# Hype Cycle for Data Center Infrastructure Technologies, 2023

Published 13 July 2023 - ID G00791197 - 94 min read

By Analyst(s): Henrique Cecci, Philip Dawson

Initiatives: I&O Platforms

Infrastructure and operations leaders must devise sustainable, secure, scalable, agile, intelligent and resilient physical infrastructure platforms at an acceptable cost. Data center infrastructure technologies help achieve this by focusing on advances that enhance business technology platforms.

**Additional Perspectives**

- Summary Translation: Hype Cycle for Data Center Infrastructure Technologies, 2023 (25 September2023)

**More on This Topic**

This is part of an in-depth collection of research. See the collection:

- 2023 Hype Cycles: Deglobalization, AI at the Cusp and Operational Sustainability

## Strategic Planning Assumptions

By 2027, 75% of organizations will respond to stakeholder pressures by implementing a data center infrastructure sustainability program driven by cost optimization, which is a major increase from less than 5% in 2022.

By 2027, 35% of data center infrastructure will be managed through a cloud-based control plane, which is an increase from less than 10% in 2022.

# Analysis

## What You Need to Know

Infrastructure and operations (I&O) leaders and data center professionals face relentless pressure to improve data center infrastructure technologies for mission-critical legacy environments, support growth initiatives, and optimize reliability and operational costs.

On-premises data center infrastructure technologies are not immune to these trends. However, these technologies are also adapting to address key new industry trends, such as:

- Requirements for new levels of computational capabilities per data center physical spaces, mainly due to intensive use of computing and artificial intelligence (AI), increasing global demand for sustainability.
- The need to accommodate the data center infrastructure advances and the drive for externally sourced services.
- I&O leaders and data center professionals growing trend to adopt new I&O operating models, automation and sustainability practices.

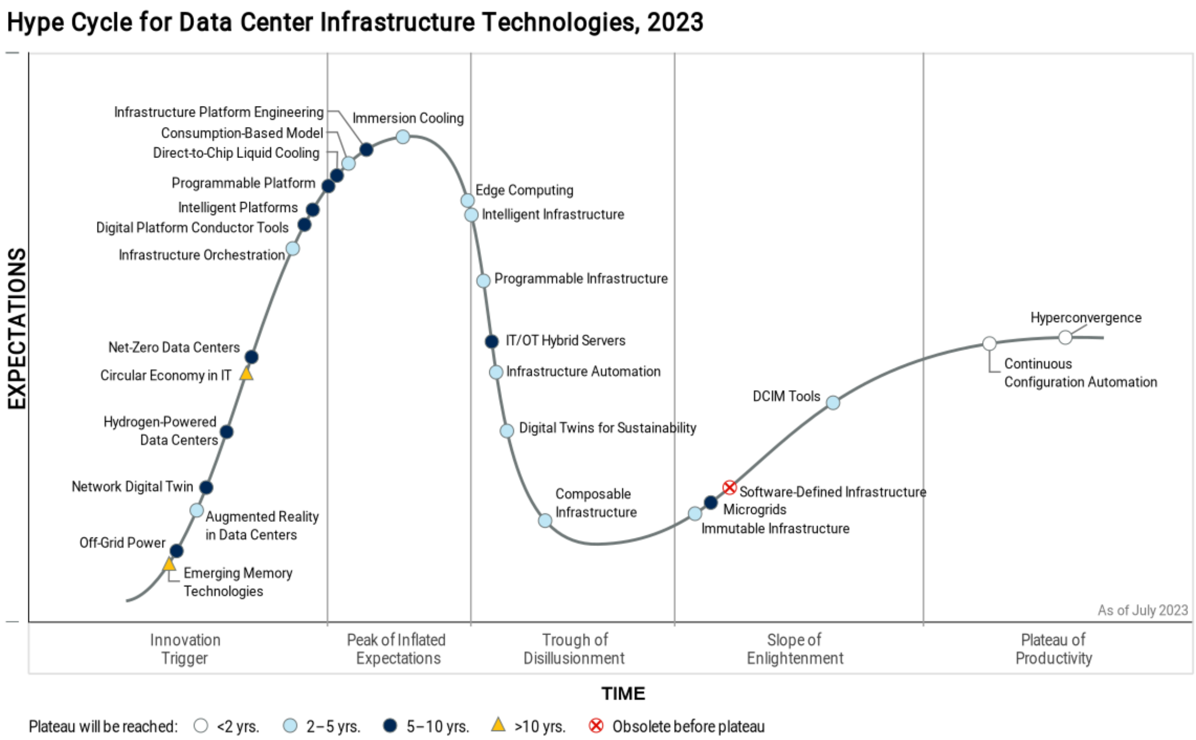
## The Hype Cycle

This Hype Cycle includes 28 discrete infrastructure tools, techniques and technologies, which enable I&O leaders to improve efficiencies, resiliency and sustainability in their data center infrastructures. These make infrastructure systems more efficient to free up funds that are then reinvested for digital business initiatives.

Although their significant growth in newer infrastructure approaches continues, mainly due to sustainability, resiliency, automation and AI initiatives, adoption has slowed for several reasons, such as shifting priorities. During the past decade, data center managers have increasingly emphasized developing and implementing a hybrid cloud strategy. In addition, external options — such as edge computing, colocation, managed service and hosting options — have improved maturity and acceptance by enterprises.

These developments have driven numerous advancements in power, cooling, processing and automation technologies. Enterprises need to take a holistic approach by assessing the available options outside the realm of “only” infrastructure technologies.

## Figure 1: Hype Cycle for Data Center Infrastructure Technologies, 2023




Gartner

## The Priority Matrix

Of the 28 technologies in this Hype Cycle, our analysis indicates that about 43% will plateau during the next five years. Hyperconvergence and continuous configuration automation are already in the mainstream adoption area, with the latter expected to be adopted by nearly 75% of organizations during the next two years.

A strong pipeline of upcoming technologies will help with the digital business and sustainability challenges that data center managers will be facing during the next five years or so. Some have moderate benefits, due to small potential market applications, such as hydrogen-power technologies. However, nearly 50% of the technologies are classified as high or transformative in benefit; as they are all relatively new technologies and approaches. Several deal with the higher heat densities expected in future data centers, including liquid cooling technologies.

**Table 1: Priority Matrix for Data Center Infrastructure Technologies, 2023**

(Enlarged table in Appendix)

| *Benefit* | *Years to Mainstream Adoption* | | | |
|---|---|---|---|---|
| ↓ | ***Less Than 2 Years*** ↓ | ***2 - 5 Years*** ↓ | ***5 - 10 Years*** ↓ | ***More Than 10 Years*** ↓ |
| Transformational | | DCIM Tools<br>Edge Computing | Digital Platform Conductor Tools<br>Network Digital Twin | Emerging Memory Technologies |
| High | Continuous Configuration Automation<br>Hyperconvergence | Composable Infrastructure<br>Consumption-Based Model<br>Digital Twins for Sustainability<br>Infrastructure Automation<br>Infrastructure Orchestration<br>Intelligent Infrastructure<br>Programmable Infrastructure | Infrastructure Platform Engineering<br>Intelligent Platforms<br>IT/OT Hybrid Servers<br>Microgrids<br>Net-Zero Data Centers<br>Off-Grid Power<br>Programmable Platform | Circular Economy in IT |
| Moderate | | Augmented Reality in Data Centers<br>Immersion Cooling<br>Immutable Infrastructure | Direct-to-Chip Liquid Cooling<br>Hydrogen-Powered Data Centers | |
| Low | | | | |

Source: Gartner (July 2023)

## Off the Hype Cycle

Because this is a brand-new Hype Cycle, no technologies have been removed. However, in previous years, Gartner published the Hype Cycle for Data Center Infrastructure, 2017, which featured several technologies that are also included in this Hype Cycle.

## On the Rise

### Emerging Memory Technologies

**Analysis By:** Tony Harvey

**Benefit Rating:** Transformational

**Market Penetration:** 1% to 5% of target audience

**Maturity:** Embryonic

#### Definition:

Emerging memory technologies, such as MRAM and ReRAM, are potential future replacements for DRAM in servers. They have the potential to provide higher density, lower power and persistence, but must show clear advantages over DRAM in order to succeed in the market.

#### Why This Is Important

Emerging memory technologies will be required to maintain future memory scalability growth. Both NAND flash memory and DRAM memory are reaching physical performance limits. The continued increase in flash memory layers (currently 232) demonstrates that scalability is no longer possible by shrinking cell sizes, and it will be increasingly difficult and costly to add more layers. NAND flash roadmaps show up to 1,000 layers, but 3D DRAM remains nascent with commercialization expected in three to six years.

#### Business Impact

Systems using emerging memory technologies enable scale-up computing systems to perform faster and handle larger analytics workloads. Alternatively, these systems can provide greater workload consolidation, reducing costs by shrinking the data center space and power required. In addition, a key impact of this technology will be to accelerate the adoption of in-memory computing architectures.

#### Drivers

There are three key technologies, although phase change memory is no longer being actively manufactured:

- Phase Change Memory — this technology is used in 3D XPoint memory from Intel and is no longer being actively manufactured.

- Magnetoresistive random-access memory (MRAM) — this is a type of nonvolatile memory (NVM) that uses the magnetic properties of controlled electrons to store data.
- Resistive random-access memory (ReRAM) — ReRAM is a type of nonvolatile memory in which the presence or absence of a conducting path within a dielectric between two electrodes is used to determine the stored state.

The advantages of these emerging memory technologies are:

- Resistance-based memory (RBM) technologies lend themselves well to scaling to finer process geometries, potentially leading to higher memory densities.
- RBM can accommodate multiple layers, another approach to increased density.
- As RBMs are persistent, they can support both system memory and storage applications.
- RBMs have lower power consumption and can simplify system design — they are persistent and do not need to be refreshed like DRAM.

### Obstacles

In order to succeed when compared to DRAM or Flash, any emerging memory technology will need to overcome the following obstacles to move to market success:

- Performance must be similar to or faster than server DRAM.
- Cost per GB must be equal to, or significantly, lower than server DRAM.
- Power consumption must be equal to or lower than DRAM.
- System designers will need to be able to integrate DRAM and the Emerging Memory Technology in the same system with minimal changes required by the host.
- Use of the technology in a system must be transparent to applications.
- Must be available from multiple sources in modules that can be used interchangeably.

### User Recommendations

- Engage with memory and server vendors by understanding their long-term plans for emerging memory in their server systems.
- Stress the need for application transparency and the need for a new technology to have significant advantage over DRAM to enable adoption by using the failure of Intel Optane.
- Ensure that vendors understand the need for a general purpose solution that, like DRAM, can be supplied by multiple vendors and used across a broad variety of CPUs and platforms.

### Sample Vendors

Avalanche Technology; Everspin Technologies; Hewlett Packard Enterprise; IBM; KIOXIA; Micron Technology; Samsung Electronics; SK Hynix; Spin Memory; Western Digital

### Gartner Recommended Reading

Emerging Tech Impact Radar: Compute and Storage

Emerging Technology Horizon for Semiconductors and Electronics, 2022

Emerging Tech Impact Radar: Semiconductor and Electronics Technologies

Emerging Tech: Compute Express Link Redefines Server Memory Architectures

2022 Strategic Roadmap for Compute Infrastructure

## Off-Grid Power

**Analysis By:** Tony Harvey, Autumn Stanish, Jason Donham, Philip Dawson

**Benefit Rating:** High

**Market Penetration:** 1% to 5% of target audience

**Maturity:** Embryonic

Definition:

Off-grid power supply and delivery is a growing option for high-power-demanding data centers and can contribute to reducing the carbon footprint of on-premises data centers. It can be used to reduce the load on already overstressed power grids as well as provide higher reliability in areas where power delivery can be unreliable.

### Why This Is Important

High-capacity data centers now require megawatts (MW) of power, with some of the larger data centers' campuses requiring hundreds of MW or more (see Google's announcement on sustainability). This is putting inordinate pressure on power generation and distribution networks, and global power prices have grown by 22% since February 2022. Organizations that need to run on-premises data centers must look for additional power capacity and locations from a network that simply cannot continue to provide them.

### Business Impact

The lack of power availability can halt or pause data center expansion. This will delay the implementation of new IT capabilities as well as increase costs for data center alternatives, such as colocation space and cloud services. By investing in off-grid power, businesses mitigate the risks of lack of power, grid reliability issues or unpredictable volatility in the wholesale energy markets.

### Drivers

- Energy costs and requirements are growing exponentially for all data centers, as there is more demand for their services.

- Distribution networks for electricity are already overstretched and the move to electrification, for cooking, heating and transport, is only exacerbating the issue. There have already been moratoriums on new data center builds in some countries, and this is expected to continue (see Dutch Call a Halt to New Massive Data Centres, While Rules Are Worked Out, DutchNews.nl).

- Sustainable energy sources, such as solar, tidal or wind, are unpredictable and must be backed up with a continuous power source.

- The new concept of small modular reactors (SMRs) offers between 5 MW and 1,000 MW of power. An SMR could be used to power a data center or a number of data centers. The first SMRs are in production in Russia and China (see IAEA Ups Support for SMRs, Nuclear Engineering International). While other SMRs have been licensed by their local nuclear regulators, they do not yet have a go-live date.

### Obstacles

- A large-scale capital investment is required for setting up an SMR, large-scale wind farm or solar array, however, data centers typically have a 15- to 20-year life span, allowing a long-term view of ROI.
- There are security and planning permission impacts in running an SMR, with additional security controls and deterrence needed, that would not normally be required except for Tier 4 data centers.
- Nuclear SMR reactors, while sustainable, have a PR problem and will generate political and planning opposition.
- Local or central government licensing must be considered, with some countries averse to nuclear power due to political pressure, or to large-scale wind and solar farms due to local resident resistance.

### User Recommendations

- Review your data center expansion plan by assessing any issues related to grid capacity, reliability, and/or energy price volatility in current and future locations.
- Prepare for off-grid generation by investigating what investments in solar, wind, SMR, fuel cell, heat or multisource power generation could be used to power your current and planned data center facilities.
- Ensure that ROI calculations include the value of off-grid generation by evaluating a data center build project for a 10- to 15-year investment period, and include the possibility of incentives for sustainable power or off-grid capability.

### Gartner Recommended Reading

The Road to a Net Zero Data Center

Quick Answer: How Do I Assess On-Site Renewable Energy Options?

Forecast Analysis: Data Center Sites, Worldwide

Toolkit for Estimating Data Center Build and Modernization Costs by Tier Level

Strategies to Plan for GHG Emissions Reduction

### Augmented Reality in Data Centers

**Analysis By:** Henrique Cecci

**Benefit Rating:** Moderate

**Market Penetration:** 5% to 20% of target audience

**Maturity:** Emerging

#### Definition:

Augmented reality (AR) technology has several applications in data centers, particularly in managing, securing, maintaining and expanding data center infrastructure, troubleshooting, and staff training.

#### Why This Is Important

AR in data centers provides an innovative way to manage, operate, and maintain data center infrastructure. It may help infrastructure and operations (I&O) teams work more efficiently and effectively. With AR, data center operators can visualize and manipulate virtual objects in real time, improving the accuracy and speed of tasks such as troubleshooting, maintenance and repairs, and can also enhance staff training and education programs.

#### Business Impact

AR in data centers can deliver improved productivity, cost savings, enhanced safety, improved training and knowledge transfer, faster issue resolution, and reduced downtime. AR provides technicians with real-time data visualization, remote guidance, and interactive 3D models to facilitate complex tasks and maintenance processes. It can eliminate the need for physical manuals and reduce the time needed for troubleshooting and repairs.

#### Drivers

Key drivers for using AR in data centers include:

- **Improved efficiency:** AR enhances the accuracy and speed of data center maintenance and repairs, reducing downtime and increasing uptime.

- **Cost savings:** AR data center technicians identify issues more quickly and accurately, reducing the need for additional labor and minimizing the risk of equipment damage.

- **Improved safety:** AR provides technicians with real-time information about potential hazards, allowing them to take proactive measures to mitigate risks and avoid accidents.

- **Enhanced training:** AR is used to provide immersive and interactive training experiences for data center technicians, enabling them to develop new skills and knowledge more quickly and effectively.

- **Improved customer experience:** AR helps data center operators provide a more engaging and informative experience for customers, improving satisfaction and retention.

### Obstacles

Typical obstacles for using AR in data centers include:

- **Technology complexity:** Implementing and maintaining an AR system is technically challenging, requiring a combination of hardware, software and connectivity.
- **Integration:** Integrating AR with existing data center management tools like existing data center infrastructure management (DCIM) or IT service management (ITSM) tools and systems is challenging and complex.
- **Cost:** The cost of implementing AR solutions is high, especially for small and medium-sized data centers.
- **Training:** AR technology requires specialized skills and training.
- **Security:** Data center security is crucial, and AR potentially introduces new security risks that must be addressed.
- **Adoption:** The adoption level of AR technology in data centers is still relatively low due to a lack of awareness and understanding of its benefits.

#### User Recommendations

- Clearly define your goals and objectives, for example, improving maintenance, reducing downtime, or enhancing employee training.
- Start with a small pilot project to test the effectiveness of AR technology before scaling up to larger and/or production deployments. Focus on integration and compatibility with existing I&O tools and systems like ITSM, DCIM or others.
- Ensure AR adoption complies with your organization's security policies, as well as with industry regulations and standards. Train staff in standardized processes and necessary skills.
- Continuously evaluate the effectiveness of AR in your data center operations and identify areas for improvement to maximize the benefits and ROI.

#### Sample Vendors

Axonom; Dell EMC; Inceptum; Nlyte; Schneider Electric; Vertiv

#### Gartner Recommended Reading

Emerging Technologies: Tech Innovators in Augmented Reality — AR Cloud

Emerging Technologies: Tech Innovators in Augmented Reality — Spatial Web

Emerging Technologies: Tech Innovators in Augmented Reality — Augmentation and Spatial Interaction Layer

### Network Digital Twin

**Analysis By:** Tim Zimmerman, Andrew Lerner

**Benefit Rating:** Transformational

**Market Penetration:** Less than 1% of target audience

**Maturity:** Emerging

Definition:

A network digital twin is a model of the behavior of campus, software-defined WAN (SD-WAN) or data center network components. It is usually delivered as software and provides a model that can be used for validation of the configuration or policies of a single network component or the entire network. It automatically synchronizes with the production network.

### Why This Is Important

The complexity of enterprise networks continues to grow, coupled with the business demands for faster and accurate delivery of network activities, such as changes. In addition, there is an increase in network updates and the limited availability of skilled resources mandates the automation of testing to reduce the costs of shadow IT and duplicate systems. A network digital twin allows for the validation of configuration and security policies made to individual components.

### Business Impact

For IT leaders, a network digital twin allows faster testing and consequent delivery of network changes with fewer personnel resources and less cost by reducing shadow IT equipment requirements. We believe a network digital twin can improve delivery times for requests by 20% across the network.

Drivers

- **Lack of time to test network component updates:** As network vendors adopt agile development processes, new versions of operating systems and applications are being delivered at a much increased pace compared to several years ago. Most IT organizations do not have the ability to completely test one version before the next version arrives.
- **Amount of configuration errors due to network complexity:** Over 80% of network problems are due to improper configuration and change management.
- **Cloud migration issues:** More than 15% of security breaches are caused by misconfigured cloud services
- **Increasing desire to improve automation:** Using an automated pipeline to deliver data center network changes.
- **Need for virtual production networks:** Organizations struggle to replicate, simulate or emulate their production networks due to cost of equipment or operational expense to keep the environment synched with production systems.
- **Training:** Organizations can use a network digital twin to train new employees in a lower-risk environment.

### Obstacles

- **Amount of physical resources required:** The compute and memory resources required to model a discrete component will require that modeling and testing be completed in the cloud.
- **Ability to model existing discrete network components:** Creating network models and test suites is complex and will require skill sets that are different from those that are currently available in IT.
- **Creating composite networks:** Connecting discrete network components for a single vendor is difficult but network digital twin tools and IT skills will need to improve before multivendor composite networks can be built.
- **Lack of standards:** Standards are not yet in place and current solutions may be proprietary/vendor specific.
- **Lack of trust:** Enterprise network teams lack trust in digital twins either due to immaturity of the technology, or their lack of experience with it.
- **Lack of awareness:** The technology is relatively new and immature.
- **Lack of a clear business case:** To justify spending on the new technology.

### User Recommendations

- Pilot a network digital twin if you're looking to roll out data center networking changes in an automated CI/CD pipeline.
- Prefer a network digital twin that is delivered as a service, offers subscription-based pricing and supports multiple vendors.
- Deploy a campus network digital twin to validate changes to existing campus network configuration, changes in security policies or application migration to the cloud.
- Calculate ROI for network digital twin saving resulting from preventing security and configuration issues.
- Educate senior management on the value of discrete and composite network digital twin modeling.

### Sample Vendors

Ericsson; Extreme Networks; Forward Networks; Intentionet; Keysight; Nokia; NetBrain

### Gartner Recommended Reading

Quick Answer: What Is a Digital Twin?

Emerging Tech: Venture Capital Growth Insights for Digital Twins

## Hydrogen-Powered Data Centers

**Analysis By:** Henrique Cecci

**Benefit Rating:** Moderate

**Market Penetration:** Less than 1% of target audience

**Maturity:** Adolescent

### Definition:

Hydrogen-powered data centers use hydrogen fuel cells as a backup or primary energy source instead of traditional sources like diesel generators or the grid. They offer benefits such as improved efficiency and reliability, low to zero carbon emissions, low maintenance requirements, modularity, and potential cost savings, as well as additional implementation costs.

### Why This Is Important

Hydrogen-powered data centers provide environmental sustainability, energy resilience, cost savings, flexibility, scalability, and improved public perception. They offer a sustainable, reliable, and possibly cost-effective alternative to traditional energy sources with the potential to play an increasingly important role in meeting our energy requirements while minimizing environmental impact.

### Business Impact

Hydrogen-powered data centers can have significant business impacts, including cost savings, improved reliability, improved environmental performance, competitive advantage, and regulatory compliance. They offer a potential long-term solution for reducing operating costs, enhancing reliability, complying with regulations, and appealing to environmentally conscious customers.

Drivers

- **Environmental sustainability:** Data centers are significant consumers of energy and can have a large carbon footprint, and hydrogen fuel cells offer a sustainable and low to zero-emissions alternative to traditional energy sources.
- **Energy resilience:** Hydrogen fuel cells can provide a reliable and independent source of energy, reducing the risk of downtime and potential revenue loss.
- **Cost savings:** Hydrogen fuel cells can potentially reduce the operating costs of data centers over time due to their higher efficiency and lower maintenance requirements.
- **Flexibility and scalability:** Hydrogen fuel cells can be modular and scalable, allowing data centers to add capacity as needed without significant infrastructure changes.
- **Improved public perception:** Adopting hydrogen fuel cells can improve public perception of data centers by demonstrating a commitment to sustainability and innovation.

Obstacles

- **Maturity of the technology:** As it's a relatively new technology organizations may have concerns regarding the safety of handling and storing large quantities of hydrogen due to its flammability.
- Higher implementation and operational costs.
- **Physical footprint:** Hydrogen fuel cells may require large amounts of surface area, which can be difficult to come by in the case of data centers located in urban or suburban areas.
- **Transportation and storage:** It may require specific infrastructure for transportation, storage, handling and distribution. For example, pipelines for high-pressure gaseous hydrogen, while road, rail and sea routes would normally use low-temperature liquid hydrogen.

#### User Recommendations

- Focus on the quantitative environmental and sustainability benefits.
- Conduct a feasibility study before investing in hydrogen-powered data centers, and assess their technical and economic viability as well its risks and challenges.
- Evaluate the supply chain including the availability of hydrogen fuel, transportation, storage and distribution infrastructure.
- Assess safety considerations for handling, storing and using hydrogen fuel cells owing to the flammable nature of hydrogen.
- Start with a hybrid approach by integrating hydrogen fuel cells into the data center's power system.
- Plan for future growth by designing for modularity and scalability.

#### Sample Vendors

Bloom Energy; bp; Cummins; ECL; Plug Power; Schneider Electric

#### Gartner Recommended Reading

Industry Insights: Green Hydrogen Accelerating Transition Toward Clean Energy

2023 Utility Trend: Green Hydrogen Expectations Are High, but So Are Challenges

### Circular Economy in IT

**Analysis By:** Simon Mingay, Sarah Watt

**Benefit Rating:** High

**Market Penetration:** 5% to 20% of target audience

**Maturity:** Emerging

Definition:

A circular economy enables IT vendors and end-user enterprises to decouple consumption of IT assets from growth in demand. Three principles form a basis of a circular economy: design out waste, keep materials in use at the highest quality for long and return materials to the environment in a way that has a positive impact. The IT sector must meet strong regulatory requirements for e-waste in many places, notably the EU, but has challenges around engaging strategically with circularity.

### Why This Is Important

IT asset life cycles are short. In a linear economy, old products become seen as waste; either lost to landfill or selectively recycled. The circular economy maximizes value. Assets are supported, repaired, upgraded, returned, refurbished and resold. Successful deployment relies on thoughtful design, new materials, reverse supply chains, the economics of residual value, repair and refurbishment partners and the application of new business models. A circular economy limits environmental impact, grows market share and creates new experiences.

### Business Impact

Positive business impacts of the circular economy include enabling new “as-a-service” business models, meeting customer needs in constrained supply environments and acting as a hedge against raw material price increases. These benefits are realized when product design, support, parts availability, warranty, usual technology obsolescence issues, are addressed, and supply chain mechanisms are in place to pull back strategic products from the market for refurbishment.

Drivers

- **Legislative Drivers:** The “right to repair” provides access to spare parts and technical information to extend product life. EU waste policy aims to ensure that high-quality resources are not lost from the economic system. Concerningly, 60% of EU household e-waste still goes to landfill (see E-Waste in the EU: Facts and Figures (Infographic), the European Parliament).

- **Supply Chain Fragility:** Fragile supply chains can be addressed through the circular economy either by meeting customer demand through second-life products or by reclaiming raw materials for manufacturing new products.

- **Climate Change:** A circular economy has the potential to reduce climate change impacts, as product-embodied energy (and emissions) is used more efficiently. By slowing the rate of material, component, *new* product consumption, demand for new raw materials (and their concomitant emissions) is reduced. Hardware vendors must undertake life cycle analysis to review the environmental impacts of end-of-life options, enabling trade-off decisions to be made (for more information, see A New Circular Vision for Electronics, World Economic Forum).

- **Enhanced Value:** The circular economy enables enterprises to offer new, more sustainable business models and products, sustainability-focused performance analytics and a differentiated sustainability narrative.

- **Innovation:** The circular economy is a catalyst for innovation by equipment vendors. Examples include modular design for repairability, design for reuse and longevity, use of innovative materials such as bio-based components, and design for disassembly (for more information, see The Global Electronics Council).

- **Cloud:** Cloud services are inherently based on resource-sharing business models and offer significant potential to dematerialize parts of the IT sector, should the cloud service providers choose to make the most of that potential.

### Obstacles

- **Mindset Shift:** A mindset shift is needed to scale the circular economy for IT.
- **Partnerships:** Scaling the circular products relies on engaging with ecosystem partners. Partner capabilities vary by market.
- **Impact:** No clear harmonized metrics or standards have been agreed on to report progress on circular activities. Comparison between enterprises is challenging.
- **Product Complexity:** Electronic devices are complex, frequently containing materials that require careful, expensive disassembly and specialized processing. Materials toxicity presents a reuse, health and safety challenge.
- **Obsolescence:** Premature obsolescence is a significant problem in the IT sector. In part due to design and vendor business model issues, and in part enterprise procurement and asset management practices.
- **Fashion:** Technology as fashion accessory, and the need for the latest, slimmest, sleekest, fastest device.
- **Security Policies:** Many enterprises have security policies that demand that media (such as hard-disk drives and SSDs) must be shredded. More sustainable software sanitization options available.

### User Recommendations

- **Select Circular Solutions:** Favor public cloud solutions. Focus on cloud providers demonstrating leading circularity practices. Select products with improved circularity performance by making use of ecolabels such as EPEAT and TCO Certified. Favor products that avoid materials of concern, such as mercury, PVC and brominated flame retardant.
- **Assess OEM Circularity Ethos:** Engage with your key OEMs, and assess the extent to which they are designing their products and business models for circularity. Favor vendors that demonstrate a strategic and systemic approach to circularity.
- **Extend Utility:** Look to extend the life of less energy-intensive devices such as phones, PCs and monitors. Increase asset utilization, and minimize overprovisioning capacity.

### Gartner Recommended Reading

Industry Insights: The Value of Circular Economies

Build a Circular Supply Chain to Unlock Growth and Improve Environmental Impact

3 Circular Economy Interventions That Will Mitigate Supply Chain Disruption

3 Accelerators to Advance the Circular Economy in Supply Chain

Quick Answer: How Will Degradable Electronics Enhance Sustainability?

### Net-Zero Data Centers

**Analysis By:** Simon Mingay, Philip Dawson, Autumn Stanish, Matthew Brisse

**Benefit Rating:** High

**Market Penetration:** 20% to 50% of target audience

**Maturity:** Emerging

#### Definition:

Net-zero data centers aim to achieve a net-zero climate target by minimizing their direct and indirect greenhouse gas (GHG) emissions, and offsetting the balance appropriately. They should be able to demonstrate world-class energy and water efficiency, circularity practices, and asset utilization.

#### Why This Is Important

Enterprises and data center/cloud service providers are under mounting pressure from customers, investors, regulators and employees to reduce GHG emissions, increase energy efficiency, and establish a credible sustainability narrative. Pressure is growing on infrastructure and operations (I&O) leaders to adopt methods to increase transparency and performance, in order to make their data center operations efficient and environmentally sustainable.

#### Business Impact

IT leaders are confronted by a significant, unrelenting, year-on-year growth in compute and storage capacity, with spiraling energy consumption and associated GHG emissions. For enterprises with energy and GHG reduction targets, this is unsustainable. To meet cost targets, organizations must make radical improvements in data center, and I&O efficiency and emissions. Net-zero data centers will be essential infrastructure for all enterprises with ambitious or net-zero GHG targets.

Drivers

- Alignment of I&O with the organization's energy and GHG targets.
- External stakeholders' (specifically customers, investors and regulators) net-zero expectations.
- The need to mitigate the liability associated with costs of energy and GHG emissions.
- The need to build resilience in the face of increased contention for electrical and renewable supply capacity.
- Increased emphasis on cost reduction in IT systems.
- Increased consumption of cloud services, moving workloads out of the enterprise data center.
- Data center consolidation, caused by office consolidation following the COVID-19 pandemic, end of economical life of data centers and refocus of business operations.
- Customers demanding GHG footprint transparency from data center and cloud service providers.

### Obstacles

- Unrelenting and significant growth in compute and storage capacities.
- Lack of a strong business case in the absence of ambitious enterprise GHG reduction goals.
- Lack of availability and cost of renewable energy, along with lack of capital to invest in power purchase agreements.
- Imprecise, complex and costly measurement, management, and mitigation of Scope 3 GHG emissions.
- The immaturity of circular economy practices and services.
- Costs of transition to more efficient cooling and HVAC systems, using technology such as immersion and free air cooling.
- Reducing water consumption.
- Lack of cost-effective low-carbon alternatives to diesel generators.

### User Recommendations

- Secure a long-term supply of renewable energy.
- Measure the success of the data center sustainability program in broader enterprise sustainability initiatives by creating realistic KPIs for GHG emissions and water consumption.
- Conduct an audit of the data center's GHG emissions, waste and water consumption to understand its footprint. Liaise with the enterprise's sustainability, energy management, real estate and legal teams to build the business case and support the program.
- Reduce the data center's emissions footprint by investing in energy efficiency measures, GHG, and water and waste reduction.
- Follow through the full green value chain and do not consider the cloud as a legitimate offset of responsibility.
- Support the development of global industry/university consortia to focus on innovative solutions and standards for low-power computation, and data communication and storage.

### Gartner Recommended Reading

Strategies to Plan for GHG Emissions Reduction

Building a Low-Carbon Energy Strategy

Ignition Guide to Building a Net-Zero Greenhouse Gas (GHG) Emissions Roadmap

Maverick Research: Net Zero Will Stall Tech Growth and Innovation

Toolkit for Estimating Data Center Build and Modernization Costs by Tier Level

## Infrastructure Orchestration

**Analysis By:** Chris Saunderson

**Benefit Rating:** High

**Market Penetration:** 5% to 20% of target audience

**Maturity:** Emerging

### Definition:

Infrastructure orchestration (IO) enables platform and I&O teams to design, deliver, operate and ensure orchestrated services across on-premises, cloud and edge deployments. IO enables templated service creation and management, spanning provisioning, Day-2 operations and integration with CI/CD, self-service portals, and API access to orchestrated services.

### Why This Is Important

Infrastructure orchestration provides strategic workflow capabilities to drive life cycle delivery and ongoing maintenance of complex deployed infrastructure. These practices and tools enable agile, iterative automation delivery and execution of the processes required via self-service and API access. This investment improves the velocity and quality of infrastructure services, improves traceability and visibility of service delivery and reduces inconsistencies from manual activities.

### Business Impact

Infrastructure orchestration drives consumer experience improvements of deploying and managing standardized infrastructure. I&O teams realize operational efficiencies through reduced manual efforts to deliver infrastructure, embedding security and compliance requirements into the delivered services, and offering cost optimization opportunities. I&O staff can transform their role into an automation-first focus and scale to meet increased business demands.

### Drivers

- **Business agility**: Organizations must increase responsiveness to meet customer needs and adapt to market and technology changes. They must be able to deliver products that meet these changing demands and requirements quickly.
- **Cost optimization:** Infrastructure teams leverage orchestration to deliver scalable, reliable and secure platforms. This helps to improve delivery efficiency, reduce human work, and reduce downtime due to change failures.
- **Value extraction**: adoption of orchestration capabilities unlocks additional value from the automation tools already implemented, enabling incident response, request servicing and other tasks to be more richly automated and consumed.
- **DevOps**: Infrastructure orchestration is a key enabler of continuous software delivery, allowing the DevOps team to automate the provisioning and management of environments.
- **Infrastructure complexity**: Increasingly complex deployment topologies require greater automation to improve the consumability of infrastructure and the ongoing maintenance of deployments,
- **Security and compliance**: Increased automation enables the implementation of security and compliance controls through orchestration and avoids any audit failures. The end-to-end visibility and traceability of the provisioning and configuration can enable continuous compliance automation of the infrastructure.

### Obstacles

- **Skill development:** Infrastructure orchestration practices and tools can be complex to implement and sustain, as they require skills beyond scripting to get maximum value. These tools leverage software engineering skills that can be challenging to find in I&O teams.
- **I&O operating models:** The organizational structure of many I&O teams is set up by domain specializations, making it hard to develop and deliver end-to-end services through orchestration. Perceptions of stability and reliability risks slow adoption.
- **Automation constraints:** To automate maintenance activities, a certain level of maturity needs to be reached within the organization. Orchestration requires that automated tasks be available to be able to realize maximum return on investment.

### User Recommendations

- Identify and catalog use cases and constraints in your delivery workflows that are injecting delay into service delivery, especially for tasks that are executed manually.
- Benchmark existing service delivery execution time and quality problems to measure against to demonstrate improvement.
- Catalog operational tasks that are being executed manually today and are candidates to develop workflows to implement.
- Identify candidate orchestration platforms to execute proof of value testing with, ensuring that the candidates can be integrated into your existing operational environment.
- Monitor implementation to identify successes and opportunities for improvement and build a success story demonstrating velocity, quality, throughput and operational improvements.

### Sample Vendors

Cloudsoft; Crossplane; Dell Technologies; env0; Itential; Morpheus Data; PagerDuty; Pliant; RackN; SpaceLift

### Gartner Recommended Reading

Innovation Insight for Continuous Infrastructure Automation

To Automate Your Automation, Apply Agile and DevOps Practices to Infrastructure and Operations

Market Guide for Infrastructure Automation Tools

Market Guide for Continuous Compliance Automation Tools in DevOps

## Digital Platform Conductor Tools

**Analysis By:** Roger Williams, Dennis Smith

**Benefit Rating:** Transformational

**Market Penetration:** Less than 1% of target audience

**Maturity:** Embryonic

### Definition:

Digital platform conductor (DPC) tools coordinate the various infrastructure tools used to plan, implement, operate and monitor underpinning technology and services for applications and digital products. They enable digital business, regardless of the environments used or who owns them. DPC tools provide a unified view of underpinning technologies and their connection to applications. This augments strategic decision making and improves the value obtained from technology investments.

### Why This Is Important

Traditional, cloud and hybrid infrastructure management tools do not inherently provide an integrated view of infrastructure across all environments. Moreover, as infrastructure and operations (I&O) leaders struggle to manage their portfolio of investments to enable composable business, optimize costs and reduce risks, they need help with filling the gaps in visibility, assurance and coordination. DPC tools promise to help close these capability gaps and are improving in their ability to do so.

### Business Impact

DPC tools deliver the following benefits not inherent in more focused infrastructure management toolsets:

- Visualizing digital platform performance across all life cycle stages — planning, implementing, operating and monitoring.

- Enabling continual optimal performance and placement of workloads in all environments — on-premises, in the cloud or at the edge.
- Ensuring tangible business value from improvement efforts across all technology architectures — compute, storage, middleware and network layers.

Drivers

- Difficulty in maintaining a coherent view of all technology infrastructure resources and their dependencies that are aligned with changes to services, applications and components, as well as the configuration of their promised performance levels.
- Lack of transparency into spending on hybrid digital infrastructure and how resource capacity aligns with actual application workload demand.
- Need to guide where workloads are processed (data center, public cloud, colocation facility, etc.) based on requirements, including capacity, cost and dependency dynamics.
- Challenges with estimating the value, efficiency, quality and compliance delivered by hybrid digital infrastructure based on aggregated data from performance analysis tools and other hybrid digital infrastructure management (HDIM) toolset data feeds.
- Desire for a single point of entry and reporting for digital platform resource requests, and routing them to appropriate HDIM tooling for fulfillment.
- Desire to reduce the level of skills and effort required within initiatives to improve operations and digital employee experiences.
- Gaps, duplication and conflicts in data to support application workload migration and business continuity goals, as well as protection of data from accidental deletion or malicious activities.
- Inability to confirm compliance of application workloads and digital platforms to identity requirements and security baselines as part of the organization’s cybersecurity mesh approach.
- Poor credibility of business cases for digital platform improvements, including: assessing business impact; measuring gaps between current and desired performance; providing oversight of improvement efforts; and validating benefits delivered.

### Obstacles

- Lack of interoperability: Tool sprawl and difficulties in integration inhibit DPC tool adoption. The technology landscape is littered with failed approaches that were intended to support data sharing between vendors.
- Lack of data credibility: The desire for a complete, accurate view of all technology as a precondition for decision making has been around for decades, yet is no closer to being realized. Customers that demand perfect data before they act, and vendors that require complete and accurate data for their tools to function properly, will continue to co-create expectations that will not be met.
- Lack of budget: DPC tools may be viewed as “overhead” that does not have a compelling business case. No one likes paying for something that does not appear to address specific pain points felt today.
- Lack of vendor commitment: Many vendors will be tempted to “DPC wash” their existing offerings and claim that these capabilities are already addressed or can be added for very little cost.

### User Recommendations

- Build a DPC tooling strategy that supports digital business ambitions by defining the management elements, environments and technology layers required to meet the organization’s infrastructure needs now and in the future.
- Address measurement and coordination gaps by working with key stakeholders to identify infrastructure value and risk and cost objectives, and by making targeted investments in integration, dependency mapping and continuous improvement capabilities.
- Plan for DPC tooling investments by determining which DPC capability aspects are needed in the short, medium and long term. Compare these capabilities to current and future vendor offerings for infrastructure management tooling that can provide initial DPC tool functionality.
- Ensure that DPC tooling investments can deliver sustained value by requiring that DPC tool marketers show how the tool will address current organizational pain points and how it will adapt to future needs as organizational requirements evolve.

### Sample Vendors

Cloudsoft; Flexera; HCLTech; IBM (Turbonomic); Oomnitza; OpsRamp; ReadyWorks; Snow Software; Virtana

### Gartner Recommended Reading

Market Guide for Digital Platform Conductor Tools

3 Steps to Improve the Reliability of Large, Complex and Distributed IT Systems by Leveraging SRE Principles

## Intelligent Platforms

**Analysis By:** Philip Dawson, Nathan Hill

**Benefit Rating:** High

**Market Penetration:** 5% to 20% of target audience

**Maturity:** Emerging

### Definition:

Intelligent platforms provide the administration composability of infrastructure and programmable API functions with automated infrastructure intelligence. They integrate compute, storage and networking assets with some or the entire application software stack, creating dedicated workload architectures. Intelligent platform vendors also include components such as application intimacy, management tools, OSs, and virtualization bought and/or consumed as a service.

### Why This Is Important

Intelligent platform solutions are differentiated against integrated system or hyperconverged infrastructure (HCI) solutions with a separate software stack purchase tied to the hardware. Pricing strategies vary greatly throughout the integrated software stack solutions as part of the shift to consumption-based infrastructure delivery. Intelligent platforms also integrate applications and business logic as bundles and partnerships.

### Business Impact

Intelligent platforms optimize:

- Delivery of workload performance or application manageability that crosses over from hardware that promises lower operational costs and increased IT agility via automated, pooled resources.
- Automation and machine learning of complete stacks, hardware administration and software programmability on top of integrated systems.
- They are stand-alone running proprietary workloads that rarely compete with each other as the software stacks set the hardware options.

Drivers

- The intelligent platform market is influenced by multiple aspects of resilience and availability across on-premises, hosting or colocation and cloud locations driving composable, programmable and intelligent functions.
- Intelligent platforms are integrated as everything as a service (XaaS), with automation and management, and differ from integrated stack systems, which are hardware-integrated dedicated appliances.
- Multiple vendors are driving the market for intelligent platforms around integrated systems, HCI, cloud and virtualization. Intelligent platforms are built from a software perspective on top of HCI rather than a traditional integrated stack system that is built as a hardware appliance around hyperconverged integrated systems (HCIS).
- Vendors such as Microsoft, Nutanix and VMware are promoting valid intelligent platform software, and the market momentum around HCI software in the cloud now creates a market for multiple hardware vendors to build software management and integration services.

### Obstacles

- Hybrid and multicloud strategies may not integrate well with integrated platforms, continuing the silo mentality of cloudlike delivery.
- Other platform as a service (PaaS) momentum is being integrated from packaged vendors such as SAP and Oracle, which are bundling integrated stack systems and distributed cloud infrastructure with application platform and database management system (DBMS) software. Here, the intelligence is with the PaaS software, not the intelligent infrastructure.
- An intelligent platform provides balanced XaaS workload performance, application optimization and integration, but this comes at the expense of greater vendor dependency, and inflexibility for future application customization and workload requirements.

### User Recommendations

- Select infrastructure software management frameworks for overlay management as well as links to cloud infrastructure. Do not implement hardware-dependent or locked-in intelligent platform frameworks and adapters.
- Define successful intelligent platform implementations by assessing data center stakeholders and other vested interests (for example, procurement) with other lines of business responsible for agreeing with SLAs.
- Automate the infrastructure requirements for cloud management platforms (CMPs) through the use of intelligent platforms as you deliver XaaS through infrastructure platforms.

### Sample Vendors

CU Coding; DataDirect Networks (DDN); Dell Technologies; Hewlett Packard Enterprise; Microsoft; Nutanix; Oracle; VMware

### Gartner Recommended Reading

How to Evolve Your Physical Data Center to a Modern Operating Model

Quick Answer: How Can I Optimize the Use of Programmable Platforms for Effective Software Delivery?

How Do I Plan for Migrating My Data Center Infrastructure Into an XaaS Model?

## At the Peak

### Programmable Platform

**Analysis By:** Philip Dawson, Bill Blosen

**Benefit Rating:** High

**Market Penetration:** 1% to 5% of target audience

**Maturity:** Emerging

#### Definition:

Programmable platforms are API-driven for delivery of applications in a cloud model by using and applying methods and tooling from the software development area to management of IT infrastructure and platform concepts. It includes resilient platform architectures and agile techniques.

#### Why This Is Important

Software engineering and infrastructure and operations (I&O) have coexisted but been separated by their differing toolsets and APIs. Modern digital businesses need their software engineering and I&O teams to deliver a cohesive platform that encompasses both application and infrastructure delivery. Programmable infrastructure delivers the underlying technical capability that enables this integration.

#### Business Impact

Greater value and agility (rather than cost optimization) is achieved via programmable platforms' ability to drive adaptive application delivery. Programmable infrastructure, API provisioning and automated processes allow faster responses to new business demands, driving service quality and freeing application delivery staff and administration staff from infrastructure functions. Programmable platforms enable a sustainable and highly responsive IT infrastructure service to the business.

#### Drivers

- Software architects and engineers are moving to modular distributed applications built on containerization, control, data and service architecture. In essence, they use the pattern of separating the application front end from the business logic back end.
- API layers are being adopted through self-service capabilities and organizing the programmable platform APIs into paved roads.

- Software engineers are reducing the cognitive load of using APIs and programmable platforms, improving developer experience, driving productivity and retention of key talent, and also improving adherence to architectural and security guardrails.
- I&O teams are moving workloads and application delivery to cloud infrastructure and platforms as in anything-as-a-service (XaaS) models. In essence, they have embraced programmable infrastructure, that is, applying software development methods, APIs and tooling to manage the control and data planes around I&O services.
- The incumbent architecture of programmable platforms is deploying modular building blocks. Updates to modules are automatically rolled out to any platform that is currently using that module rather than updating each platform discreetly, leveraging efficiency, scale and management dependencies.
- Platform engineering principles guide the programmable platform buildout using templates, APIs and automation to simplify the usage and adoption of programmable infrastructure. The platforms being built must respond to the developers pain points and ease adoption of the most used functions. Agile product ownership is the best model to build feedback loops between the developer programmable platform teams and the platform's users and communities.

### Obstacles

- The boundaries between programmable platforms and infrastructure are at best emerging. I&O teams must be conscious of how their architectures and deployments interface with the programmable infrastructure at the control plane and data plane.
- The architecture and order of APIs across and up the stack and across layers need planning and integration, increasing lock-in. I&O and software teams conceptualize these layers as part of one programmable platform with four distinct layers: the application presentation front end, the business logic back-end functions, the control or management, and data tier repository. Moreover, programmable platforms are restricted in both topology and maturity, which drives clients toward platform as a service (PaaS) and cloud delivery.
- Programmable platform governance is limited by the lack of standardization of the APIs. In balance, established mature APIs drive engagement or interface between the layers managed through well-defined and structured APIs.

### User Recommendations

- Use platform engineering principles to improve developer experience and ease the cognitive load of programmable platforms. Before considering programmable platforms, balance the aversion to vendor specific services lock-in which is prevalent in PaaS and XaaS.
- Deliver platform engineering principles by providing developer platforms to abstract and address the complexity of the APIs.
- Design programmable platform control plane APIs to not only monitor and manage consumption and provision, but also provide governance and compliance guardrails.
- Use APIs for SLAs, chargeback and consumption models led by the drive of standardization and automation of cloud delivery with programmable platforms.

### Sample Vendors

Avesha; CU Coding; Microsoft; Oracle; SAP; Silk

### Gartner Recommended Reading

Quick Answer: How Can I Optimize the Use of Programmable Platforms for Effective Software Delivery?

Adopt Platform Engineering to Improve the Developer Experience

A Software Engineering Leader's Guide to Improving Developer Experience

## Consumption-Based Model

**Analysis By:** Jeff Vogel, Philip Dawson

**Benefit Rating:** High

**Market Penetration:** 5% to 20% of target audience

**Maturity:** Early mainstream

### Definition:

A consumption-based sourcing model strategy for hybrid cloud on-premises data center storage, compute and networking infrastructure is an acquisition, deployment and support model that includes a cloud-like pay-for-use and platform services model optimized for predictable usage.

### Why This Is Important

The consumption-based model provides IT operations with an on-premises cloud-like operating model for storage, compute and networking. It eliminates capital expenditure (capex) financing, simplifies capacity planning and optimizes asset usage to actual workload use, effectively aligning asset costs-to-value. It has brought a whole new way of procurement sourcing and asset consumption, with pay-as-you-use and as-a-service platforms becoming the preferred deployment methodology for storage and compute.

### Business Impact

A consumption-based sourcing model and services strategy will:

- Shift responsibility for maintenance and support costs to vendors investing in AI for IT operations (AIOps) to automate IT administration.
- Preserve cash by avoiding upfront capex in exchange for strategic priorities.
- Shift IT and finance resource budget cycles to a services-based platform delivery model.
- Provide more flexible and agile IT operations aligned with business demands.

### Drivers

Infrastructure and operations (I&O) leaders are embracing cloud-native hardware and software consumption models as a strategy to replace owned, on-premises infrastructure and to lower data center operations' costs. This trend is driven by:

- The need for a more flexible cloud-like operating model for on-premises infrastructure.
- The massive growth of enterprise data that makes capacity planning difficult and upfront purchasing for three to five years of growth expensive and impractical.
- Prolonged procurement lead time increases due to persistent supply shortages.
- The need for an application-aware services delivery model.
- The preference for operating expenditure (opex) to capex with cloud-like benefits, while avoiding risks or costs associated with moving mission-critical workloads to the public cloud.

- The need for a more cost-effective, flexible and efficient sourcing strategy that aligns with business demands.
- The need to augment IT budget priorities to redirect investments to develop cloud-native platform skills that support business growth initiatives.
- The shift from exiting the life cycle management of infrastructure assets in the long term to freeing up IT resources.

### Obstacles

A consumption-based sourcing model may:

- Be more expensive than capex financing.
- Be organizationally challenging to implement.
- Be unsuitable for IT operations that have a more stable and predictable growth and variability in forecast demand or lean toward sweating assets.
- Require minimum-usage commitment levels that can't be justified regardless of what is actually consumed.
- Require three- to five-year contracts with vendor-centric services.
- Lack the skills or culture alignment to shift from sourcing products to platform SLA services.
- Not take into account long-term supply chain price fluctuations during the contract period, when declining hardware costs or supply constraints are considered.
- Conflict with financial asset depreciation and amortization schedules or corporate balance sheet objectives.
- Conflict with established industry accounting standards and operational norms.
- Software licensing terms may be incompatible with the use of consumption based hardware.

#### User Recommendations

- Adopt a cloud operating model as a platform services strategy to shift to ITOps-as-a-service to increase productivity and flexibility.
- Organize and implement a joint team approach to include I&O, vendor management and finance to establish a strategic sourcing strategy.
- Rightsize and align IT I&O resources to a consumption-based platform model to free up resources to focus on business priorities.
- Assess the economics and requirements against a range of vendor consumption programs before committing.
- Ensure that contract terms match financial requirements, accounting for capex versus opex, and that contracts include appropriate end-of-term options, such as book value buyout.
- Address licensing options and term constraints as they pertain to usage.
- Link consumption-based costs to specific usage level requirements along with remediation terms to enforce minimum levels.
- Retire legacy technical debt and onerous support fees, and modernize systems and processes.

#### Sample Vendors

Cisco; Dell Technologies; Hewlett Packard Enterprise; IBM; Lenovo; NetApp; Pure Storage

#### Gartner Recommended Reading

Market Guide for Consumption-Based Models for Data Center Infrastructure

Competitive Landscape: Consumption-Based Model for On-Premises Infrastructure

Quick Answer: How Can I Use Storage as a Service to Reduce IT Spend?

### Direct-to-Chip Liquid Cooling

**Analysis By:** Henrique Cecci

**Benefit Rating:** Moderate

**Market Penetration:** 5% to 20% of target audience

**Maturity:** Early mainstream

### Definition:

Direct-to-chip (D2C) liquid cooling is a cooling technology that involves circulating coolant liquid over heat-generating components, such as CPUs, GPUs and memory modules, to draw off heat through cold plates or evaporation units. Compared to traditional air cooling, D2C liquid cooling is highly efficient in dissipating heat, potentially leading to increased efficiency, cost savings, improved reliability, space optimization and greater sustainability.

### Why This Is Important

The latest CPUs and GPUs have much higher thermal density properties compared to older architectures. Moreover, server manufacturers are incorporating more CPUs and GPUs into each rack to keep up with the increasing demand for high-performance computing and AI applications. However, traditional air cooling systems are struggling to cope with the cooling requirements of these high-density racks in a sustainable and efficient manner but with D2C liquid cooling this is achievable.

### Business Impact

Direct-to-chip (D2C) liquid cooling can bring multiple benefits to businesses, including improved energy efficiencies, additional cost savings, increased sustainability and improved reliability. By lowering operating temperatures, liquid cooling can boost computing systems' performance and energy efficiency, leading to significant cost savings in the long run. Additionally, it can optimize physical space utilization and enable the use of high-performance computing (HPC).

### Drivers

- The growing demand for HPC and the increasing use of artificial intelligence (AI) and machine learning (ML) applications or high-density computing environments where traditional air cooling methods are insufficient.
- Requirements for lower energy consumption and associated costs.
- Data center space optimization and performance optimization.
- Lower noise level. Because liquid cooling systems don't rely on fans, they can operate more quietly than air-cooled systems.
- Sustainability and environment-friendly data center operations.

#### Obstacles

- High initial capital investment costs.
- Specialized expertise required to design and implement the system.
- Potential risks associated with leaks or system failures.
- Compatibility concerns with existing IT infrastructure.
- Regulatory requirements.
- Environmental impact concerns.
- Businesses may need time to realize the long-term cost savings of liquid cooling and overcome the early costs of using this technology.

#### User Recommendations

- Choose the right coolant by selecting one with good thermal conductivity, low viscosity and low electrical conductivity.
- Ensure proper flow rate. For example, maintaining the flow rate between 0.5 to 1.5 gallons per minute per kilowatt of heat.
- Maintain cleanliness and regularly replace the coolant, this is essential to maintain optimal performance and prevent any damage to the components.
- Take into account the design of the system, the cooling system should be designed to provide efficient cooling while minimizing the risk of leaks and damage to the electronics.

#### Sample Vendors

Asetek; Chilldyne; CoolIT Systems; Fujitsu; Huawei; Iceotope; JetCool Technologies; Rittal North America; Schneider Electric; STULZ

#### Gartner Recommended Reading

Market Guide for Servers

Emerging Tech Impact Radar: Compute and Storage

### Infrastructure Platform Engineering

**Analysis By:** Hassan Ennaciri, Paul Delory

**Benefit Rating:** High

**Market Penetration:** 20% to 50% of target audience

**Maturity:** Adolescent

**Definition:**

Infrastructure platform engineering is the discipline of building internal software products that present IT infrastructure to users or other platforms in an easily consumable way. Infrastructure platforms are self-service tools that allow nonexpert users to deploy and manage infrastructure themselves while I&O retains governance, security and compliance. Infrastructure platforms are often used as the foundation of higher-order, self-service layers such as internal developer platforms.

### Why This Is Important

Digital enterprises are pressured to innovate and deliver products faster to meet customer needs. This requires adopting new operating models and modern practices to deliver scalable, reliable platforms that enable faster product delivery. Infrastructure platform engineering provides automated delivery of curated secure, reliable and scalable infrastructure services that can be available via self services or APIs and reduce the effort and cycle time for users to request and access the products.

### Business Impact

Infrastructure platform engineering abstracts the complexity of the digital infrastructure to deliver platforms that continuously evolve to meet customer needs. It is an agile approach necessary to enable software products’ value streams to meet customer needs and expectations. It also provides on-demand, fast access to environments, services and tools that improve customer experience and productivity.

Drivers

- **Business agility and innovation:** Digital businesses are required to be responsive to customers' needs and changing market conditions. They must have the ability to quickly deliver products that meet these changing demands and requirements.

- **Cost optimization:** Infrastructure platform engineering teams leverage automation to deliver scalable, reliable and secure platforms. This helps to improve efficiency, reduce resource cost due to manual work and reduce downtime due to change failures. Standardizing tools and platforms also optimizes resource utilizations and reduces cost incurred in tool proliferation.

- **Digital infrastructure and platform complexity:** Public cloud IaaS and PaaS deliver extensive capabilities and are designed to be consumable by developers, but most enterprises need additional governance and management that is best delivered by a platform engineering team.

- **Improve developer experience and productivity:** Infrastructure platform engineering abstracts complexity from developers and provides them with quick access or self-service in the environments they need to develop and test their software. Services can be made via an internal developer portal (IDP) such as Backstage, Calibo or Humanitec.

- **Compliance and security:** Infrastructure platform engineering automates and integrates compliance and security controls into software delivery pipelines, improving the organization's security posture and reducing the burden from developers.

### Obstacles

- **Confusion:** There is a lot of hype and confusion about platform engineering and what it means. Many vendors are defining it to help sell their products, causing uncertainty with teams trying to adopt it.
- **Cultural:** This operating model is a new, modern approach that requires a shift in how teams work and collaborate, which is the hardest obstacle to overcome for many organizations.
- **Lack of skills:** Infrastructure platform engineering requires software engineering and specialized skills that may not exist in the organization.
- **Structure of traditional I&O operating models:** The organizational structure of many I&O teams is set up by domain specializations, making it hard to develop and deliver end-to-end services.
- **IT service management approaches:** The current approaches are process-heavy and rely on tickets and handoffs.
- **Complexity:** Successful implementation of infrastructure platform engineering is challenging because it requires new roles and involvement from many stakeholders.

### User Recommendations

- **Start small and evolve:** Define initial goals and objectives of the platform by understanding common user needs and delivering viable products that continuously evolve to meet those needs.
- **Build a dedicated team with the right skills:** Successful infrastructure engineering practice requires dedicated teams with diverse skills in infrastructure platforms and software engineering.
- **Identify and fill critical roles such as platform owner and platform architect.** Acquire new talent with the required technical skills, the right mindset and strong interpersonal skills. Develop existing resources by provisioning continuous learning opportunities.
- **Adopt a product mindset:** Thread platform users as customers and ensure that you talk to them and continuously get their feedback to meet their existing needs as well as anticipate their future needs. Enable users and reduce the level of effort required to use the platform products.

### Gartner Recommended Reading

Adopt Platform Engineering to Improve the Developer Experience

Top Strategic Technology Trends for 2023: Platform Engineering

Innovation Insight for Internal Developer Portals

Quick Answer: How Can I Optimize the Use of Programmable Platforms for Effective Software Delivery?

Guidance Framework for Implementing Cloud Platform Operations

## Immersion Cooling

**Analysis By:** Jeffrey Hewitt, Philip Dawson

**Benefit Rating:** Moderate

**Market Penetration:** 1% to 5% of target audience

**Maturity:** Adolescent

### Definition:

Immersion cooling is a type of data center server cooling system that immerses server boards in a nonconductive heat transfer liquid, typically built using an immersion container in a dense, closed system. Immersion systems deliver well-above-average power efficiency, enabling compute systems to run at high performance while requiring less floor space.

### Why This Is Important

Immersion cooling shifts power from cooling to computing, potentially doubling the compute density in power-constrained locations. It allows servers to operate in constrained environments such as 5G network control nodes and Internet of Things (IoT) edge servers, and significantly increases the efficiency of enterprise servers by reducing the need for air cooling.

### Business Impact

Immersion cooling systems enable enterprises to deliver on sustainability and deploy higher levels of compute capability to strategic locations than is possible with conventional air-cooled racks. Key applications include data centers in facilities with limited space, factory automation, edge data centers and data centers in remote or unattended locations. Immersion cooling is well suited to the small remote data centers that will support 5G mmWave deployments.

### Drivers

- **Immersion cooled systems are smaller, quieter and more efficient than traditional rack systems.** Their initial value will most likely come from outside the data center, where they enable higher compute density at higher energy efficiency and lower noise. Although the capital cost of the system is typically higher because of the mechanical and cooling infrastructure involved, there are environments where these systems outperform any alternative.

- **Immersion cooling can recover expensive server space and power costs.** For an enterprise constrained to operate servers in expensive spaces, there is often a floor power budget that covers both equipment and cooling. An immersion cooling system could improve power efficiency by 40%, theoretically, enabling 67% more energy for computing within the same power budget. These systems may also recover floor space, and are so quiet that they need no sound baffling.

- **Immersion cooling enables edge servers to operate in otherwise hostile locations.** Medium-scale edge computing nodes or wireless telecom nodes often operate under the thermal, spatial and power constraints of a remote server bunker, pole or closet. Shipboard or truck-based mobile data centers also benefit from these space and power efficiencies. For certain GPU-centric small-scale supercomputing tasks, these systems represent a practical on-premises solution. Isolation of the components also facilitates their use in locations with high levels of particulate pollutants like dust.

### Obstacles

- **Data centers must be replumbed for immersion cooling.** Immersion cooling requires redundant plumbing for the warm water loop. It is most efficient when used with a passive heat exchanger.
- **Cooling fluids require special handling.** Immersion systems use vegetable oil or fluorocarbons. Oil-based systems require that staff handle and bag oil-coated boards for repair or replacement. This requires skills normally not possessed by IT administrators. Fluorocarbon systems operate with robotic, sealed pods as all fluids have to be recovered and contained.
- **Nonstandard compact server motherboards deliver the best economics.** To achieve the floor space reductions that liquid cooling offers, most systems require smaller motherboards as many systems are horizontal, rather than vertical. Vertical systems can use standard motherboards and are well suited to environments such as 5G closets, where the vertical form factor is a better fit.

### User Recommendations

- **Evaluate immersion cooling for environments where power and space are expensive.** Immersion cooled systems are significantly smaller, quieter and more energy efficient than traditional data center racks. These systems will prove cost-effective in space- and power-constrained environments.
- **Fully comprehend how to communicate the use of vegetable oil in environments.** Despite marketing terms, the cooling fluids are primarily vegetable-derived and present manageable fire and mess risks. Understanding how to present and defuse these issues will be important in review meetings.
- **Plan to use immersion cooling for larger edge server deployments.** Edge systems inevitably face power, cooling and space constraints. Immersion cooling significantly eases cooling design by eliminating the need to pass air over components. Immersion cooling also improves reliability through lower overall operating temperatures, and exclusion of oxygen from contacts.

### Sample Vendors

Asperitas; Green Revolution Cooling; Iceotope; Immersion Systems; QCooling; Submer; TMGcore

### Gartner Recommended Reading

Predicts 2023: Edge Computing Delivery and Control Options Extend Functionality

Unlock the Business Benefits of Sustainable IT Infrastructure

## Edge Computing

**Analysis By:** Bob Gill, Philip Dawson

**Benefit Rating:** Transformational

**Market Penetration:** More than 50% of target audience

**Maturity:** Early mainstream

### Definition:

Edge computing describes a distributed computing topology in which data storage and processing are placed in optimal locations relative to the location of data creation and use. Edge computing locates data and workloads to optimize for latency, bandwidth, autonomy and regulatory/security considerations. Edge-computing locations extend along a continuum between the absolute edge, where physical sensors and digital systems converge, to the "core," usually the cloud or a centralized data center.

### Why This Is Important

Edge computing has quickly become the decentralized complement to the largely centralized implementation of hyperscale public cloud. Edge computing solves many pressing issues, such as sovereignty, unacceptable latency and bandwidth requirements, given the massive increase in data produced at the edge. The edge-computing topology enables the specifics of Internet of Things (IoT), digital business and managed distributed IT solutions.

### Business Impact

Edge computing improves efficiency, cost control, and security and resilience through processing closer to where the data is generated or acted upon, fostering business opportunities and growth (e.g., customer experience and new real-time business interactions). Earliest implementations succeeded in enterprises that rely on operational technology (OT) systems and data outside core IT, such as the retail and industrial sectors.

Drivers

- Growth of hyperscale cloud adoption has exposed the limits of extreme centralization. Latency, bandwidth requirements, the need for autonomy and data sovereignty or location requirements may be optimized by placing workloads and data closer to the edge, rather than centralizing in a hyperscale data center.
- Data growth from interactive applications and systems at the edge often cannot be economically funneled into the cloud.
- Applications supporting customer engagement and analysis favor local processing for speed and autonomy.
- IoT is evolving from simply reporting device status to using edge-located intelligence to act upon such status, bringing the benefits of automation and the creation of immediately responsive closed loop systems.
- Edge computing's inherent decoupling of application front ends and back ends provides a perfect means of fostering innovation and enhanced ways to do business. For example, using technologies such as machine learning and industrial sensors to perform new tasks at locations where business and operational events take place, or at the point of interaction with a retail customer, can drive significant business value.

#### Obstacles

- The diversity of devices, software controls and application types all amplify complexity issues.
- Widespread edge topology and explicit application and networking architectures for edge computing are not yet common outside vertical applications, such as retail and manufacturing.
- Edge success in industrial IoT applications and enhancing customer experience in retail are well-understood, but many enterprises still have difficulty understanding the benefits, use cases and ROI of edge computing.
- A lack of broadly accepted standards slows development and deployment time, creating lock-in concern for many enterprise users.
- Edge physical infrastructure is mature, but distributed application management and orchestration challenges are still beyond most vendor-supplied component management offerings. The tasks of securing, maintaining and updating the physical infrastructure, software and data require improvement before management and orchestration can mature.

#### User Recommendations

IT leaders responsible for cloud and edge infrastructure should:

- Create and follow an enterprise edge strategy by focusing first on business benefit and holistic systems, not simply focusing on technical solutions or products.
- Position edge computing as an ongoing, enterprisewide journey toward distributed computing, not simply individual isolated projects.
- Establish a modular, extensible edge architecture through the use of emerging edge frameworks and design sets.
- Accelerate time to benefit and derisk technical decisions through the use of vertically aligned systems integrators and independent software vendors that can implement and manage the full orchestration stack from top to bottom.
- Evaluate “edge-as-a-service” deployment options, which deliver business-outcome-based solutions that adhere to specific SLAs while shifting deployment, complexity and obsolescence risk to the provider.

**Gartner Recommended Reading**

Market Guide for Edge Computing

5 Top Practices of Successful Edge Computing Implementers

## Intelligent Infrastructure

**Analysis By:** Philip Dawson, Nathan Hill

**Benefit Rating:** High

**Market Penetration:** 20% to 50% of target audience

**Maturity:** Emerging

**Definition:**

Intelligent infrastructure is built from simple, repeatable infrastructure building block components, integrated and managed in a standardized, automated manner. It optimizes infrastructure resources for application consumption through infrastructure machine learning (ML) and tuning as software overlays through an automated software intelligence plane.

### Why This Is Important

Intelligent infrastructure encapsulates generative AI and ML into the infrastructure configuration. Building on the capabilities of virtualization, it adds the dynamic hardware composition capability of a composable infrastructure to deliver a hardware configuration that is optimized for a specific application. Intelligent infrastructure additionally adds or feeds the generative AI/ML automation functions to the intelligence plane.

### Business Impact

Intelligent infrastructure is an innovation in delivering automated optimized systems for application delivery. It builds on earlier innovations, including converged, hyperconverged, software-defined and composable infrastructures, helping deliver hybrid cloud-like infrastructure on-premises or with a provider. It feeds off the application API-led programmable infrastructure that tunes infrastructure through system calls and requests, which improves application and infrastructure integration.

### Drivers

- IT leaders now recognize that cloud infrastructure, cloud platforms and cloud-native applications drive the overall composable, programmable and intelligent infrastructure journey.
- Cloud delivery and edge expansion are fueling the standardization of infrastructure, design and architecture and the expansion of the three areas to edge and Internet of Things (IoT) locations beyond remote offices/branch offices (ROBOs).
- Adding generative AI and automation on top of this infrastructure composition capability ensures that infrastructure is always optimized for the application load.
- In intelligent infrastructure, the "control plane" is enhanced with automation driven by infrastructure analytics ML, to become an automated "intelligence plane."

### Obstacles

- The intelligence plane automates infrastructure and workload provisioning to application consumption. Intelligent infrastructure should not be tied to hardware features, but rather software functions.
- As with software-defined and composable infrastructures, traditional system vendors often tie intelligent infrastructure to hardware-related features, which can propel lock-in.
- Cloud management platforms are used as overlays for cloud migrations. Intelligent infrastructure has to adapt to hybrid cloud and multicloud delivery, delivering client value whether on-premises, with a provider or public cloud through anything as a service (XaaS).

### User Recommendations

- Select infrastructure solutions based on their ability to meet the current business requirements while still offering the flexibility to exploit the integration and automation of intelligent infrastructure innovations to be delivered over the next five years.
- Increase agility and business alignment by integrating application, asset management and sourcing information into the infrastructure intelligence and control planes as a drive to platform- and infrastructure-driven consumption models.
- Prepare for the evolution of application delivery and workload provisioning by incorporating intelligence/ML infrastructure functions with intelligent fabrics into your future system requirements.

### Sample Vendors

Cisco; CU Coding; Hewlett Packard Enterprise; IBM; Intel; Microsoft; Tintri; VMware

### Gartner Recommended Reading

How to Evolve Your Physical Data Center to a Modern Operating Model

Market Guide for Servers

Quick Answer: How Can I Optimize the Use of Programmable Platforms for Effective Software Delivery?

## Sliding into the Trough

### Programmable Infrastructure

**Analysis By:** Philip Dawson, Nathan Hill

**Benefit Rating:** High

**Market Penetration:** 5% to 20% of target audience

**Maturity:** Adolescent

#### Definition:

Programmable infrastructure is the concept of using and applying methods and tooling from the software development area to management of IT infrastructure. This includes, but is not limited to, APIs, immutability, resilient architectures and agile techniques.

#### Why This Is Important

Programmable infrastructure ensures optimal resource utilization, while driving cost efficiencies. A continuous delivery approach requires continuous insight and the ability to automate application responses. Moving to an API-driven infrastructure is the key first necessary step to enabling anti-fragile and sustainable automation through programmatic techniques.

#### Business Impact

Greater value (rather than cost reduction) is achieved via programmable infrastructure's ability to drive adaptive automation — responding faster to new business infrastructure demands, driving service quality and freeing staff from manual operations. Programmable infrastructure reduces technical debt with investment and enables a sustainable and highly responsive IT infrastructure service to the business.

#### Drivers

- Programmable infrastructure strategies are applied to private cloud, hybrid cloud and infrastructure platforms as well as public cloud. Demand for programmable infrastructure grows as heterogeneous infrastructure strategies are embraced.
- Programmable infrastructure is needed to manage the life cycle of infrastructure delivery from provisioning, resizing and reallocation to reclamation, and in the case of external resources, manage elasticity and the termination of consumption.

- Programmable infrastructure is needed to optimize and reduce the dependency on the infrastructure life cycle. More importantly, it enables the desired (performance, cost, speed) infrastructure provisioning and orchestration in line with business demands.

### Obstacles

- The ongoing cost of refreshing API-enabled infrastructure components on-premises after initial implementation adds financial pressure to organizations.
- Applying automation to existing monolithic infrastructure components fails due to the lack of platform agility and vendor lock-in.
- While APIs enable integration across different infrastructure platforms, the lack of open APIs/API compatibility across vendor platforms creates a siloed mentality.
- The implementation of programmable infrastructure is hampered by the early adoption of it within infrastructure and operations (I&O), and the shortage of skilled software engineering resources to comprehensively exploit it (especially in web technologies such as HTTP and JSON to develop these APIs).

### User Recommendations

- Deploy a programmable infrastructure to further abstract application from infrastructure delivery and pursue an agile digital business outcome.
- Implement a programmable infrastructure by investing in infrastructure automation tools and continuous delivery (example vendors for these markets are listed below, but no single vendor or platform can enable an organizationwide programmable infrastructure strategy) leading to API-led programmable platforms.
- Invest in infrastructure and DevOps, and modernize legacy IT architectures to implement an API-driven infrastructure.
- Examine reusable programmable infrastructure building blocks leveraging programmable infrastructure strategy built on repeatable and available skills from providers.

### Sample Vendors

Amazon Web Services; CU Coding; Google; IBM; Microsoft; Oracle; Quality Technology Services; RackN; Tencent; VMware

### Gartner Recommended Reading

Market Guide for Servers

Predicts 2023: XaaS Is Transforming Data Center Infrastructure

Quick Answer: How Can I Optimize the Use of Programmable Platforms for Effective Software Delivery?

## IT/OT Hybrid Servers

**Analysis By:** Tony Harvey

**Benefit Rating:** High

**Market Penetration:** 1% to 5% of target audience

**Maturity:** Adolescent

### Definition:

Information/operational technology (IT/OT) hybrid servers are edge devices that interface, collect and process data from OT systems that provide real-time control of physical systems and industrial processes. They are designed to operate with a higher resilience to shock, vibration, humidity and temperature than typical data center servers. Industrial communications interfaces — such as CAN bus, Modbus or Profinet protocols, as well as wireless or 5G technology — may also be included.

### Why This Is Important

IT/OT hybrid servers allow the data created by OT systems to be processed in real time to optimize the process under control. Connections to IT networks allow hybrid IT/OT servers to collect and transmit data. This data can then be used for training AI/ML models to deliver further efficiencies and provide insight into manufacturing and production capacity and scheduling.

### Business Impact

IT/OT hybrid servers help enterprises realize the potential of the large data pool that is generated by OT systems. The ability to use this data will generate new cost efficiencies and innovations in manufacturing and industrial control processes. Enterprises that successfully integrate IT/OT hybrid servers into their digital transformation strategy will lower their costs and deliver new services to market faster. Enterprises that do not adopt them, however, may find themselves left behind.

### Drivers

- Businesses need real-time analysis and decision making based on capturing data that allows the optimization of industrial processes and assets to reduce costs and increase quality.
- By using near-real-time reporting of manufacturing, operations and production data, businesses will be able to provide more predictability in order cycles and a better usage of components.
- Equipment breakdowns can cause line stoppages, which drive manufacturing costs up. By enabling the collection and analysis of device monitoring, IT/OT servers enable predictive maintenance to prevent these issues.
- Regulatory and compliance requirements mandate that certain datasets should be processed and stored at edge locations, which requires the deployment of appropriate systems on-site. Further, latency and bandwidth limitations at these sites further stress the need for on-site systems.
- Organizations are collecting OT data to enable AI/ML training and digital-twin-model building.
- There is a need for specialized servers that can meet the environmental requirements for industrial sites.

### Obstacles

- Industrial enterprises are cautious about the security risk of using IT and network connectivity systems in industrial process control, where failure could result in loss of life or significant property damage.
- IT and OT are separate groups with different cultures and different risk perceptions. The differences between these groups must be managed for any successful implementation.
- Businesses grapple with the complexity of defining what data must stay at the edge versus what data should be transmitted to and subsequently processed in the cloud.
- Budgeting for IT/OT hybrid servers can be difficult because there is an overlap between OT and IT systems.
- Management solutions designed to operate at large scale across a wide geographic range with highly variable connectivity characteristics are very immature.
- Standard IT equipment will not meet the harsh environmental requirements of industrial locations. Further, there could also be issues with electronic noise and interference.

### User Recommendations

- Create an integrated IT/OT group that has full responsibility for these solutions, reducing the disconnects related to technology, management and budgeting.
- Reduce the risk of conflicts between the teams by aligning the IT & OT groups across architecture, governance, security and software management, and infrastructure, support and software acquisition.
- Develop a blended IT/OT culture that mixes the rigor and risk awareness of the OT engineering mindset with the flexibility and tolerance for change that is inherent in an IT mindset.
- Embed safety, security and risk training, foster awareness and include talent in hybrid IT/OT teams to ensure that systems are designed with safety and security in mind.
- Remove budget conflicts by defining upfront the budget sources for ongoing support, maintenance and dependencies across the entire combined IT/OT environment.

### Sample Vendors

Dell Technologies; Hewlett Packard Enterprise; Lenovo; Schneider Electric

### Gartner Recommended Reading

As IT and OT Converge, IT and Engineers Should Learn From Each Other

Survey Analysis: IT/OT Alignment and Integration

When Does a CIO Need to Be Involved in OT?

2022 Strategic Roadmap for IT/OT Alignment

How IT Standards Can Be Applied to OT

## Infrastructure Automation

**Analysis By:** Chris Saunderson

**Benefit Rating:** High

**Market Penetration:** 20% to 50% of target audience

**Maturity:** Mature mainstream

### Definition:

Infrastructure automation (IA) enables DevOps and infrastructure and operations (I&O) teams to deliver automated infrastructure services across on-premises and cloud environments. This includes the life cycle of services through creation, configuration, operation and retirement. These infrastructure services are then made available through platform delivery, self-service catalogs, direct invocation and API integrations.

### Why This Is Important

IA delivers velocity, quality, efficiency and reliability, with scalable, declarative approaches for deploying and managing infrastructure. These tools integrate into delivery pipelines targeting deployment topologies that range from on-premises to the cloud, and enable infrastructure consumers to build what is needed when they need it. Once deployed, IA provides day-2 and beyond operational automation, and extends to provide policy compliance and enforcement capabilities.

### Business Impact

Implementing and maturing IA services will enable:

- **Agility** — continuous infrastructure delivery and operations
- **Productivity** — version-controlled, declarative, repeatable, efficient deployments
- **Cost improvement** — reductions in manual effort expended via increased automation
- **Risk mitigation** — compliance driven by standardized configurations
- **Collaboration** — delivering environments that product teams need with security, cost and compliance requirements baked in.

### Drivers

I&O leaders must automate delivery through tool and skills investments to mature beyond simple deployments. The target should be standardized platforms that deliver the systemic, transparent management of platform deployments. This same discipline must be applied to the operation of these deployed platforms, ensuring that efficient operations (including automated incident response) can be achieved. IA tools deliver the following key capabilities to support this maturation:

- Multicloud/hybrid cloud infrastructure delivery
- Support for immutable and programmable infrastructures
- Predictable delivery enabling automated operations
- Self-service and on-demand environment creation
- Integration into DevOps initiatives (continuous integration/delivery/deployment)
- Resource provisioning, including cost optimization capabilities
- Operational configuration management efficiencies
- Policy-based delivery and assessment/enforcement of deployments against internal and external policy requirements
- Enterprise-level framework to enable maturing of automation strategies

- Skills and practice development inside infrastructure teams, enabling agile and iterative development and sustaining of services

### Obstacles

- The combination of tools needed to deliver IA capability can increase tool count and complexity.
- Software engineering skills and practices are required to get maximum value from tool investments.
- IA vendor capability expansion overlaps and confuses the tool landscape, resulting in over-investment.
- Steep learning curves can cause developers and administrators to revert to familiar scripting methods to deliver required capabilities.

### User Recommendations

- Identify existing IA tools in use to catalog capabilities, identify use cases and document overlaps to aid decision making.
- Assess existing internal IT skills to incorporate training needs that more fully enable IA, especially for an automation architect role to coordinate standards development and implementation.
- Baseline how managed systems and tooling will be consumed (e.g., engineer, self-service catalog, API or on-demand).
- Integrate security and compliance requirements into scope for automation and delivery activities.
- Develop an IA tooling strategy that incorporates current needs and near-term roadmap evolution.

### Sample Vendors

Amazon Web Services; HashiCorp; Microsoft; Perforce; Pliant; Progress; Pulumi; RackN; Upbound; VMware

### Gartner Recommended Reading

Market Guide for Infrastructure Automation Tools

Innovation Insight for Continuous Infrastructure Automation

To Automate Your Automation, Apply Agile and DevOps Practices to Infrastructure and Operations

How to Start and Scale Your Platform Engineering Team

## Digital Twins for Sustainability

**Analysis By:** Alfonso Velosa

**Benefit Rating:** High

**Market Penetration:** 1% to 5% of target audience

**Maturity:** Adolescent

### Definition:

Digital twins for sustainability are software proxies that mirror the state of a thing such as an asset, person, organization or process to meet business and sustainability objectives. The three types of digital twins for sustainability are discrete, composite and organizational, which serve different types of decisions. Building blocks include model, data, unique one-to-one association and monitorability. These are built-in software like analytics, GIS, CRM and Internet of Things (IoT).

### Why This Is Important

Digital twins for sustainability help enterprises meet sustainability outcomes — such as improved visibility and documentation, energy and waste reduction, and new business models — by creating virtual representations of unclear, difficult to achieve and time-consuming activities/tasks. Specific examples include improving supply chain decisions via better supply and demand visibility, and reducing unplanned downtime by monitoring equipment state.

### Business Impact

Enterprises, from oil production to manufacturing to real estate, are implementing digital twins for sustainability to:

- Acquire visibility into things, such as equipment or purchasing processes, that otherwise would have led to waste or duplication of effort.

- Assess, simulate and optimize operations by improving the design to eliminate inefficiency and reduce energy or resource consumption.
- Improve employee safety and training certification.
- Contribute to sustainability documentation requirements.

Drivers

- Enterprises are using digital twins for sustainability to address the increasing pressure to improve resource and energy efficiency, and increase transparency, across their operations, supply chains and wider ecosystems.
- Enterprises are pushing hard to reduce greenhouse gas (GHG) emissions across their operations and supply chain. Asset-intensive sectors, such as oil and gas, manufacturing, and aviation, are using digital twins for sustainability to integrate emissions data into business process optimization.
- Leading OEMs are deploying digital twins for sustainability to help add energy efficiency or align to their enterprise customers' sustainability objectives.
- Leading-edge enterprises are implementing digital twins for sustainability to feed simulation models to predict benefits from emissions reduction or cost optimization.
- Technology providers, from large cloud vendors to startups, are identifying potential ways to serve customers' sustainability goals using digital-twin-enabling product portfolios. In particular, they are developing libraries of templates to demonstrate energy reduction, minimize waste or reduce emissions.

#### Obstacles

- Enterprises often lack an understanding of what they are trying to achieve, the full potential of and the success metrics for their investments in digital twins for sustainability. This limits investment into initiatives that can take advantage of digital twins for sustainability.
- Currently, only a few enterprises have cross-functional fusion teams like business, finance, operations and IT, required to develop objectives to achieve digital twins for sustainability.
- Digital twins for sustainability challenge most enterprises technically due to the blend of operational and information technologies needed.
- Pricing remains an art, and most vendors focus on their technology differentiation, despite enterprises looking for business value and sustainability outcomes.
- Standards remain emergent across a range of digital twins for sustainability integration, metadata and other technical issues.

#### User Recommendations

- Co-create the digital twins for sustainability strategy with the enterprise business unit to identify opportunities and challenges.
- Establish clear success metrics, identify sponsors, super-users and budget support, and build a roadmap that starts small and scales up.
- Avoid digital twins for sustainability projects that lack a business sponsor, as this is key to success and lack of internal sponsorship will waste IT resources.
- Identify IT organization technology, governance and skills gaps, and build a plan to resolve them.
- Develop an architectural, security and governance framework to manage large numbers of discrete, composite and organizational digital twins for sustainability.
- Select vendors that demonstrate the ability to deliver both strategic sustainability objectives and tactical business results, based on their existing portfolio of technologies and libraries of prebuilt templates, for digital twins for sustainability.

### Sample Vendors

Accenture; Atos; Capgemini; Cognite; Envision Digital; NTT; Tata Consultancy Services; Xemelgo

### Gartner Recommended Reading

2023 Oil & Gas Trend: Digital Twins Expansion

Quick Answer: What Is a Digital Twin?

Emerging Tech: Tool — Digital Twin Business Value Calculator

Life Cycle Management of Software-Defined Vehicles: Step 3 — Vehicle Digital Twin 2.0

## Composable Infrastructure

**Analysis By:** Tony Harvey, Paul Delory, Philip Dawson

**Benefit Rating:** High

**Market Penetration:** 20% to 50% of target audience

**Maturity:** Early mainstream

### Definition:

Composable infrastructure uses an API to create physical systems from shared pools of resources. The implementation connects disaggregated banks of processors, memory, storage devices and other resources by a hardware fabric. However, composable infrastructure software can also aggregate or subdivide resources in traditional servers or storage.

### Why This Is Important

Servers, storage and fabrics are traditionally deployed as discrete products with predefined capacities. Individual devices, or resources, are connected manually and dedicated to specific applications, making the system inflexible and expensive to change and scale. Composable infrastructure replaces this with a pool of components that can be dynamically assigned as needed, increasing agility, easing capacity planning and reducing costs.

### Business Impact

Stranded hardware resources that are underutilized represent significant costs in IT. The composable infrastructure enables hardware resources to be aggregated from a pool of components via APIs to dynamically match the infrastructure to the needs of the workload. This increases component utilization, reduces hardware overprovisioning, decreases costs, and improves IT responsiveness to the business's requirements.

### Drivers

- Compute Express Link (CXL) provides the necessary capabilities to disaggregate and pool memory and I/O as well as providing a standardized set of APIs to manage the disaggregated hardware.
- Hyperscale cloud vendors are moving toward composable designs utilizing CXL to increase hardware utilization and reduce the costs of stranded hardware.
- Test and development environments benefit from composability, where infrastructure with varying characteristics must be repeatedly deployed, deconstructed and redeployed.
- Multitenant environments benefit from composable infrastructure by allowing a pool of hardware to be dynamically configured, assigned, reconfigured and reassigned based on tenant requirements.

### Obstacles

- Current composable implementations are limited in that pooled resources are restricted to using hardware from a single vendor.
- Existing composable infrastructures are limited to just composing storage and I/O, limiting the use cases.
- A proliferation of vendor-specific APIs and a lack of off-the-shelf software for managing composable systems are also headwinds to widespread adoption.

### User Recommendations

- Deploy composable infrastructure when the workload or use case demands that infrastructure must be resized and administered frequently or when composability increases the use of packaged standardized high-cost components.
- Replace existing infrastructure to obtain composable infrastructure only if you have sufficiently mature automation tools and skills to implement composable features and yield financial or business benefits.
- Verify that your infrastructure management software supports composable system APIs or that you have the resources and skill sets to write your own management tools.

### Sample Vendors

Cisco; Dell Technologies; GigaIO; Hewlett Packard Enterprise; Intel; Liqid; Western Digital

### Gartner Recommended Reading

Market Guide for Servers

Emerging Tech: Compute Express Link Redefines Server Memory Architectures

Emerging Tech Impact Radar: Compute and Storage

2022 Strategic Roadmap for Compute Infrastructure

## Climbing the Slope

### Immutable Infrastructure

**Analysis By:** Neil MacDonald, Tony Harvey

**Benefit Rating:** Moderate

**Market Penetration:** 5% to 20% of target audience

**Maturity:** Early mainstream

#### Definition:

Immutable infrastructure is a process pattern (not a technology) in which the system and application infrastructure, once deployed, are never updated in place. Instead, when changes are required, the infrastructure and applications are simply updated and redeployed through the CI/CD pipeline.

#### Why This Is Important

Immutable infrastructure ensures the system and application environment, once deployed, remains in a predictable, known-good-configuration state. It simplifies change management, supports faster and safer upgrades, reduces operational errors, improves security, and simplifies troubleshooting. It also enables rapid replication of environments for disaster recovery, geographic redundancy or testing. This approach is easier to adopt with cloud-native applications.

#### Business Impact

Taking an immutable approach to workload and application management simplifies automated problem resolution by reducing the options for corrective action to, essentially, just one — repair the application or image in the development pipeline and rerelease. The result is an improved security posture and a reduced attack surface with fewer vulnerabilities and a faster time to remediate when new issues are identified.

#### Drivers

- Linux containers and Kubernetes are being widely adopted. Containers improve the practicality of implementing immutable infrastructure due to their lightweight nature, which supports rapid deployment and replacement.

- The GitOps deployment pattern, which emphasizes continuously synchronizing the running state to the software repository, has become an effective way to implement immutable infrastructure in Kubernetes-based, containerized environments.
- Infrastructure as code (IaC) tools (including first-party cloud provider IaC tools) have increasingly integrated configuration drift detection and correction, improving the practicality of implementing immutable infrastructure across an application's entire stack and environment.
- Interest in zero-trust and other advanced security postures where immutable infrastructure can be used to proactively regenerate workloads in production from a known good state (assuming compromise), a concept referred to as "systematic workload reprovisioning."
- For cloud-native application development projects, immutable infrastructure simplifies change management, supports faster and safer upgrades, reduces operational errors, improves security, and simplifies troubleshooting.

### Obstacles

- The use of immutable infrastructure requires a strict operational discipline that many organizations haven't yet achieved, or have achieved for only a subset of applications.
- IT administrators are reluctant to give up the ability to directly modify or patch runtime systems.
- Applying the immutable infrastructure pattern is most easily done for stateless components. Stateful components, especially data stores, represent special cases that must be handled with care.
- Implementing immutable infrastructure requires a mature automation framework, up-to-date blueprints and bills of materials, and confidence in your ability to arbitrarily recreate components without negative effects on user experience or loss of state.
- Many enterprise applications are stateful applications deployed on virtual machines. These applications are oftentimes commercial off-the-shelf and are not designed for fully automated installation when redeployed.

### User Recommendations

- Reduce or eliminate configuration drift by establishing a policy that no software, including the OS, is ever patched in production. Updates must be made to individual components, versioned in a source-code-control repository, then redeployed.

- Prevent unauthorized change by turning off all administrative access to production compute resources. Examples of this might include not permitting Secure Shell or Remote Desktop Protocol access.

- Adopt immutable infrastructure principles with cloud-native applications first. Cloud-native workloads are more suitable than traditional on-premises workloads.

- Treat scripts, recipes and other codes used for infrastructure automation similar to the application source code itself, as this mandates good software engineering discipline.

- Include immutable infrastructure scripts, recipes, codes and images in your backup and ransomware recovery plans as they will be your primary source to rebuild your infrastructure after an infection.

#### Sample Vendors

Amazon Web Services; Google; HashiCorp; Microsoft; Perforce; Progress; Red Hat; Snyk; Turbot; VMware

#### Gartner Recommended Reading

Comparing DevOps Architecture to Automate Infrastructure and Operations for Software Development

2022 Strategic Roadmap for Compute Infrastructure

To Automate Your Automation, Apply Agile and DevOps Practices to Infrastructure and Operations

Innovation Insight for Continuous Infrastructure Automation

Market Guide for Cloud-Native Application Protection Platforms

### Microgrids

**Analysis By:** Ethan Cohen

**Benefit Rating:** High

**Market Penetration:** 5% to 20% of target audience

**Maturity:** Early mainstream

### Definition:

A microgrid is a self-sufficient group of interconnected electrical loads and distributed energy resources that can operate as a stand-alone power system or be connected to the grid to provide optimization optionality. Microgrids commonly range in size from 100 kilowatts (kW) to 10s of megawatts (MW) and can connect to and disconnect from the grid. Microgrids can operate in both grid-connected and island modes based on technical and economic criteria to optimize energy cost and availability.

### Why This Is Important

Microgrid uses include rural electrification, residential or community power networks, commercial, industrial, municipal, hospital and military base power grids. Microgrids leverage traditional and renewable generation sources. Microgrids offer a compelling alternative to traditional energy generation and distribution, using Internet of Things (IoT) to enable integrated control of distributed power generation assets.

### Business Impact

Microgrids impact utility generation, distribution and energy retailing domains. They are becoming more important as utilities create new energy ecosystems and expand energy services offerings. Microgrids are also examples of energy technology consumerization, challenging the traditional business model of utility-provisioned energy delivered as a cloud service. By facilitating consumer integration into the energy market, microgrids are contributing to consumer energy management and the energy delivery infrastructure's geodesic transformation.

### Drivers

Microgrids offer advantages to utilities and customers by improving energy efficiency, reducing transmission and distribution losses, improving reliability, reducing environmental impact and providing a more cost-efficient electricity infrastructure compared to the traditional electricity distribution grid, as they also:

- Provide local options regarding the choice of electricity generation source and supply, such as distributed renewable energy sources (particularly those that are for energy storage).

- Enable energy customers to self-provision for operational resilience and collaborate or partner with utilities to achieve specific outcomes.
- Support renewable energy and energy efficiency through a viable approach to local grid modernization while incorporating local distributed energy supplies and storage technologies to meet the specific needs of their constituents while networking with the main grid.
- Deliver benefits to utilities by supporting the central grid in handling sensitive loads and the variability of renewables locally and supplying ancillary services to the bulk power system.

### Obstacles

- Microgrids do not have the same economies of scale and the coincident load factor of the centralized grid.
- The commercial integration of microgrids into energy markets will require a platform for the energy-sharing, resource-sharing and market-sharing economy.
- Technical constraints inhibit the integration of microgrids into distribution grids, including specific elements like dual-mode switching functionality, reliability, power quality and protection.
- Central electricity network operation impacts for microgrids require new utility systems, such as distributed energy resource management systems and advanced distribution management systems, which can be costly and complex to deploy.
- Microgrids must have mechanisms to regulate voltage and frequency in response to changes in load and system disturbances. This is because all power in microgrids comes from distributed generation resources and controllable loads within the microgrid, which typically requires investment in operational technology (OT) to perform distributed control.

### User Recommendations

As microgrids progress into mainstream utility, CIOs should:

- Observe market developments in microgrid use cases, and evaluate what kinds of offerings might be advanced to develop new revenue, enhance resilience and improve energy provisioning.

- Enable the utility to quickly and thoroughly evaluate microgrid development and/or operation by developing minimum viable products for microgrid cases. Despite the significant promise and industry excitement about the concept, relatively few fully commercialized state-of-the-art microgrids have been deployed by utilities in many regions.
- Advance computational capabilities by improving physical models that leverage machine learning and automation.
- Dedicate some investment to a microgrid design authority to improve microgrid operations reliability, security and self-healing capabilities in intelligent grid operation for electricity distribution.

#### Sample Vendors

Alencon Systems; Ameresco; Generac Power Systems; NRG Energy; PowerSecure; Schneider Electric; Siemens; Veritone; Yokogawa Electric Corporation (PXiSE Energy Solutions)

#### Gartner Recommended Reading

2023 Utility Trend: Orchestrate Flexible Resources to Maintain Power System Operational Integrity

### Software-Defined Infrastructure

**Analysis By:** Philip Dawson

**Benefit Rating:** Low

**Market Penetration:** 20% to 50% of target audience

**Maturity:** Obsolete

#### Definition:

Software-defined infrastructure (SDI) enables abstraction of the physical infrastructure, with its services exposed via APIs enabling greater levels of automation, policy-based orchestration and reuse. SDI includes software-defined data center, network, storage, compute and SD edge infrastructure.

### Why This Is Important

Software-defined is the further abstraction of software from hardware. It enables businesses to be more agile and flexible by enabling programmatic control of the infrastructure through software interfaces. SDI combines compute (SDC), network (SDN) and storage (SDS), but SDI also extends to non-data-center infrastructure, with the use of either software-defined monitoring devices or machines.

### Business Impact

While data center SDI is embedded in other data center initiatives, such as cloud and hyperconverged infrastructure, SDI is now focused on key verticals operating in multiple edge locations, such as retail, manufacturing, retail banking, distribution and utilities. It also continues to extend Internet of Things (IoT), non-data-center SDI and SDS storage initiatives for new IT and software-defined WAN (SD-WAN) operations and functions.

### Drivers

- SDI data center infrastructure is well-covered with compute (SDC), network (SDN, now obsolete), edge (SD-WAN), and storage (SDS), but SDI also extends to non-data-center infrastructure with the use of monitoring devices or machines that are software-defined.
- SDI reaches beyond and between software-defined data centers (SDDCs), leveraging SDI benefits and features for new multimode applications and edge and/or IoT endpoints.
- In 2023, SDI's continued presence of hype is enabled through the use of sensors and adapters that are abstracted through software, stretching SDI to the edge, IoT and operational technology (such as retail point of sale [POS]), rather than traditional, IT-driven SDI through a data center or cloud.
- Key verticals operating in multiple, geographically distributed locations, such as retail, manufacturing, retail banking, distribution and utilities, are extending IoT and non-data-center SDI initiatives for new edge and IoT operations and functions.

### Obstacles

- SDI is now tied to extending data center vendor technology, not interoperability.
- SDI overlaps other integrated systems taxonomy, like hyperconvergence, as it drives cloud to data center and edge adoption.
- SDI continues releasing vendor-specific silo technology (not heterogeneous and service-driven) and, hence, it continues to be obsolete as multivendor interoperability standards and technology silos persist, limiting SDI integration between vendors.
- SD-WAN segmentation is driving SDI to the edge and is architecturally different from SDN, which is focused more on data center infrastructure convergence.

### User Recommendations

- Include the integration and measurement of non-data-center edge infrastructure, as SDI initiatives roll out tied to SD-WAN and edge initiatives.
- Focus on core IT SDI for compute, network, storage and facilities, but expand the impact of SDI on IoT, edge computing, remote office/branch office (ROBO) and other operational technologies.
- Anticipate SDI to be tied to a specific vendor or technology silo, such as SDS storage and SD-WAN network hardware or virtualization software. Be cautious not to commit to a vendor's SDI without realizing the specific area of lock-in.

### Sample Vendors

IBM (Red Hat); Intel; Microsoft; VMware; Wipro Enterprises

### Gartner Recommended Reading

Predicts 2023: XaaS Is Transforming Data Center Infrastructure

How Do I Plan for Migrating My Data Center Infrastructure Into an XaaS Model?

## DCIM Tools

**Analysis By:** Henrique Cecci

**Benefit Rating:** Transformational

**Market Penetration:** 5% to 20% of target audience

**Maturity:** Mature mainstream

### Definition:

Data center infrastructure management (DCIM) tools of third-generation monitor, measure, manage and sometimes control data center resources and energy consumption of both IT-related equipment (such as servers, storage and network switches) and facilities infrastructure components (such as PDUs and CRACs) with focus on optimization opportunities and sustainability.

### Why This Is Important

Unlike general building management system tools, DCIM tools are used to optimize power, cooling and physical space within data centers. These tools support near-real-time monitoring of power and environmental factors, resource management and reporting/visualization, and they usually include other capabilities as well.

### Business Impact

The business impact of DCIM can be significant as it may include:

- **Improved energy efficiency:** It helps to identify and eliminate inefficiencies in power and cooling.
- **Better capacity planning:** By providing more accurately forecast capacity needs.
- **Enhanced compliance:** By supporting regulatory requirements and industry standards related to energy efficiency and environmental impact.
- **Simplified management**: Including IT equipment, power and cooling infrastructure, and environmental conditions.

### Drivers

- **Cost savings:** DCIM helps to identify underutilized or inefficient resources, reducing the need for additional hardware purchases and lowering operating costs.
- **Sustainability:** DCIM enables more efficient use of power and cooling resources, reducing energy costs and environmental impact.
- **Capacity planning:** DCIM provides insights into resource availability, allowing for better capacity planning and avoiding costly downtime due to capacity constraints.
- **Compliance:** DCIM helps organizations meet regulatory requirements and industry standards related to data center operations.
- **Improved decision making:** DCIM provides real-time data and analytics, enabling informed decision-making and facilitating proactive problem-solving.

### Obstacles

- **Complexity:** DCIM solutions can be complex and may require significant training.
- **Integration:** DCIM solutions must be integrated with existing IT systems and other I&O tools.
- **Implementation costs:** The initial investment required to adopt DCIM solutions can be significant, including software licenses, hardware, sensors and implementation costs.
- **Data quality:** Data collection and management processes must be well-defined and executed effectively to ensure data quality.
- **Vendor lock-in:** DCIM solutions are often provided by a single vendor, which can lead to vendor lock-in and limit flexibility and choice for organizations.

### User Recommendations

When adopting DCIM tools I&O leaders should implement or evaluate the following recommendations:

- Define goals and requirements to align with your needs.
- Evaluate multiple DCIM tools for functionality, scalability and integration.
- Ensure compatibility with existing infrastructure and systems.

- Plan for accurate data collection from various sources.
- Choose a scalable tool that accommodates future growth.
- Train and involve staff from relevant teams.
- Establish standardized data center processes aligned with the tool.
- Utilize analytics and reporting capabilities for data monitoring and analysis.
- Regularly update and maintain the DCIM tool for security and reliability.
- Evaluate the ROI by measuring benefits and costs.

**Sample Vendors**

Device42; FNT Software; Nlyte Software; Schneider Electric; Sunbird Software

**Gartner Recommended Reading**

Market Guide for Digital Platform Conductor Tools

To Maximize the Value of Data Centers, Combine DCIM Tools With Other Sources

## Entering the Plateau

### Continuous Configuration Automation

**Analysis By:** Chris Saunderson

**Benefit Rating:** High

**Market Penetration:** More than 50% of target audience

**Maturity:** Mature mainstream

#### Definition:

Continuous configuration automation (CCA) tools enable infrastructure administrators and developers to automate the deployment, configuration and operation of systems and software. They support the definition, deployment and maintenance of configuration states and settings. Most CCA tools have an open-source heritage, and most offer commercial support.

#### Why This Is Important

CCA tools are critical to delivering operational efficiency. They enable automation and DevOps initiatives by deploying and managing infrastructure elements, associated software and configuration changes as code. In combination with infrastructure automation tools, CCA tools form the core of infrastructure-as-code (IaC) capabilities. They also enable the automation of Day 2 operations of deployed systems, expanding their reach into networking, containers, compliance and security use cases.

#### Business Impact

By enabling automated deployment and configuration of systems, settings and software programmatically, organizations realize:

- **Agility improvements:** CI/CD enablement for infrastructure and operations (I&O) services.
- **Productivity gains:** Repeatable, declarative, version-controlled infrastructure deployment/operation.
- **Cost optimization:** Reductions in manual interventions by skilled staff.
- **Risk mitigation:** Compliance assessment and remediation using standardized processes Improved change success and reliability are also realized.

### Drivers

- Organizations need a broader set of deployment and automation functions beyond configuration management, including IaC, patching, application release orchestration (ARO), configuration assessment and auditing (e.g., for regulatory or internal policy compliance) and orchestration of operational tasks.
- CCA provides an automation framework that enables deployment automation and Day 2 operational automation of changes, compliance, and response to incidents or problems.
- CCA tools are essential for I&O administrators to mature from task-based scripting to a more structured approach to automation and delivery.
- There is a need for repeatable, declarative, standardized deployments to be made available to end users or I&O administrators that enable quick delivery of infrastructure to meet end-user needs and I&O policies and baselines.

### Obstacles

- Confusion around the capabilities and overlaps of automation tools causing conflict and overinvestment.
- Developers and administrators may use CCA in a silo, further inhibiting enterprisewide adoption.
- IT skill sets hinder the adoption of these tools, requiring source code management and software engineering skills to make full use of capabilities.
- The growing use of IaC tools has created confusion about the role of CCA tools; however, CCA tools are necessary to deliver effective and efficient IaC.

### User Recommendations

- Clarify the role that CCA tools fulfill in their toolchain and make selections based on the tasks that are in scope.
- Evaluate the availability of content against organizational use cases. Prioritize CCA tools that provide out-of-the-box content that addresses current pain points and accelerates time to value.
- Include both professional services for enablement and training requirements in cost evaluations. Costs associated with CCA tools extend beyond just the licensing cost.

- Expect to invest in training beyond tool implementation to fully realize the benefits of these tools.
- Guard against developers and administrators reverting to known imperative scripting methods to complete specific tasks in place of using CCA capabilities.
- Maximize the value of CCA tool investments by ensuring that your organization's culture can embrace CCA tools strategically and automate toil (for DevOps and I&O leaders).

### Sample Vendors

Inedo; Perforce (Puppet); Progress; Red Hat; Rudder; VMware

### Gartner Recommended Reading

Market Guide for Infrastructure Automation Tools

To Automate Your Automation, Apply Agile and DevOps Practices to Infrastructure and Operations

Innovation Insight for Continuous Infrastructure Automation

## Hyperconvergence

**Analysis By:** Philip Dawson, Jeffrey Hewitt

**Benefit Rating:** High

**Market Penetration:** More than 50% of target audience

**Maturity:** Mature mainstream

### Definition:

Hyperconvergence combines storage, computing and networking into a single system that reduces data center complexity and increases scalability. Multiple servers can be clustered together to create pools of shared compute and storage resources (or nodes), designed for convenient consumption. Delivery models include physical and virtual appliances, reference architectures, as a service or public cloud.

### Why This Is Important

Infrastructure and operations (I&O) leaders seeking a cost-effective solution with a single management interface that excludes proprietary, external hardware controller-based storage should consider hyperconvergence as a viable option. Possible use cases include virtual desktop infrastructure (VDI), edge/Internet of Things (IoT), hybrid cloud and cloud-native.

### Business Impact

Hyperconvergence enables IT leaders to be responsive to new business requirements in a modular, small-increment fashion, avoiding the large-increment upgrades typically found in three-tier infrastructure architectures. It is of particular value to midsize enterprises that can standardize on hyperconvergence and to the remote sites of large organizations that need cloudlike management efficiency with on-premises edge infrastructure.

Drivers

- Hyperconvergence provides simplified management that decreases the pressure to hire hard-to-find specialists. Adoption is greatest in dynamic organizations with short business planning cycles and long IT planning cycles tied to hybrid cloud delivery. The hyperconverged infrastructure (HCI) market is now trifurcating, focusing on the data-center-led “hybrid cloud” management use case with cloud-native applications, the VDI use case and the “edge/IoT” remote management use case.
- Hyperconvergence leads to lower operating costs, especially as it supports a greater share of the compute and storage requirements of the data center.
- Nutanix, an early innovator in hyperconverged integrated system (HCIS) hardware appliances, has largely shifted to a Hyper Converged Infrastructure (HCI) software revenue model and continues to increase its number of OEM relationships and partners.
- Larger clusters are now in use, and midsize organizations are considering hyperconvergence as the preferred alternative for on-premises infrastructure for block storage.
- Hyperconvergence vendors are achieving certification for more demanding workloads, including Oracle and SAP, and end users are beginning to consider hyperconvergence as an alternative to integrated infrastructure systems for some workloads.
- As more vendors support hybrid and public cloud deployments, hyperconvergence is a stepping stone toward public cloud agility as suppliers are expanding hybrid cloud deployment offerings for cloud-native applications.
- A number of niche hyperconvergence suppliers offer scale-down solutions to address the needs of remote office/branch office (ROBO) and edge environments.

### Obstacles

- Applications designed for scale-up architectures (as opposed to scale-out ones) are unlikely to meet cost or performance expectations when deployed on hyperconverged infrastructure.
- The acquisition cost of hyperconvergence may be higher, and the resource utilization rate lower than for three-tier architectures.
- While HCI has somewhat matured from a hypervisor compute and storage function, software defined in networking is split between the obsolete software-defined networking (SDN) and networking around software-defined WAN (SD-WAN), driving edge deployments.
- For large organizations, hyperconverged deployments will remain another silo to manage.

### User Recommendations

- Implement hyperconvergence for hybrid cloud infrastructure and cloud-native applications when agility, modular growth and management simplicity are of greatest importance.
- Establish that hyperconvergence requires alignment of compute, network and storage refresh cycles; consolidation of budgets; operations and capacity planning roles; and retraining for organizations still operating separate silos.
- Test the impact on disaster recovery and networking under a variety of failure scenarios, as solutions vary greatly in performance under failure, their time to return to a fully protected state and the number of failures they can tolerate.
- Ensure that clusters are sufficiently large to meet performance and availability requirements during single and double node failures, and require proofs of concept to reveal any performance anomalies.

### Sample Vendors

Cisco; Dell; Microsoft; Nutanix; Sangfor; Scale Computing; StorMagic; VMware

### Gartner Recommended Reading

Market Guide for Full-Stack Hyperconverged Infrastructure Software

Gartner Peer Insights 'Voice of the Customer': Hyperconverged Infrastructure Software

# Appendixes

## Hype Cycle Phases, Benefit Ratings and Maturity Levels

### Table 2: Hype Cycle Phases

(Enlarged table in Appendix)

| *Phase* ↓ | *Definition* ↓ |
|---|---|
| *Innovation Trigger* | A breakthrough, public demonstration, product launch or other event generates significant media and industry interest. |
| *Peak of Inflated Expectations* | During this phase of overenthusiasm and unrealistic projections, a flurry of well-publicized activity by technology leaders results in some successes, but more failures, as the innovation is pushed to its limits. The only enterprises making money are conference organizers and content publishers. |
| *Trough of Disillusionment* | Because the innovation does not live up to its overinflated expectations, it rapidly becomes unfashionable. Media interest wanes, except for a few cautionary tales. |
| *Slope of Enlightenment* | Focused experimentation and solid hard work by an increasingly diverse range of organizations lead to a true understanding of the innovation's applicability, risks and benefits. Commercial off-the-shelf methodologies and tools ease the development process. |
| *Plateau of Productivity* | The real-world benefits of the innovation are demonstrated and accepted. Tools and methodologies are increasingly stable as they enter their second and third generations. Growing numbers of organizations feel comfortable with the reduced level of risk; the rapid growth phase of adoption begins. Approximately 20% of the technology's target audience has adopted or is adopting the technology as it enters this phase. |
| *Years to Mainstream Adoption* | The time required for the innovation to reach the Plateau of Productivity. |

Source: Gartner (July 2023)

Table 3: Benefit Ratings

| ***Benefit Rating*** | ***Definition*** |
| --- | --- |
| *Transformational* | Enables new ways of doing business across industries that will result in major shifts in industry dynamics |
| *High* | Enables new ways of performing horizontal or vertical processes that will result in significantly increased revenue or cost savings for an enterprise |
| *Moderate* | Provides incremental improvements to established processes that will result in increased revenue or cost savings for an enterprise |
| *Low* | Slightly improves processes (for example, improved user experience) that will be difficult to translate into increased revenue or cost savings |

Source: Gartner (July 2023)

**Table 4: Maturity Levels**

(Enlarged table in Appendix)

| *Maturity Levels* ↓ | *Status* ↓ | *Products/Vendors* ↓ |
|---|---|---|
| *Embryonic* | In labs | None |
| *Emerging* | Commercialization by vendors<br>Pilots and deployments by industry leaders | First generation<br>High price<br>Much customization |
| *Adolescent* | Maturing technology capabilities and process understanding<br>Uptake beyond early adopters | Second generation<br>Less customization |
| *Early mainstream* | Proven technology<br>Vendors, technology and adoption rapidly evolving | Third generation<br>More out-of-box methodologies |
| *Mature mainstream* | Robust technology<br>Not much evolution in vendors or technology | Several dominant vendors |
| *Legacy* | Not appropriate for new developments<br>Cost of migration constraints replacement | Maintenance revenue focus |
| *Obsolete* | Rarely used | Used/resale market only |

Source: Gartner (July 2023)

## Recommended by the Authors

Some documents may not be available as part of your current Gartner subscription.

Understanding Gartner's Hype Cycles

Create Your Own Hype Cycle With Gartner's Hype Cycle Builder

Predicts 2023: XaaS Is Transforming Data Center Infrastructure

Market Guide for Consumption-Based Models for Data Center Infrastructure

Market Guide for Digital Platform Conductor Tools

Market Guide for Servers

Market Guide for Integrated Systems

Competitive Landscape: Consumption-Based Model for On-Premises Infrastructure

Emerging Tech: Top Semiconductor Technology Trends Impacting Data Centers for 2023

Emerging Tech Impact Radar: Compute and Storage

**Table 1: Priority Matrix for Data Center Infrastructure Technologies, 2023**

| *Benefit* | *Years to Mainstream Adoption* | | | |
|---|---|---|---|---|
| ↓ | ***Less Than 2 Years*** ↓ | ***2 - 5 Years*** ↓ | ***5 - 10 Years*** ↓ | ***More Than 10 Years*** ↓ |
| Transformational | | DCIM Tools<br>Edge Computing | Digital Platform Conductor Tools<br>Network Digital Twin | Emerging Memory Technologies |
| High | Continuous Configuration Automation<br>Hyperconvergence | Composable Infrastructure<br>Consumption-Based Model<br>Digital Twins for Sustainability<br>Infrastructure Automation<br>Infrastructure Orchestration<br>Intelligent Infrastructure<br>Programmable Infrastructure | Infrastructure Platform Engineering<br>Intelligent Platforms<br>IT/OT Hybrid Servers<br>Microgrids<br>Net-Zero Data Centers<br>Off-Grid Power<br>Programmable Platform | Circular Economy in IT |
| Moderate | | Augmented Reality in Data Centers<br>Immersion Cooling<br>Immutable Infrastructure | Direct-to-Chip Liquid Cooling<br>Hydrogen-Powered Data Centers | |
| Low | | | | |

Source: Gartner (July 2023)

## Table 2: Hype Cycle Phases

| *Phase* ↓ | *Definition* ↓ |
| --- | --- |
| *Innovation Trigger* | A breakthrough, public demonstration, product launch or other event generates significant media and industry interest. |
| *Peak of Inflated Expectations* | During this phase of overenthusiasm and unrealistic projections, a flurry of well-publicized activity by technology leaders results in some successes, but more failures, as the innovation is pushed to its limits. The only enterprises making money are conference organizers and content publishers. |
| *Trough of Disillusionment* | Because the innovation does not live up to its overinflated expectations, it rapidly becomes unfashionable. Media interest wanes, except for a few cautionary tales. |
| *Slope of Enlightenment* | Focused experimentation and solid hard work by an increasingly diverse range of organizations lead to a true understanding of the innovation's applicability, risks and benefits. Commercial off-the-shelf methodologies and tools ease the development process. |
| *Plateau of Productivity* | The real-world benefits of the innovation are demonstrated and accepted. Tools and methodologies are increasingly stable as they enter their second and third generations. Growing numbers of organizations feel comfortable with the reduced level of risk; the rapid growth phase of adoption begins. Approximately 20% of the technology's target audience has adopted or is adopting the technology as it enters this phase. |
| *Years to Mainstream Adoption* | The time required for the innovation to reach the Plateau of Productivity. |

| ***Phase*** ↓ | ***Definition*** ↓ |
| --- | --- |

Source: Gartner (July 2023)

## Table 3: Benefit Ratings

| ***Benefit Rating*** ↓ | ***Definition*** ↓ |
| --- | --- |
| *Transformational* | Enables new ways of doing business across industries that will result in major shifts in industry dynamics |
| *High* | Enables new ways of performing horizontal or vertical processes that will result in significantly increased revenue or cost savings for an enterprise |
| *Moderate* | Provides incremental improvements to established processes that will result in increased revenue or cost savings for an enterprise |
| *Low* | Slightly improves processes (for example, improved user experience) that will be difficult to translate into increased revenue or cost savings |

Source: Gartner (July 2023)

### Table 4: Maturity Levels

| *Maturity Levels* | *Status* | *Products/Vendors* |
|---|---|---|
| *Embryonic* | In labs | None |
| *Emerging* | Commercialization by vendors<br>Pilots and deployments by industry leaders | First generation<br>High price<br>Much customization |
| *Adolescent* | Maturing technology capabilities and process understanding<br>Uptake beyond early adopters | Second generation<br>Less customization |
| *Early mainstream* | Proven technology<br>Vendors, technology and adoption rapidly evolving | Third generation<br>More out-of-box methodologies |
| *Mature mainstream* | Robust technology<br>Not much evolution in vendors or technology | Several dominant vendors |
| *Legacy* | Not appropriate for new developments<br>Cost of migration constraints replacement | Maintenance revenue focus |
| *Obsolete* | Rarely used | Used/resale market only |

Source: Gartner (July 2023)