certain AI workloads, such as chatbots or voice assistants, impose strict latency requirements for acceptable user experiences. Optimizing AI models to speed response is vital, and may require careful workload placement to minimize latency.
- **Overcoming complexity:** AI workloads (especially GenAI) can be complex, and may require specialized software, hardware, and expertise for best results. This applies especially to applications like computer vision and natural language processing (NLP).

Thus, there are stringent server requirements at work for the brave new world of AI workloads. A flexible, capable architecture that works well for data- and compute-intensive tasks, even for the most demanding workloads, is what enables GenAI to deliver real value. As you'll learn in the following sections, the [Intel 5th Gen Xeon powered HPE ProLiant Gen11 server family](#) covers all those bases, and meets those foregoing needs, with dispatch and careful design—and does so at the edge, in the cloud, and in the data center.

Generative AI (GenAI) Defined

GenAI employs a model that can create text, images, sound, videos, or other data using machine learning to gain its intelligence from training on huge volumes of the kinds of data it'll ultimately be asked to emit. GenAI models learn patterns, structure, and characteristics from input training data and then create new data with similar elements in response to user queries or prompts. This can be text, images, video, or speech, depending on the training data and intended uses that GenAI models seek to support.



Many businesses (and business owners) hang onto legacy systems longer than they should. In fact, keeping outdated technology in service can fall short in providing necessary services and innovation, and reduce productivity and business value. That’s how the benefits of replacing outmoded technology more than offset the risks involved in retaining legacy systems.

Making GenAI Work for Your Business

GenAI provides methods to create—that is, to generate—new high-quality, high-level content in response to user requests (also called “prompts” or “queries”). Indeed, GenAI facilities to produce usable text, images, audio, video, animation and simulations already exist, and keep getting better. Businesses can deploy this kind of content on demand for a variety of purposes from training, to customer service, to sales support, to automate routine IT and other departmental tasks, and more. Such uses can boost productivity, and free up human resources for high-value tasks (creating new offerings, opportunities, markets, and more).

Training a model involves determining relationships between values in training data sets, and recording or encoding those determinations to establish the model’s parameters and behaviors. Inference occurs when a model generalizes from such parameters and behaviors to interpret new data it receives for inspection and analysis. The quality of that data, the interpretation and insights that GenAI supplies, also enables GenAI to add business value.

When an AI model is handling input data, it processes those inputs as they arrive in real time. A user query is compared with the information it has established during the training phase for its parameters and behaviors. Responses depend on the tasks involved, where the goal is to calculate and deliver actionable insights or results.

Because inferencing can involve expenditure of significant resources (compute, storage, power, cooling, time, and so forth), optimizing speed and efficiency during inferencing helps ensure the best user experiences and minimizes costs and consumption. This goes straight to the bottom line, and provides improvements in speed, quality, and accuracy that add value and productivity. That explains the relentless push in AI systems to speed up inferencing, reduce costs, and find efficiencies at all levels (hardware, platforms and frameworks, and application software).

Common Use Cases for GenAI

Businesses and enterprises are increasingly using GenAI to build applications, including chatbots, digital assistants, analysis and prediction tools, and developer aids of all kinds. While the possibilities are many and incredibly varied, common use cases for GenAI-based models include the following:

- **Customer service and support:** Because AI agents can listen to voice inputs, understand the underlying needs and wants that drive them, and recommend products or services, or help solve problems, enterprises can put such chatbots to ready and productive use for employees, customers, and partners. The more routine or tedious work comes off the shoulders of support staff, freeing them up to deal with more valuable, interesting, and useful activities.
- **Employee, user, or customer information navigators:** Generative AI can assist humans in finding information about benefits, insurance, and other complex offerings among which they must (or want to) choose. It can help them find and understand information resources, get better results from systems and tools, manage selection and deployment of programs, and more.

- **Creative outputs and ideas:** Generative AI helps content and other creators formulate new concepts, designs, images or videos, and all kinds of text content, based on textual descriptions (“prompts”) that draw from training data, models, and ongoing user interactions. Creators can be more productive as they can benefit from AI-based help at every step in the creative process from brainstorming, to creation, to refinement, to release and feedback collection.
- **Data analysis and interpretation:** Generative AI excels at assembling and synthesizing data to understand customer or user behavior, profiles and needs to generate content or insights for better marketing, communication, and services delivery. It holds out the possibility of fabulous user or customer experiences based on a real understanding of general and narrowly focused needs, wants and behaviors.
- **Code creation:** Generative AI helps automate coding tasks based on deep understanding of existing examples and models, common workflows across the development lifecycle, and extremely flexible, customized code that fits an organization’s security, governance, and compliance requirements and internalizes business goals and ethics.

The HPE ProLiant Gen11 server family uses the latest 5th Gen Intel Xeon processors. Each of the foregoing use cases draws on advanced AI inferencing, analysis and outputs, actions or recommendations designed to deliver optimal and speedy performance that’s also energy-efficient.

Why HPE and Intel Excel at GenAI Workloads

When it comes to GenAI computing, more is not simply better—abundant but affordable resources and capability are essential for providing positive user experiences and manageable TCO. HPE ProLiant Gen11 servers deliver more and faster graphical capabilities, enabling organizations to innovate using the advanced GPU accelerators and the ultra-scalable architecture that GenAI workloads need. By deliberate design, the HPE ProLiant Gen11 server family powered by Intel supports up to one-third higher GPU density, with increased flexibility for GenAI workloads at the edge (chatbots, data analysis for IoT and other local resources, loss prevention based on video surveillance, and other latency-sensitive uses) and in the data center (model creation and training, coding, creative support, and more).

Organizations that deploy HPE ProLiant Gen11 compute also benefit from a more capable resulting IT infrastructure. These servers bring higher efficiency, improved scalability, and better economics to speed and enhance business outcomes while lowering TCO. Higher density also means more workloads run in the same rackspace (server footprint) for improved ROI and better performance for GenAI workloads.

Using HPE ProLiant Gen11 servers with Intel 5th Gen Xeon processors also helps organizations control costs through rightsizing and scaling on-demand. Indeed, GenAI workloads run faster on HPE ProLiant Gen11 clusters with predictable, completely visible costs. You can scale capacity up or down on-demand—using a next-generation architecture with continuous monitoring—to right-size or even add burst capacity on-site. All this adds up to improved GenAI performance and capability.

Even better, HPE makes the operating experience intuitive (just like the cloud). You can simply use a single set of controls from edge to cloud via the Intel powered HPE ProLiant Gen11 cloud computing

experience. This enables fully digital business operations and transformation, with global visibility and insight through a singled unified console with HPE GreenLake for Compute Ops Management. Organizations can swiftly and painlessly automate tasks for efficient deployment, simplified support and lifecycle management. This helps GenAI workloads integrate seamlessly into existing enterprise operations.

Finally, the HPE ProLiant Gen11 Server family with the latest Intel Xeon processors embodies Trusted Security by Design. It is built to protect your infrastructure, workloads, and data from threats to hardware and third-party software. It does this through a trusted edge-to-cloud security approach that's based on HPE's silicon root of trust, trusted supply chain, zero trust hardware and software technologies. That means reduced security risks and exposures, and better data protection, for your advanced and data-heavy GenAI workloads.

In the next chapter, we switch gears from surveying the business considerations and requirements for GenAI to take a look at how businesses and organizations can extract value from using GenAI.

CHAPTER 7

Getting Value from GenAI

GenAI has been setting the headlines ablaze, as well as firing a multitude of business strategies and projects. Organizations, both public and private, are seeking earnestly to get ahead of the curve when it comes to GenAI. For those not already in the know, GenAI is an area of artificial intelligence (AI) that uses leading-edge machine learning (ML) techniques to generate content or data. What gets generated depends on the specific training data that drives ML. These days, that includes text, images, animation, 3D models, video and more. From a carefully curated and managed collection of training data meant to model some reality sufficiently to make it accurate and convincing, generated content is meant to be useful, engaging, and meaningful.

Compute matters tremendously for making GenAI successful, because the models that drive it are huge and complex. This requires enormous amounts of compute horsepower, when a model may include millions of parameters (or more) and involve billions upon billions of data objects for training up front. Then, when a model is determined to be sufficiently accurate and representative of reality, it ingests even larger volumes of real-world data to generate content in keeping with its carefully trained-up view of reality. That view keeps

changing and getting refined over time as it becomes more accurate and informed by the data it handles and creates, and as the input stream changes to reflect changes in the real world.

GenAI models come with a variety of value-adds they can bring to organizations that deploy them. These include the following and are closely tied to the kind of data used for training and expected by way of output:

- **Automated content creation:** With the right training data to drive it, GenAI can produce text, graphics, videos animations, and more. Be it for summaries, reports, help files, and so forth, AI can handle routine or repetitive tasks that opens an opportunity for people to be more productive and pursue innovation. Likewise, GenAI can produce images, clip art, diagrams and timelines, plus other visual media that saves human time and effort.
- **Improved customer support:** GenAI drives chatbots that can use customer history (individual and in the aggregate) to provide status, answer questions, and speed customers through quality interactions and encounters. Human agents can focus on complex topics, solve problems, and build better customer relationships.
- **Optimize workflows:** GenAI can help with automating data entry, assembling more efficient delivery routes, organizing pick and pack assignments, and tracking task progress and completion. This is how GenAI-driven solutions can reduce manual effort, steer faster completions, and boost productivity.
- **Drive innovation:** GenAI offers insights and recommendations that can do more than optimize and improve existing processes and methods. From the huge volumes of data it ingests, GenAI can identify patterns to suggest new approaches to existing products or services and cultivate new ideas. These can lead to new products (or new product categories), suggest marketing strategies, and help elaborate new business models.

Judicious application and careful use of GenAI can provide organizations with better ways to do what they already do, and help them find, develop, and deliver innovations and new opportunities.

The HPE ProLiant Gen11 server family has solutions engineered specifically for GenAI f, at the edge and in the data center, and in the cloud. It offers breakthrough performance, and ultra-scalable architectures with more graphics capabilities than ever before available. This open-ended architecture lets you grow your business at scale with adherence to industry standards, with the flexibility to incorporate and accommodate new technologies, and innovate wherever your apps and data live.

Harness GenAI and Make It Your Own

There's a big learning curve involved in putting GenAI to work in an enterprise or organization. That's why HPE's approach is to help enterprises to be able to better take advantage of a number of common use cases for which GenAI solutions are ready to jump right into AI inferencing. A pre-configured solution saves developers the time and effort involved by utilizing a pretrained model (e.g., Llama-3, Nemotron-3 Mistral, and other open source models) with ready-to-deploy toolsets, frameworks, and optimized infrastructure with necessary GPU performance. This enables enterprises to skip the training phases which is considered computational-intensive and costly while enabling them to focus on fine-tuning and utilizing RAG with inference to apply domain knowledge for GenAI applications (e.g., edge- or data center-oriented).

HPE also provides a broad range of server platform choices that include 5th Gen Intel Xeon processors that range from pay as you go offerings to standalone rack-mounted 1U and 2U configurations suitable for on-premises data centers or edge deployments. Indeed,

all these can work together seamlessly through managed services and capacity—including scale-up and scale-out—to support a demand-driven, consumption-based payment model for excess or peak demand capacity. This comes in the context of a single, coherent management process that handles assets and resources alike from edge to data center to cloud, as needed. Then, organizations can take the fullest possible advantage of what HPE and Intel have to offer by way of servers, services, management, control, and flexibility.

Are You Ready to Benefit from GenAI?

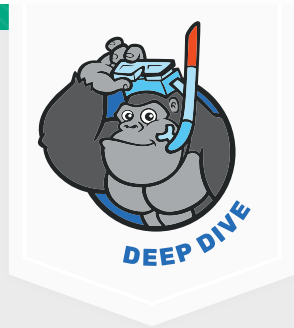
Indeed, organizations must explore and evaluate infrastructure options to get the most out of generative AI deployments. In practice this means identifying and piloting cases at the edge, in the data center, and in the cloud. Some deployments make most sense in the data center for reasons related to security, IP protection, and sensitivity of data, algorithms, or models.

Other situations may require edge deployments to optimize end-user experience by minimizing latency and response times: This is very much the case for computer vision where GenAI uses live security camera feeds as the basis for actions, notifications, or evidence collection. This gives them the flexibility and capacity to accommodate wide swings in uptake and use, or geo-graphical flexibility when latency and customer/user proximity are both important (e.g., in a “follow the sun” pattern that tracks daylight hours around the globe).

Generative AI is both interesting and challenging because it requires understanding AI workloads and how best to handle them. GenAI in particular, and AI inferencing in general, can require tons of research, trial deployments, and mid-course adjustments to get things “just right.” For organizations on the fast track, it’s worth noting that HPE’s various GenAI solutions ship ready-to-run on a variety of HPE

AI/GenAI Are Compute Intensive!

Though it might seem that complex models trained up on reams and reams of data is the only way to practice AI, it's just one sub-field in a large, complex undertaking that seeks to use computers to build representations of things so they can be understood, changed around, refined, or extended. But no matter where AI pops up, count on it to be resource-intensive, for compute. That's why you'll find the most demanding GenAI applications running in data center server clusters all the way up to the biggest high-performance supercomputers.



ProLiant Gen11 servers with 5th Gen Intel Xeon CPUs and appropriate NVIDIA edge- or data center-oriented GPUs. They're also pre-trained for specific use cases, too, so ramp-up and training times are seriously shortened.

In particular, organizations are likely to face certain specific challenges when working toward GenAI deployments:

- 1. Leveraging private data securely:** Intel-powered HPE ProLiant Gen11 takes a zero trust approach from the silicon level all the way up to applications and services under the GreenLake Compute Ops Management umbrella, backed up with a secure supply chain and a never-trust/always-verify approach to firmware, access controls, and more.
- 2. Optimize infrastructure for maximum AI advantage:** this means providing high performance as and when it's needed, with the ability to manage latency and meet stringent customer experience and SLA requirements.

- 3. Stand up and run infrastructures quickly and securely:**
HPE's pay as you go model means management comes as a natural consequence of accessing and using its tools: there's no set-up, no staging, and no added complexity. Hook it up and put it to work!
- 4. Organizations can work from the pre-trained models**
to shorten the development cycle with optimized tools, platforms, and so forth. HPE's pre-defined use cases represent real value for organizations, with a shorter, smoother time-to-value backed up by expert consulting services to further shorten that interval.

Other Key HPE ProLiant Gen11 Factors: Business-Critical for Today's Enterprise Customers

Beyond its robust support for GenAI and a set of specific use cases, HPE ProLiant Gen11 servers powered by Intel all share other important characteristics in common. Broadly speaking, these fall under the headings of secure lifecycle management and “as-a-service” offerings based on HPE GreenLake Flex Solutions. Each of these is discussed here:

- HPE ProLiant Gen11 employs **security by design** that extends all the way from a silicon root of trust, with firmware and controller checks at start-up and runtime to make sure systems remain tamper-free. Even construction and delivery are covered, thanks to the company's secure supply chain, available globally through the [HPE Server Security Optimized Service for HPE ProLiant](#)). In fact, HPE puts zero trust to constant use as well, so that all access requests are constantly checked and verified (“never trust—always verify”). Through HPE

GreenLake and its Compute Ops Management (COM) facility, HPE provides security protection for software at the OS, platform, and application layers, with APIs to provide organizations and partner with direct, secure access as well.

- **Lifecycle management** comes to HPE ProLiant Gen11 servers, courtesy of various consoles and management tools. These include the server-at-a-time Integrated Lights-Out (iLO) management facility, but also HP OneView and especially the veritable “single pane of glass” that HPE GreenLake COM provides for servers at the edge and in the data center ... and beyond.
- **Pay as you go offerings** via HPE GreenLake Flex Solutions bring organizations a veritable cornucopia of cloud-based capability. Thus, they can extend their own platforms at the edge and in the data center using consumption-based IT solutions built upon standardized, centrally managed IT modules that create extensible IT infrastructure building blocks in the cloud. Consumption-based pricing, plus predictability and reliability, gives organizations what they need to scale out GenAI solutions as and when they need extra capacity or capability.

Organizations can put Intel-powered HPE ProLiant Gen11 to work in either centralized data centers or distributed edge locations with great results and simple, unified, automated management (through GreenLake COM). This works especially well for cloud-native workloads, so organization can scale up and scale out, take advantage of infrastructure as code, and achieve agility in fast-moving, fast-changing usage situations.

Why HPE and Intel Can Help Your Advance Your Initial GenAI Moves

HPE ProLiant Gen11 servers with Intel 5th Gen Xeon processors are engineered for peak performance especially for compute-intensive applications like GenAI. That means advanced graphics rendering, data acceleration, large language models, computer vision, and speech and language processing all work well on such servers. They can bring compelling economic returns to organizations, whether GenAI apps run in the data center or at the edge.


HPE's other inferencing solutions include specific elements tailored for different AI applications, with Intel Xeon Scalable processors and NVIDIA edge or data center GPUs to handle such workloads. Available offerings focus on:

- Computer vision AI at the edge, in a compact design for up to 4 edge-oriented GPUs, plus pre-packaged applications for working with cameras in real time for loss-prevention, smart spaces, and safety.
- Natural Language Processing (NLP) provides optimized support for applications that include speech AI and conversational AI, also supported through data center-oriented GPUs, coupled with other optimized HPE ProLiant Gen11 servers with Intel Xeon processors.

Overall, HPE ProLiant Gen11 servers with Intel 5th Gen Xeon processors offer wide-ranging, highly capable and efficient solutions for enterprises seeking to stake out their spot in the arena of GenAI (and similar applications). With enhanced security, reliability, and industry-leading performance as well, these servers excel at handling demanding, compute-intensive AI workloads. Dig deeper into the economics, sustainability, energy efficiency, and supply chain,

and you'll understand how HPE ProLiant Gen11 Intel-powered servers enable organizations to remain agile, manage IT challenges, and focus on long-term innovation and market development.

How HPE ProLiant GenAI Solutions Meet Challenges



By design, HPE ProLiant Gen11 solutions with Intel 5th Gen Xeon CPUs for GenAI are carefully designed. They're built to leverage proprietary data so organizations can obtain unique and lasting value from that data. GenAI solutions will help organizations obtain and use the insights, predictions, or recommendations that get generated.

HPE ProLiant Gen11 solutions, including HPE GreenLake and its COM, lets an organization own and control its entire AI compute stack from end to send. That makes HPE ProLiant Gen11 AI solutions tuned, secured, and certified for edge or data center deployments, where the organization owns and runs everything. Expert consulting and management are available from HPE to assist (or take over) as circumstances dictate. This gives the organization complete control over its own IP.

HPE ProLiant Gen11 servers with Intel 5th Gen Xeon CPUs and NVIDIA GPUs are purpose-built and certified to support an ultra-scalable, ultra-efficient, and ultra-secure platform for AI workloads, across a broad range of use cases. This ensures organizations get the performance they need to keep users happy and provide a worthwhile return on their technology investments.

HPE ProLiant Gen11 offers AI solutions that are tested and optimized by HPE and Intel and their use case partners. Three families of use cases help eliminate the cost and risk typically needed during the training phase, enabling enterprises to fast track computer vision, GenAI, and natural language processing. This means organizations

spend less time researching, sifting through alternatives, and building up their GenAI environments. That translates into shorter time-to-value and a quicker payoff on the buy-in.

Finally, HPE ProLiant Gen11 with Intel 5th Gen Xeon CPUs can transform the operating model for IT with HPE GreenLake. This lets organizations access advanced HPE ProLiant servers in a cloud-like, consumption-based environment. In turn, this adds the advantages of on-premises devices to the flexibility, scalability, and accessibility of the cloud. A service-focused approach supports global management at the edge, in the data center, and in the cloud with pay-as-you-go pricing to manage costs and ever-shifting demand. Organizations can use as much or as little technology as they need, and make sure it pays for itself to meet financial targets.

In the next chapter, we'll walk through the steps and considerations involved when evaluating and prioritizing pilot GenAI projects. We'll also explain how HPE and Intel can help with hardware, software, and consulting services to make pilots happen in short and affordable order.

CHAPTER 8

Build Your First GenAI App with HPE ProLiant Gen11 Intel Xeon Servers

As we've already mentioned, GenAI is reshaping the way enterprises think, innovate, and work across the entire business landscape. Enterprises of all kinds and all sizes are customizing large language models with their own data to support a wide range of GenAI applications from new customer experiences to enhancing employee productivity. This enables them to better handle their unique business needs and situations and empowers AI applications across all aspects of their operations—not just in IT, but in procurement and supply, HR, sales and marketing, manufacturing and logistics, and more.

Building a first GenAI application comes with plenty of interesting challenges. IT operations must learn how to work with hardware and software, specific technology stacks and various frameworks, platforms, and models to support AI operations. HPE ProLiant servers with Intel 5th Gen Xeon processors provide an efficient and high performant foundation that enables enterprises to start small and grow large while maintaining control over data and costs.

Choosing a First GenAI Application

Some organizations start modestly, with one or perhaps two GenAI projects to get the ball rolling. This gives developers, operations, and stakeholders a chance to learn about the technology and what it can do, hopefully through some easy wins to gain interest in and support for further, more significant efforts.

Given the right strategy and approach, research¹ shows that a pilot can launch in as little as a calendar quarter (give or take a month). That strategy can be outlined in the form of a six-step plan, as follows:

- 1. Identify initial use cases:** Look for repetitive, data-driven, or predictable step-oriented tasks like content creation, summarizations and task automation. These are most likely to benefit from applying GenAI to speed completion and boost productivity. Try to start with something already known, where you can work to see measurable results
- 2. Enlist stakeholder support:** Team up with a department or another sponsor to find an AI project to make somebody's job easier. It might be automating a manual task, or combining multiple individual tasks done in sequence into a single task. Work with your partners to agree on what to deliver, then do just that. Get executive sponsorship and buy-in to make sure your project goes forward, fits the overall digital strategy, and is recognized for value.

¹ Gartner Report: [How to Pilot Generative AI](#), July 10, 2023

- 3. Gather and prepare quality data:** Data is what drives AI and determines its relevance and usefulness. Before building any GenAI model, you should collect and groom data to inform the model for proper quality, accuracy, and quantity. Look at the process, document incoming data flows, identify people involved, and learn which systems, applications, and processes touch that data across your organization. This helps illuminate the data lifecycle, and where and how it's stored.
- 4. Find SMEs:** Subject matter experts (SMEs) are the people who regularly work with and best understand the data and operations that you're trying to automate or improve. They're the ones who really know what's going on. Typically, they'll be highly competent and busy. Indeed, only a single data scientist or GenAI expert is plenty for pilots as they work with IT staff/developers and SMEs to define and prioritize use cases, prepare data, build and train simple models, and get ready to pilot their efforts. HPE consulting services can help with staffing if you have no local expert on tap.
- 5. Operationalize a GenAI model:** When the preceding steps are done it's time to deploy into production. When this happens, elements to monitor, manage, govern, and analyze the model's results are needed to make sure it's working and doing what it needs to do. Early on, a full-blown management system may not be necessary (though it will be required once you go past the pilot stage). Instead, get your experts to manage the model from end to end. Make sure it's up-to-date, check that there's no model drift, make sure that responses or results make sense and make a (helpful) difference. This shows an understanding of AI, exposes its value, and makes it have genuine value.
- 6. Select your partner:** HPE. Once you've built a pilot or three and get things into production, you will better understand the systems, workflows, software, and data that models need. This means it's now time to

find products and platforms to help meet your GenAI needs for the years ahead (typical planning horizons for GenAI run three to five years). This helps speed model development and training into production. HPE can also provide a unified data foundation that spans hybrid cloud environments from the edge to the data center and into public and private clouds. HPE even offers AI and data transformation services to help companies get started on this journey and find their way to successful outcomes.

Corresponding HPE ProLiant Gen11/12 Servers and GPU Options

Intel and HPE have partnered up to architect specific solutions platforms for typical cases at the edge and in the data center: Each of two possible server packages is described in the sections that follow; both fall under HPE GreenLake for Compute Operations Management (COM) for simplified and automated control of your compute infrastructure in the data center or at the edge.

HPE PROLIANT DL320 GEN11 SERVER FOR AI AT THE EDGE

This 1U single socket server offers the kind of compact, energy efficient yet capable server package that organizations need to position themselves at the network edge for a class of applications where minimizing latency is key. These include loss prevention, site security and surveillance, IoT access, factory floor automation, and more.

Key elements of the DL320 architecture include:

- 5th Gen Intel Xeon processor
- Up to 2TB DDR5 RAM

- Up to 10 small form factor SSDs for solid state storage
- Either two double-wide or 4 single-wide GPUs for AI workloads, i.e. up to 4 NVIDIA L4 GPUs
- Choice of networking interfaces (ask your HPE Partner or sales rep for details)

HPE PROLIANT DL380A GEN11 SERVER FOR THE DATA CENTER

This 2U dual socket server offers support for up to two high-end 5th Gen Intel Xeon scalable processors. These rack-mounted servers are ideal for data center deployment and offer an ideal combination of small footprint and energy efficiency for maximum use of cubic volume. They are ideal for resilient, available and capable access to AI, compute and graphics-intensive workloads.

Key elements of the HPE ProLiant DL380a Gen11 architecture include:

- 5th Gen Intel Xeon scalable processors
- Up to 3TB DDR5 RAM
- Up to 8 EDSFF E3.S 1TB SSDs
- Up to 4 double-wide or 8 single-wide GPUs for the most demanding AI workloads, i.e. up to 4 NVIDIA H100 or L40S GPUs
- 4-port GbE OCP3 adapter or 10GbE 2-port adapter (other options available in some markets—ask your HPE Partner or sales rep for details)

HPE PROLIANT COMPUTE DL380A GEN12 FOR THE DATA CENTER

This 4U dual socket server offers support for up to two ultra-efficient Intel Xeon 6 processors, plus up to eight NVIDIA H200 NVL Tensor Core GPUs. These scalable, rack mounted servers aim squarely at large-scale deployment of GenAI-based workloads for scale-up

fine-tuning and massive inference workloads. It can also deal with mixed workloads at scales associated with High-Performance Computing, including simulations, weather forecasting, high-end 3D modeling, protein folding and genomics, and more. The overall emphasis is on maximum performance, efficiency and reliability at scale for production data center use.

Key elements of the HPE ProLiant Compute DL380a Gen12 architecture include:

- Up to 2 Intel Xeon 6 processors
- Up to 4TB of DDR5 RAM
- Up to 8 NVIDIA H200 NVL Tensor Core GPUs
- Choice of networking interfaces (ask your HPE Partner or sales rep for details)
- Up to six dedicated and redundant power supplies for GPUs for efficient, reliable performance

At the edge or in the data center, HPE ProLiant Gen11 and HPE ProLiant Compute Gen12 servers with Intel processors offer tremendous value, terrific energy efficiency, and best use of rack space for capability delivered.

HPE Services Does AI for Customers

Don't forget that HPE also offers an [AI Services](#) arm as part of its HPE Services operations. This group of experts can help make the most of AI's potential, which it promises but sometimes fails to deliver. HPE AI Services can find, provide and support the right expertise, data, tools, technology, and partnerships to make sure your AI efforts come out on the winning side. HPE Services personnel can handle the whole lifecycle, from planning, to implementing, and deploying the right

AI solutions. They'll bring their data science, ML engineering, and ML operations (ML Ops) experience and knowledge to bear so you can assess and capture business expectations, meet technology and performance requirements, and prepare data for training and deployment safely and securely.

FROM GENAI TO KEY BUSINESS WORKLOADS

GenAI offers some of the most interesting and exciting opportunities for organizations to make effective and profitable use of a modernized infrastructure around. Certainly, markets have been exploding with new platforms, models, chatbots, and AI-driven applications for everything from customer service to user account and application set-up, provisioning and maintenance. Behind the scenes, there's a strong and growing requirement to do more, faster and to bring as many execution and memory resources to bear as possible.

Indeed, modern workloads are nothing if not extreme in their requirements for compute, memory, storage, and networking resources. In the next section of this guide, we'll dig into the requirements and resources needed to support key business workloads, including those related to data analytics, extremely large databases, and VMware virtualization using VMware vSAN and the Express Storage Architecture (ESA). Along the way, you'll gain a better understanding of just how HPE ProLiant servers powered by Intel Xeon Scalable processors are tailor-made to handle such workloads—and more.

CHAPTER 9

Delivering Data Analytics

Modern business has seized on data as a key source for insight, innovation, and understanding of how things work, where they are going, and what customers want. If some data is valuable, more data is still better, and maximum acquisition, analysis, and use of data is best of all. That helps to explain why computing resources focus relentlessly on acquiring and using ginormous data volumes, a lot faster, and with as much efficiency, capability, and capacity as technology allows.

What does this mean, when it comes to processing massive, complex workloads with panache and dispatch? Consider these characteristics typical in modern business workloads:

- Billions of values and parameters, enormous datasets, and complex analyses that often include trillions of operations to calculate and transform values for modeling, simulation, or rendering.
- Extreme processing requirements from CPUs and other silicon adept at processing large, high-volume, high-precision floating point calculations and operations (e.g., GPUs and related software for parceling out highly parallelized processing tasks).

- Extravagant demands for data storage and transfer, from the terabyte to the exabyte scale (that represents from 10^{12} to 10^{18} data items, any of which can involve gigabytes of data: that takes us to the zettabyte level or 10^{27}).

Types of workloads with such characteristics include data analytics and modeling, 3-D imaging and modeling, complex simulations (weather forecasting tools like the U.S. and European hurricane/storm prediction toolsets; protein folding; and genomics), video rendering and editing, cryptocurrency mining, and the software that supports cloud-scale virtualization, orchestration, monitoring, and management (e.g., public clouds such as AWS, Microsoft Azure, Google Cloud Platform, Oracle Cloud, IBM Cloud, and so forth; and SaaS or private cloud platforms like HPE GreenLake, VMware, SAP, SAS, Cisco Cloud, and others).

KEY BUSINESS CHALLENGES FOR MODERN BUSINESS WORKLOADS:

Here's a sample of the issues that businesses and organization face as they seek to run modern, resource-intensive business workloads like the ones described in the previous paragraph:

- Data analytics requires scope and space for unthinkable quantities of data.
- Seamless integration is needed between data and applications on-premises and in the cloud, with scale-up and scale-out for on-demand cloud bursting, redundancy and failover, content delivery service staging, and more.
- Data remains incredibly valuable, and increasingly constrained by rules and regulations around privacy, confidentiality, sovereignty, and more. Governance can be tricky when multiple sets of rules apply, depending on context and location.

- Powerful, automated management and monitoring is needed to drive all aspects of big data from model development and ongoing analytics to refinement and lifecycle stages.

In short, just as there's a huge amount of stuff going on, with lots of moving parts and a shifting landscape of rules, regulations, and best practices, there's a lot to think over, and to handle, to make things work safely, securely, productively, and efficiently.

In this chapter, you'll obtain insights and information you need to modernize and maximize your data solutions, analytics, and visualizations. This begins with a discussion of workload characteristics and associated resource requirements, with an emphasis on various kinds of unstructured data typically ingested for data analysis, often in NoSQL databases, Hadoop, and Cloudera.

As we drill down into the kinds of workloads that qualify as demanding, resource intensive, and worth accelerating to expose and use the insights they can deliver, data analytics nearly always comes out at or near the top of this category. What is it that makes data analytics so important, if not vital, to businesses and organizations? In two words: data and the interpretations or actions that its analysis encourages and enables.

Above all, there are two fundamental characteristics one can find in all data analytics workloads, without having to look either very hard or very far—namely:

- **Huge data volumes:** Global climate modeling, healthcare and genomics, financial services (fraud detection, risk management, and algorithmic trading), and large-scale telecommunications (deliver best QoS, manage bandwidth, create new communications tools and technologies). These kinds of use cases already involve zettabyte scale for data analytics today.

- **Access to standard platforms and tools:** Given the volumes of data and the complexity of analytics needed, companies and organizations turn to industry standard platforms and toolsets to help them gear up and get running quickly, securely, and accurately. Among others, HPE ProLiant servers with Intel Xeon CPUs support the Hadoop Distributed File System (HDFS), Ranger (an Apache security administration tool for Hadoop ecosystems), Atlas (an Apache data governance and metadata management tool), Hive (an Apache data warehouse infrastructure tool), and Hue (an open source SQL assistant for databases and data warehouses that can query, explore, and visualize data).



Coursera defines data analytics as “... the collection, transformation, and organization of data in order to draw conclusions, make predictions, and drive informed decision making.” The online learning and professional development company is careful to distinguish data analysis—which involves “extracting meaning from data”—from this activity, because it involves both data science (using data to theorize and forecast) and data engineering (building robust and powerful systems that can handle the huge volumes of data through which data analytics chews routinely).

Thus, data analytics involves analysis techniques from pure and applied mathematics, statistics, and computer science to tease out and assemble insights from large data sets. That explains why Coursera says that data analytics “includes everything from simply analyzing data to theorizing ways of collecting data and creating the frameworks needed to [store that data].” It’s a big, multi-disciplinary field that covers lots of ground—and data!

Thus, HPE ProLiant servers with Xeon CPUs can not only ingest and accommodate the data types and volumes, including unstructured data, that data analytics needs, it can also support the infrastructure and tools needed to set up, control, secure, and manage such data. This even includes the means to assure that the data is subject to governance, best practices, and proper protection. Indeed, that covers data lineage (tracking sources of origination and changes), auditing (tracking access and use), and search (tracking queries and data provided in return).

The work that goes on to support data analytics only underscores the idea that data engineering is a serious job, and involves careful, thoughtful approaches to ingesting, using, and storing data across the entire lifecycle. This involves a variety of activities that businesses and organizations find useful, including:

- **Developing and serving up predictive models:** Once the data scientists create such models, it's up to the engineers to make them available and usable. HPE's support for the preceding tools and platforms makes this safe and straightforward.
- **Setting up a data mart:** A task that involves data engineers, database admins, business analysts, and interested stakeholders, a data mart provides a way to access the data sources, data types, and reporting or prediction/forecast outputs to meet business needs.
- **Consuming data from services:** This means extracting data from source systems such as databases, flat files or external APIs (which talk to other systems), transforming that data to fit the data mart schema (usually means clean-up, transformation, validation, deduplication and more).
- **Making data available for consumption:** Loading cleaned-up, validated data that's ready for internal analytics systems and services to access and use on demand.

The Private Cloud Base

HPE and Cloudera have worked together to design hardware and software configurations to deliver optimal computing performance for Hadoop-based systems and applications. Built around HPE ProLiant Gen 10 and 11 servers, HPE offers a [Reference Architecture](#) for Cloudera's Data Platform Private Cloud Base (PCB) that's been tested and vetted as "ready-to-run" with this popular data analytics platform and service.

Available consoles and instrumentation in this reference architecture also make it easy for businesses and organizations to manage Hadoop clusters and configurations, and to monitor cluster performance to deliver optimal user experiences within this environment. Management capabilities encompass managing installations, upgrades, maintenance workflows, encryption, access controls, data replication, and more. This ensures that all aspects of setup, use, maintenance, security, and recovery are covered and ready for prime time. The reference architecture also supports the entire collection of Apache elements for the Hadoop ecosystem. In addition to the items already mentioned (HDFS, Ranger, Hive, and Atlas), you'll also find support for YARN, MapReduce, Hadoop Common, Impala, Pig, HBase, Spark, Oozie, Ozone, Zookeeper, and more.

The HPE ProLiant servers powered by Intel Xeon CPUs deliver great capability and value by balancing compute, memory, and network bandwidth. See the [Reference Architecture](#) document for specifics on HPE ProLiant makes and models, which includes the HPE ProLiant DL380 (Storage/Compute Blocks) and HPE ProLiant DL360 (Control Block).

This concludes Chapter 9, which provides some details on how HPE ProLiant servers help businesses handle the most demanding workloads—including use cases like 3-D models, simulations, financial services, genomics, and more—with ease and dispatch. It's time to move onto the next waypoint. Chapter 10 will explain how those same servers can tame (and expedite) even the most monstrous of database services and applications.

CHAPTER 10

Supporting Monster Databases

Managing databases across an enterprise involves more than “mere” database design, provisioning, and upkeep. It also means managing licenses and instances, and trading on-premises costs and controls against one or more private and public clouds (including vendor-specific clouds like those from Microsoft for SQL Server and Oracle for its massive database portfolio). This can be demanding for servers not just because of the gargantuan data volumes involved, but also because of the tricks and traps inherent in setting up and managing distributed database environments across multiple sites, and on private and public clouds. Things can get overly interesting in a hurry for those not prepared to embrace complexity, and to accommodate the crazy big and demanding workloads on which modern distributed database environments thrive and prosper.

Reduce Costs and Improve Performance



Optimizing user experiences and managing databases in a hybrid cloud environment is easier with HPE, thanks to the HPE Database Performance Profiler (DPP). This tool lets organizations understand

and optimize their database workloads and run them on the right infrastructure mix to reduce costs and improve performance. The HPE DPP enables you to:

- Gain insights into workload and usage patterns.
- Consolidate server infrastructure and increase compute capability on fewer, well-placed CPUs.
- Optimize database licensing to reduce overall associated costs.
- Manage server and storage footprint.

The HPE ProLiant DL360 Gen11 server is ideal for high-volume, low-latency online transaction processing (OLTP) workloads. Indeed, its Intel CPUs (2 processors, each with up to 64 cores) support millions of transactions per minute on Microsoft SQL Server 2022 without swamping those processors. High-performance NVMe drives in the [HPE Alletra](#) storage array enables this configuration to deliver the high transaction rates with very low latency.

The HPE ProLiant DL360 Gen11 Server powered by Intel Xeon Scalable processors is highly scalable and compute-dense, which makes it ideal for high-performance database workloads. It is a rack-optimized, 1U 2P dense solution that can provide exceptional compute capability, upgraded high-speed data transfer, and memory depth at 2P compute levels. A silicon root of trust anchors server firmware to an HPE-specific ASIC, to create an immutable fingerprint for the CPU and system components that must match exactly with what's discovered before the server is allowed to boot. Visit HPE's [ProLiant DL360 product page](#) and click the Specifications tab for a complete list of its components, storage, software, power, and other capabilities, which include up to 20 EDSFF EXes NVMe SSD, plus 2 M.2 NVMe SSDs for resilient boot/system coverage.

Microsoft SQL Server 2022 Feature highlights

Businesses and organization seeking database satisfaction have lots of reasons to give Microsoft SQL Server 2022 a second—and even, a third—look. These include the following outstanding DBMS features:



- Intelligent query processing and query store simplifies and speeds database interactions, especially repeat maneuvers.
- Ordered clustered columnstore index speeds queries, offers better compression, and supports faster data load times.
- Improved memory management in large memory servers also boosts performance.
- Enhanced spinlock algorithms deliver improved scalability (improved concurrency and better use of CPU resources, faster access to data structures, and better performance under load (reduced contention and latency).
- Instant file initialization brings faster file operations, improved performance, and better handling for transaction logs.

All these changes, and others not covered here, mean that SQL Server 2022 runs faster and more reliably at scale. For more information, please consult Microsoft Learn's What's New in SQL Server 2022.

HPE Alletra 6000 is a perfect storage solution for high-performance, high-volume database access. This array lets IT add on-demand, as-a-service storage to its existing choices. Alletra supplies intelligent, resilient storage with no single point of failure to provide guaranteed, 6-nines availability. Back this up with recovery SLAs for fast,

integrated app-aware backup and recovery on-premises and in the cloud, including major DBMS environments (Microsoft SQL Server, Oracle DB and MySQL, and more).

Alletra is AI-driven to predict and prevent disruptions to storage, services and VMs. It includes predictive support automation with automated level 1 and level 2 support services, with direct access to critical resources via automated escalation. Best of all, HPE Alletra works as a consumable, data infrastructure as-a-service through HPE GreenLake.

Now that you've learned how HPE ProLiant servers help businesses and organizations make the most of massive databases for inquiry, analysis, transaction processing, analytics, and so forth, it's time to step further down the trail. In Chapter 11, you'll appreciate how those same HPE ProLiant servers help improve things for those out-fits where virtualization happens through VMware, in ways both big and small.

CHAPTER 11

Use Virtualized VMware Storage

In a rather more circumscribed use case, some organizations and businesses may want to make use of HPE ProLiant Servers for VMware vSAN ReadyNodes. If they're already on the VMware bandwagon, and looking for high-performance, high-quality, and high-reliability solutions, they can run on HPE ProLiant servers powered by Intel. These operate as a trusted, high-performance platform, using the vSAN storage kernel. Together with 10 and 25 GbE-based networking (options up to 100 GbE are available) these servers have the speed and connectivity to handle the high levels of network traffic necessary to support vSAN. Configurations built around ProLiant Gen11 servers powered by Intel are already listed in the [VMware Compatibility Guide aka VCG](#). All such HPE ProLiant servers work through the new VMware vSphere Foundation or the VMware Cloud foundation. **FIGURE 3** shows the relevant ProLiant Server models, based on the latest Xeon generation (appears as "Sapphire-Rapids-SP").

All these various units also include new DDR5 RAM, which offers higher speeds and more concurrent memory channels for access to further boost server performance and speed I/O along. Host Bus Adapter (HBA) options for these servers use all-flash SAS or SATA drives for those who want them in vSAN Original Storage Architecture

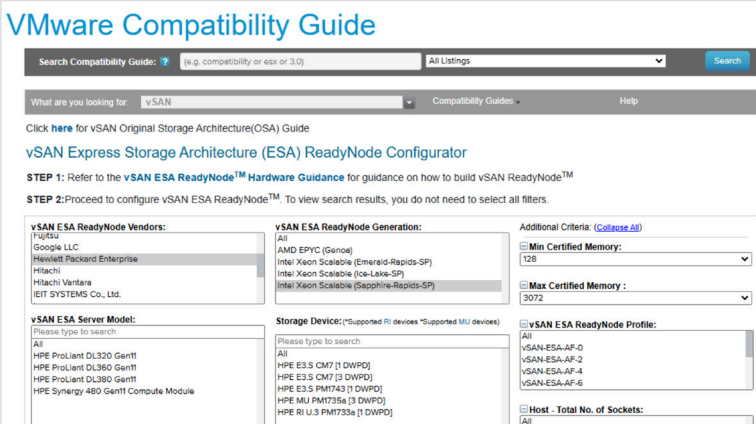


FIGURE 3: Relevant HPE ProLiant Gen11 model numbers for VMware vSAN include DL320, DL360, and DL380, as shown

(OSA). Other options are available for all-NVMe builds to design newer-technology VMware vSAN Express Storage Architecture (ESA) ReadyNodes. Better yet, ESA ReadyNodes from HPE meet compliance for storage and platform compatibility performance standards for operation in VMware vSAN. The VMware Compatibility Guide is the best tool to steer you to the exact configurations that are pre-vetted for such use.

Benefits of VMware vSAN ReadyNodes

Many benefits of vSAN ReadyNodes come from these turnkey solutions to help businesses accelerate their journey to implement software-defined storage (SDS). Buyers who choose the HPE ProLiant

Gen11 servers powered by Intel Xeon CPUs will understand they can benefit from certain characteristics, including:

- Easy to order and faster time to market (HPE is a single-source vendor and controls server builds from inception through delivery, choosing from prescribed configurations speeds time to market and deployment).
- These HPE ProLiant Gen11 servers powered by Intel Xeon CPUs are designed to be simple to install, set up, and manage, with a single coherent and consistent management UI for physical and virtual environments.
- Lower costs come from highly efficient HPE ProLiant servers powered by Intel Xeon CPUs that deliver high energy efficiency and industry-leading performance.
- Solid security from HPE ProLiant Gen11 servers powered by Intel Xeon CPUs deliver a 99.8% detection rate for attacks and exposures, with 100% malware protection across all apps, clouds, and devices. Such design boosts network security sufficiently to require 60% fewer traditional firewalls for equivalent protection.

Bottom line: the [VMware Compatibility Guide](#)'s validated HPE server configurations get a joint recommendation from HPE, Intel, and VMware to run vSAN ESA with no hitches, glitches, or gotchas.

Maximizing VMware vSAN Original Storage Architecture

Also known as OSA, the vSAN Original Storage Architecture offers a dramatically simplified system for hypervisor-converged storage, as shown in **FIGURE 4**. This diagram depicts how VMware VMs can interact with one or more different storage facilities, through a shared, virtualized vSAN shared datastore with access to multiple and various storage arrays.

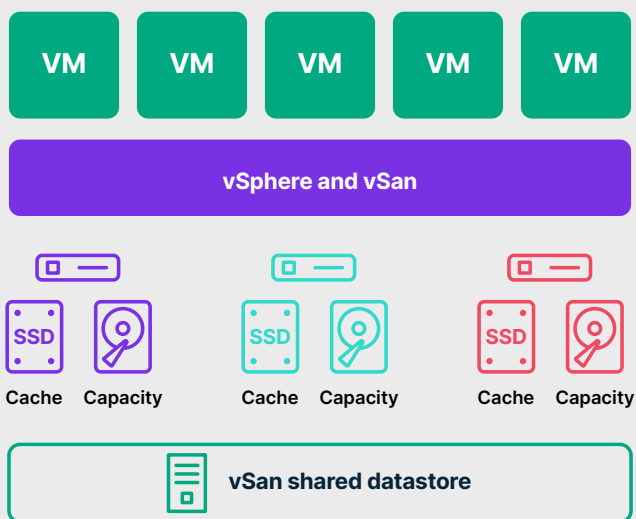


FIGURE 4: vSAN Original Storage Architecture goes from shared datastore, through storage arrays, into vSphere and vSAN to service VMs

Drawing on the simplified diagram, basic OSA characteristics and features grant vSAN storage many valuable and useful characteristics, including:

- Software-defined storage is embedded in VMware vSphere. It's been thoroughly proven for eight years, and an SDS market leader for even longer than that.
- Works on both vSphere 7.x and 8.x versions to cover the vast majority of use cases and business needs.
- Network shared storage is available through both cache and capacity tiered drives, so that applications or services can choose to emphasize performance or capacity depending on user requirements and SLAs.

- Operated through a storage policy-based management (SPBM) framework, OSA lets businesses draw upon security, access controls, retention, protection, and other data management and governance practices and policies.
- OSA works with both hybrid (SSD/HDD) and all-flash (SAS and NVMe) storage devices, so businesses can match storage types to meet application and service requirements.
- HPE ProLiant Gen11 servers powered by Intel Xeon CPUs are certified VMware vSAN ReadyNode servers. Buyers may elect to DIY VMware vSAN with their own components, but this approach saves time and is guaranteed to work.
- Works with Intel-based HPE ProLiant Gen 10, Gen10 Plus, and Gen11 servers, where the newest models support both Gen Intel Xeon CPUs.

The primary reason businesses and organizations use VMware vSAN to deliver enterprise-class, high-performance storage to virtualized applications is because OSA (and its successor, ESA) brings seamless vSphere integration along with the full VMware stack. For those who use VMware vSphere this brings numerous benefits to such an implementation approach, including:

- No other storage products get embedded within the vSphere kernel.
- vSphere is the only software needed.
- Enabling vSAN only requires turning on vSphere DRS or vSphere HA.
- All vSAN storage gets managed (along with everything else) via vCenter or the vSphere web client.

Advancing to Express Storage Architecture (ESA)

Validated, Intel-powered preconfigured server options are defined in [the VMware Virtual SAN Compatibility Guide](#). The newest vSAN Express Storage Architecture aka ESA—the one shown in the **FIGURE 4** snippet from the VCG—requires VMware vSphere 8.x, and only works with vSAN ReadyNodes (or emulated counterparts). ReadyNodes use only single-tier NVMe with no cache or capacity storage segments. They bring buyers better TCO than OSA and no performance loss using RAID 5 vs. RAID 1 at higher storage utilizations. Indeed, ESA offers the exact same vSAN experience as OSA. For many organizations, VMware vSAN works best on HPE ProLiant Gen11 servers using the latest Intel Xeon Gen processors (note that Models DL320, DL360, and DL380 all appear in **FIGURE 4** as well).

When comparing vSAN OSA to ESA the following observations are worth considering:

- ESA switches from capacity and cache arrays to all-NVMe storage (see **FIGURE 5**)
- ESA TCO can be up to [40% lower](#) than OSA.
- ESA architecture provides RAID 5/6 capacity with better performance than OSA RAID 1 (usually positioned as the “fastest” form of disk mirroring).
- ESA RAID 6 vs. OSA RAID 1 increases usable storage capacity (same number of drives), or can reduce costs by requiring fewer NVMe for the same capacity (smaller number of drives).
- ESA performance boost combined with capacity increase supports more VMs in the same hypervisor space.

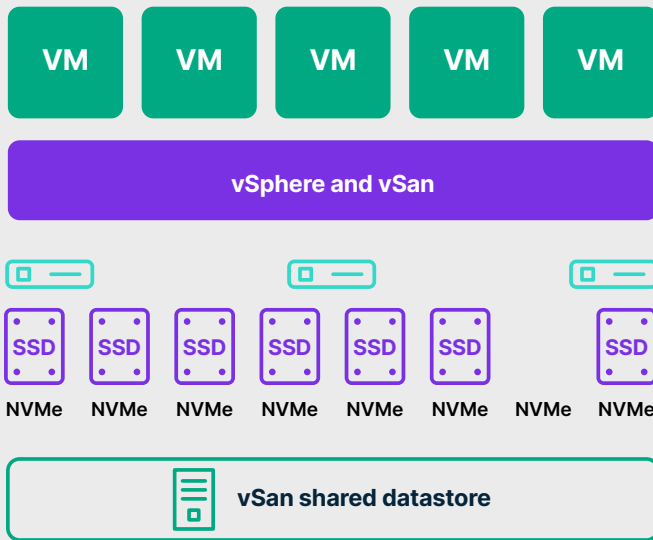


FIGURE 5: The ESA architecture switches out OSA's capacity and cache arrays for an all-NVMe approach

The key takeaway is that vSAN ESA delivers better performance and the same capacity as OSA, but with significantly fewer hosts that are required for an equivalent OSA configuration. This reduces costs for power, space, hardware, and licensing. Furthermore, ESA delivers 4 times better compression, triples buffer capacity, and handles more and heavier workloads over time. ESA architecture is identical to OSA, except for the all-NVMe layer at the middle shown in **FIGURE 5**.

Indeed, this explains why all the items shown in the VMware Compatibility Guide in **FIGURE 3** are ESA compliant. There's simply no reason to buy anything else today, especially given that ESA is backward-compatible with OSA for environments where the two storage architectures must interoperate.

Modern Management Mindset

Crunching through huge datasets, handling models with tens of billions of parameters, and managing huge, complex virtualized infrastructures complete with applications and services is the stuff of legend, and some of the key reasons for IT modernization. But what really brings IT into the modern age is a formal discipline known as computer operations management (aka compute ops management, or even COM). This brings HPE ProLiant servers powered by Intel Xeon Scalable processors into a modern management mindset, starting with silicon and software-based support for remote access and control, even for servers that may be unavailable (because they've crashed, or their network connections aren't working).

In the next and final section of this eBook, you'll dig into HPE Integrated Lights-Out (iLO) management. It not only supports secure remote access for individual HPE ProLiant servers, it also supports remote management APIs to enable on-premises management via OneView, and enterprise-wide management via HPE GreenLake (through its Compute Ops Management facility). Overall, these management elements enable HPE ProLiant Server to partake of and participate in modern, global compute operations management for organizations from end to end, in the data center, at the edge, and in the cloud.

CHAPTER 12

Introducing Modern Compute Operations Management

In today's modern organizations, compute operations management (aka COM) leverages advanced technologies to get things done more quickly and efficiently. Such leveraging embraces artificial intelligence, big data, automation, and cloud computing to streamline and optimize how compute resources get organized, accessed, monitored, and managed. These leading-edge technologies are specifically intended to support more data-driven, relevant, and customer-oriented operations alongside a holistic and global view of the entire enterprise from end to end, in data centers, at the edge, and in the cloud.

Typical components or aspects within a COM environment include at least the following:

- Centralized management platforms provide a proverbial “single pane of glass” for monitoring and management, with clear visibility into distributed compute environments that include private and public clouds, plus on-premises elements
- Automation of routine everyday tasks, real-time monitoring and proactive, intelligent alerting work to reduce manual effort, improve server and system health, and ensure compliance with regulatory and other information standards and policies

- A deliberately holistic approach to compute management means organizations can scale operations efficiently as they deliver high performance and maintain stringent security standards

In general, modern COM seeks to protect and secure its systems, data, and assets while making them accessible and available to authorized users, systems, and connections.

In this chapter, you'll find a basic primer on COM, along with discussions of key related HPE and Intel technologies useful for supporting and securing a modern digital enterprise. That means as you go down this trail, you'll also find discussions of key HPE and Intel technologies to protect, preserve, and manage HPE ProLiant servers powered by Intel Xeon Scalable processors. Likewise for key supporting software, systems, and infrastructure that serve to establish and maintain COM across an entire enterprise including cloud-based assets and resources.

In practice, Compute Operations Management may be reduced to its initials—namely, COM. But it's more than just an acronym. It's an outright digital discipline that seeks to define and employ a set of conceptual and physical models, within a body of best and accepted practices. Ultimately, the goal of COM is to position all elements of an IT infrastructure into a well-understood framework, which plays into a lifecycle model through which technology is continuously evaluated, chosen, deployed and maintained and finally, replaced with newer and presumably more capable alternatives. Thus, COM is best understood as a process that involves computing within a framework of practices and procedures used according to a management discipline. Indeed, COM unites users, IT staff, operations and security professionals, and business stakeholders (including executives) across the whole enterprise.

The COM Breakdown, Part 1: What Is Compute?

Simply put, compute is an abstract service that's made available to authorized users who wish to acquire, access, analyze and manipulate data. It's easy to over-simplify compute as the sum collection of computing devices in an enterprise. But compute also embraces resources beyond computation and includes memory that applications and operating systems use to perform tasks, storage from which (and into which) data comes and goes, and the networking links and infrastructure that brings users, resources, and services or applications together.

The COM Breakdown, Part 2: What About Operations?

In modern organizations, the IT operations function—which may be called IT Operations, or abbreviated as ITOps—seeks to ensure, secure and maintain smooth functioning of the IT infrastructure, and other key related capabilities, all managed with one eye on tools and technologies and the other on business goals and strategies.

To maximize and optimize return on IT investments, ITOps covers these key roles in modern organizations:

- **System Administration:** Managing and maintaining IT systems, including servers, databases, and networks.
- **Service Management:** Ensuring that IT services are delivered efficiently and effectively to meet business needs.
- **Security Management:** Protecting IT systems and data from cyber threats and ensuring compliance with security policies.

- **Software and Hardware Deployment:** Installing, updating, and managing software and hardware components.
- **Network Operations:** Monitoring and managing network performance, troubleshooting issues, and ensuring connectivity.
- **Support and Maintenance:** Providing technical support to users and maintaining IT infrastructure to ensure reliability and performance.

Operations plays into compute by organizing and marshalling its resources and functions, to make sure that systems and services are properly configured, correctly secured, and kept up-to-date as new threats emerge, new tools replace old ones, and new technologies get adopted. In some organizations this may also involve a development organization, and incorporate related functions such as need analysis, software design and architecture, development and test, deployment and maintenance, user support, and more.

The COM Breakdown, Part 3: Why management Is Key

In the COM context, management is best understood as the driver who keeps the wheels turning as they should be, and what makes sure the organization's IT capabilities work as they should, when they should, across the enterprise. In the COM triad, compute describes what gets used and consumed, and operations describes the roles and capabilities that fall beneath the IT umbrella. Management brings all the pieces together, and defines processes, procedures, policies, events and schedules, and governance for what gets managed (compute) and what kinds of tasks and activities are involved (operations). It also addresses the key notion of access control, which keeps track

of who's authorized to do what in the enterprise, and associates tools for authentication to prove one's identity as part establishing and maintaining proper security.

Compute Operations Management à la HPE and Intel

FIGURE 6 shows how COM figures into a complex IT environment. Individual users (labeled Administrator, Operator, and Observer at left) go through an Authentication facility to obtain access to

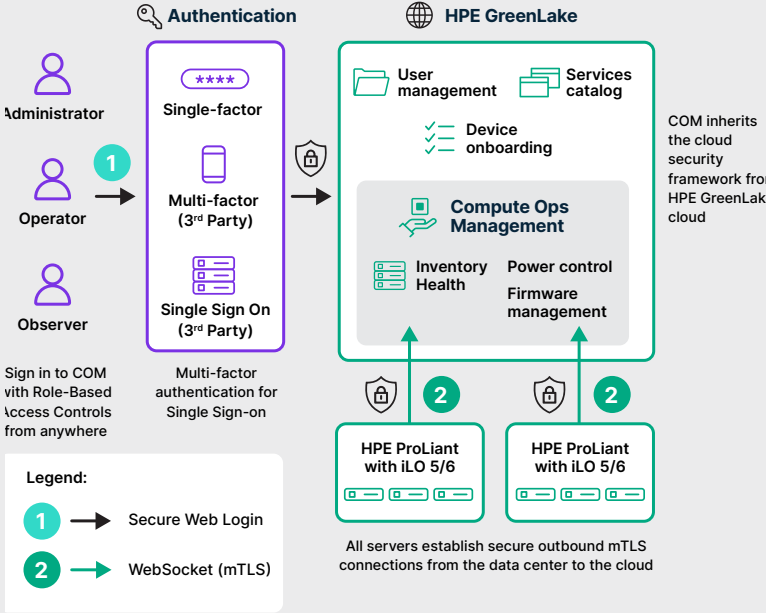


FIGURE 6: Within an HPE environment, both users (left) and servers (below, at multiple sites) access its cloud-based HPE GreenLake Compute Operations Management capabilities

COM. This might mean a typical account–password login, though multi-factor authentication is highly recommended and nearly ubiquitous. (via third-party tools such as Okta, Onelogin, Google Authenticator, MS Entra, and so forth, or native HPE Single Sign-On, and other third-party SSO facilities such as CyberArk, Autho and more). On the other hand, servers and other network devices use secure, certificated based Transport Layer Security (TLS) to establish valid identities and interact with management tools such as inventory, health, power control, and firmware management.

In the next chapter, you’ll leave the generalities of Compute Ops Management behind and dig into how HPE’s ProLiant servers use demonstrably secure identity information, secure firmware inside a silicon root of trust that creates a fingerprint for 4th and 5th Gen Intel Xeon Scalable processors that must be matched exactly before the server will boot. This establishes and maintains server security from build-out, to transport and delivery, to configuration and set-up, and on into the useful, productive parts of the lifecycle. Keep reading for the next riveting adventure on this important trail. Further down the trail (in Chapter 15, in fact), you’ll learn a great deal more about HPE Compute Ops Management in the context of HPE GreenLake, HPE’s cloud-based hybrid cloud solution that lets organizations bring a public cloud experience to all its computing services, even within on-premises data centers.

CHAPTER 13

Starting from Silicon: HPE Integrated Lights Out (iLO) Management

As is so often the case, HPE's purpose-built server management technology goes mostly by its acronym. It's called HPE Integrated Lights Out, but it is universally known as iLO. Indeed, iLO provides the foundation for per-server management, but also includes the interfaces through which sophisticated management platforms and tools can interact with HPE ProLiant servers powered by Intel Xeon Scalable processors. That's what's shown at the bottom of Figure 1 in the preceding chapter, in fact (and names the Mutual TLS or mTLS connections, in which both parties authenticate to each other using TLS certificates, to establish bona fides and set up a secure, encrypted connection between COM on one side, and a ProLiant Server (or site) on the other side.

What makes iLO both valuable and interesting is that it keeps working even if network connections go down or become unavailable or non-responsive. iLO includes separate hardware and so-called out-of-band (OOB) connections that let COM interact with, and even cold start, an iLO-equipped server using a separate remote connection (which may be some kind of wireless or telephone link, a separate network link, and so forth).

Out of band (OOB) Terms and Capabilities

The cognomen comes from iLO's ability to establish and provide remote access to HPE ProLiant servers powered by Intel Xeon Scalable processors, even when no network connection is available. That's the basis for the "lights-out" designation, which uses OOB connections to reach out and interact with devices as needed. They even work when the lights are out at a remote location and provide an alternate path into networks and devices for troubleshooting when the normal paths are blocked or MIA.



Basic Characteristics of iLO

Integrated Lights Out provides a bevy of useful capabilities designed to let admins interact with servers even when normal communications channels—or the servers themselves—aren't working. These include the following key characteristics which offer life-saving (or server-saving) opportunities galore:

- iLO out-of-band management lets admins interact with Intel Xeon-powered HPE ProLiant servers remotely to configure, monitor and update them, even if they are offline or inaccessible
- iLO offers important management facilities that include remote console access, KVM over IP, and a wide range of automated management tasks for configuration, set-up, firmware and software updates, and more

- iLO's primary impetus is to simplify and streamline server setup, troubleshooting, and firmware updates, and it delivers on that promised capability in spades

All in all, iLO provides secure and efficient server operations under even the most trying of circumstances.

Why Do Organizations Need iLO?

First and foremost, iLO opens avenues to 24x7 uptime so organizations can maintain functionality and deliver mission critical services. Indeed, iLO provides a means to get to servers even if the network is down, or a server shuts down or crashes for any reason. This access then extends to key IT assets such as routers, firewalls, servers, switches, power, and telecom gear through an iLO connection. Thus, admins can investigate, repair, and bring their networks back to life. iLO provides key technologies to bridge such gaps as and when they're needed.

Where Does iLO Get Its Value?

iLO is built into all HPE ProLiant servers powered by Intel Xeon Scalable processors and includes scalable licenses with mobile app features to support IT staff anywhere at any time. iLO uses its own proprietary, discrete microprocessor chip, with separate power supply and communications links. That's how it remains available even if a server fails outright. iLO can reach out to diagnose what's wrong, and then fix things quickly and efficiently even if the server is powered down. At the same time HPE iLO provides a toolbox of capabilities to help automate and simplify server provisioning, troubleshooting, and firmware updates.

Introducing HPE Intelligent Provisioning

The server deployment and maintenance facility that's embedded in HPE ProLiant servers powered by Intel Xeon Scalable processors is firmware-based. Thus, it requires no physical media to access and run. Furthermore, HPE Intelligent Provisioning software available through iLO speeds server deployment and maintenance through the following capabilities:

- Step-by-step deployment assistance
- System configuration via guided, profile-driven or scripted pathways that integrate seamlessly with standard IT processes and practices
- iLO recognizes when system software is out-of-date, and prompts download and install of current updates

iLO Provides Critical Connection Capabilities

The same remote access, KVM and automation supports built into iLO are also available to external management tools and platforms. In the next sections, you'll learn about integrated management platforms that interoperate with HPE ProLiant servers powered by Intel Xeon Scalable processors. In the following chapter, you'll learn more about an entire, cloud-based COM facility built into the company's on-demand, cloud-based HPE GreenLake platform. It provides global holistic monitoring, management, automation capabilities for the entire enterprise, on-premises and into private and public clouds.

iLO runs on the server side to provide a point of entry and control on an individual device. In the next chapter, we shift our focus from a server-at-a-time view to one of HPE's best-known and still widely used management consoles—namely, HPE OneView. You'll learn more about important and ongoing uses of this toolset that should keep OneView running and gunning for many years to come.

CHAPTER 14

HPE OneView

HPE OneView is a powerful IT infrastructure management software especially designed to simplify and streamline management, monitoring, and automation of on-premises HPE ProLiant servers powered by Intel Xeon Scalable processors. HPE OneView offers the following capabilities:

- On-premises Infrastructure management software
- Provides composable solutions across compute, storage, and networking resources
- Supports HPE Synergy and dark sites

For the record, HPE Synergy is a composable, software-defined infrastructure (includes virtual switches, networks, servers, storage, and more) designed for hybrid cloud environments. It supports fluid pools with physical and virtual elements for compute, storage and network fabric that can be arranged into any configuration to handle a full range of workloads. In the same vein, a dark site describes a pre-production environment used to test and validate code (and code changes) before deployment into live production environments. That makes a dark site something for IT teams to safely test updates, configurations and new deployments without impacting production systems and real users.

OneView Remains Alive and Well

Contrary to the widespread belief among some customers, HPE OneView is neither obsolete nor on its way out the door. In fact, OneView is absolutely the right tool for HPE customers using Synergy, or those who operate dark sites as part of their development and test environments. Indeed, OneView provides a way to bring such things together within a single, unified management console and a shared set of tools, visualizations, automation and more.

HPE OneView remains an active, supported go-forward solution. Based on a conservative view over our product timelines, HPE intends to support and develop OneView for the foreseeable future. Though third-party sources may occasionally—and incorrectly—label OneView as a legacy tool, it remains an important, even vital part of HPE's Compute Operations Management strategy and toolset.

Indeed, a significant majority of HPE's current major compute accounts have standardized on HPE OneView. They use OneView daily across various mixtures of rack mounted and HPE Synergy servers. HPE plans to keep OneView going to serve existing and new customers, especially those who use Synergy and/or dark sites as part of their IT infrastructures. For those reasons, the company will keep investing in new, value-added OneView features and functions. Indeed, HPE is committed to main OneView as the preferred management platform for Synergy and dark site situations as far as the company's planning horizon extends.

What HPE OneView Brings to the Party

OneView lets authorized administrators use a single, consistent set of consoles, tools, and automation to interact with on-premises and

cloud-based Intel Xeon-powered HPE ProLiant servers and related resources. It also consolidates and coordinates alarms, alerts, notifications, and workflows (including updates, configurations, and more) across all servers and resources under its purview. In particular, OneView offers users the following benefits:

- A unified interface for managing compute, storage, and networking resources, allowing IT administrators to deploy and manage infrastructure more efficiently
- Administrators can create and manage server profiles and templates, automate routine tasks, and ensure consistent configuration across the entire infrastructure
- Real-time monitoring and alerting, helping to proactively address potential issues before they impact operations

OneView Also Leads into HPE Compute Ops Management

Even better—as you’ll learn more in this next chapter—HPE OneView integrates with automation tools and cloud management platforms, such as HPE Compute Ops Management, thereby enabling seamless orchestration from core to cloud for the whole enterprise. This includes a veritable cornucopia of capabilities, such as:

- **Manage Software-Defined Infrastructures:** Through Synergy, organizations can create virtual networks, servers, storage and entire infrastructures. OneView makes them easy to define, manage and maintain wherever they’re housed (on-prem or in the cloud).
- OneView supports powerful automation to deploy infrastructures faster, simplify lifecycle operations and increase productivity (for more info, see [Infrastructure automation made simple](#))

- OneView helps make infrastructure programmable, so that infrastructure elements can be composed like code, using special scripts that automate and increase agility. This helps make IT tasks predictable and ensures compliance.
- OneView lets organizations connect from core to cloud: It can provision turnkey private cloud infrastructures, or integrate with preferred cloud platforms (e.g., AWS and Azure). This simplifies project outcomes and ensure timely, high-quality delivery.
- Admins or developers can work with the OneView API directly, to define, refine and optimize workflows, automation, infrastructure compositions, and more to improve productivity and speed deployment.
- A single, cohesive OneView Dashboard brings all the pieces together, so admins can view and interact with infrastructures and dark sites using a consistent set of views, tools, automation and more.
- Beyond HPE's implementation and platform, a plethora of OneView Partner Integrations is also available (and worth perusing for the likes of VMware and Microsoft).

CASE IN POINT: BRAZIL'S PORTO ITAPOA

Porto Itapoa is a major Brazilian container terminal facility that's recently updated its data and computing infrastructure. That modernization effort boosted its container handling capability and overall efficiency, for a 67% boost in daily gate transactions. That puts the company on track to meet its ambitious goal to process 2 million containers yearly, raise port efficiency rankings, and attract more vessels—and business.

To obtain the level of performance and availability it needed, Porto Itapoa implemented HPE Alletra 9060 storage and HPE Synergy 480 infrastructure servers across its IT domain. That included ERP, berth management, and external portals, with support from HPE Services

and channel partner Sercompe. Its new deployment ensures constant system availability, accelerates access to portals and data, and improves container handling with faster berth planning and decision making. These improvements come in no small part from the 100% data availability that HPE Alletra 9060 delivers with fast, consistent performance and minimal latency even at scale. HPE OneView makes managing all the Synergy and Alletra components involved an absolute snap!

ABOUT HPE COMPUTE OPS MANAGEMENT, ONEVIEW ADDITION

HPE offers an add-on subscription service for OneView organizations. It aggregates HPE OneView connections into the HPE Compute Ops Management centralized, cloud-based console. This facility readily

Engaging with HPE consulting Services for OneView—and Beyond

HPE's global Consulting Services arm offers end-to-end support for organizations using OneView, iLO, and even GreenLake Compute Ops Management. Its offerings let organizations choose as much—or as little—help and support as they may need. That includes design, deployment and implementation services for these platforms, and ongoing or on-demand technical support and troubleshooting services. Organizations can also ask HPE Consulting services about related training and education offerings, available on-site, online, or at scheduled training sites. (Please note: the OneView Global Dashboard Essentials class is a particular favorite among HPE customers.)



supports multi-site management interaction, including reporting and analytics for HPE OneView Appliances and other connected infrastructure devices. For organizations moving into or already heavily invested in the cloud, this provides a great means to bring Synergy and dark site resources under the same general COM umbrella. This offering is called [HPE Compute Ops Management – OneView Edition](#). With the right mix of platforms, it's well worth looking into.

Though OneView supports various methods for cloud access and integration, it is primarily an HPE ProLiant server-focused management console and toolset. Options and opportunities really open up when you look at HPE's cloud-native management platform HPE Compute Ops Management. By no coincidence whatsoever, that naturally leads us into Chapter 15 where we dig into HPE Compute Ops Management in some detail.

CHAPTER 15

HPE Compute Ops Management

HPE Compute Ops Management is a cloud-native management platform. It's designed to simplify and automate management of compute resources wherever they may reside, on-premises or in a private or public cloud. HPE Compute Ops Management provides a single, centralized cloud-based console for all connected assets and resources. Admins can use it to access, monitor and manage servers across a wide range of distributed environments (see **FIGURE 7**). With a subscription to the OneView Edition for this platform, such management even extends into Synergy infrastructures and dark sites.

HPE Compute Ops Management Features



Key HPE Compute Ops Management features include the following:

- Real-time monitoring for health, events, alarms, alerts, logs and more
- Automated lifecycle management, with a large collection of pre-defined, customizable workflows for routine tasks, on- and offboarding, user equipment handling and more

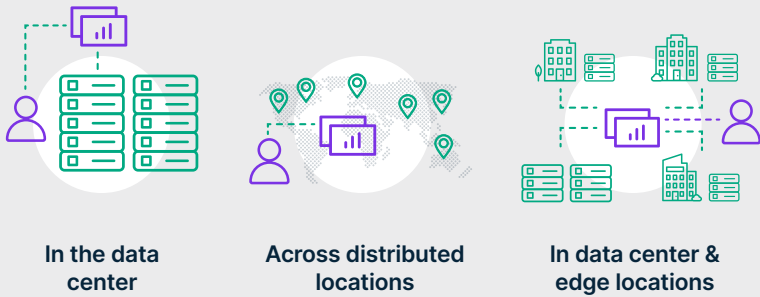


FIGURE 7: HPE Compute Ops Management allows teams to manage everywhere from anywhere

- Demonstrably secure operations based on Mutual Transport Layer Security (mTLS), security certificates, and tightly integrated identity, authentication and access control services

By design, HPE Compute Ops Management allows organizations to unify and aggregate compute management across the entire organization. It uses AI and automation to reduce manual effort, improve IT efficiency, and to enforce and ensure compliance with rules, regulations, and applicable mandates, with policy enforcement and governance integration:

- A single global console with secure Single Sign-On (SSO) unifies server management from edge to cloud to data center
- As with other HPE management Tools, it uses underlying iLO access for fast, simple server provisioning and management
- Works wherever a server is situated, from edge to cloud and into data centers

- Provides the same, consistent single console for remote access, monitoring, and management
- IT can focus on managing infrastructure and services, rather than learning and managing complex toolsets
- Single sign-on provides access to a common platform that embraces security, compliance, elasticity, and data protection
- Superior as-a-service experience lets organizations manage their entire compute landscapes, with access to automated management, secure access, provisioning and setup, updates and maintenance, and more

Key Elements in HPE Compute Ops Management:

The most important aspects of HPE Compute Ops Management come from its global, holistic, end-to-end enterprise coverage. Because of its accommodating design and flexible capabilities, it handles servers and infrastructures wherever they are. That means for organizations that adopt HPE Compute Ops Management:

- Server management gets unified without boundaries
- Organizations can easily use APIs to automate server management across the lifecycle and the enterprise
- Cloud-based access uses highly secure channels for management and other services (growing list, more available soon)

CASE IN POINT: KIMLEY-HORN

For those not already in the know, Kimley-Horn is a U.S.-based planning, surveying, engineering and design consulting company. It's a project-driven business, with about half its efforts devoted to ongoing special projects, and the other half to incoming limited-duration projects. The company operates around 100 offices around the country, each with at least one server. The home office IT department is responsible for those servers but works with local IT staff on location to maintain its gear.

This highly distributed situation makes remote access vital, and poses certain challenges when servers get deployed, or their firmware needs an update. Before modernization, server changes or updates involved substantial data transfers and coordinated installation. It could also require as much as 5 hours of downtime during scheduled maintenance intervals. Kimley-Horn adopted HPE Compute Ops Management specifically to simplify and unify managing complex infrastructures. But it also benefited greatly from automating essential admin and upkeep tasks.

A particularly helpful benefit of HPE Compute Ops Management came from scoping updates for individual servers or sites. Instead of transferring huge data files to cover all bases, the system would download only the necessary files and data for updates scheduled through the system. Though Kimley-Horn is not completely finished with its deployment of HPE Compute Ops Management, it's convinced that new alerting tools and detailed dashboards should provide more and better information about overall server health. They say that because HPE listens to its customers and responds to their feedback, their solutions should help Kimley-Horn manage its servers faster, more easily, and with greater reliability. See this [case study](#) for all the details.

Benefits of HPE Compute Ops Management

In Chapters 12 through 15, you've seen just how the right kinds of server management can benefit your organization. They provide improved server uptime, guaranteed access, automate routine tasks and responses, and provide real visibility into and control over your digital estate. It's not too much to say that Compute Ops Management is the foundation upon which modernization rests, and the approach that enables organization to stay secure, be productive, and respond quickly and well to ever-changing customer demands and market conditions.

Working Toward Modernization with HPE

Now that we've reached the end of this guide, please understand that HPE supports modernization in many ways through its ProLiant Servers with Intel processors, and through its various management platforms and services—namely OneView and HPE GreenLake (including the HPE Compute Ops Management facility therein). There are a lot of pieces and parts involved, to be sure. But you should also know that all these parts work together to improve what your organization can do, and how you can boost user and IT productivity thanks to secure deployment, maintenance, updates, and more. That includes profound support for GenAI and all kinds of extreme, demanding, and intense workloads wherever they're needed: at the edge, in the datacenter, or in the cloud.

Indeed, HPE and Intel have your modernization needs covered from silicon all the way into the private and public cloud (or multiples thereof, as your needs and circumstances may dictate). To learn more about HPE's modernization portfolio, built around the ProLiant Server family, consult any or all of these terrific resources:

- [HPE Consulting Services](#) covers all pieces of the modernization puzzle, including security, GenAI, workload planning, modeling, deployment and management, plus HPE OneView and HPE GreenLake
- HPE Integrated Lights-Out [homepage](#), [demo](#), and [explainer](#)
- HPE OneView [homepage](#), with copious information about and demo for its GreenLake Compute Ops Management plug-in. See also [Infrastructure Management Made Simple](#).
- HPE Compute Ops Management: [homepage](#), [explainer video](#), and [demo video](#)

Please visit [HPE.com](https://www.hpe.com) and use its “Let’s connect” online chat facility if you have any questions or would like to initiate a sales call.

ABOUT HPE



Hewlett Packard Enterprise

HPE is the global edge-to-cloud company built to transform your business. How? By helping you connect, protect, analyze, and act on all your data and applications wherever they live, from edge to cloud, so you can turn insights into outcomes at the speed required to thrive in today's complex world. <https://www.hpe.com/us/en/about.html#ourpurpose>

ABOUT INTEL



Intel is an industry leader, creating world-changing technology that enables global progress and enriches lives. Inspired by Moore's Law, we continuously work to advance the design and manufacturing of semiconductors to help address our customers' greatest challenges. By embedding intelligence in the cloud, network, edge and every kind of computing device, we unleash the potential of data to transform business and society for the better. To learn more about Intel's innovations, go to [Intel.com](https://www.intel.com).

ABOUT ACTUALTECH MEDIA



ActualTech Media, a Future company, is a B2B tech marketing company that connects enterprise IT vendors with IT buyers through innovative lead generation programs and compelling custom content services.

ActualTech Media's team speaks to the enterprise IT audience because we've been the enterprise IT audience.

Our leadership team is stacked with former CIOs, IT managers, architects, subject matter experts and marketing professionals that help our clients spend less time explaining what their technology does and more time creating strategies that drive results.

If you're an IT marketer and you'd like your own custom Gorilla Guide® title for your company, please visit actualtechmedia.com.