2022 Planning Guide for Cloud and Edge Computing

By Paul Delory, Lydia Leong, Tony Iams, Matthew Brisse, Douglas Toombs, Angelina Troy, Stanton Cole, Fintan Quinn, Mohini Dukes, Marco Meinardi

Gartner®

2022 Planning Guide for Cloud and Edge Computing

Published 11 October 2021 - ID G00753853 - 74 min read

By Analyst(s): Paul Delory, Lydia Leong, Tony Iams, Matthew Brisse, Douglas Toombs, Angelina Troy, Stanton Cole, Fintan Quinn, Mohini Dukes, Marco Meinardi

Initiatives: Cloud Computing for Technical Professionals

Cloud adoption continues to accelerate as businesses seek agility, flexibility and new sources of competitive advantage. I&O technical professionals focused on cloud must design a scalable, resilient and secure architecture. This Planning Guide explores six key technical planning trends for 2022.

Overview

Key Findings

- Businesses need the flexibility and scalability of cloud services to respond in real time to rapidly shifting and unpredictable economic conditions. Organizations that can successfully do this will have a competitive advantage over rivals.

- Almost every IT organization is using public cloud, but rarely in an optimized way. In 2022, new advancements make it possible to improve cloud deployments and extend them to previously unsuitable use cases.

- Hybrid and multicloud models remain popular, but they increase complexity and cost. Kubernetes has become the de facto standard for cross-cloud orchestration and a pillar of cloud architectures.

- The lack of cloud skills has reached crisis levels. This is often the real barrier to cloud adoption. Businesses are missing out on opportunities for cost savings and superior technology because their IT organizations lack the skills to support cloud services.

Recommendations

As a technical professional focused on cloud computing in 2022, you should:

- Design cloud infrastructure for maximum speed and business value, even at the expense of technical optimization. Give business units the flexibility they need; you can optimize later.

- Build resilience into your applications, rather than your infrastructure. In the cloud, resilience should be a function of the application code, not the result of a patchwork of infrastructure technologies.

- In multicloud architectures, prioritize a primary strategic cloud provider, and favor its native tools. Don't move to a secondary provider or introduce third-party tools, until you need capabilities that your primary cloud provider doesn't have.

- Define a formal cloud strategy. You need a workload placement rubric, a portability framework, pervasive automation, cost controls, disaster recovery and an exit strategy. Your cloud strategy should also include a talent enablement program that will help to bridge skills gaps.

Cloud and Edge Computing Trends

Download All Graphics in This Material

Businesses worldwide are preparing for 2022, which will be a year of economic uncertainty, anxiety and threat. In the year ahead, they need the flexibility and scalability of cloud services to respond in real time to rapidly shifting and unpredictable conditions. The COVID-19 pandemic, whose aftereffects seem poised to continue well into 2022, exposed that many enterprises were not as agile as they thought. IT organizations faced unforeseen, urgent demands on their infrastructure. These included virtual desktops for a suddenly remote workforce, migration of data out of closed offices, scaling infrastructure to handle unprecedented growth in online orders and replacing software designed for in-person customer service with new code. Not surprisingly, many IT organizations hurriedly deployed cloud services, some for the first time.

Cloud adoption accelerated rapidly in 2021, and Gartner predicts it will accelerate further in the years to come. Businesses that can successfully exploit cloud will have a competitive advantage. It might even determine whether they survive.

Nearly every enterprise already makes at least some use of the public cloud; to avoid cloud entirely is almost inconceivable. But many cloud implementations are poor. Too often, deployments are ad hoc, implementations are inconsistent, costs are poorly controlled or governance is incomplete (or nonexistent). Moreover, there remain some barriers to cloud adoption. Issues around data sovereignty, data gravity, latency and cloud economics, among others, have forced certain workloads to remain on traditional infrastructure. In the year ahead, new cloud capabilities and business-focused architectures will address these common shortcomings and knock down some of the remaining barriers to cloud adoption. But cloud technical professionals must learn new skills and embrace new ways of working.

> **In 2022, the cloud will see the emergence of disruptive new technologies and a renewed focus on business value. These will serve to improve cloud implementations, encourage cloud adoption and tie cloud services ever closer to the bottom line.**

First, cloud deployments must prioritize bottom-line business outcomes above technical considerations. Don't delay implementation or upgrades of cloud services on technical grounds. Give business units the flexibility they need; you can optimize later.

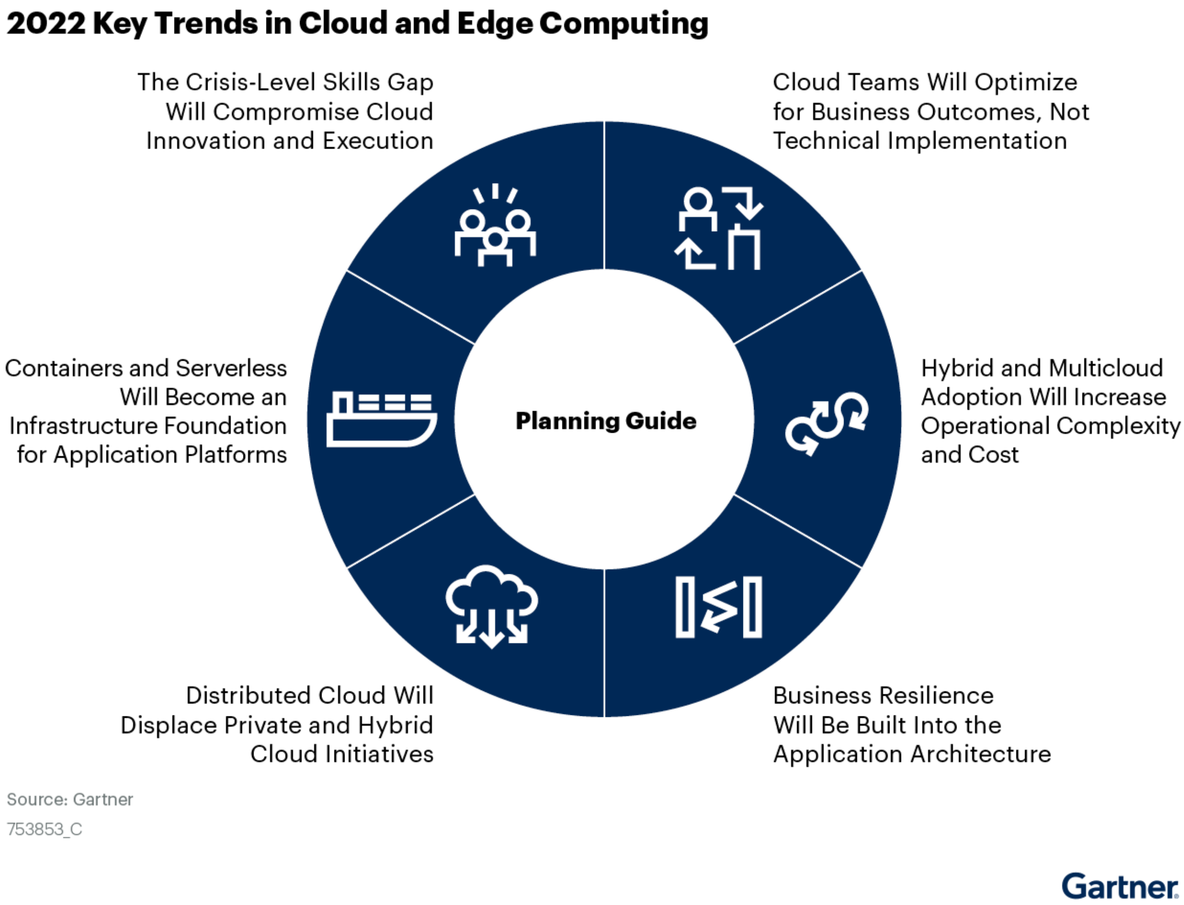
Second, the distinction between infrastructure and application is becoming increasingly blurred. Many capabilities are best implemented in the application code itself, rather than baked into the infrastructure. For example, build business continuity and disaster recovery (BC/DR) into application code, not infrastructure. Also, use serverless technology — which includes not only cloud functions, but also hosted containers — to deploy code rapidly without needing to manage the underlying infrastructure.

Third, use the new crop of “distributed cloud” offerings to bring native cloud services on-premises. Even if you can't move a workload, it can still use real, native cloud services. The major public cloud providers now sell turnkey on-premises infrastructure. In 2022, this market is still emerging, but there is a very real possibility that in the future you will not build your on-premises infrastructure, but rather acquire it from a cloud provider.

Lastly, but perhaps most importantly, technical professionals must grow their cloud skills. In many IT organizations, this is the real barrier to cloud adoption. Many cloud projects have been aborted because technical professionals lacked the skill to deploy and manage them effectively. Cloud skills are among the most in-demand (and lucrative) on today's job market. Improving yours is both highly beneficial and an excellent career move.

In 2022, cloud and edge computing will be shaped by six long-term trends, which must inform the decisions of cloud technical professionals. Figure 1 illustrates these trends.

Figure 1: 2022 Key Trends in Cloud and Edge Computing


2022 Key Trends in Cloud and Edge Computing
The Crisis-Level Skills Gap Will Compromise Cloud Innovation and Execution
Cloud Teams Will Optimize for Business Outcomes, Not Technical Implementation
Containers and Serverless Will Become an Infrastructure Foundation for Application Platforms
Planning Guide
Hybrid and Multicloud Adoption Will Increase Operational Complexity and Cost
Distributed Cloud Will Displace Private and Hybrid Cloud Initiatives
Business Resilience Will Be Built Into the Application Architecture
Gartner

Source: Gartner
753853_C

The remainder of this research is a detailed exploration of these six technical planning trends:

1. Cloud teams will optimize for business outcomes, not technical implementation.

2. Hybrid and multicloud adoption will increase operational complexity and cost.

3. Business resilience will be built into the application architecture.

4. Distributed cloud will displace private and hybrid cloud initiatives.

5. Containers and serverless will become an infrastructure foundation for application platforms.

6. The crisis-level skills gap will compromise cloud innovation and execution.

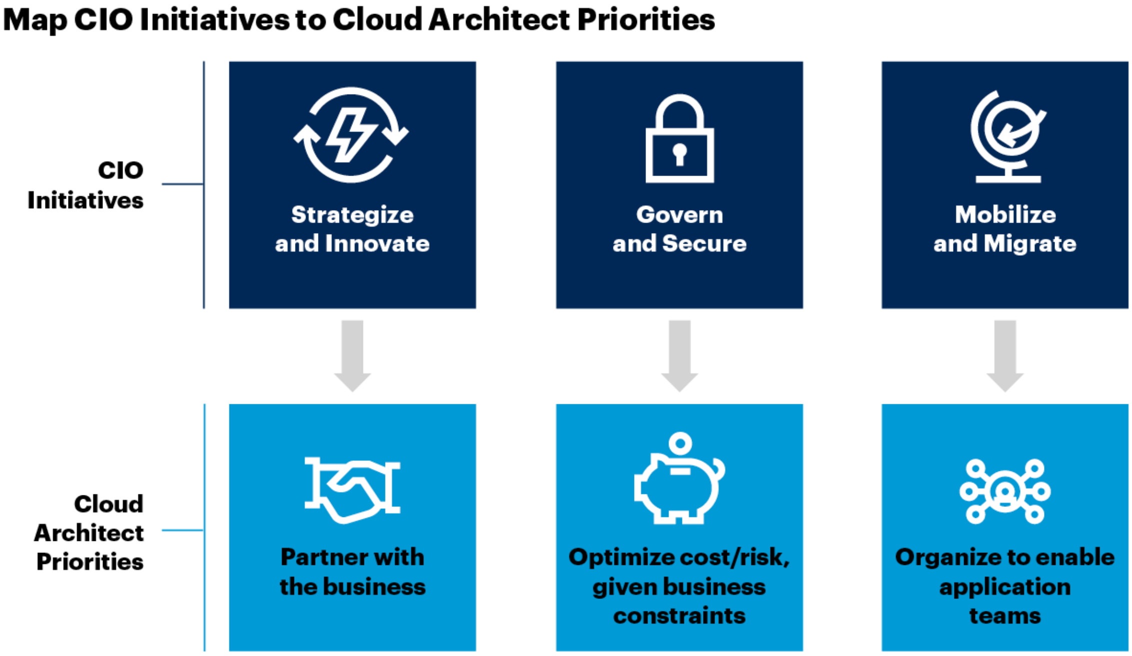
Cloud Teams Will Optimize for Business Outcomes, Not Technical Implementation

The business climate of 2022 will be one of uncertain conditions and global variability. Businesses need to be adaptable. They must respond to changes quickly, making hard decisions under conditions of uncertainty. Moreover, the COVID-19 pandemic has led to lasting changes: The shift to hybrid work and digital touchpoints is likely permanent (see Forecast Analysis: Remote and Hybrid Workers, Worldwide). In this climate, businesses that are able to rapidly seize opportunities — including short-term opportunities — created by this shifting landscape will be more successful. Technical professionals may not always consider the business implications of the work they do. But cloud services do have an important role to play in making the business they serve more agile and, therefore, more likely to succeed.

Certainly, many business leaders see cloud services this way. But, too often, cloud technical professionals do not. We often focus on the minutiae of engineering and architectural details at the expense of the more important consideration: the cloud as a source of business value. In this climate, it is imperative you do not let debates over architectures or optimal workload placement derail the quick transition to cloud services — even if that means making uncomfortable compromises in initial implementation quality. To be sure, cloud technical professionals cannot abdicate their responsibilities to keep cloud-driven business functions safe and highly available. But, beyond that, the pursuit of technical perfection cannot take precedence over the pursuit of business-driven innovation and experimentation.

First and foremost, cloud technical professionals need to support executive-driven initiatives for cloud acceleration. See Leverage Cloud for Business Innovation and Transformation for the CIO perspective. Figure 2 illustrates how CIOs think about cloud services as a business-enabling platform for the organization and how technical professionals can best embody that.

Figure 2: Map CIO Initiatives to Cloud Architect Priorities


Map CIO Initiatives to Cloud Architect Priorities
CIO Initiatives
Strategize and Innovate
Govern and Secure
Mobilize and Migrate
Cloud Architect Priorities
Partner with the business
Optimize cost/risk, given business constraints
Organize to enable application teams

Source: Gartner
753853_C

Cloud technical professionals have a critical role to play in achieving these three CIO-level priorities:

- **Strategize and innovate:** Cloud teams must be capable of partnering with the business to deliver solutions to business problems. They must bring a business mindset, not merely technology skills, to this work. Technical professionals must be able to ideate with the business, making business leaders aware of business innovations that could be enabled by new cloud technical capabilities.

- **Govern and secure:** The emphasis on faster business outcomes means that cloud teams may be pressured to take implementation “shortcuts” or to accept larger risks in return for potentially faster delivery. This does not mean cloud teams can abrogate their responsibility to lay the proper cloud foundations or that they should accept a bottomless pit of technical debt. Cloud teams must implement adaptable governance frameworks that have the flexibility to handle different implementation demands and risk profiles. Cloud teams should not seek the lowest cost and risk but, rather, optimize cost and risk based on the business needs.

- **Mobilize and migrate:** Cloud teams must be organized for cloud success and to support the transformation of the organization as a whole. Cloud-enabled business outcomes are likely to be supported by serving the needs of business application owners or application teams that want to deploy new applications or migrate existing applications into the cloud.

Cloud technical professionals must meet the business's need for urgency in as technically responsible a fashion as possible. In 2022, therefore, cloud technical professionals must:

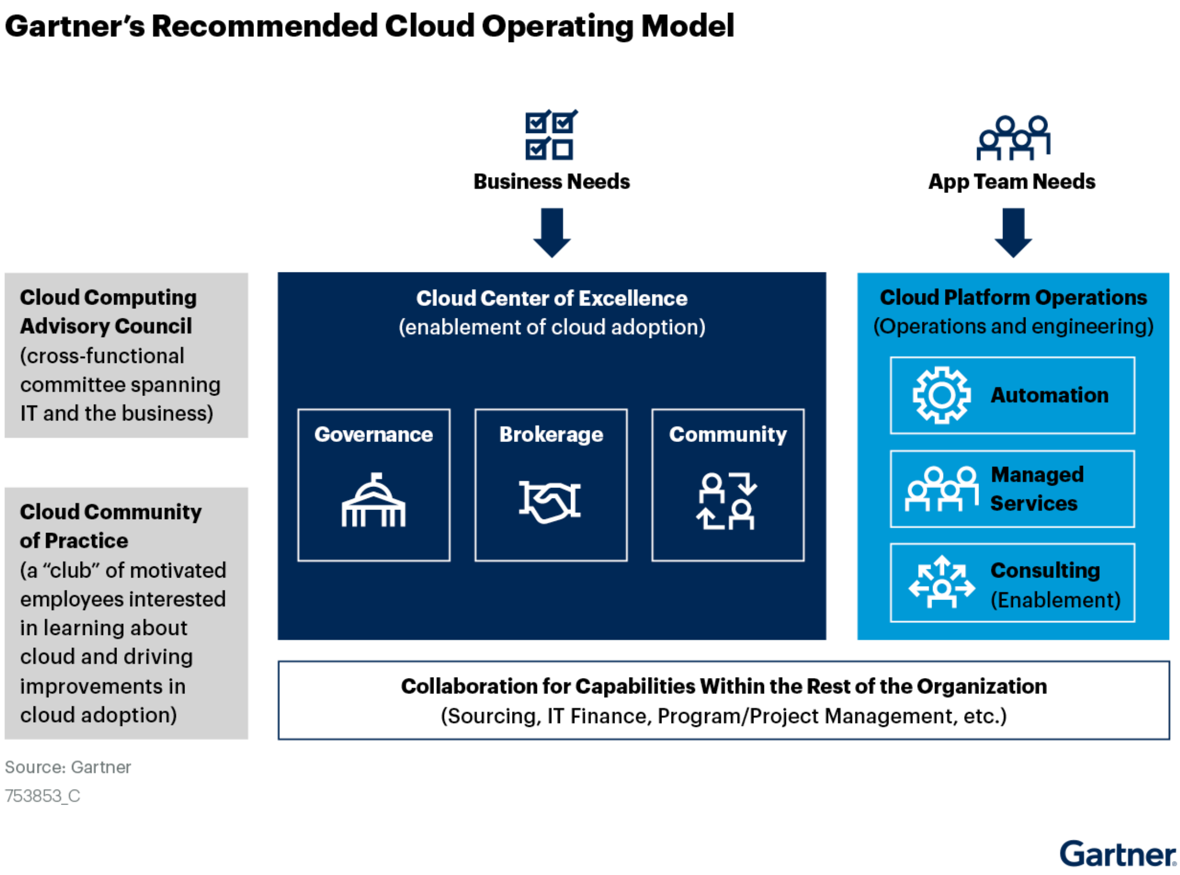
- Develop a strategic cloud operating model.
- Prioritize the cloud adoption framework.
- Mitigate risks created by suboptimal cloud adoption.

Planning Considerations

Develop a Strategic Cloud Operating Model

Successful journeys to cloud computing require collaboration across a wide variety of functional areas within the organization. Collaboration is necessary not just across IT teams, but also between IT and the business. Gartner recommends the creation of formal organizational structures to grow and mature cloud adoption, as illustrated in Figure 3.

Figure 3: Gartner’s Recommended Cloud Operating Model



Gartner’s Recommended Cloud Operating Model

Business Needs

App Team Needs

Cloud Computing Advisory Council (cross-functional committee spanning IT and the business)

Cloud Community of Practice (a “club” of motivated employees interested in learning about cloud and driving improvements in cloud adoption)

Cloud Center of Excellence (enablement of cloud adoption)

Governance

Brokerage

Community

Cloud Platform Operations (Operations and engineering)

Automation

Managed Services

Consulting (Enablement)

Collaboration for Capabilities Within the Rest of the Organization (Sourcing, IT Finance, Program/Project Management, etc.)

Source: Gartner
753853_C

The key organizational elements, as described in How to Build a Cloud Center of Excellence (Part 1 — Designing for Cloud Adoption Success), are:

- **Cloud computing advisory council:** Many organizations create a cross-functional “cloud computing committee” near the start of their cloud journey. Gartner recommends that this effort be formalized as a cross-functional, part-time committee that is tasked with being a sounding board for cloud strategy and policy.
- **Cloud community of practice (CCOP):** A community of practice is an practitioner-led organization where members who share a common interest collaborate to better themselves and their organization. The cloud COP is effectively a “club” that focuses on sharing cloud knowledge and experiences, and is a useful way to discover best practices, drive cloud evangelism and improve cloud skills in the organization.

- **Cloud center of excellence (CCOE):** The CCOE is an enterprise architecture function that drives the cloud strategy, governs and brokers the use of cloud computing in the organization, and guides the cloud community in its transformational efforts. It is staffed by cloud architects and serves the key functions of enabling the business with cloud solutions, and setting policy and direction for cloud implementation efforts.

In addition, you will need a cloud operations function that is responsible for cloud implementation and day-to-day operations. See Comparing Cloud Operations Approaches for an exploration of possible ways to implement this function. If you choose to implement this function yourself rather than outsourcing it to a cloud managed service provider (MSP), the pattern that Gartner sees as most frequently successful is “cloud platform operations.” This pattern closely imitates the delivery patterns of high-quality cloud MSPs.

Benefits:

- **Automation:** Automation engineers develop and integrate a collection of software tools that provide a foundation for cloud governance and operations, together with standard orchestration templates and scripts that serve as baseline patterns for cloud-related automation.
- **Service management:** Cloud operators perform service management tasks that cannot be automated, such as cloud-related incident management, problem management, system administration, security operations and other support activities.
- **Consulting services:** Cloud engineers implement cloud architecture and engineering projects. This includes consultative assistance to application teams and other technical end users that need help building, automating and securing cloud solutions, including infrastructure as code.
- **Operational maturity:** Establishing and maturing these fundamental capabilities — for cloud architecture, governance, brokerage, transformation and operations — is vital for driving mature, repeatable, scalable, risk-managed and cost-efficient cloud adoption.

Recommended research:

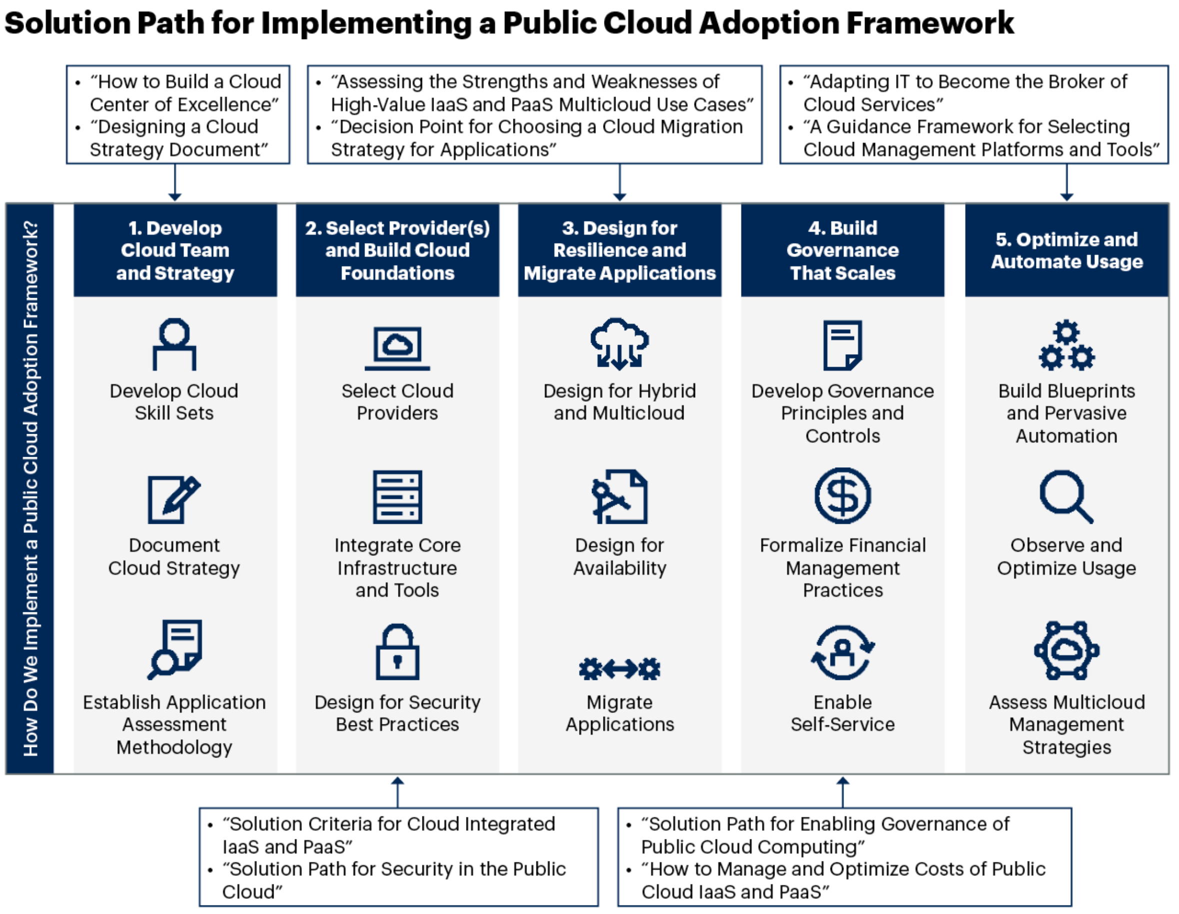
- Comparing Cloud Operations Approaches

- Guidance Framework for Implementing CloudPlatform Operations
- How to Build a Cloud Center of Excellence (Part 1 — Designing for Cloud Adoption Success)
- How to Build a Cloud Center of Excellence (Part 2 — Implementing the Foundations for Cloud Adoption Success)

Prioritize Your Cloud Adoption Framework

Infrastructure, applications and data will continue to proliferate everywhere. They would have done so even without a global pandemic, but the lasting changes brought about by COVID-19 have made distributed infrastructure even more important. Now, IT organizations must move beyond simply adopting cloud for some use cases and start making hard decisions about permanently shutting down some (or perhaps all) of their on-premises infrastructure capacity. Doing so must start with a well-planned, structured approach for cloud adoption, designed to avoid potential complexities and common pitfalls. The starting point for developing this process is Gartner's Solution Path for Implementing a Public Cloud Adoption Framework, which provides the high-level cloud adoption framework illustrated in Figure 4.

Figure 4: Solution Path for Implementing a Public Cloud Adoption Framework


Solution Path for Implementing a Public Cloud Adoption Framework
• "How to Build a Cloud Center of Excellence"
• "Designing a Cloud Strategy Document"
• "Assessing the Strengths and Weaknesses of High-Value IaaS and PaaS Multicloud Use Cases"
• "Decision Point for Choosing a Cloud Migration Strategy for Applications"
• "Adapting IT to Become the Broker of Cloud Services"
• "A Guidance Framework for Selecting Cloud Management Platforms and Tools"
How Do We Implement a Public Cloud Adoption Framework?
1. Develop Cloud Team and Strategy
Develop Cloud Skill Sets
Document Cloud Strategy
Establish Application Assessment Methodology
2. Select Provider(s) and Build Cloud Foundations
Select Cloud Providers
Integrate Core Infrastructure and Tools
Design for Security Best Practices
3. Design for Resilience and Migrate Applications
Design for Hybrid and Multicloud
Design for Availability
Migrate Applications
4. Build Governance That Scales
Develop Governance Principles and Controls
Formalize Financial Management Practices
Enable Self-Service
5. Optimize and Automate Usage
Build Blueprints and Pervasive Automation
Observe and Optimize Usage
Assess Multicloud Management Strategies
• "Solution Criteria for Cloud Integrated IaaS and PaaS"
• "Solution Path for Security in the Public Cloud"
• "Solution Path for Enabling Governance of Public Cloud Computing"
• "How to Manage and Optimize Costs of Public Cloud IaaS and PaaS"

Source: Gartner
744641_C

Gartner

A cloud adoption framework enables organizations to rapidly and methodically step through the adoption of cloud services because they can leverage knowledge acquired between cloud projects. Adhering to a framework also allows businesses to scale the use of cloud services more quickly, while ensuring that they retain governance and control. In other words, employing a cloud adoption framework improves efficiency and reduces risk.

Benefits:

- **Governance:** The adoption framework defines cloud computing-related policies. Policy is created in collaboration with a cross-functional team and enforced by a mixture of tools and organizational processes. This approach provides appropriate risk management as well as financial management.

- **Exit strategy:** A good cloud adoption framework includes an exit strategy. This way, the organization has a predefined method of extracting its valuable assets from a cloud service or provider.
- **Scalability:** With a proper cloud adoption framework in place, cloud infrastructure and services can scale predictably and reliably, while maintaining both governance and financial controls.
- **Portability:** A generalized framework for cloud adoption is transferable between cloud services and providers, easing the transition to hybrid and multicloud architectures.

Recommended research:

- Solution Path for Implementing a Public CloudAdoption Framework
- Guidance Framework for Initiating Public Cloud IaaS Adoption Through Pilot Projects
- Designing a Public Cloud Exit Strategy
- How to Successfully Perform Lift-and-Shift Migrations to Public Cloud IaaS

Mitigate Risks Created by Suboptimal Cloud Adoption

We have already discussed why, in the present climate, it will often be necessary to forgo robust technical planning and optimization of cloud computing to deliver business value as quickly as possible. This is the (perhaps unfortunate) reality of operating cloud infrastructure in 2022. But this does not mean that IT organizations can ignore the inherent risks of public cloud adoption. Rather, they will need to mitigate risks retroactively, applying risk mitigation strategies to public cloud infrastructure and services that are already in production.

Many IT organizations are already taking this approach, though perhaps not intentionally. Often, organizations subconsciously accept certain risks and work on mitigation only once their consequences are fully exposed. For example, an IT organization may strengthen its cloud governance only after its sensitive data has been displayed by mistake on a public storage space. Or, the organization may rush to buy a cloud cost management solution only after it has received an unexpectedly high bill at the end of the month.

In 2022, technical professionals should identify risks as early as possible and develop mitigation strategies that help reduce them to an acceptable level. Identifying mitigation early in your cloud adoption program will help you prioritize the development of operational processes that will speed up cloud adoption while still protecting your organization.

Here are the major risks associated with cloud computing, with compensating controls for each:

- **Supplier:** Risks such as vendor lock-in and lack of trust in a cloud vendor can be mitigated with the development of an exit strategy, a multicloud strategy, the scrutiny of the cloud provider's certifications and the use of abstraction layers.
- **Availability:** The unavailability of applications, service disruptions and data loss can be mitigated with the implementation of highly available, multicloud and hybrid architectures. Such architectures allow organizations to replicate data and make applications available outside of a single cloud provider's domain.
- **Confidentiality:** Risks of breaches in the confidentiality of data can be mitigated through governance, the use of security tools, role-based access control, encryption and data anonymization.
- **Compliance:** Not being able to meet the requirements of a regulatory framework presents an "out of compliance" risk for organizations operating in regulated industries. Meeting regulatory requirements when using cloud computing is possible but requires the careful application of processes, self-auditing and compliance checks.
- **Overspending:** Cloud bills are issued based on provisioned resources and usage. Such flexibility can disguise a risk of overspending, especially for organizations that are used to dealing with fixed-cost data centers with yearly budget cycles. Mitigating overspending risks is possible by developing financial management processes that allow organizations to forecast, track and reduce cloud costs on an ongoing basis.

Benefits:

- **Increased confidence:** A well-thought-out risk mitigation strategy shows end users and executives that you've thought through potential challenges, giving them increased confidence both in the infrastructure and in your skills.

- **Early intervention:** Identifying mitigation early in your cloud adoption program will help you prioritize the development of operational processes that will speed up cloud adoption while protecting your organization.
- **Speed with safety:** Most importantly, effective retroactive risk mitigation means the infrastructure can move at the speed of the business, while still maintaining an acceptable (if imperfect) level of governance.

Related research:

- A Guidance Framework for Managing Vendor Lock-In Risks in Cloud IaaS
- Performing Effective Security Risk Assessments of Public Cloud Deployments
- How to Manage and Optimize Costs of Public Cloud IaaS and PaaS

Hybrid and Multicloud Adoption Will Increase Operational Complexity and Cost

In 2020, nearly 80% of cloud buyers surveyed by Gartner that cited use of public, hybrid or multicloud indicated that they worked with more than one cloud provider. [1] By 2022, many organizations will have fully stabilized and operationalized at least one major public cloud infrastructure as a service (IaaS)/platform as a service (PaaS) provider. These organizations will look to add a second cloud provider for additional application use cases and/or to mitigate single-vendor risk concerns. This may entail the use of the secondary provider's primary public regions, obtaining a "distributed cloud" on-premises capability or potentially both.

However, each new platform tends to increase overall operational complexity and costs. While cost tracking and financial management has been reasonably well-addressed by third-party software vendors (see Gartner's CloudScores comparison of public cloud third-party cost optimization tools), overall operational complexity still remains a challenge.

Many organizations hope to find "one tool" that can seamlessly provide multicloud governance, monitoring, asset tracking and security, for example. In reality, there is no one tool in the market that will adequately address most organizations' needs in both breadth and depth. Broad tools tend to be very shallow in terms of their capabilities, whereas deep tools tend to have a narrow focus. (Gartner's Solution Criteria for Cloud Management Tools can help organizations assess third-party tools in a number of distinct functional areas.) For now, most organizations will need to accept that proper operations management may require a blend of provider-supplied management tools, third-party vendor tools and custom in-house developed management tools.

Once an IT organization has a well-managed multicloud environment, it will need to define a strategy for cloud workload placement. This means deciding which providers are preferred and which are strategic, and then developing a cloud workload placement framework that matches workload needs with the right-fit cloud provider.

In 2022, cloud technical professionals must:

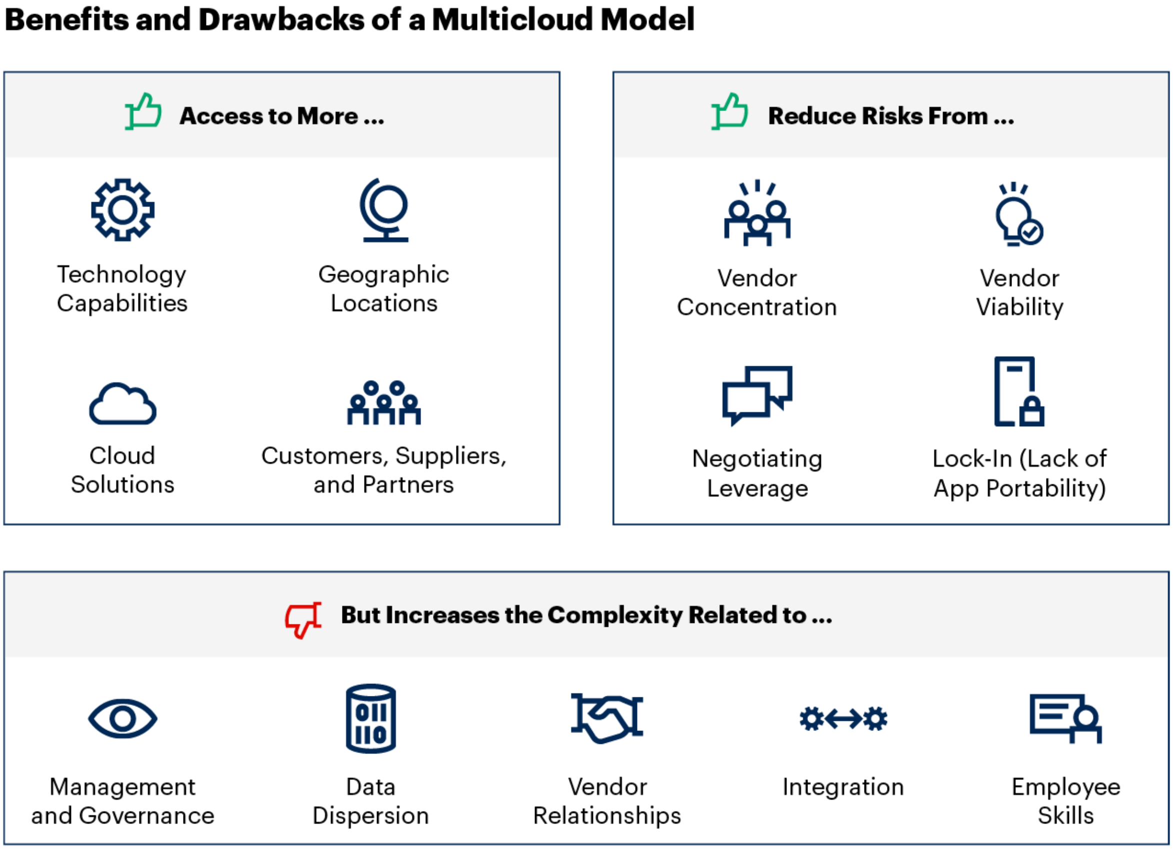
- Reduce cost and complexity by prioritizing a primary strategic provider.
- Exploit each provider's native capabilities to the fullest.
- Rearchitect networks to integrate across cloud providers.

Planning Considerations

Reduce Cost and Complexity by Prioritizing a Primary Strategic Provider

Multicloud strategies improve flexibility, but also increase complexity and cost. Although they deliver similar services, cloud providers have proprietary characteristics that require a clear understanding of their differences. This duplication of efforts is costly, and organizations must make conscious decisions to pursue a multicloud strategy. Furthermore, being accidentally multicloud — without a deliberate workload placement strategy — results in multicloud sprawl, chaos and further unnecessary costs. The benefits and drawbacks of a multicloud approach are summarized in Figure 5.

Figure 5: Multicloud Benefits and Drawbacks



Source: Gartner
733273_C

In 2022, cloud technical professionals should:

- **Prioritize a primary strategic provider:** Organizations should begin their cloud journey with just a single cloud IaaS+PaaS provider and invest in becoming highly competent with that provider. When the organization has business requirements that cannot be met using that provider, the organization can then add additional providers in an orderly fashion driven by specific business needs. Some organizations may also designate a secondary strategic cloud provider if they have disparate needs. For instance, there might be a desire to select one provider for new initiatives, and another for migrating existing applications.

- **Adopt a workload placement policy:** Creating an explicit cloud workload placement strategy will prevent further multicloud sprawl and drive more orderly adoption. Designate a primary strategic cloud provider — and possibly a secondary strategic cloud provider, if you have two significantly disparate needs — and invest in deep support for them. All other providers should be tactical and their use limited to requirements that cannot be met by a strategic provider. Create a cloud workload placement policy that provides guidance for where to place applications, based primarily on data and integration affinities. Seek a "good enough" technical fit rather than the best possible technical fit.

- **Evaluate key differences between approved cloud providers:** Invest in appreciating key differences between cloud providers' characteristics and implementation details. Assessing differences in capabilities (see Gartner CloudScores' scorecard for the Cloud Integrated IaaS and PaaS market) will help develop a proper capability-based placement policy. Knowing the differences in service implementations will help normalize the management and governance processes and maximize consistency between providers.

Benefits:

- **Well-governed sourcing of capabilities:** Relying on a strategic provider and only approving exceptions when that provider cannot offer a necessary capability gives the organization reduced complexity and cost. At the same time, it leaves open the flexibility to source additional capabilities from other providers when required by business needs.

- **Simpler implementation model:** A single-provider focus allows a simpler operational model, more consistent application architectures and greater ease of integration.

- **Focused skills:** Concentrating on a single provider (or very few) means staff has more expertise and experience with the provider (or providers), leading to improved implementation quality.

- **Higher commitment-based discounts:** The more an organization's spending is concentrated in a single provider, the larger the volume-based commitment discounts it will receive.

Related research:

- Comparing Cloud Workload Placement Strategies

- Designing a Cloud Workload Placement Policy Document
- Key Services Differences Between AWS, Azure and GCP: Governance and Policy Management

Exploit Each Provider's Native Capabilities to the Fullest

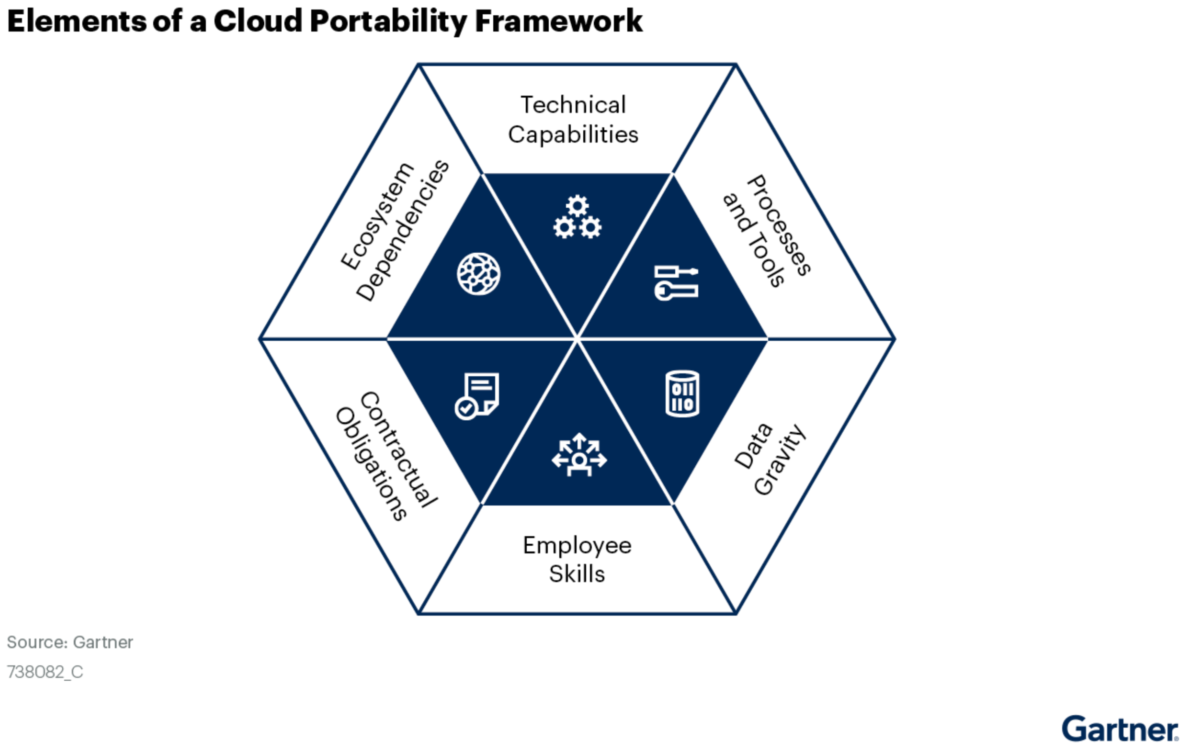
In 2022, regulators — and enterprise risk managers — will increasingly scrutinize cloud resilience, continuing a trend toward a heightened awareness of overall organizational resilience that has stemmed from the pandemic. This is causing cloud architects to consider anew the hard problem of cloud portability — whether between cloud providers or between the cloud and on-premises environments.

Cloud benefits are maximized when organizations thoroughly exploit the provider's native capabilities. It is these capabilities that typically deliver the differentiated value that attracts customers to the cloud. Cloud provider services are not commodities and will not become so. Even relatively fungible infrastructure elements, such as cloud networking, still differ widely among providers.

Furthermore, the problems of cloud provider lock-in cannot simply be abstracted away. IT organizations often seek to do this by introducing third-party tools that promise multicloud capabilities. Third-party tools do often provide additional value, but this is typically not through portable abstractions. Indeed, the third-party capability often creates a new point of dependency — the third party itself.

Cloud provider "lock-in" is created by dependence on technical capabilities, processes and tools, employee skills, and ecosystem dependencies (such as independent software vendor [ISV] licensing and support, managed service providers, and system integrators). Contractual obligations can prevent cost-effective shifts to another provider or repatriation on-premises. Data gravity makes it costly and complex to move. A cloud portability framework needs to address all these elements intelligently, as illustrated in Figure 6.

Figure 6: Elements of a Cloud Portability Framework



Source: Gartner
738082_C

Cloud portability should be treated as an application portability concern, rather than a problem with infrastructure design or vendor contracts. Long-lived strategic applications may need to outlive a given cloud provider or set of technologies.

When cloud portability is of concern, cloud architects should guide application teams to architect for contextual independence. They should identify cloud service dependencies, encapsulate dependencies in well-defined interfaces with as much transparency and as little overhead as possible, and reduce dependencies that do not drive business or technical benefits. Approach implementing abstraction layers with caution. Neither Kubernetes nor containers in general significantly improve most aspects of portability — contrary to vendor hype. Use of a PaaS framework can help abstract some infrastructure dependencies from application developer awareness, but this is not always a good fit for an application's architecture and technical requirements.

Benefits:

- **Maximizes ecosystem effects:** ISVs and other service providers in the ecosystem frequently deliver unique cloud-provider-specific benefits even when the vendor has multicloud support.

- **Maximizes third-party tools:** If you purchase a multicloud management tool, it is likely that it will not have the same depth of features for all supported providers. You would be denying yourself many of the benefits of such a tool if you didn't use provider-specific features. Most such tools do not successfully abstract away provider differences.
- **Minimizes cost and complexity:** Most organizations find full multicloud portability results in unsustainable cost, significant reductions in developer productivity and an unacceptable slowdown in digital business efforts.

Related research:

- A Guidance Framework for Managing Vendor Lock-In Risks in Cloud IaaS
- Assessing Kubernetes for Hybrid and Multicloud Application Portability

Rearchitect Networks When Integrating Across Multicloud Environments

When adopting clouds, most organizations build their cloud networks according to the designs (and the limitations) imposed by each cloud provider's offerings. This means each cloud provider's network is built and managed differently, and network expertise is needed in each cloud provider. To counter this, some organizations have lifted and shifted virtual versions of their data center network devices. However, lifting-and-shifting is suboptimal. It prevents organizations from reaping the major benefits of public cloud infrastructure, such as agility, elasticity, metered billing and self-service.

Instead, technical professionals should rearchitect multicloud networks to remove dependencies on individual cloud providers' tools and legacy mindsets driven by data center networks. The architectural options abound. What is most important is to establish a network architecture based on workload requirements, instead of the features available in a given cloud provider's offerings.

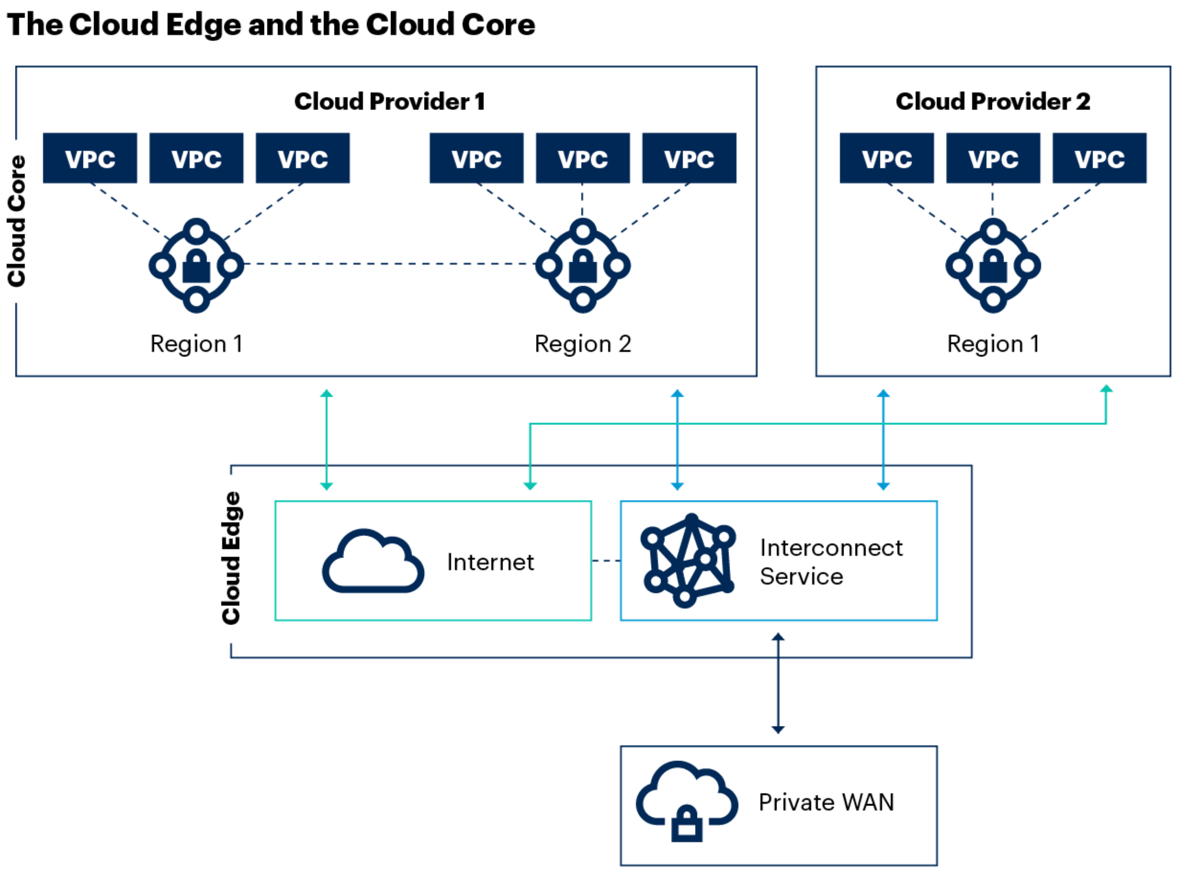
A proper multicloud network architecture has two overlapping but distinct layers:

- **The cloud edge:** This layer connects customers, partners and enterprise networks to cloud providers. This topic is generally well-understood. The constructs are widely deployed and well-documented. There is some innovation here, however, and the topic is still relevant.

- **The cloud core:** This layer hosts the customer's cloud workloads and provides network services. The cloud core space is less mature than the cloud edge's, and its mysteries are less well-known to end users.

Figure 7 illustrates the differences between cloud edge and cloud core. The next section focuses on the cloud core.

Figure 7: The Cloud Edge and the Cloud Core



Source: Gartner
753853_C

Cloud Core

In the cloud core, the virtual private cloud (VPC) is the basic building block. Each cloud provider offers its own style of VPC, which is slightly different from other providers' VPCs. There is no reason to reinvent the VPC, but, as organizations deploy multiple VPCs in each cloud region, the question of provisioning, managing and routing traffic across VPCs, and across clouds, quickly comes to the fore.

Key concerns for managing the cloud core include:

- Provisioning
- Routing across VPCs
- Visibility
- Day 2 operations and troubleshooting
- vNIC and subnet security controls
- Load balancing

Each of these domains can be treated independently. For example, an organization might choose to use the cloud provider's native load balancing and visibility offerings in each cloud. But the same organization might also choose a third-party classical routing solution such as Arista Networks, Cisco, Juniper or VMware; or a more cloud-aware solution such as Alkira or Aviatrix. Moreover, they might use Terraform for provisioning, but then implement each cloud provider's native security control management tools paired with a third-party cloud security posture management tool.

In 2022, technical professionals should take a cloud-provider-agnostic approach to architecting the cloud core. Evaluate only the cloud provider's native tools and third-party tools that were purpose-built for the cloud. Let the demands of each application or workload dictate which tool you choose. Use third-party offerings to provide advanced features or consistency across clouds, only when the provider's native tools cannot do this. But make the decision on a domain-to-domain basis.

Benefits:

- **Time to value:** A best-of-breed approach delivers the fastest time to value. In some cases, this means choosing the default cloud provider's offering because of the ease of integration; in others, this means choosing a third-party platform that obviates the need to learn and navigate multiple independent provider network stacks.
- **Consistency:** Third-party tools offer better consistency across clouds, but incur additional cost and (potentially) tool lock-in.
- **Troubleshooting:** Third-party tools can often help with visibility and operational optimizations in multiple cloud providers. This can decrease mean time to repair.

Related research:

- Solution Path for Evolving to Next-GenerationEnterprise Networks
- How to Architect Network Connectivity Across Multiple IaaS Cloud Providers and Regions
- How to Architect Your Network to Optimize Internet Performance and Reliability

Business Resilience Will Be Built Into the Application Architecture

Disaster recovery (DR) has always been centrally important in technology design. Historically, reliability has been the responsibility of operators, who ensured it by building DR capabilities into the infrastructure. But, increasingly, reliability is a feature of the application code itself.

In 2022, cloud technical professionals should build resilience at the application layer, not the infrastructure layer.

The focus of IT resilience must now shift to resilient application architectures, rather than infrastructure failure domains. In today's BC/DR plans, redundant API calls are more important than storage replication; and distributed build pipelines are more important than site failovers. For this continuous application architecture, public cloud infrastructure is the preferred execution venue. In modern public clouds, managed services are already highly resilient and scalable. They can replace the complexity and expense of do-it-yourself DR.

However, this method is not easily mastered. Application resilience places a premium on your organization's ability to execute automated implementation and release. While the level of effort is high, the overall added capabilities will have the positive effect of allowing your highly resilient applications to better adapt to the unknown threats faced by critical tier applications. Forward-looking organizations have already begun the shift to continuous application architectures, and, in the coming year, more should follow their lead.

The spotlight on IT resilience is only growing brighter, thanks to the increasing frequency of cybersecurity threats, especially ransomware; the continuing organizational disruptions of the COVID-19 pandemic; and other natural disasters and catastrophes. The time has come to move forward with modern IT resilience.

In 2022, cloud technical professionals should:

- Focus on IT resilience rather than individual service continuity.
- Build application-level resilience for containers and Kubernetes.
- Redesign isolated recovery environments (IREs) for defense-in-depth against new ransomware threats.

Planning Considerations

Focus on IT Resilience Rather Than Individual Service Continuity

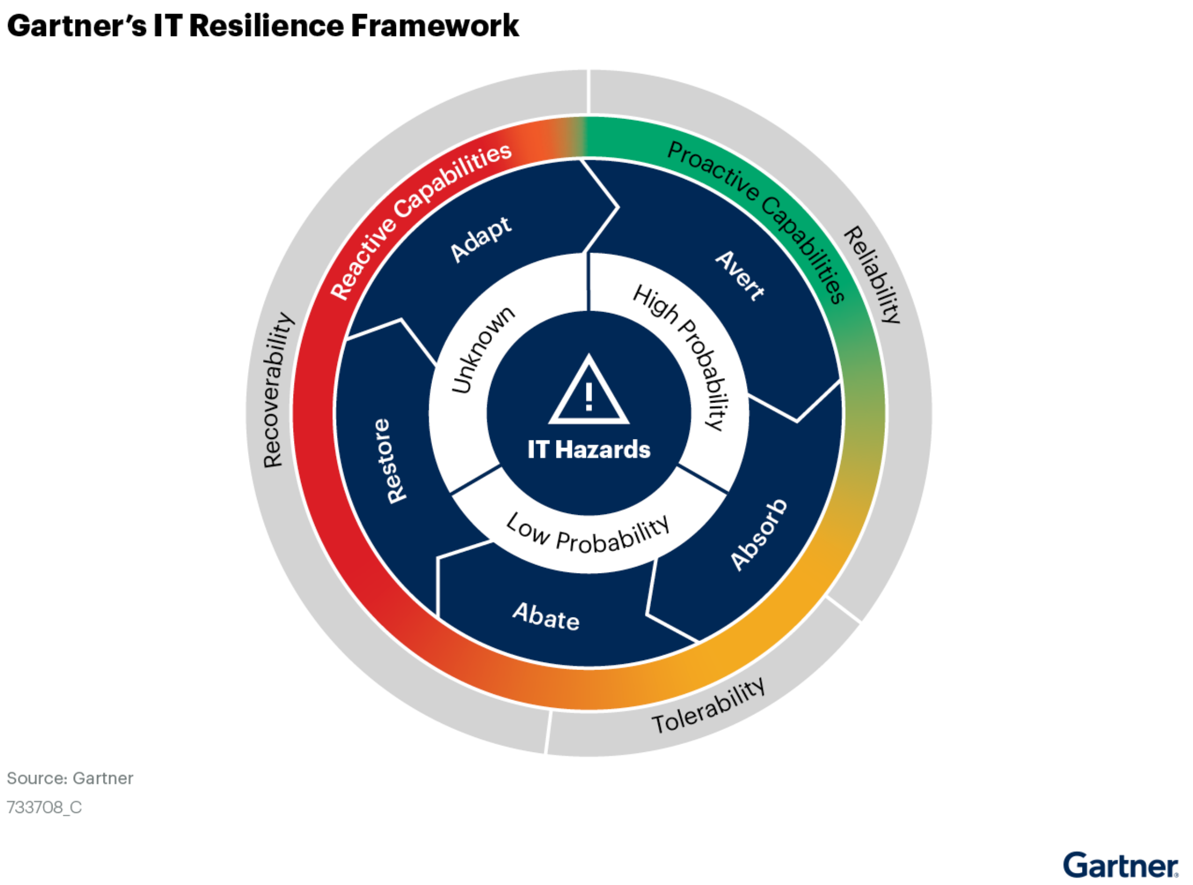
Technical professionals may focus on a single service or domain, but the IT organization exists to provide positive business outcomes. Individual systems and services are integral to ensuring the success of a line of business or the performance of a particular application, but they are not the totality of a business need. In the realm of IT resilience, it is not sufficient to design and plan for recovery or to merely protect data. Recovery and data protection are not outcomes; your goal should be true IT resilience.

Resilience is a business differentiator. If your competitors suffer through delays and downtimes, while your business carries on, then your IT services have created an opportunity to showcase the superiority of your product. Successfully navigating natural and human-made disasters with efficiency and consistency highlights your company's operational excellence. Your outcome is improving the broader business.

Definition of IT Resilience

IT resilience is the organization's ability to avert, absorb, abate, restore and adapt to known and unknown IT-related hazards such as application defects, performance thresholds, security vulnerabilities, single point of failures and service provider outages and to continuously improve end-to-end posture so systems are reliable, tolerable and recoverable (see Figure 8).

Figure 8: IT Resilience: Reliable, Tolerable and Recoverable



Building a culture that properly focuses on IT resilience takes deliberate effort. A significant barrier to successful adoption is the separation of operational, DR and product teams. Emerging roles and teams dedicated to resilience or site reliability engineering (SRE) practices can foster collaboration and eliminate the organizational silos. While a full-fledged SRE practice is beyond the capabilities of many IT organizations today, the SRE mindset can still assist tactically across functional domains.

As with all programs, use metrics to track the success of the initiative. However, operational metrics are not necessarily helpful in determining its overall effectiveness. Instead, measure success by business value, risk posture and cost. Tracking the time between significant events, such as mean time between failures (MTBF) or between threat identification and risk assessment, ties back to risk posture directly. Meanwhile, a business impact analysis (BIA) can measure business value and overall cost.

In 2022, cloud technical professionals should seek quick, iterative wins for the business. For example, you might use IT resilience metric data to identify key gaps, such as out-of-date BIAs or known but unaddressed IT hazards. As the program matures, leverage momentum to expand into other areas identified as priorities during risk assessment events. By the end of 2025, 30% of enterprises will establish new roles focused on IT resilience and boost end-to-end reliability, tolerability and recoverability by at least 45%.

Benefits:

- **Better business outcomes:** A focus on IT resilience means direct improvements to customer experience via gains in risk assessment and remediation and can lead to improvements in the bottom line. These techniques align business drivers and relevance to IT services.
- **Improved technical operations:** This also represents better prioritization of IT effort spent toward improving resilience and recoverability. Communication across operational, DR and product teams improves, as does proactive identification of IT hazards.
- **Risk mitigation:** This makes risk management a business decision instead of just an IT decision. These techniques also promote better monitoring and control over assets and services.

Related research:

- IT Resilience — 7 Tips for Improving Reliability, Tolerability and Disaster Recovery
- Assessing Site Reliability Engineering (SRE) Principles for Building a Reliability-Focused Culture
- How to Design a Data Protection Strategy for On-Premises and Cloud IaaS

Build Application-Level Resilience for Containers and Kubernetes

With cloud-native adoption continuing to explode, the Kubernetes platform is now the foundation for containerized applications both on-premises and in the public cloud. While Kubernetes takes away a lot of the pain of ensuring high availability and scalability of application services, these benefits do not extend to stateful data, making data management of Kubernetes applications a critical priority.

While it is possible to protect stateful data in Kubernetes using traditional infrastructure-driven techniques, merely backing up and replicating stateful data won't protect it against data corruption, accidental deletions, infrastructure failure and, of course, ransomware. Instead, IT organizations need a Kubernetes-centric data protection strategy. To be clear, this is also required in the public cloud, despite the common misconception that cloud services will protect data and provide resilience in the event of a regional outage. This is simply not the case.

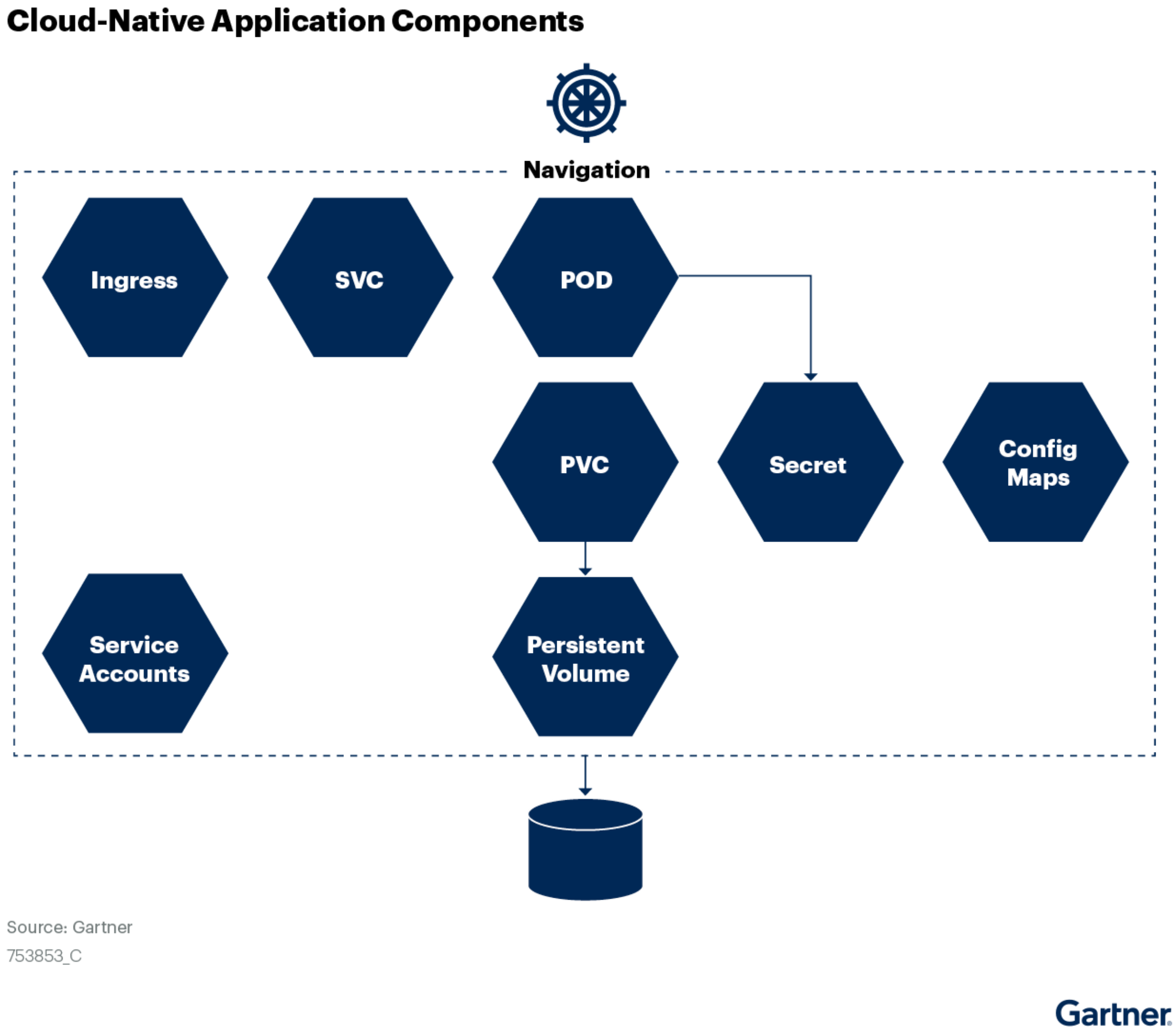
> **Protecting stateful data in Kubernetes means capturing an entire application and its dependencies in such a way that an orchestrated restore in an alternate location or platform can be achieved.**

Containerized applications run on vastly different architecture in a Kubernetes cluster than (traditional) applications running on virtual machines. Traditional data protection that targets specific virtual machines (VMs), servers or disks will not work, since Kubernetes uses its own placement policies to distribute components across all servers with different applications often residing on the same server. Containers are often rescheduled dynamically for better load balancing and performance and new applications brought on line via continuous integration (CI)/continuous deployment (CD) at any time. The makeup of a cloud-native application is constantly shifting and requires an approach to data protection that will capture applications as a whole including data on persistent volumes.

Now in the container world, there are hundreds of objects — such as secrets, configurations and disks — which only a backup solution for containerized applications will understand. Indeed a Kubernetes-native backup solution will need to handle mass amounts of components in large K8s clusters and understand the relationships that exist between applications, stateful data and related Kubernetes configuration items.

Figure 9 illustrates some of the components in a trimmed-down version of a stateful application. In the real world of microservices, applications are broken up into hundreds of components.

Figure 9: Cloud-Native Application Components



Source: Gartner
753853_C

The adoption of dynamic provisioning via CI/CD pipelines adds another challenge to the backup strategy. But, with the right tools in place, DevOps and infrastructure teams can combine forces to implement a backup system that integrates with such CI/CD pipelines with the ability to discover new and changed applications and instantly provide protection. Kubernetes deployments with continuous deployment workflows spanning multiple clusters, and with more complex architectures consisting of multiple data services, will only become more widespread.

In 2022, cloud technical professionals need to formulate a Kubernetes backup blueprint. Despite what they may believe, most IT organizations do not have such a blueprint today. You can start by identifying gaps in protection. You must define the recovery point objective (RPO) and recovery time objective (RTO) of each Kubernetes cluster running across all platforms. Root out any overly customized and/or ad hoc backup systems that may be in place now. Then, catalog the resources and persistent data that comprise each application. Armed with this information, you can begin to design a backup strategy capable of keeping pace with rapidly evolving Kubernetes deployments.

Benefits:

- **Better cross-domain integration:** Having a Kubernetes-native backup strategy will integrate development and infrastructure teams, and incorporate data protection into these CD workflows, protecting configuration changes and data.
- **Ability to restore workloads to multiple platforms:** A true Kubernetes-native strategy will allow orchestrated recovery and interoperability across all platforms, a critical capability for hybrid and multicloud implementations.
- **Protection from new and growing security threats:** With Kubernetes clusters now available on a wide variety of hosted platforms, it is easier to deploy containerized applications. These can quickly become mission-critical applications that are often left exposed due to bad practices. Having solid backups is a must to recover from all failure scenarios, including supply chain and security threats.

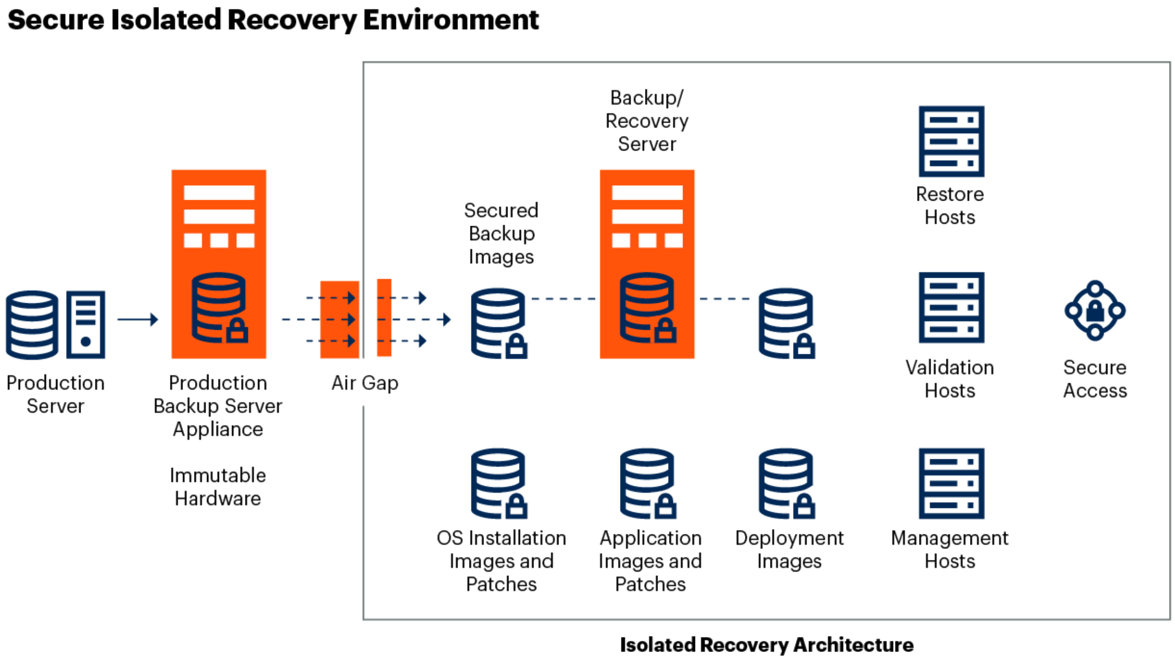
Redesign IREs to Defend Against New Ransomware Threats

Ransomware attacks continue to rise globally, and the threat actors behind them have become increasingly sophisticated in their methods. The criminal actors behind ransomware attacks tend to be highly organized, operating with a high degree of technical sophistication and business acumen. Most organizations’ backup services are not designed or configured for recovery from ransomware. Perpetrators will often target the backup servers, or backup images/repositories themselves, to prevent restores from backups. The last line of defense is to implement an IRE.

An IRE is a dedicated, isolated secure recovery environment with tools to verify and recover data from an immutable backup copy. It does not replace traditional backup and DR systems. Rather, an IRE constitutes an additional, protected environment independent of production. Since hackers might still have access to the production environment, the IRE functions as a safe place to conduct restore activities. An immutable data vault (IDV) is an immutable, air-gapped, tertiary independent copy of the backup data in a secure environment. IREs and IDVs safeguard backups against ransomware and insider attacks.

Figure 10 provides an overview of an IRE providing a secure environment where recovery can take place and the possibility of malware reinfiltrating production during recovery operations can be eliminated. Figure 10 also highlights other key items that may be required in the event of system rebuilds, such as OS installation images, and patches and application binaries.

Figure 10: Secure Isolated Recovery Environment


Secure Isolated Recovery Environment
Production Server
Production Backup Server Appliance
Immutable Hardware
Air Gap
Secured Backup Images
Backup/ Recovery Server
Restore Hosts
Validation Hosts
Secure Access
OS Installation Images and Patches
Application Images and Patches
Deployment Images
Management Hosts
Isolated Recovery Architecture

Source: Gartner
724116_C

Existing BC/DR orchestration technologies can provision an isolated environment that may be suitable for some recovery scenarios, but not all. An IRE is the most secure recovery vehicle — a separate environment with dedicated systems and no network access to production. All but the very largest organizations can afford to build an IRE environment. Many choose simply to reserve space and designate infrastructure to be used in the event of an attack. Others may select a backup or archive appliance as an IRE to help reduce costs. An IRE does not need to be built prior to its use, but its construction should be integrated in your well-tested cyberattack recovery plans and playbooks.

In 2022, cloud technical professionals should add IREs to their cyber recovery plans. At a minimum, ensure that all the pieces are in place so that a secured IRE can be quickly assembled when needed. Remember that IREs are measured as much by the amount of data they were able to recover as by their positive impact on RTO and RPO.

Benefits:

- **Immutability:** IREs and IDVs protect both production and backup systems from being attacked, modified or deleted by both outsider and insider attacks by leveraging an additional secure immutable copy of the backup data.

- **Forensics:** IREs can be used to store compromised data, which is useful for internal troubleshooting and root cause analysis, but also for law enforcement investigations, if any.

- **Prevents reinfecting systems:** IREs can help prevent malware from being reintroduced into production during recovery operations. They provide network and storage isolation with dedicated network and network services, such as Active Directory, DNS, DHCP and NTP.

Related research:

- Designing and Implementing a Ransomware Defense Architecture

- How to Recover From a Ransomware Attack Using Modern Backup Infrastructure

- Innovation Insight for Leveraging Isolated Recovery Environments and Immutable Data Vaults to Protect and Recover From Ransomware

- Survey Analysis: IT Disaster Recovery Trends and Benchmarks

Distributed Cloud Will Displace Private and Hybrid Cloud Initiatives

A large proportion of Gartner clients describe their environments as “hybrid cloud” or “private cloud,” with varying definitions and interpretations. IT organizations build private and/or hybrid clouds for many reasons: regulatory requirements, data gravity, the momentum of legacy infrastructure, limitations of staff skill sets, real-life deadlines and priorities, and more. However, Gartner clients almost unanimously cite complexity and manageability as among the highest, if not the highest, barrier to successful implementation. Clients expect, or hope, hybrid and private clouds match the ease of use, service consumption, and integration opportunities of public cloud. But do-it-yourself cloud solutions rarely meet these expectations.

Distributed cloud is an answer to this mismatch. Clients want cloud services running in their data centers, and that is precisely what solutions such as AWS Outposts, Google Anthos and Microsoft Azure Stack aim to provide. These products deliver the same native public cloud services to smaller and more local infrastructure, while remaining under the same management schema. These cloud services will be distributed to locations able to meet hybrid and private cloud needs while they retain the advantages of classic public cloud consumption.

Definition of Distributed Cloud

Distributed cloud is the distribution of public cloud services to different physical locations while ownership, operation, governance, updates and evolution of the services remain the responsibility of the originating public cloud provider. Such public cloud services often include necessary hardware and software.

Distributed cloud is an emerging market, and solutions are evolving at a rapid pace. To date, adoption is limited to niche applications with ideal benefits and requirements. In 2022, these solutions will improve and broaden their capabilities to find new use cases and verticals. As an answer to the chief difficulties of hybrid and private initiatives, distributed cloud will accelerate and emerge as a standard mode of operation for many organizations.

In 2022, cloud technical professionals should:

- Enable hybrid architectures with distributed cloud.
- Rethink connectivity between locations and devices.

- Transform the data center into an edge location.

Planning Considerations

Enable Hybrid Architectures With Distributed Cloud

Today, the typical IT organization has some infrastructure in the public cloud, some in data centers and some at the edge. This brings hybrid cloud architectures to the fore. Adoption of public cloud services is nearly universal, so every data center is now a hybrid data center. Connectivity to and portability between on-premises infrastructure and public cloud is the most important design consideration.

Many IT organizations have attempted to build their own private and hybrid clouds, usually with OpenStack. Most never found real success, and many technical professionals became deeply discouraged. But a new generation of turnkey hybrid cloud infrastructure has emerged. Today, IT organizations can simply buy a hybrid cloud, rather than building it themselves. There are two main approaches:

1. **Cloud-inspired** approaches are typically driven by traditional data center vendors, and based on hyperconverged infrastructure. They aim to provide a full stack of software-defined infrastructure — virtual compute, storage, networking and cloud management — on-premises, with common management connectivity to public cloud services. These include VMware Cloud Foundation, Nutanix Enterprise Cloud, and Microsoft Azure Stack HCI.

2. **Cloud-enhanced** approaches are typically driven by the cloud providers themselves, with the goal of making native services that you normally use in a public cloud provider region available somewhere else. Many are expressly PaaS solutions, providing the hardware and software needed to run public cloud services on-premises. These include AWS Outposts and Microsoft Azure Stack Hub.

In either case, the goal is for the product to be a unified cloud control plane for both public and private infrastructure, while making native public cloud services available in data centers and at the edge. Hyperscale public cloud vendors and traditional data center infrastructure vendors are now competing directly with one another. The major public cloud vendors are pushing into the data center, while the leading HCI vendors are expanding their products into the public cloud. Table 1 summarizes the offerings available on the market today. Both the hyperconverged infrastructure vendors and the hyperscale public cloud vendors offer both styles.

Table 1: A Summary of the Offerings on the Market Today

Vendor	Cloud-Inspired	Cloud-Enhanced
Amazon Web Services (AWS)	VMware-on-Outposts	AWS Outposts
Microsoft	Azure Stack HCI	Azure Stack Hub
Nutanix	Enterprise Cloud	Clusters
VMware	Cloud Foundation	VMware Cloud (Various)

Source: Gartner (October 2021)

Together, these architectural styles represent a new type of IT infrastructure. They are on-premises infrastructure delivered as a cloud service or cloud services managed as on-premises infrastructure, depending on which vendor's sales pitch you're listening to. Distributed cloud represents a fusion of traditional and cloud infrastructure, services, platforms, and operating models. In 2022, it is imperative that cloud technical professionals seriously explore the new future.

Benefits:

- **Shifts labor:** With distributed cloud, you don't need to spend time and effort building an infrastructure; you'll simply acquire it from a vendor. With the time they save, technical professionals can shift their efforts to more valuable pursuits.
- **Unified management:** Any distributed cloud approach will provide a common operating model for both on-premises and cloud infrastructure, breaking down a silo that still persists in many IT organizations today.
- **Complexity and cost reduction:** Both distributed cloud styles also offer tantalizing opportunities to remove complexity from the infrastructure stack and to reduce the costs of both software and skilled personnel.

Related research:

- Solution Criteria for Hyperconverged InfrastructureSoftware

- A Guide to Distributed Cloud: The Next Frontier of Cloud Computing
- Four Types of Cloud Computing Define a Spectrum of Cloud Value

Rethink Connectivity Between Locations and Devices

The emerging distributed cloud and edge infrastructure paradigms require a distributed network that optimizes for latency and bandwidth consumption while being reliable and secure. These are long-standing, bedrock principles of good network design. But distributed edge models make them more important than ever because:

- **Data sources are distributed.** Applications, users and end devices are all sources of data, which can live anywhere. A distributed cloud prioritizes the effects of proximity to improve services to these data sources.
- **Edge is defined by location of data sources.** Prioritizing location requires edge compute to be flexibly applied anywhere on the network to process the data. This is a major change from the classical understanding of the network edge as the point where one network domain transitions into another.

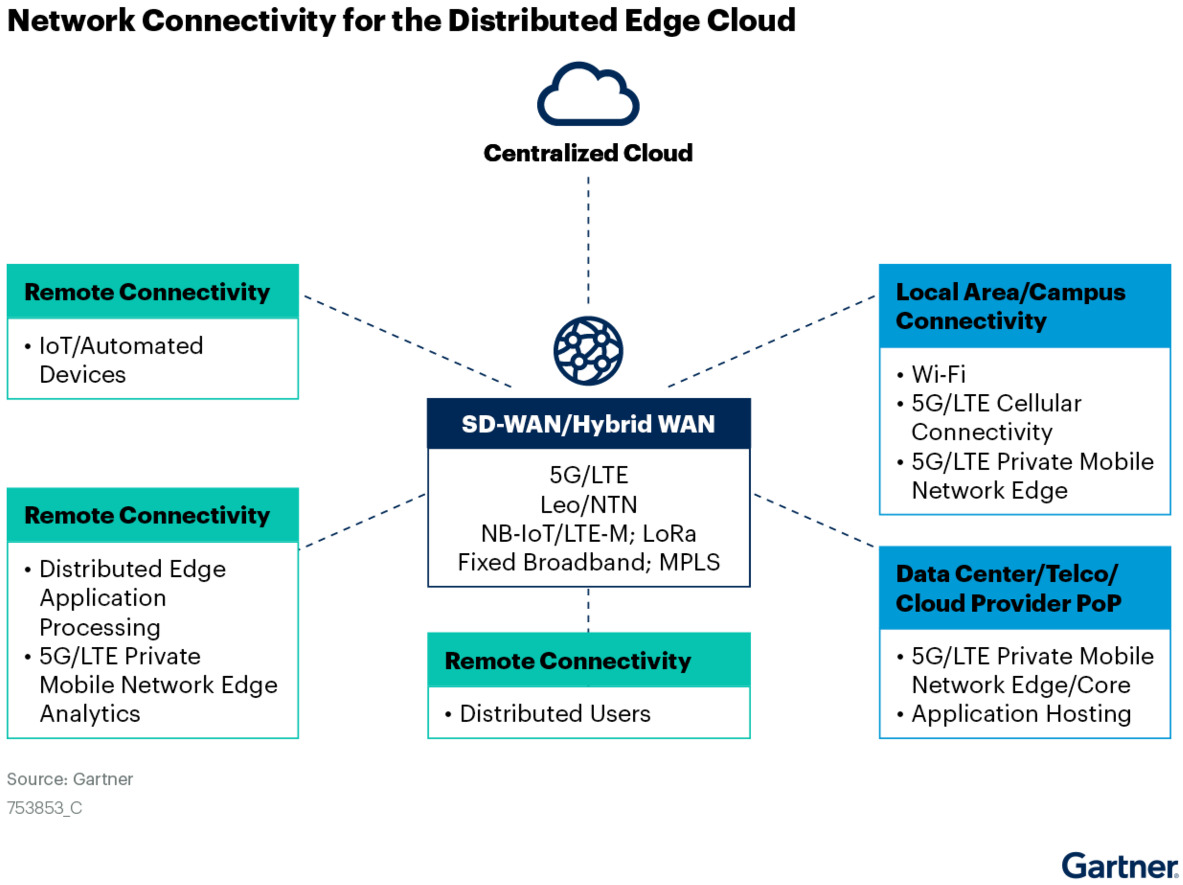
Distributed cloud and edge models push the limits of classical approaches to network architecture. In 2022, therefore, technical professionals must reorient their network architectures according to these principles:

- **Connectivity follows data sources:** Distribute the network resources to the edge, alongside the servers, so that data can be processed as close to the source as possible.
- **Connectivity is an array of multiple access technologies:** For example, Wi-Fi is an optimal option for corporate environments, and Wi-Fi 6 promises higher bandwidth. 5G/LTE can be used in the WAN or for in-building low-latency local connectivity. For industrial or smart city environments consisting of applications such as computer vision, LTE or 5G connectivity used in conjunction with edge compute is more beneficial. Similarly, Internet of Things (IoT) devices at remote sites may be connected over the WAN using mobile technologies, such as LTE-M, NB-IoT or LoRa, or even fixed broadband. Low earth orbit (LEO) satellite links and nonterrestrial networks are emerging as potential options for data sources at locations with little to no connectivity.

- **SASE is required:** A secure access service edge (SASE) framework preserves the identity and context of the connection, thus converging network and security. The security perimeter is distributed, while control and management are centralized. This approach blends well with SD-WAN, which can extend network access based on zero trust across locally and regionally distributed or centralized infrastructure. The path to SASE begins with extending existing security capabilities, then building new security measures based on the integration of policy orchestration between the network and security components.
- **You are an internal service provider:** Consider your network a connectivity platform that can be customized to support many different usage scenarios. You can deploy and operate networks in-house, or source them by using network as a service (NaaS) offerings or a managed network service (MNS).

Figure 11 illustrates modern network connectivity for the distributed cloud and edge.

Figure 11: Network Connectivity for the Distributed Edge Cloud



Network Connectivity for the Distributed Edge Cloud

Centralized Cloud

SD-WAN/Hybrid WAN

5G/LTE
Leo/NTN
NB-IoT/LTE-M; LoRa
Fixed Broadband; MPLS

Remote Connectivity

- IoT/Automated Devices

Remote Connectivity

- Distributed Edge Application Processing
- 5G/LTE Private Mobile Network Edge Analytics

Remote Connectivity

- Distributed Users

Local Area/Campus Connectivity

- Wi-Fi
- 5G/LTE Cellular Connectivity
- 5G/LTE Private Mobile Network Edge

Data Center/Telco/ Cloud Provider PoP

- 5G/LTE Private Mobile Network Edge/Core
- Application Hosting

Source: Gartner
753853_C

Benefits:

- **Flexibility:** A fundamental objective of a modernized network architecture is to emphasize users, applications and devices at any location, unrestricted by traditional topologies, boundaries and connectivity options.

- **Zero trust:** Modern network architectures secure distributed infrastructure through zero-trust models.

- **Business-driven agility:** The service provider mindset emphasizes enabling business innovation, thinking in ecosystems rather than vendors, especially for 5G/LTE or nonterrestrial network technologies.

- **User experience:** Modern networks are highly automated and elastic, scaling out, up or down, as needed. This improves application performance, ensures reliability and supports business innovation.

Related research:

- Architecting a Reference Framework for 5G PrivateMobile Networks

- How to Architect In-Building Connectivity With Cellular Technologies

- Assessing 5G Mobile Technology for Organizations

Transform Your Data Center Into an Edge Location

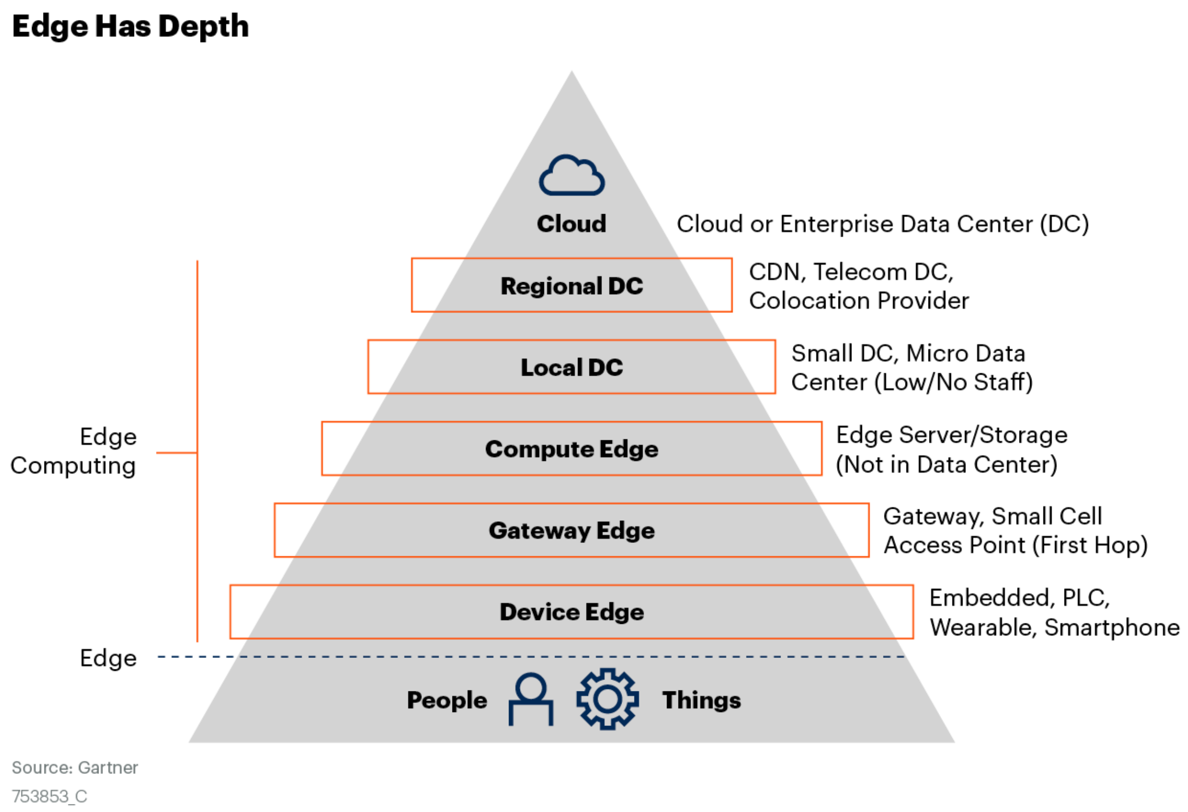
Gartner clients increasingly see the corporate data center as unnecessary, housing only technical debt. They often ask whether it is time to shut down data centers and “get out of the data center business.” Some will indeed do so, but, for many IT organizations, on-premises infrastructure is advantageous or required. These organizations will still seek to reduce their investment in data centers by consolidating, improving density and modernizing. Colocation facilities are an attractive alternative to owning data centers. For all these reasons and more, data centers are transforming into niche use cases, often relegated to supporting only non-x86-based workloads and legacy applications that can’t cost-effectively be ported.

Increasingly, on-premises infrastructure is found not in dedicated, expensive data centers, but in edge locations. Indeed, your data is probably already distributed and will only become more so. In 2022, more than 50% of enterprise-generated data will be created and processed outside the traditional data center or cloud. [3] In other words, more than half of the data your organization generates will originate at edge locations. This means the data center must transform into an edge location — a peer of both cloud and remote-office infrastructure, rather than the central warehouse of all applications and data.

In this emerging world of generalized distributed computing, the key criterion is location. The variety of locations where your services are delivered must be evaluated in the context of how services are delivered to customers. You must select, orchestrate and manage these locations to optimize customer experience, which includes network consumption, latency sensitivity, privacy considerations and resiliency of services.

Figure 12 illustrates the place of the data center in the hierarchy of the edge. Remember that the edge is defined by a relation to a core infrastructure, but that core infrastructure does not have to be identical for every edge use case or architecture. The data center itself may be the core of one use case but the edge of another (see Figure 12). Edge computing has depth — it can take place embedded in equipment, in gateways that connect people and things to the internet, in aggregated processing located on-premises or nearby, or in a local data center. Most edge computing solutions will leverage many of these layers, in addition to the back-end enterprise or cloud data center.

Figure 12: Edge Has Depth


Edge Has Depth
Cloud
Cloud or Enterprise Data Center (DC)
Regional DC
CDN, Telecom DC, Colocation Provider
Local DC
Small DC, Micro Data Center (Low/No Staff)
Compute Edge
Edge Server/Storage (Not in Data Center)
Gateway Edge
Gateway, Small Cell Access Point (First Hop)
Device Edge
Embedded, PLC, Wearable, Smartphone
Edge Computing
Edge
People
Things
Source: Gartner
753853_C
Gartner

Not everyone is ready to embark on integrating a distributed cloud model into their enterprise architecture, but new paths to edge computing have emerged. These range from “build your own stack” to buying a stack from a system integrator, to business-outcome-based “edge computing as a service” solutions. These second two paths reduce complexity, risk and time to deployment. In the long run, data is the most important consideration. The combination of edge data aggregation and analytics, and cloud access for scale-up capabilities such as analytics and model training, creates new demands for edge infrastructure.

Benefits:

- **Cost savings:** Making a strong business case for traditional data centers will become more difficult over time as more applications migrate out of the data center and the cost of supporting legacy applications increases. But delivering fast, consistent service to customers does not require a data center.

- **Supports emerging technologies:** Edge computing augments and expands the possibilities of today's primarily centralized, hyperscale cloud model, supports the systemic evolution and deployment of the IoT, and supports entirely new application types, enabling next-generation digital business applications.

- **Improved latency:** This is often the main reason for deploying workloads to the edge, but moving processing and storage closer to users and "things" that are the source of data generation can also address concerns such as bandwidth, data privacy and autonomy.

- **Improved security:** A wider attack surface is an advantage in defending against distributed denial of service (DDoS), ransomware and cyberattacks. Attacks can be segmented at the edge, while the data traffic back to the core is minimized.

Related research:

- 4 Steps to Successful Edge Computing Deployments
- Deploy Leaner AI at the Edge: Comparing Three Architecture Patterns to Enable Edge AI
- 2021 Strategic Roadmap for Edge Computing
- Infographic: Understanding Edge Computing
- The Technical Professional's Guide to Edge Infrastructure

Containers and Serverless Will Become an Infrastructure Foundation for Application Platforms

In the beginning, public cloud IaaS was delivered exclusively via virtual machines. But, today, new virtualization methods are taking hold, including containers and serverless computing. Gartner noted this trend in last year's 2021 Planning Guide for Cloud and Edge Computing, and it has only accelerated since. As cloud computing principles become more embedded in application development and infrastructure operations, containers and serverless will become increasingly attractive deployment vehicles for code.

First, Kubernetes-based container as a service (CaaS) platforms promise to provide standard abstractions and a common control plane for managing containerized applications. This is critical for hybrid and multicloud use cases. They can still impose some points of lock-in with a particular cloud service. But, when appropriate design principles are applied, the use of containers can make applications portable and operations consistent between on-premises infrastructure and different cloud services (see Assessing Kubernetes for Hybrid and Multicloud Application Portability).

Second, serverless computing approaches based on function platform as a service (fPaaS) have grown into a powerful tool for developing applications with event-driven architectures. These are best deployed for use cases such as service integration, cloud operations and IoT data processing. Moreover, the leading public cloud fPaaS offerings continue to refine their ability to orchestrate containers with "serverless" approaches. This allows end users to offload the responsibility for deploying and managing the infrastructure to run containers to the cloud service provider.

Containers and serverless computing permit resource consumption to be tailored more precisely to the actual requirements of applications than did earlier virtualization methods. This improves agility, automation, efficiency and cost optimization for infrastructure. Container management platforms based on Kubernetes help to implement DevOps because the Kubernetes orchestrator provides a standard framework for I&O and application development to collaborate on deploying and operating containerized applications at scale. The self-service capabilities in Kubernetes platforms also give product teams the ability to deploy services without involving I&O. The use of serverless computing eliminates operating overhead because operators do not have to preprovision cloud resources or configure autoscaling for applications. Serverless computing also helps to optimize the cost of cloud infrastructure because users are only billed for the resources consumed at the time of application execution.

In 2022, cloud technical professionals should:

- Optimize applications for cloud-native architecture with containers and serverless.
- Use container self-service platforms for DevOps and other use cases.
- Leverage provider-native container services to optimize cost and capacity.

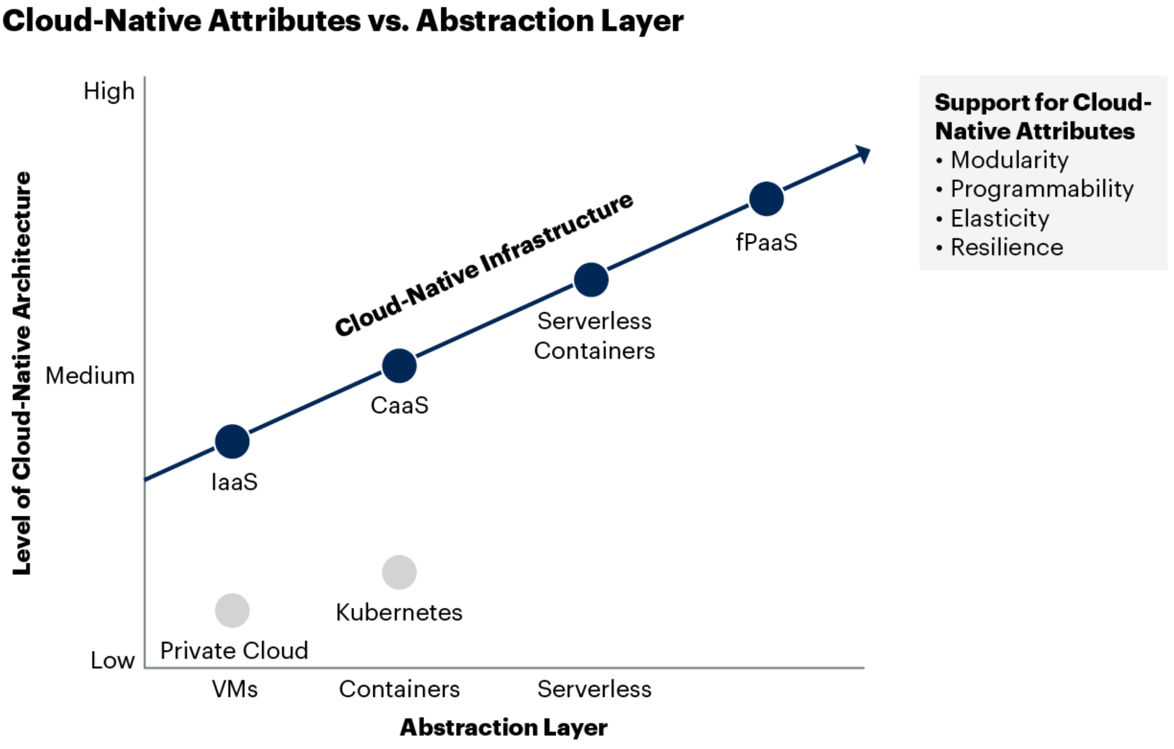
Planning Considerations

Optimize Applications for Cloud-Native Architecture With Containers and Serverless

Although the term “serverless” has been most closely associated with fPaaS platforms such as AWS Lambda and Azure Functions, serverless computing can be applied at multiple layers of abstraction, including native Kubernetes services in the cloud with a serverless data plane to run containers. Serverless platforms automatically perform all necessary provisioning to execute the code required to perform a task. When consumed in the public cloud, serverless container platforms enable truly on-demand consumption because there are no idle resources or orphaned VMs or containers. Applications and services are billed based on the number of container executions, so users are charged only for the capacity they use. As a result, serverless computing has low entry costs and delivers very high efficiency.

Figure 13 illustrates the possible abstraction models for compute resources in cloud-native applications.

Figure 13: Cloud-Native Attributes Versus Abstraction Layer



Source: Gartner
728982_C

Applications that are designed with cloud-native architecture aren't optimized to run in traditional IaaS based on VMs. They require a higher degree of service discovery, programmability, automation, observability, robust network communications and security. Containers and serverless platforms are based on runtime mechanisms that work at higher levels of abstraction than VMs. At these levels of abstraction, it becomes possible to match applications and services to resources at finer granularity than possible with VM-based approaches.

Benefits:

- **Cloud-native:** Organizations pursue solutions with cloud-native architecture when they have strategic goals of increasing software velocity, enabling developer agility, maximizing application scalability, implementing application-centric resilience and reducing technical debt.
- **Speed of delivery:** Containers provide consistent application packaging and streamlined configuration management. Most container images are based on open-source software, but the use of containers by independent software vendors (ISVs) to deliver commercial off-the-shelf (COTS) applications is also increasing.
- **Modernization:** While containers were initially used mainly for development of new applications based on cloud-native architecture, they are now increasingly being considered as a method to help modernize traditional monolithic applications.
- **Optimized cost and scalability:** The improved granularity enables orchestration to be carried out more precisely, and it enables tighter integration between the processes for developing applications and processes for iterative deployment.

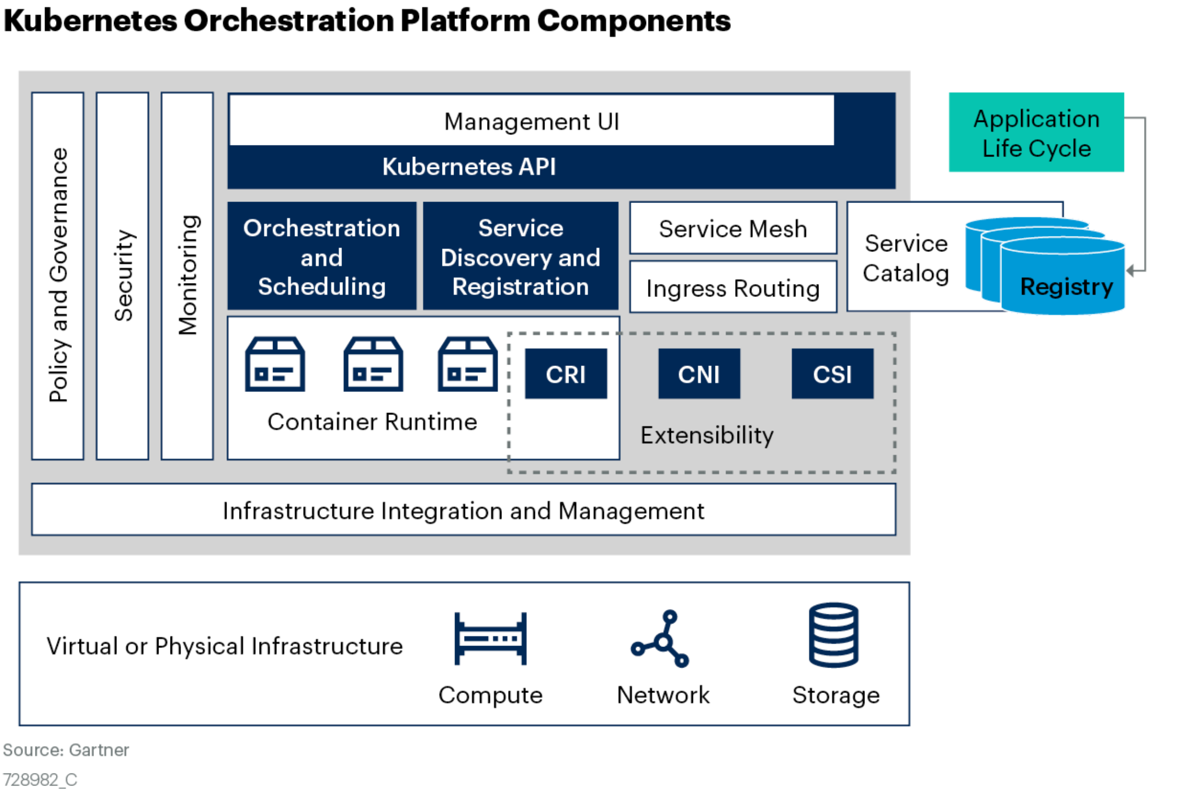
Related research:

- How to Modernize Your Application to Adopt Cloud-Native Architecture
- Decision Point for Selecting Virtualized Compute: VMs, Containers or Serverless

Use Container Self-Service Platforms for DevOps and Other Use Cases

Kubernetes has become the industry's leading container orchestration platform, but the Kubernetes open-source project itself provides only a subset of the functionality required for a complete container platform. Kubernetes includes the basic functions needed to orchestrate containers, but it leaves room for cloud service providers to add value with add-on tooling and platform innovation. Figure 14 shows the various components of a Kubernetes-based container orchestration platform, with the dark blue components representing the core functions of Kubernetes.

Figure 14: Kubernetes Components in a Container Orchestration Platform



The easiest way to deploy a container orchestration platform in production is to use a managed CaaS platform in the public cloud, which obviates the need for I&O teams to build and support their own container orchestration platform. In 2022, cloud technical professionals should deploy Kubernetes using a CaaS platform in the cloud unless they have specific compliance or functional requirements that prevent them from using Kubernetes outside of on-premises infrastructure.

A variety of solutions are available to use Kubernetes as a CaaS in public cloud services. These solutions include:

- **Public cloud Kubernetes services** operated by public cloud IaaS providers, such as Amazon Elastic Kubernetes Service (Amazon EKS), Google Kubernetes Engine (GKE) and Microsoft Azure Kubernetes Service (AKS). Red Hat also offers managed versions of Red Hat OpenShift Container Platform for AWS and Azure, which are jointly engineered and supported with each of these cloud platforms.
- **Self-managed Kubernetes distributions** such as Docker Enterprise, D2iQ Konvoy, Red Hat OpenShift Container Platform, Rancher or VMware Tanzu Kubernetes Grid, deployed on public cloud IaaS.
- **SaaS-based solutions** that operate Kubernetes control planes for orchestrating containers on-premises or in the cloud, such as Platform9 and Giant Swarm.
- **DIY approaches** based on upstream Kubernetes code.

Choose one of the first three approaches for deploying Kubernetes in the cloud. The DIY approach may have been common with early adopters, but it requires more engineering resources to sustain in production than most enterprise clients would want to invest in the effort.

Gartner is also seeing some interest from clients in lifting and shifting legacy applications into containers to take advantage of containers' portability and improvements to life cycle management. However, users should be particularly careful when considering the containerization of legacy applications. They should ask whether containerization is solving a packaging and deployment problem, or a scale and platform problem. If the goal is only to improve the packaging and deployment problem of legacy applications, they have to make sure they understand the additional complexity that containers introduce and the possible alternatives.

Benefits:

- **Enables DevOps:** Containers and Kubernetes help to implement DevOps because they provide a standard framework for I&O and application development to collaborate. The self-service capabilities in Kubernetes platforms also give product teams the ability to deploy services without involving I&O.
- **Enables platform ops:** When multiple teams want to practice DevOps using a shared container management platform, the IT organization should embrace platform ops. Thus, the team responsible for the container management platform focuses specifically on providing DevOps teams with tools.

- **Scalability:** The long-term value of adopting platform ops will be demonstrated in the ability to smoothly scale the number of teams practicing DevOps.

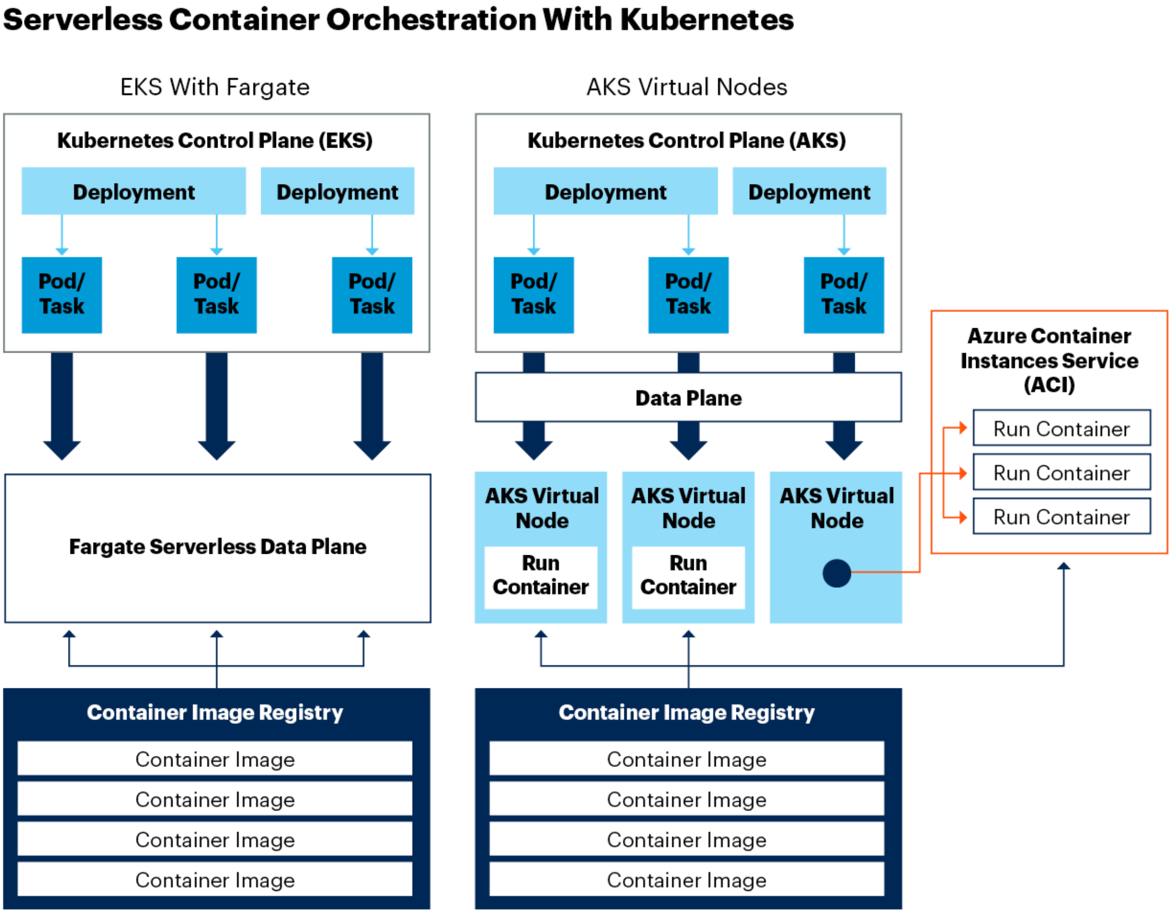
Related research:

- Solution Criteria for Public Cloud Kubernetes Services
- Using Platform Ops to Scale and Accelerate DevOps Adoption
- How to Prepare for Containers and Kubernetes

Leverage Provider-Native Container Services to Optimize Cost and Capacity

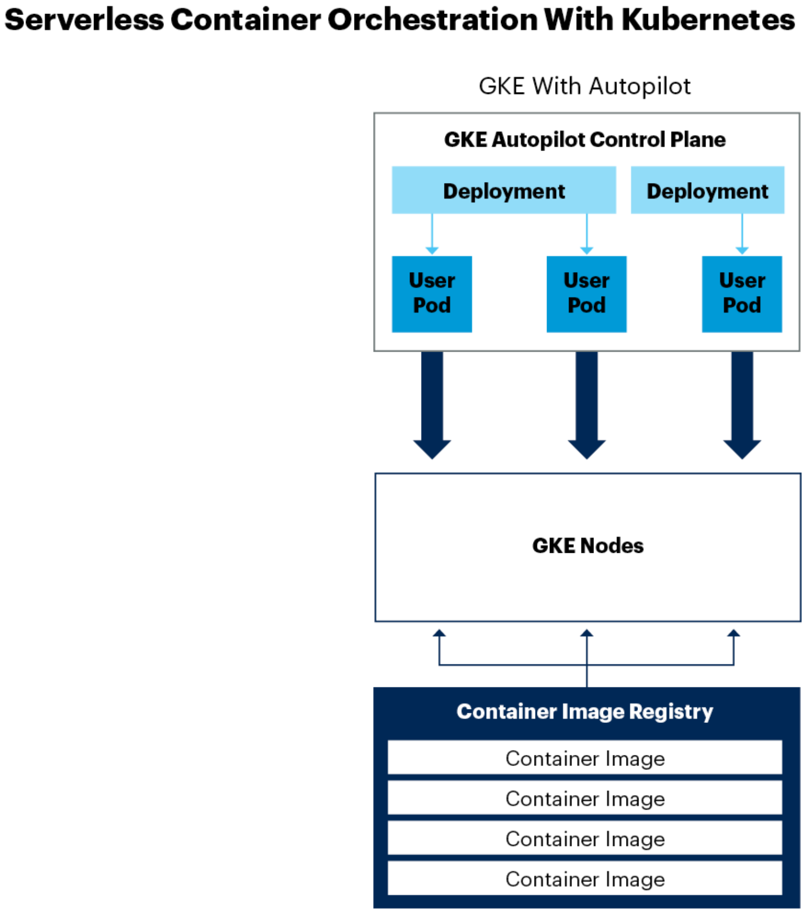
Public cloud services offer a variety of choices for deploying containers. Some now offer serverless container platforms where the cloud provider builds and operates the Kubernetes control plane, removing that burden from the end user. These services provide a compromise betweenCaaS and fPaaS, aka function as a service (FaaS) abstractions. They go beyond managed Kubernetes services, which still require the user to configure worker nodes for the clusters. With serverless container platforms, infrastructure resources are provisioned automatically, on-demand. Figures 15 and 16 illustrate the architecture of a serverless container platform.

Figure 15: Serverless Container Orchestration With Kubernetes



Source: Gartner
753853_C

Figure 16: Serverless Container Orchestration With Kubernetes



Source: Gartner
753853_C

Serverless container platforms, such as AWS App Runner, Azure Container Instances (ACI) and Google Cloud Run, all enable containers to be deployed without requiring administrators to provision resources. Each service that offers serverless containers for Kubernetes has its own unique approach for implementing the capability:

- Amazon Elastic Kubernetes Service (Amazon EKS) integrates Kubernetes clusters with AWS Fargate, which removes the need to provision and manage servers for Kubernetes worker nodes so that users only pay for the resources required to run containers.
- Azure Kubernetes Service (AKS) can offload the responsibility to manage worker nodes with its virtual node feature, which deploys pods inside Azure Container Instances that are automatically and elastically configured in response to changing workload requirements.

- Google Kubernetes Engine (GKE) in autopilot mode provisions and manages all of a cluster's underlying infrastructure. This gives developers an optimized cluster with a hands-off experience so they do not have to calculate the amount of compute capacity that is required by containers.

In 2022, cloud technical professionals should use serverless options for Kubernetes in the public cloud whenever they want to fully exploit the cloud platform's infrastructure for running containers with the Kubernetes orchestration method, but do not have the engineering resources to deploy and operate clusters themselves.

Benefits:

- **On-demand consumption:** Serverless container platforms enable truly on-demand consumption because there are no idle resources or orphaned VMs or containers. Applications and services are billed based on the number of container executions, so users are charged only for the capacity they use.
- **Management simplicity:** Cloud service providers can simplify the Kubernetes experience and reduce its operating burden by operating the Kubernetes data plane as a service on behalf of customers.
- **Compatibility:** Unlike FaaS and fPaaS offerings, these platforms preserve compatibility with open-source Kubernetes. Furthermore, because the containerized OS appears like a standard OS, they require no code changes to implement.
- **Cost optimization:** Because of easy scalability, serverless containers deliver low entry costs that grow on demand. Because they deploy individual containers, usage will be highly efficient, with fewer wasted resources.

Related research:

- Solution Scorecard for Amazon Elastic Kubernetes Service
- Solution Scorecard for Microsoft Azure Kubernetes Service
- Solution Scorecard for Google Kubernetes Engine

The Crisis-Level Skills Gap Will Compromise Cloud Innovation and Execution

Although public cloud infrastructure is hardly new, many veteran I&O staff still lack expertise with these technologies. Indeed, the typical I&O department still harbors some public cloud antagonists. In fairness, IT organizations have often underestimated how steep the skills ramp can be for veteran technical professionals who are learning public cloud for the first time. There are new skills to be acquired, new roles to navigate and new responsibilities to take on. Not every data center engineer is obstinately opposed to learning public cloud; some are just overworked and underresourced. Whatever the reason, technical professionals have not acquired cloud skills fast enough to satisfy the growing demand for cloud services.

> **In many IT organizations, the lack of cloud skills has reached crisis levels.**

Some cloud initiatives have already foundered for want of skilled engineers and architects. A lack of skills can delay or curtail a cloud project — or even cause it to fail outright. For example, in a cloud migration, refactoring the application is usually preferable to a lift-and-shift migration. Rewriting the application to use the cloud provider's native services often improves reliability, performance and cost. But, this is impossible without real, domain-specific expertise in the cloud provider's offerings. If no one present in the IT organization has sufficient knowledge, then the only choice may be to pursue a suboptimal, lift-and-shift migration — or to forgo the public cloud entirely. There are many such stories of workloads that cannot take full advantage of public cloud resources for want of skilled personnel to execute the necessary changes. As public cloud continues to grow rapidly, there are likely to be many more.

It may not be feasible to hire outside experts to fill these gaps. Gartner's analysis reveals some curious trends in the IT job market. There are labor shortages at the top of the job market because there aren't enough people with modern skills. At the bottom of the market, however, there is often a surplus of people with legacy skills for which demand is declining. Incumbent IT employees express increased optimism about the job market and higher expectations for the reward of switching jobs, yet fewer report they actually intend to switch jobs. Thus, the labor market for those in-demand skills is tight and growing tighter. (This is good news for technical professionals, but may be bad news for leadership.)

As it grows more difficult to hire skilled personnel — and as they command ever-higher salaries — IT organizations will need to grow public cloud skills internally. This is a job for technical professionals as much as for managers.

In 2022, I&O technical professionals must:

- Prioritize Kubernetes and DevOps skills development.
- Build a talent-enablement program.
- Find creative ways to bridge the skills gap.

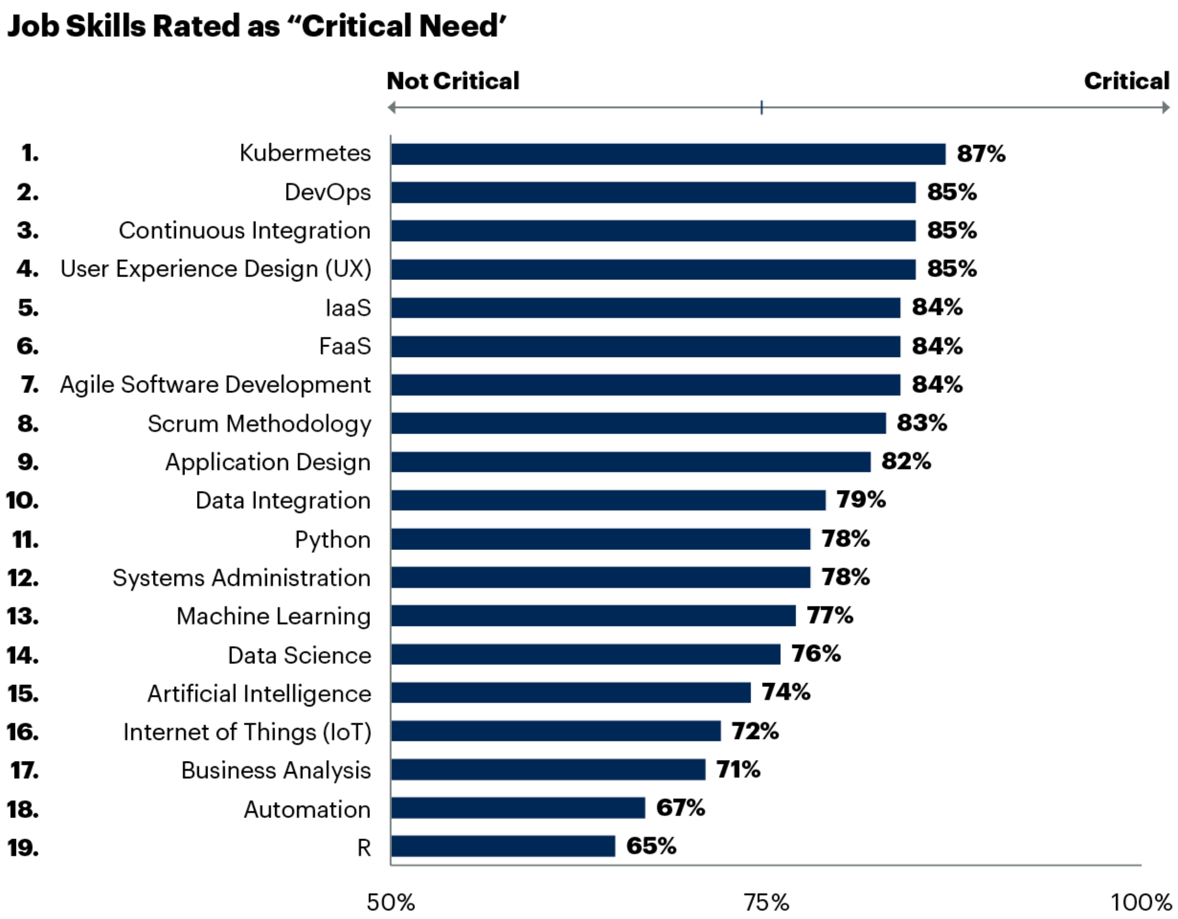
Planning Considerations

Prioritize Kubernetes and DevOps Skills Development

Among all the skills required to operate modern cloud computing environments, Kubernetes and DevOps have pride of place. Kubernetes has become the industry-standard substrate for cloud portability. Its orchestration capabilities allow a container to be instantiated anywhere, making it the ideal code delivery vehicle across public and private clouds. Useful abstractions enable further automation and permit managing the infrastructure as declarative, state-focused configurations. Meanwhile, DevOps techniques have become the gold standard for infrastructure automation and modern infrastructure management. DevOps pipelines are required in order to take full advantage of Kubernetes. Whether in the public cloud or on-premises, Kubernetes and DevOps have become pillars of modern operating models.

So, it comes as no surprise that Kubernetes and DevOps are the skills in highest demand for I&O technical professionals. Gartner's TalentNeuron tracks job postings, wage data and hiring scale to identify the most-sought-after skills in the IT labor market. In a recent survey, Kubernetes and DevOps topped that list — along with continuous integration, which is closely related. Figure 17 illustrates the 19 skills rated as "Critical Needs" in our 2020 IT Skills Roadmap (from which this figure is reproduced).

Figure 17: Job Skills Rated as “Critical Needs”



Source: Gartner 2020 IT Skills Roadmap [G00717934]
753853_C

Other core cloud management skills, such as automation, IaaS and PaaS, also rank among the most-in-demand skills.

Benefits:

- **Widely applicable**: Kubernetes and DevOps skills are highly transferable. They are increasingly important to cloud computing, yes, but also to many other specialties. These skills will be useful regardless of your role or focus area.
- **Future-proof**: As Kubernetes gains additional traction, these skills will only grow more valuable. At a time when many IT skills are declining in usefulness, these will hold their worth into the foreseeable future.

- **Salary premium:** Our TalentNeuron data continues to show that technical professionals with these skills command higher salaries than those without. Learning Kubernetes and DevOps is a very good career move.

Related research:

- New Roles and Skills for I&O Professionals in DevOps
- 2020 IT Skills Roadmap
- IT Workforce Report 1Q21: Emerging Labor Market Optimism and Implications

Build a Talent Enablement Program

A talent enablement program (TEP) is a foreign concept to most IT organizations. They have traditionally focused on narrow skills specialization, rather than equipping their staff with the tools to develop a broader set of emerging technical skills. Without understanding how skills requirements are likely to change, staff will struggle to develop in line with the organization's needs. To address this challenge, organizations must create dedicated, purposeful talent enablement programs. These are designed to promote and develop the necessary skills and to cultivate the necessary roles within the IT organization.

A TEP can help:

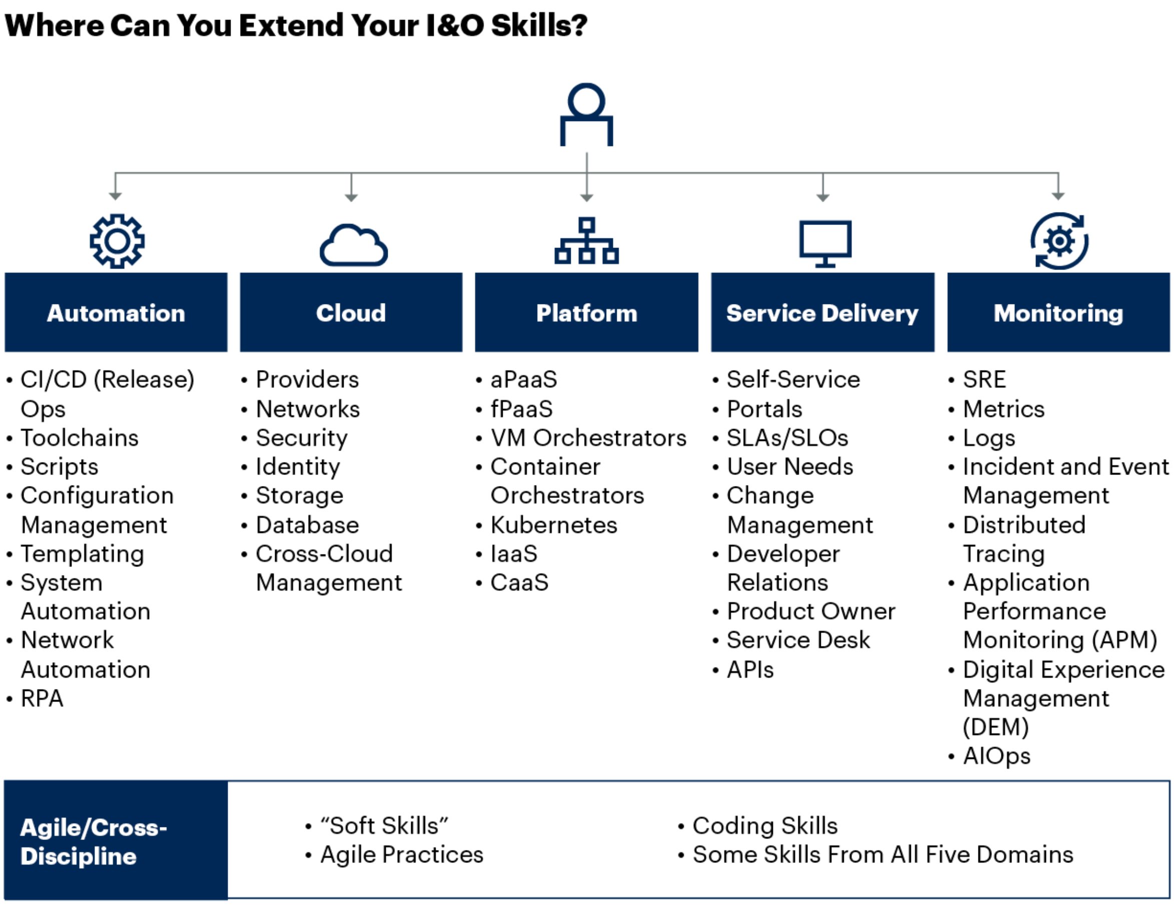
- Refine and define skills the organization needs, now and in the future.
- Improve recruiting efforts.
- Direct technical professionals to the most critical skills they need to improve on or expand into.

> Don't become obsolete. As your organization transitions into new modes of working, such as Agile and/or DevOps, you must ensure that you are personally positioned to participate in either cross-functional or specialist support roles.

These important benefits are not usually delivered by traditional learning or training regimes. For example, effective skills coaching is often difficult for managers, who are even more likely to lack cutting-edge engineering or architecture skills. A TEP speeds up learning new skills and adopting new technologies. Investing in these programs benefits both the IT organization and the individual technical professional.

Your goal as a technical professional must be to enhance your skills in the following areas: agile/cross-discipline practices, automation, cloud operations, monitoring, platform operations and service delivery. The new roles are cloud architect, cloud engineer, automation architect/automation engineer and platform engineer. The new roles and skills are needed to support the evolving responsibilities of I&O. Figure 18 shows the six areas with examples of specific skills in each.

Figure 18: Areas Where I&O Professionals Should Extend Their Skills



Where Can You Extend Your I&O Skills?

Automation
- CI/CD (Release) Ops
- Toolchains
- Scripts
- Configuration Management
- Templating
- System Automation
- Network Automation
- RPA

Cloud
- Providers
- Networks
- Security
- Identity
- Storage
- Database
- Cross-Cloud Management

Platform
- aPaaS
- fPaaS
- VM Orchestrators
- Container Orchestrators
- Kubernetes
- IaaS
- CaaS

Service Delivery
- Self-Service
- Portals
- SLAs/SLOs
- User Needs
- Change Management
- Developer Relations
- Product Owner
- Service Desk
- APIs

Monitoring
- SRE
- Metrics
- Logs
- Incident and Event Management
- Distributed Tracing
- Application Performance Monitoring (APM)
- Digital Experience Management (DEM)
- AIOps

Agile/Cross-Discipline
- “Soft Skills”
- Agile Practices
- Coding Skills
- Some Skills From All Five Domains

Source: Gartner
720642_C

These skills are best developed as part of a systematic talent enablement program, which identifies and prioritizes the highest-value skills to the IT organization. It provides opportunities for direct mentorship and hands-on experience that are not delivered by classroom training or individual research. The TEP does not replace other learning methods or certifications, which remain important. Rather, it provides an important real-world interface between theoretical knowledge and practical experience. Therefore, in 2022, technical professionals should advocate for TEP initiatives within their IT organizations.

Benefits:

- **Addresses organizationwide skills gaps:** A TEP can identify which skills are most relevant to the IT organization overall, then map their existing capabilities against it. This directs learning investments to the most impactful areas.
- **Transparency:** Historically, technical professionals decided which training they were interested in and pursued it. A TEP should publish a technical skills roadmap that highlights the most critical roles and skill sets, making clear to the technical professional what he or she should learn.
- **Practical, hands-on experience:** The TEP should enable real-world knowledge transfer and job sharing among peers. The cloud offers a unique opportunity in this area, with labs and services that are easy to set up and tear down.
- **Apprenticeships:** Relatedly, the TEP should include a formal apprenticeship program. This is an opt-in mentoring program that pairs domain novices with domain experts. It is intended to build a collaborative environment in which technical professionals help one another based on their areas of deep expertise. It is not uncommon for a technical professional to be an apprentice in one area and a skilled craftsperson in another.
- **Environment of positive change:** At a time when technical professionals are confronted with vast changes in technology, a TEP creates a sense of stability and progress. It helps technical professionals feel more comfortable in their jobs and confident in their value to the organization.

Related research:

- Solution Path for Transforming Your I&O Work and Career Strategies
- Assessing Online Learning Platforms for Technical Skills Development

- Jump-Start Your Cloud Skills With These Objective-Driven Assignments

Find Creative Ways to Bridge the Skills Gap

There are far more businesses attempting to accelerate their cloud journey than there are skilled employees on the job market. Four main approaches to staff shortages exist, but each can be problematic:

- **Training:** Encouraging existing employees to expand and deepen their skills is always welcome. However, a training-focused approach often means that cloud projects will be delayed significantly, until cloud skills are brought up to par. Furthermore, while cloud skills can be learned, acquisition of basic skills requires an investment in training and the willingness to allow people to learn and make mistakes in the “real world” environment of the business. Acquiring true expertise takes several years.
- **Hiring:** Finding new employees with the necessary skills has become increasingly difficult, especially now that the pandemic has made remote work feasible. Many employers in high cost-of-living areas are willing to hire cloud talent that lives anywhere, driving up cloud salaries in lower cost-of-living areas.
- **Augmenting:** Bringing in temporary personnel can be an effective short-term solution to address immediate, critical needs. However, this is best suited to filling junior- and mid-level roles that consist largely of executing defined technical tasks. Skilled contractors with senior-level cloud expertise are in high demand and command commensurately high fees. Moreover, there is often inadequate knowledge transfer from consultants to permanent staff.
- **Outsourcing:** This is the most common way to bridge skills gaps. In Gartner’s 2020 Cloud End-User Buying Behavior Survey, 65% of responding cloud decision makers and advisors indicate that their organizations work with external service providers such as cloud MSPs or outsourcing providers. [6] But technical professionals may worry this is a slippery slope that leads to a loss of control and an ever-increasing dependence on the outsourcer. Furthermore, outsourcing may tempt the organization to not aggressively drive cloud skills acquisition for employees.

Fortunately, there are many ways to bridge the skills gap that do not involve the trade-offs inherent in the traditional methods. For example, organizations can look to their existing cloud **community of practice (CoP).** The cloud CoP can be a valuable place to recruit internal candidates for cloud roles. Employees who desire to enter a cloud role can enroll in cloud provider training courses, which are available online for free or for a low cost. Such courses typically lead to a certification exam. Once employees achieve a technical certification, they can be assigned to a cloud project on a trial, temporary basis, while they continue to perform their existing job duties. If they do well on that project, they can be assigned to additional trial projects. After three to five projects, they may apply to transfer into a full-time cloud-related role. This allows employees to demonstrate that they are motivated to acquire the new skill set and to "try out" the new role. At the same time, it enables the organization to verify that employees are qualified and suited to that role before transitioning them to that role full-time.

Pairing is also an effective way of growing cloud skills. Indeed, this is one of the most effective uses of the staff augmentation staffing model. You can pair a skilled external cloud architect with an employee that does not yet have cloud skills but has solid general architecture skills and a good relationship with business stakeholders. That pairing can be used to develop the internal employee into a full-fledged cloud architect. While the employee is also likely to need formal technical training, the pairing can allow the employee to execute some cloud projects with assistance. This is particularly valuable when establishing a cloud center of excellence (CCOE), since a cloud architect in the CCOE needs a deep understanding of the business needs and existing IT environment.

Insourcing tasks that have previously been outsourced to third parties can also help to build internal cloud skills. Many cloud MSPs have a migrate-operate-handoff plan that involves operating side-by-side with the customer over a period of one to three years, allowing the customer to gradually assume responsibility as employees gain the necessary skills.

In 2022, technical professionals should expect to encounter any or all of these models, and be prepared to work effectively in any of them.

Benefits:

- **Flexibility:** IT organizations get a broader range of options, which allows them to choose the best model for their unique circumstances.

- **Employee versatility:** In many cases, the change in the nature of cloud roles goes beyond a mere shift in required technical skills. Often, a "versatilist" mindset, along with"soft skills" such as self-starting, communication and business understanding are required by the new roles.

- **Alignment with personnel directives:** Cloud architects must support CIO and HR initiatives to improve recruiting and retention of cloud talent. These models are well-understood in the HR world and can easily snap into existing frameworks.

- **Architectural improvements:** The insourcing model is particularly useful for aligning external providers with internal business policies and controls. Internal architects may lack the necessary pragmatic cloud knowledge to make good policy decisions. Combining the employee's existing skills and relationship with the external architect's technical knowledge can be very effective.

- **Objectivity and independence:** Another problem with outsourcing cloud expertise is that it may be difficult to trust that the outsourced expertise is being delivered in an objective fashion that is at least somewhat removed from the assisting entity's self-interest. The insourcing and pairing models address this directly.

- **Reduced lock-in:** All of this is an excellent way to reduce dependencies on third parties in the cloud space.

Related research:

- Magic Quadrant for Public Cloud Infrastructure Professional and Managed Services, Worldwide

- Using a Digital Talent Management Framework to Future-Prodthe IT Workforce

- Solution Path for Transforming Your I&O Work and Career Strategies

Setting Priorities

Most organizations will not be able to tackle every planning consideration described in this document within a single year. Therefore, organizations should set priorities based on their specific needs and consider using the following priority-setting framework.

For all organizations:

- Advocate a cloud-first approach, and build a multiyear cloud maturity plan to advance the use of cloud services across the business.

- Continuously evaluate moving workloads to public clouds. Conduct migrations opportunistically: Wait for a natural breakpoint in the applications' life cycle.

- Determine whether your IT organization would benefit from distributed cloud solutions and, if so, which model you might choose. Your preferences for best-of-breed versus integrated systems, and centralized versus decentralized models, will inform your decision.

- Formalize your multicloud strategy. If you have not done so already, begin the process of onboarding a second strategic IaaS or PaaS provider. Look for a common foundation of capabilities — networking, identity, security and billing.

- Develop a cloud management strategy, including hybrid and multicloud management. Prefer your providers' native services, but invest in third-party tools to bridge any capability gaps.

- Audit public cloud workloads regularly. Verify them against your guidelines and guardrails for cloud service consumption. Automate remediation of policy violations.

- Transform the IT organization into a broker of cloud services. This transformation will require new expertise, processes, tools and, perhaps, people.Evangelize the benefits of formal training and talent enablement programs. Although cloud technical professionals are generally not able to execute these initiatives on their own, they influence IT leadership.

For organizations just starting out with cloud computing:

- Build a CCOE to help guide and mature cloud adoption. Staff this CCOE with cloud architects capable of helping both IT and the business adopt cloud computing in an effective fashion.

- Plan to adopt multiple clouds, even if you do not do so immediately. Begin the process of selecting and onboarding a second strategic IaaS or PaaS provider early to build up cross-vendor skill sets and processes for long-term success.

- Appoint a cloud architect to oversee the cloud adoption journey. Choose a candidate with a breadth of skills, including both hard, technical and soft communication skills. Empower that cloud architect with the responsibility of shepherding the organization through cloud adoption and with the ability to effect change within the organization.

- Adopt a cloud decision framework that incorporates preferences for high-abstraction cloud services. Prioritize SaaS solutions to free IT from management burdens, and advocate for PaaS where feasible.

- Document your cloud strategy. Identify the benefits and risks of cloud computing for your organization. Document what aspects of the cloud you will and will not use, and why. For more information about this topic, see Designing a Cloud Strategy Document.

- Create a cloud decision framework to help select the best location to host applications and data. If that location turns out to be the public cloud, the framework should also provide guidance about whether IaaS, PaaS or SaaS is the right deployment model.

- Choose a strategic, blended IaaS and PaaS strategy, and select two strategic providers that you will maintain a relationship with for several years.

- Augment data center backup services with tools that support recovery of cloud-based data. Include a DR strategy when migrating workloads to the public cloud or developing new cloud-native applications.

- Implement distributed cloud infrastructure to provide a hybrid cloud that enables your business to consume cloud services at scale. You'll need to integrate networks, identity, data and services across multiple cloud providers.

- Take advantage of serverless containers in the cloud to come up to speed with container orchestration as quickly as possible, without having to develop the extensive skills needed to operate a particular container orchestration platform.

Evidence

Solution Path for Transforming Your I&O Work and Career Strategies

[1] **2020 Gartner Cloud End-User Buying Behavior Survey:** This study was conducted to understand how technology leaders approach buying, renewing and using cloud technology. The research was conducted online from July through August 2020 among 850 respondents from midsize to large organizations (over $100 million in revenue) in the U.S., Canada, the U.K., Germany, Australia and India. Industries surveyed include energy, financial services, government, healthcare, insurance, manufacturing, retail and utilities. All organizations were required to currently have cloud deployed.

Respondents were involved, either as a decision maker or a decision advisor, in new purchases, contract renewals or contract reviews for one of the following cloud types in the past three years:

- Public cloud infrastructure as a service (IaaS)
- Public cloud platform as a service (PaaS)
- Public cloud software as a service (SaaS)
- Private cloud infrastructure
- Hybrid cloud infrastructure
- Multicloud infrastructure

Respondents were also required to work in IT-focused roles, with a small subset also in procurement roles.

The study was developed collaboratively by Gartner analysts and the Primary Research Team.

Disclaimer: Results of this study do not represent global findings or the market as a whole, but reflect sentiment of the respondents and companies surveyed.

[2] How to Prevent Toxic Data Center Assets From Limiting Digital Transformation.

[3] Tech Providers 2025: CSP-Offered Composable Edge and 5G Services Will Enable Enterprise Agility and Growth

[4] **2019 Gartner Edge Computing Survey:** Results presented are based on a Gartner study conducted to further understand the buying process for edge technology. The research was conducted online from June through July 2019 among respondents located in the United States and Germany. Respondents surveyed were employees of organizations with more than 50 employees and $10 million in revenue. Two hundred respondents were primarily focused in the following three industries: healthcare, manufacturing and retail. Other industries were included at a lower incidence. Respondents were also required to work for organizations that were currently implementing or planning to implement edge computing by the end of 2020, and be personally involved in decisions related to edge computing.

The results of this study are representative of the respondent base and not necessarily the market as a whole.

[5] Hype Cycle for Edge Computing, 2021

[6] Tech CEO Cloud MSPs Are Crucial Partners for Hyperscale Cloud Providers' Growth

Strategy and Success

Document Revision History

2021 Planning Guide for Cloud and Edge Computing - 9 October 2020

2020 Planning Guide for Cloud Computing - 7 October 2019

2019 Planning Guide for Cloud Computing - 5 October 2018

2018 Planning Guide for Cloud Computing - 29 September 2017

2017 Planning Guide for Cloud Computing - 13 October 2016

2016 Planning Guide for Cloud Computing and Virtualization - 2 October 2015

Recommended by the Authors

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

Solution Path for Implementing a Public Cloud Adoption Framework

Designing a Cloud Strategy Document

How to Build a Cloud Center of Excellence (Part 1 — Designing for Cloud Adoption Success)

The Cloud Engineer: Skills Guidance for Modern Technical Professionals

Jump-Start Your Cloud Skills With These Objective-Driven Assignments

Assessing Kubernetes for Hybrid and Multicloud Application Portability

Enhancing Operations Automation With Serverless Computing

The Many Faces of Private Cloud

How to Design a Data Protection Strategy for On-Premises and Cloud IaaS

Decision Point for Selecting Single or Multicloud Workload Deployment Models

Table 1: A Summary of the Offerings on the Market Today

Vendor	Cloud-Inspired	Cloud-Enhanced
Amazon Web Services (AWS)	VMware-on-Outposts	AWS Outposts
Microsoft	Azure Stack HCI	Azure Stack Hub
Nutanix	Enterprise Cloud	Clusters
VMware	Cloud Foundation	VMware Cloud (Various)

Source: Gartner (October 2021)

Actionable, objective insight

Position your IT organization for success. Explore these additional complimentary resources and tools for I&O and IT leaders:

Connect With Us

Get actionable, objective insight to deliver on your most critical priorities. Our expert guidance and tools enable faster, smarter decisions and stronger performance. Contact us to become a client:

U.S.: 1 855 811 7593

International: 44 (0) 3330 607 044

Become a Client

Learn more about Gartner for IT leaders
gartner.com/en/it

Stay connected to the latest insights in 🐦 ▶

Attend a Gartner conference
View Conference

Gartner®