Futurum
RESEARCH

# OCP Global Summit Unveiled Time to Online as a Newly Critical Measure

## Power Dominated Messaging at OCP Global Summit, but Progress on Readiness Highlighted the Importance of Time to Online for AI Capacity

Analyst(s): Guy Currier
Publication Date: November 3, 2026
Document #: AINGC202610

## What You Need to Know

- The 2026 Open Compute Project (OCP) Global Summit, held October 12–15 in San Jose, experienced its first sell-out, drawing 15,000 attendees with 3,000 more on the waitlist; OCP will move the event to the Moscone Center in San Francisco next year. OCP continued to frame its "chip-to-grid" scope, highlighting its AI Computing Continuum project for AI beyond the hyperscale data center. HPE and Intel presented this project's first server specification, AI-Native Edge MHS 1.0, at the show.

- The data-center CPU showed more AI clawback progress. AMD introduced Helion, an Open Rack v3 design for agentic AI with a CPU-only configuration of up to 32,768 EPYC cores, and NVIDIA contributed its Vera Rubin rack and compute tray designs to OCP in version 1.2 of the MGX specification alongside a 256-CPU Vera rack. Intel showed Xeon 6+ racks exceeding 46,000 cores at 133.45 kW (as well as its D-series inference GPU, formerly code-named Crescent Island). HPE reported seeing CPU-to-GPU ratios moving from one-to-four toward one-to-one as agentic AI spreads.

- Google, Meta, Microsoft, and NVIDIA contributed a specification for AI data-center behavior during grid disturbances, and OCP touted its new Energy Project for generation and storage, which kicked off in July of this year.

- Multiple vendor pitches at the conference featured messaging about speeding paths to productive capacity – the headline of this report. Supermicro positioned its Data Center Building Block Solutions around time to online, and Meta showcased its use of tent-like structures as data-center buildings, enabling the company to bring capacity online in only three to five months.

- Microsoft, Marvell, and Oracle introduced the Tessera open architecture for hardware security modules (HSMs) that enables Marvell's HSM to interoperate with Azure's integrated HSM without exporting keys. Lattice and AMI launched the Secure Control

and Management Plane (SCMP) workstream that encompasses silicon, firmware, software, and networking.

## The Futurum View

The OCP Global Summit 2026 was OCP's biggest yet. Much of its keynote messaging, from OCP and from vendors, predictably went to power: how to get it, how to deliver it from the grid to the chip, and how much more of it AI will require. These are real, pressing issues on several fronts. But the most useful new lesson from San Jose was the importance of a moment in the implementation of infrastructure that sits between power availability and the conventional delivery of business value. Once a system has power, installed capacity has to be cooled, cabled, lit, validated, and handed to platform teams, and *that* interval, time to online, is critical to success, especially in AI. That is one of three themes that ran through the conference: the three "time-to" clocks (time to power, time to online, and time to value); grid-to-chip convergence; and a deepening heterogeneity in systems that makes open standards a strong lever for infrastructure buyers in their goals of ensuring trust in products and control of their estates. This report draws on keynotes, announcements, and sessions from OCP, AMD, NVIDIA, Intel, HPE, Arm, Meta, Microsoft, Oracle, Google, Marvell, Lattice, AMI, Supermicro, Eaton, and Cisco at the October 12–15 Summit in San Jose; it does not evaluate individual products' performance claims or vendors' financial outlooks.

## Recommendations

For organizations planning AI and other modern infrastructure builds over the next year, three moves follow from the summit:

1. Add a Clock for Time to Online: Measure the interval from system installation to productive use, especially for every AI pilot and provider commitment. From the start, count as value the learning and experience gained in the organization.

2. Assess Open Specifications and Their Versions for Products: For any rack, component, integrated system, or solution using open specifications, ask about degree of compliance to weigh portability risks against engineered value.

3. Comprehend Power and Cooling in the Reference Architecture: Facility requirements belong in AI systems designs before the hardware is ordered, especially as high-end power and liquid cooling technologies migrate out of hyperscale.

## Time to Online Should Get More Attention

Time to power gets plenty of coverage right now, and it deserves it. Time to online – the step in the "time-to" chain before time to value – gets far less. For AI, in particular, time to online matters more than it first appears. AI is still early in its maturity curve, however advanced it may seem, so much of AI's value arrives through experiments, trials, and the often-unmeasured


©2026 The Futurum Group, LLC. All rights reserved.

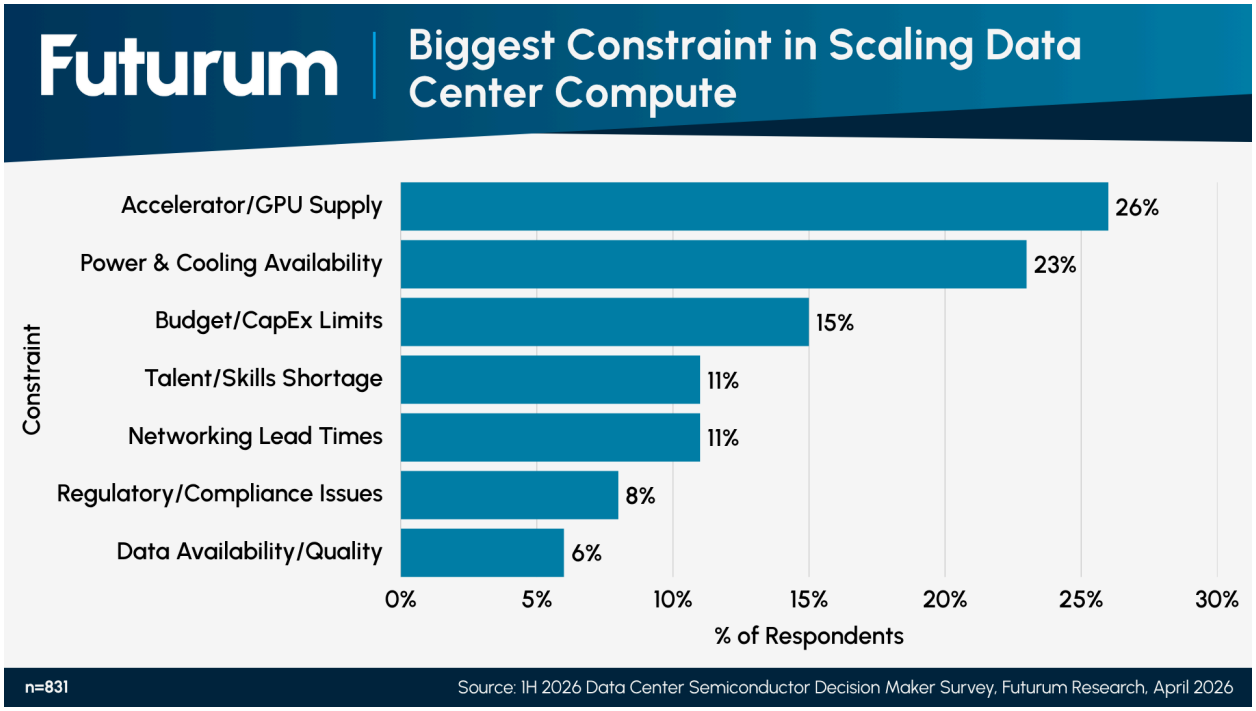
organizational learning that comes only from doing. Every week that installed capacity sits idle is a week of that learning lost.

Presenters on stage at the conference each described this issue in their own terms. Meta's Santosh Janardhan described how, when someone asks him for a data center, his answer used to be "you should have talked to me two years ago." He described Meta's use today of tent-like structures that come online in months rather than years. HPE's Trish Damkroger called industry-wide lead times one of the biggest current challenges, and Lattice CEO Ford Tamer said silicon development cycles are shrinking sharply, putting pressure on lights-on cycles.

Facilities, system, and silicon advancements each run on their own clock, and the slowest one sets the pace, defining the time to online.

Futurum's own research points the same way: In Futurum's Data Center Semiconductors Decision Maker Survey, 26% of organizations named accelerator or GPU supply and 23% named power and cooling availability as their single biggest constraint in scaling data center compute (Figure 1), and networking lead times were the top constraint for another 11%.

Figure 1: Biggest Constraint in Scaling Data Center Compute




*Source: 1H 2026 Data Center Semiconductor Decision Maker Survey, Futurum Research, April 2026*

Potential time-to-online bottlenecks aren't only a neocloud and hyperscaler issue. When the enterprise buys capacity from a provider, whether cloud or colo, time to online shapes the capacity, regions, and pricing that provider can offer; part of any quoted lead time is the provider's own readiness. Build for yourself, even at only the scale of an AI experiment, and the same clock applies: hardware can sit for months waiting on circuits, cooling retrofits, and facilities approvals. Those stalls must be tracked and cannot be worked around because they


©2026 The Futurum Group, LLC. All rights reserved.

mark where adoption can fail before experience and learning, so critical in these still-early days of AI, can even begin.

## Grid-to-Chip Moves Deeper Into Standards at OCP Global Summit

Grid-to-chip convergence, in which energy supply, power conversion, cooling, fabric, compute, and storage are designed together, is fairly new outside of embedded systems on the one hand and hyperscale design and operations on the other. OCP announced its grid-to-chip strategic initiative only last year, a significant extension of OCP's work. This year showed its first implementation fruits, largely in power delivery. In his Day 2 keynote address, OCP CEO George Tchaparian said plainly that "the energy is becoming the bottleneck," making it a hot target for standards development.

Oracle's Pradeep Vincent said grid interface requirements are still being set site by site and expressed the company's support for OCP's emerging standardization of them. NVIDIA's Ian Buck argued that AI factories can be enabled to *contribute* to the grid rather than only draw on it, and Eaton's Peter De Bock, in describing 800 VDC as an architecture rather than a product, implied its integration into rack-scale products as one part of a whole.

Grid-to-chip products are only starting to emerge, and no product or service was presented as using the new grid specification. The Energy Project's work has just begun, and neither the enterprise edge nor the telco edge was particularly visible on the show floor either. Standards tend to move first, but a vendor's or provider's grid-to-chip capabilities and standards use are worth asking about directly because of how density, efficiency, and resilience move downstream to customers as cost and risk.

## All Heterogeneity All the Way Down

This isn't your granddaddy's heterogeneity, which fits together different server, storage, and networking brands (along with the cloud platforms you needed). This is an increase in the variety of vendors, functions, and formats that runs down to the chip level and up to the whole infrastructure estate – driven by the need to support specialized workloads. AI has accelerated this "deepening" of heterogeneity by raising demand for components tailored to one or another part of the AI service supply chain in order to utilize data, power, and real estate more efficiently. "There's no one size fits all when it comes to compute," AMD's Lisa Su said in her keynote.

When diversity in formats and functions is added to diversity in vendors, and at so many levels of infrastructure, the potential for complication increases. In a heterogeneous environment, "complexity goes up exponentially," according to Lattice's Tamer. Open infrastructure standards defray both the difficulty and the cost, however, and the Summit added to standards at multiple levels: NVIDIA's MGX 1.2 contribution and the Open Data Center Specification v0.7 for racks and facilities, the USFP pluggable optic from NVIDIA and Cisco for interconnect, the


©2026 The Futurum Group, LLC. All rights reserved.

Tessera open hardware security module (HSM) architecture from Marvell, Microsoft, and Oracle, and the launch of the Secure Control and Management Plane (SCMP) workstream.

The AI Computing Continuum project, announced last year, is a further, significant example. As OCP's response to AI's rapid spread across a bewildering variety of infrastructure, from the core to the far edge and every gradation between, the AI Computing Continuum will extend existing open standards to encompass regional neoclouds, colocation, on-premises sites, telco points of presence, and industrial locations. Its first server specification, AI-Native Edge MHS 1.0, takes the modular hardware standards built for the core and extends them to edge limits on power, thermals, and footprint. "Edge" can mean a branch office to a telco but a drone, sensor, or wearable to other enterprises. Defining it by operating conditions – placement, service resilience, power, connectivity, and temperature – rather than by location or form factor will matter a great deal to the ongoing development and usefulness of these specifications.

## The Benefits of Open Standards Come by Degree

The value of a standard being open is the reduction of risk and cost in vendor and product decisions and the increase of control over IT systems. These benefits accrue to the degree to which vendors and products comply; the use of a specification is not itself a guarantee. A product can comply fully or only partly; or perhaps more typically, differentiation of a compliant product may lie most in areas not yet covered by the specification.

Compliance with an open standard is not neutrality, either. A vendor that contributes a specification is working to shape the industry around its own IP. The real test, as Futurum's Mitch Ashley put it about USFP, is whether other vendors build to it.

## What to Watch

- Time-to-Online Claims: System integrators, colocation providers, cloud service providers, and neoclouds that provide time-to-online commitments alongside those for power, capacity, and price are building offers that meet the moment. Such providers will make readiness a competitive term; those that don't will leave it buried in lead times.
- Grid Specifications in Practice: The new grid specification is hyperscaler-focused, and the Energy Project's work has only just begun. The first cloud services and architectures built on them, and how utilities respond, will demonstrate the promise of grid-to-chip in driving speed to market and service resilience for end users.
- First Products Built to AI-Native Edge MHS 1.0: Whatever the degree to which new products comply with the specification and differentiate in ways outside of its latest version, its use will demonstrate progress toward consistent and dependable core-to-edge AI operations. That will, in turn, prove the AI Computing Continuum concept and vision of unifying fundamental elements of these growing use cases.


©2026 The Futurum Group, LLC. All rights reserved.

- **Designs Becoming Open Standards:** NVIDIA's MGX 1.2 and AMD's planned Helion contribution bring processor vendor architectures to OCP. Which vendors and vendor-partner ecosystems build to and support them, and how security and management standards such as Tessera and SCMP integrate, can either increase choice and control over the estate or continue to require divisions and silos.

See more on the 2026 OCP Global Summit website.

## Other Insights from Futurum

Cisco-Validated SONiC NOS Highlights Cisco's Open Networking Message at OCP

Arm AGI CPU Posts 2x the Agentic Work Per Rack of AMD's EPYC 9755 at OCP 2026

Cisco's USFP Brings 400G-per-Lane Design to NVIDIA's Optical Roadmap

AMD Helion Makes the CPU Rack a First-Class Citizen of the AI Data Center


©2026 The Futurum Group, LLC. All rights reserved.

# About Us

## About the Authors

Guy Currier, Research Director, Ecosystems, Channels, & Marketplaces, Futurum Research, is a tech market analyst with decades of field experience describing technologies, their business and community value, and how they are evaluated and acquired. He began his technology career as a research director at Ziff Davis, with stints at PC Magazine, eWeek, and CIO Insight, and later spent years at Dell in marketing, product, and technical marketing groups, working closely with customers as they adopted and used IT products. Guy is based in Austin, Texas, and can be reached at gcurrier@futurumgroup.com.

## About The Futurum Group

Every day, The Futurum Group's analysts, researchers, and advisors help business leaders worldwide anticipate tectonic shifts in their industries and leverage disruptive innovation. Unlike traditional analysts, The Futurum Group works not only in analysis and research but also takes that insight and knowledge even further, engaging all the way through the go-to-market process.

Futurum Research provides in-depth research and insights on global technology markets using advisory services, custom research reports, strategic consulting engagements, digital events, go-to-market planning, and message testing. It also creates, distributes, and amplifies rich media content that all stakeholders read, watch, and listen to.

See more details on The Futurum Group at futurumgroup.com.

## Copyright & Use License

### Copyright Notice

Copyright ©2026 by The Futurum Group, LLC. All rights, including that of translation into other languages, are specifically reserved. No part of this publication may be reproduced in any form, stored in a retrieval system, or transmitted by any method or means, electrical, mechanical, photographic, or otherwise, without the express written permission of The Futurum Group futurumgroup.com. United States copyright laws and international treaties protect this publication. Unauthorized distribution or reproduction of this publication, or any portion of it, may result in severe civil and criminal penalties and will be prosecuted to the maximum extent necessary to protect the publisher's rights.

### License Notice

This document may be distributed within the licensed organization only.

The following acts are prohibited:

Transmittal to others outside your immediate organization, including partners, resellers, external consultants, etc., in any media format

Posting on a website that is accessible to others outside your immediate organization

The possession or use within an unlicensed organization

### Limitation of Liability Notice

The information contained herein has been obtained from sources believed to be reliable. The Futurum Group shall have no liability for errors, omissions, or inadequacies in the information contained herein or for interpretations thereof. The reader assumes sole responsibility for the selection of these materials to achieve their intended results. The opinions expressed herein are subject to change without notice.