Hype Cycle for Cloud Computing, 2023

Published 20 July 2023 - ID G00790443 - 147 min read

By Analyst(s): David Smith, Ed Anderson

Initiatives: I&O Platforms

Most organizations today use cloud computing, which continues to generate hype as well as technology innovations and trends. Infrastructure and operations leaders must treat the hype about cloud computing pragmatically, as innovations emerge and cloud use becomes ubiquitous.

Additional Perspectives

- Summary Translation: Hype Cycle for Cloud Computing, 2023
  (25 September2023)

More on This Topic

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

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

Analysis

What You Need to Know

Despite the broad and mainstream use of cloud computing, hype surrounding the benefits and limitations of cloud persists. Most organizations use cloud computing and are increasing investments in new cloud-based initiatives.

Cloud adoption is expanding to support evolving business outcomes and use cases such as generative AI, connected vehicles and remote healthcare. Likewise, the demand for cloud capabilities is expanding to support distributed scenarios including edge, new application models, industry-specific requirements, operational needs, and new features and functionality.

Other priorities and outcomes, such as cloud sovereignty, sustainability, industry cloud capabilities, application portability and new service consumption models, drive further technology innovation and vendor differentiation.

Cloud computing has fulfilled its potential, particularly when used to fully embrace cloud operating principles. The COVID-19 pandemic demonstrated cloud models' effectiveness in helping organizations respond to unexpected changes. It proved the models' flexibility and adaptability.

IT leaders should use this Hype Cycle to track technology trends, or what appear to be trends based on marketing, confusion and misunderstandings. They should do so to learn about cloud computing's continuing evolution, assess risks and guide their strategic planning processes to gain the full cloud benefits.

The Hype Cycle

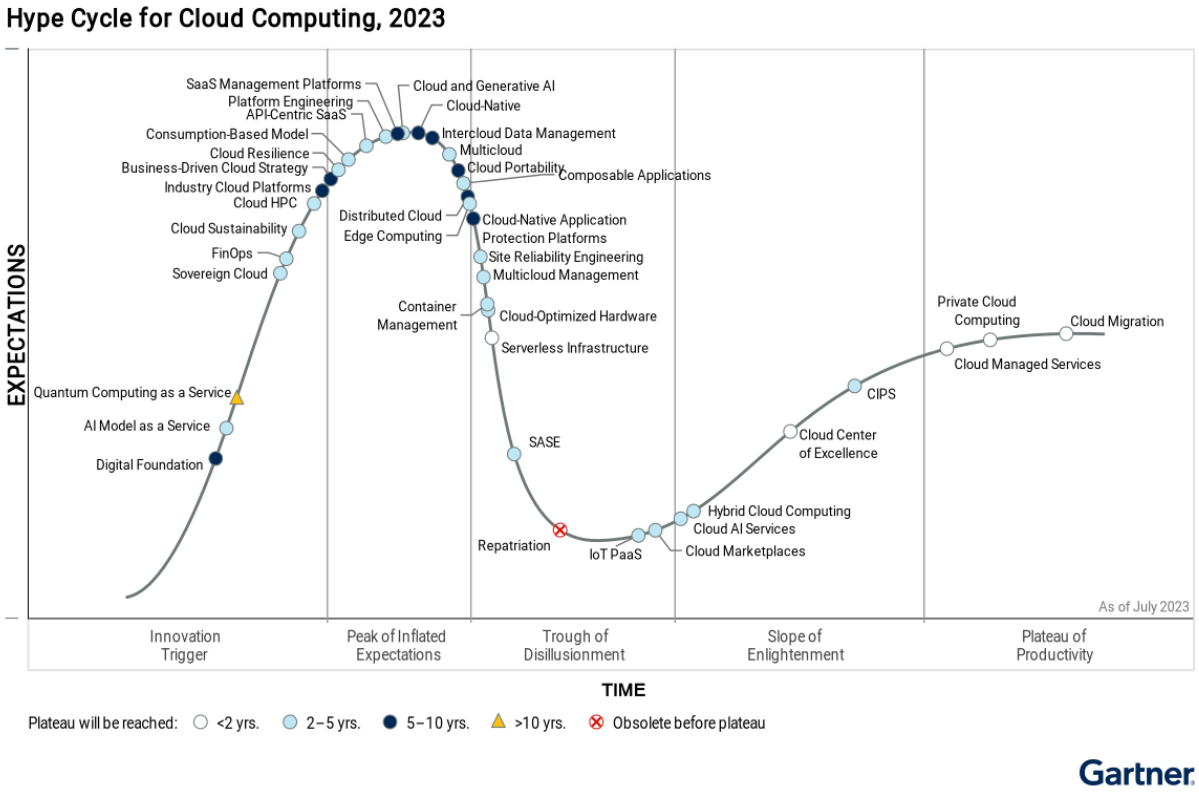
The shift to the cloud continues, and the cloud market remains the fastest-growing IT market. Cloud computing supports mainstream operations for many organizations and serves as a critical part of digital foundations. It is often a vehicle for IT modernization and the desired business outcomes associated with digital transformation. Modernized businesses often adopt cloud operating principles as their preferred style of operations.

Organizations are investing in cloud-native applications, operating models and architectures to gain the full value of the cloud. Cloud-native thinking drives many cloud decisions, along with application modernization initiatives. The influence of the cloud on IT operating models extends through IT operations management, security and networking. It also extends through business functions, such as procurement, where cloud marketplaces are becoming more prominent in technology purchasing.

Business requirements are making organizations examine cloud-based innovations such as AI, the Internet of Things (IoT), telco clouds driven by 5G, and edge computing. Providers will deliver innovations such as quantum computing as a service via the cloud to further support business growth initiatives.

Despite its overall maturity, cloud computing still attracts industry attention as a delivery vehicle for proven IT capabilities and as an innovation foundation for the capabilities that businesses across industries and throughout the world require.

Figure 1: Hype Cycle for Cloud Computing, 2023



The Priority Matrix

The Hype Cycle for Cloud Computing represents a diverse collection of high-impact technologies driving growth and disruption across markets. The dynamic nature of cloud computing causes some cloud technologies and concepts to speed through the Hype Cycle. The transformational nature of cloud-based solutions spawns an ever-increasing collection of innovations, as the innovations at the Hype Cycle's Innovation Trigger show.

Many cloud computing technologies and concepts are two to five years from mainstream adoption, and they will remain impactful. Other technologies have reached mainstream adoption and form the foundation for the next wave of innovations.

The impact of cloud and cloud-related technologies is high and often transformational. Organizations built on a cloud foundation will embrace transformational change more quickly and effectively than organizations bound to traditional IT environments, which is why modernization initiatives always include cloud computing.

Table 1: Priority Matrix for Cloud Computing, 2023

(Enlarged table in Appendix)

Benefit	*Years to Mainstream Adoption*			
	Less Than 2 Years	*2 - 5 Years*	*5 - 10 Years*	*More Than 10 Years*
Transformational	Serverless Infrastructure	Cloud and Generative AI Edge Computing Platform Engineering SASE Site Reliability Engineering	Business-Driven Cloud Strategy Industry Cloud Platforms	
High	Cloud Center of Excellence Cloud Managed Services Cloud Migration	AI Model as a Service API-Centric SaaS CIPS Cloud AI Services Cloud HPC Cloud Resilience Cloud Sustainability Composable Applications Consumption-Based Model Container Management FinOps Hybrid Cloud Computing IoT PaaS Multicloud	Cloud-Native Cloud-Native Application Protection Platforms Cloud Portability Digital Foundation Distributed Cloud Intercloud Data Management	
Moderate	Private Cloud Computing	Cloud Marketplaces Cloud-Optimized Hardware Multicloud Management Sovereign Cloud	SMP	Quantum Computing as a Service
Low				

Source: Gartner (July 2023)

Off the Hype Cycle

On the Rise

Digital Foundation

Analysis By: David Smith, David Cearley

Benefit Rating: High

Market Penetration: 1% to 5% of target audience

Maturity: Emerging

Definition:

A digital foundation is a broad set of technologies and services that can be used to build, assemble, deliver and manage digital solutions, services and experiences. Along with cohesive principles, processes and governance, these are the means to enable the business to deliver value to customers.

Why This Is Important

Business leaders want a business-outcome-focused approach to delivering flexible, dynamic and transformative digital solutions. However, the underlying technologies to support this endeavor are often rigid, poorly connected and inflexible. A properly architected digital foundation built on a set of core principles and governance models can address these deficiencies and accelerate digital transformation. Foundations are made up of modular digital assets and other aspects including data, policies and processes.

Business Impact

Meaningful transformation requires bold, cross-enterprise investments — in other words, “big bets” on modern business-enabling composable technology capabilities. Digital foundations enable multidisciplinary teams to get results faster and improve the experience of building solutions. They lessen the burden of focus on underlying capabilities and allow teams to dedicate time to customer value. Creating a digital foundation where the entire stack is a series of flexible and interconnected platforms improves business resiliency and adaptability.

Drivers

- The speed and agility requirements of business continue to grow — as do the complexities required to deliver digital solutions.

- The ability to share and scale is needed as more teams require access to the same or similar capabilities. This enables economies of scale and value to justify creating capabilities shared by multiple teams.
- There are more moving parts in delivering solutions than ever before. This places a burden on teams to either build or somehow procure a delivery system in addition to the actual software they are trying to produce.
- Advances in technologies such as cloud computing, artificial intelligence, automation and analytics — as well as digital foundations themselves — have provided many ready-made building blocks for building digital foundations.
- The digital foundation builds on a related concept of “platform engineering” which focuses on creating a platform that allows building digital solutions to focus on business value while the platform deals with all the complexities of the underlying infrastructure. Digital foundation extends the concept such that everything is a platform including componentized applications, AI services and more.

Obstacles

- Building and delivering digital solutions requires solid skills in software engineering, product management and modern infrastructure, all of which are in short supply.
- People can have unrealistic expectations for digital foundations. Previous efforts may have suffered from “if you build it they will come” thinking and may have insisted on components that do not meet the needs of the teams responsible for building the solutions.
- There may be organizational issues over who is responsible for selecting, designing and managing the digital foundation and the life cycles surrounding it. New ways of working are needed to facilitate more cooperation.
- Legacy technologies and solutions often hold you back. Embracing a full digital foundation model will require significant changes in architecture, governance and technology implementations.

User Recommendations

- Build digital foundations from broad, composable sets of technologies and services that can create and manage digital solutions, services and experiences. Composable thinking is a key part of the mindset.

- Focus not only on the technology, but also on the principles, processes and governance required to enable the business to deliver value to customers.
- Use proven technology capabilities and services, including cloud computing and other shared capabilities.
- Don't expect to buy a foundation. A foundation should utilize mostly off-the-shelf capabilities or use co-created and shared capabilities. Some assembly will be required in order to meet your needs.
- Build a roadmap and establish achievement of a digital foundation as an aspirational goal. Isolate legacy applications and grow the digital foundation to support new and updated digital solutions.

Gartner Recommended Reading

Use Collaborative Decision Making to Drive Mixed Portfolios' Digital Outcomes

Infographic: The Evolving Digital Foundation

How to Improve CRM by Architecting Your Customer Technology Platform

Infographic: Why Digital Ecosystems Are the Future of Business

AI Model as a Service

Analysis By: Rajesh Kandaswamy

Benefit Rating: High

Market Penetration: 20% to 50% of target audience

Maturity: Early mainstream

Definition:

"AI Model as a Service" is an AI-based model offered as a consumable service by cloud providers. The underlying AI models are from the cloud providers themselves, other tech companies or open-source initiatives. As foundation models and other AI models proliferate, cloud providers seek to offer such models to their clients. This includes both direct access to pretrained models and the ability to fine-tune such models for custom use.

Why This Is Important

Models are key building blocks in artificial intelligence. They vary in purpose and technique. A subset of AI models such as the foundation models (e.g., large language models such as GPT-3.5, which powers ChatGPT, and PaLM 2, which powers Bard) are pretrained on a large corpus of data using extensive computing resources. These models have a variety of uses, but are prohibitively expensive for most enterprises to train and manage. Cloud providers have the infrastructure and capital to invest in and offer such AI model as a service offerings.

Business Impact

A variety of AI model “as a service” offerings will emerge from cloud providers, providing an easy entry point for most enterprises to leverage AI through pre-trained models, fine-tuning such models or integrating them with other applications. The scope for where AI model as a service can be used is vast and includes:

- Technical tasks such as entity extraction, image classification and code generation
- Business tasks such as marketing content generation, financial application and scientific research

Drivers

Few underlying drivers caused the emergence of AI model as a service and will propel its growth further:

- The growing capabilities of foundational models are expanding their potential across a variety of horizontal and vertical applications for companies of all sizes and kinds. This leads to a proliferation of models from many sources, including the cloud providers themselves. This proliferation makes it hard for customers to invest — an opportunity that cloud providers are only eager to capture.
- The use of GenAI in technology products and business is poised to grow, but the costs of utilizing the models are likely to become a serious concern for most businesses. Cloud economies of scale promise users optimized efficiencies and will increase their use.
- Beyond foundational models, AI tooling and nonfoundational models continue to grow. Cloud providers are investing in these complementary services to capture as much AI revenue as possible through these investments.

- There has been significant investment in startups that are leveraging AI capabilities in their own technical and business offerings, directly or through partnerships. These startups typically do not possess the capital or resources to manage their own infrastructure and rely on cloud providers for models and other AI services.
- Enterprises have been shifting large chunks of their IT to cloud providers over the past few years and are operationally ready to use cloud-based services in many aspects of their IT. Such enterprises have a cloud-first approach for most technology investments and will apply the same for their AI investments.
- Automated agents driven by generative AI can propel growth as enterprises start to leverage other models as part of their processes and workflows.

Obstacles

Although AI model as a service plays a key role in the growth of AI in enterprises, a good part of its growth is predicated on the growth of generative AI itself. These obstacles can hinder growth:

- Many solutions, especially foundation models, still suffer from inaccuracies, hallucinations, bias and lack of explainability.
- Model security, data privacy, IP issues due to the data models are trained on, and other areas are murky.
- Lack of standards and incompatibilities between various models can inhibit growth.
- The use of AI-based models is still fairly new. AI-based models do not always easily fit in within today's products, processes and services. Maximizing the use of AI demands rethinking of processes or even the advent of new business and operating models. This shift is beginning, but may take years due to a few reasons. These reasons include maturity of technologies, the need for new processes and business models, and a delay as older assets are fully depreciated.

User Recommendations

- Evaluate different offerings for a given need rather than defaulting to your current cloud provider's recommendations.
- Ensure that you think strategically (i.e., prepare for potentially transformative changes), but act tactically (i.e., experiment, but make no long-term commitments constraining future change in a fast-moving space).
- Develop a bimodal strategy in what services your IT teams can use. One strategy should allow for faster experimentation with a variety of models, but with restricted use of data in sandboxes. The other should be geared toward wider enterprise use with more controls and robustness.
- Perform a detailed analysis of the security, privacy, accuracy, explainability and IP protections and involve internal compliance, security and legal teams.

Sample Vendors

Alibaba Cloud; Amazon Web Services; Google; Hugging Face; IBM; Microsoft; OpenAI; Oracle

Gartner Recommended Reading

Quick Answer: What is GPT-4?

Magic Quadrant for Cloud AI Developer Services

Quantum Computing as a Service

Analysis By: Chirag Dekate, Mark Horvath, Matthew Brisse

Benefit Rating: Moderate

Market Penetration: Less than 1% of target audience

Maturity: Embryonic

Definition:

Quantum computing as a service (QCaaS) provides enterprises with access to quantum computing systems and associated services that enable them to explore enterprise-relevant use cases and devise quantum algorithms for highly specialized sets of problems. QCaaS provides vendors with access to their own technologies, and some cloud service providers offer QCaaS that supports access to various quantum computing implementations, vendors and solution approaches.

Why This Is Important

The torrid pace of innovation in quantum computing systems means that on-premises quantum systems are impractical for most users today, given their limited utility and rapid aging. QCaaS enables enterprises to derisk quantum strategies and leverage cloud services to access, test, validate and utilize diverse quantum technologies. QCaaS environments enable enterprises to focus on exploring a variety of use cases and devising quantum algorithms, as opposed to negative ROI, on-premises acquisitions.

Business Impact

Enterprises pioneering quantum initiatives are focusing on five key applications: optimization, simulation, search, linear systems and security-related use cases. QCaaS enables enterprises to explore different types of quantum systems and accelerate quantum skills development in a relatively low-risk environment. QCaaS continues to evolve in maturity. However, these environments are not ready for production use cases, primarily due to the limited scale of underlying quantum systems.

Drivers

- Many scientific problems are unsolvable using traditional computing technology. QCaaS offers access to quantum computing technologies for organizations pursuing solutions to computationally hard problems, without the risks and costs associated with dedicated systems that are likely to age faster, given the pace of innovation in the industry.
- Rather than acquiring expensive quantum systems on-premises, enterprises can minimize cost, complexity and time to value by using QCaaS-based quantum computing services.
- Some leading cloud service providers offer access to diverse quantum systems, simulators, resource estimators and high-performance computing (HPC) for hybrid workflows, simplify identity and data management and offer streamlined pricing across diverse quantum providers. In some cases, this approach can simplify exploration of quantum technologies and significantly lower risk.
- Continued scaling of underlying quantum computing systems and implicit advancement of the field (including scalable error correction schemes) is seminal to the evolution and eventual success of QCaaS.
- The ability to address the growing set of use cases beyond the traditional five — optimization, simulation, search, BQP and security — will be essential to create virtuous business cycles.

Obstacles

- A lack of ROI, limited applicability and the inability to demonstrate value creation are key business obstacles limiting enterprise investments in quantum.
- A lack of sufficient scale in underlying quantum computing systems powering QCaaS limits the scale of applications that can be explored or run. Current classical approaches deliver better, more impactful results than any quantum alternative.
- Quantum computing systems continue to be nascent in maturity, with more than half a dozen different ways of representing qubits and organizing systems to deliver error correction and scaling. Quantum technologies that now look promising may not be the ones that deliver value in the future.
- There remains a lack of skills to leverage QCaaS effectively, including the development of applications to fully exploit quantum computing capabilities.

User Recommendations

- Leverage QCaaS to devise quantum initiatives: Avoid acquiring on-premises quantum systems. The rapid pace of innovation in quantum technologies means that most on-premises systems will be obsolete faster, as newer systems and scalable technologies come online. QCaaS minimizes the risk associated with these dynamics.
- Select single-provider QCaaS for specialization and value creation: Direct QCaaS capabilities enabled by quantum vendors can provide highly specialized access to quantum systems, while derisking your strategies. Engage in this approach if your main goal is value creation and scaling.
- Select multiquantum system QCaaS for exploration and broader enterprise cloud strategy integration: Some CSPs offer access to multiple quantum providers, enabling enterprises to evaluate diverse technologies and simplified integration to existing cloud practices.

Sample Vendors

Google; IBM; Origin Quantum; Oxford Quantum Computing; PASQAL; Quandela; Quantinuum; Rigetti Computing; Xanadu

Gartner Recommended Reading

Cool Vendors in Quantum Computing

Infographic: How Use Cases Are Developed and Executed on a Quantum Computer

Preparing for the Quantum World With Crypto-Agility

Sovereign Cloud

Analysis By: Rene Buest, Gregor Petri, Neville Cannon

Benefit Rating: Moderate

Market Penetration: 5% to 20% of target audience

Maturity: Adolescent

Definition:

Sovereign cloud is the provision of cloud services within a jurisdiction meeting data residency requirements and operational autonomy. It is intended to ensure that data, infrastructure and operations are free from control by external jurisdictions, and protected from foreign government influence and access.

Why This Is Important

The public sector and commercial organizations increasingly depend on cloud services, leading to greater demand for control and autonomy, and hence more offerings to comply with regulations. In many jurisdictions, data residency, data protection and privacy laws are increasing. Laws combined with increasing geopolitical tensions, different economic ideologies, and proliferating cybersecurity risks from a variety of directions, including state actors, are steadily elevating interest in sovereignty.

Business Impact

Concerns about the sovereignty of data, infrastructure and operations hosted in foreign-owned cloud service offerings have led to newly announced sovereign cloud offerings. Legislative mandates are being applied to limit the ability of organizations to use nondomestic vendor services. These impact buying decisions and investments organizations are willing to make in cloud offerings. As a result, end users could find themselves in a fragmented market without access to the resources they need to support their digital business initiatives.

Drivers

- Governments and commercial organizations are becoming increasingly aware of and concerned about their dependence on foreign cloud infrastructure providers and SaaS offerings. Reasons for this are souring of geopolitical relationships, tighter privacy and data protection regulations, data sovereignty, data control and operational autonomy, as well as technological sovereignty and independence.
- The market for digital and cloud technology and services is dominated by U.S. and Chinese technology and service providers. As a result, all non-U.S. and non-Chinese organizations and companies mainly have to access foreign services and technology to build and run digital business models. Hence, data is being stored within nondomestic cloud and digital service providers, which creates political uneasiness.
- As digital services become increasingly important and critical, cloud customers and regional trade bodies worry about retaining control over their data and infrastructure to stay compliant with local regulations as well as their operational autonomy.
- Some more regulated industries and governments are particularly concerned by the U.S. and Chinese legal frameworks that might allow these governments to access cloud-stored data under specific circumstances.
- Businesses increasingly depend on technology platforms they don't control. Although the risk of deplatforming remains small, the growing number of platforms enterprises use and the businesses' growing dependence on platforms increase the consequences of deplatforming.

Obstacles

- In the short to medium term, the market dynamics make it almost impossible for domestic cloud providers to present a viable alternative to hyperscale cloud offerings, as the capabilities of hyperscaler far exceed most domestic cloud offerings. Considerable technical obstacles exist if domestic clouds are expected to deliver the maturity and level of scalability, reliability and functionality of hyperscale offerings.
- Too few skilled engineers exist to replicate the design capabilities of hyperscale cloud offerings to build comparable domestic cloud offerings. With lower levels of skills being available, security and operational maturity will be compromised, potentially leading to greater security and failure risks.
- An increasing number of announcements of sovereign cloud offerings from global cloud providers hit the market, all based on various approaches and delivery models.
- Individual governments each defining their own requirements for sovereign cloud offerings may lead to compliance regimes that break public cloud scale and innovation.

User Recommendations

- Subject proposals for the sovereign cloud to the same level of risk assessment that current cloud computing offerings are subjected to. Do not assume that the sovereign cloud conveys any additional security measures in itself.
- Differentiate between different sovereign cloud approaches by type of workload, data and infrastructure when making cloud deployment decisions. Doing so, classify various delivery models between the cloud provider approaches to meet sovereignty requirements. Make sure to establish a consistent, repeatable and defensible process when assessing sovereign offerings.
- Explore evaluating locally provided cloud services for workflows that can be provided locally and leverage third-party solutions to protect data and ensure it is compliant with local requirements.
- Assess any considered sovereign cloud offerings against long-term viability, also in case legal requirements change or global offerings start to directly cater to national sovereignty requirements.

Sample Vendors

Bleu (joint venture of Capgemini and Orange); Delos Cloud; Google Cloud + T-Systems; Microsoft; Oracle; S3NS; Whale Cloud Technology

Gartner Recommended Reading

What Are the Different Provider Approaches to 'Sovereign Cloud' Demands?

What We Are Hearing About Cross-Border Data Transfers

Product Manager Insight: Three Cloud Deployment Models to Address Your Customers' Key Sovereignty Requirements

Quick Answer: Is the Risk of Relying On the New E.U.-U.S. Privacy Framework Too High for Organizations?

Three Critical Use Cases for Privacy-Enhancing Computation Techniques

Cloud Sustainability

Analysis By: Ed Anderson

Benefit Rating: High

Market Penetration: 5% to 20% of target audience

Maturity: Emerging

Definition:

Cloud sustainability is the use of cloud services to achieve sustainability benefits within economic, environmental and social systems. As such, cloud sustainability refers to both the sustainable operation and delivery of cloud services by a cloud service provider, as well as the consumption and use of cloud services by organizations and individuals to achieve sustainability outcomes.

Why This Is Important

Cloud sustainability is a key digital technology supporting organizations in their use of technology to achieve their sustainability ambitions. Cloud computing models are well-suited to deliver sustainability benefits because of their ability to operate at scale using a shared services model, which results in efficient use of computing resources. Hyperscale cloud data centers can be physically located near renewable energy sources further extending their potential to lessen environmental impact.

Business Impact

Increasing attention and focus on environmental and social issues is motivating organizations to improve their sustainability posture. Pressure from customers, investors, partners, regulators, employees, and the public at large is motivating organizations to establish sustainability goals and to demonstrate sustainability outcomes. Cloud computing has great potential to improve sustainability outcomes through efficient operations and the delivery of cloud-based technology innovations.

Drivers

- Sustainability is a rising imperative for organizations across all industries and in all countries and regions around the world. Although sustainability encompasses environmental, social and economic factors, environmental sustainability receives the most attention.
- Corporate climate and decarbonization commitments are typically cascaded to individual business functions, including IT. Consequently, IT organizations are looking at all possible ways to implement such strategies, including cloud sustainability initiatives.
- Market data shows that customers, investors, regulators, citizens and employees increasingly value organizations with demonstrable commitments to sustainability.
- Sustainability investments correlate with operational efficiency. Most organizations operating in an increasingly sustainable fashion also recognize other benefits such as reduced spending on energy, reductions in waste and improvements in water use.
- Cloud providers, being among the world's largest data center operators, show strong commitments to cloud sustainability and are making demonstrable progress toward delivering sustainable cloud service offerings.
- Regulatory and legislative mandates for sustainability are increasingly common across regions and industries. The use of cloud services and other digital technologies will help organizations comply with future regulatory reporting requirements.

Obstacles

- Sustainability definitions, metrics and reporting standards are inconsistent, varying by region and industry. Defining, tracking and reporting sustainability performance is complex for most organizations.
- Cloud providers claim to have made great strides in offering sustainable cloud solutions, but these claims are often difficult to verify and contribute to potential "greenwashing." The lack of sustainability reporting standards makes it difficult to interpret and validate provider claims.
- Achieving cloud sustainability outcomes is a shared responsibility between the cloud provider and the customer. Cloud providers must demonstrate sustainable cloud operations, and cloud consumers must employ sustainability practices in their use of cloud services.
- Renewable energy is a key enabler of cloud sustainability and yet there is insufficient capacity to generate and store the energy required to meet the needs of the world's cloud service offerings.

User Recommendations

- Establish internal sustainability goals including specific metrics and sustainability outcomes by doing a materiality assessment to determine which sustainability outcomes are most important to your organization.
- Determine the role cloud sustainability will play in the achievement of sustainability outcomes. Build internal credibility for cloud sustainability by ensuring that the sustainability benefits of specific cloud service offerings are independently validated.
- Engage relevant executives and other internal stakeholders proactively that are tasked with creating and achieving sustainability goals. Establish credible metrics for measuring and reporting cloud sustainability outcomes.
- Look to cloud providers and other experts, including IT service providers, for best practices in operating and consulting cloud services in a sustainable manner.

Sample Vendors

Alibaba Cloud; Amazon Web Services; Google; IBM; Microsoft; Oracle; Salesforce; SAP; Scaleway; VMware

Gartner Recommended Reading

Executive Leadership: Sustainability Primer for 2023

Quick Answer: How Green Are Public Cloud Providers?

Build an Environmental Cloud Sustainability Strategy

Make Sure Technology Helps More Than Hurts Sustainability

Sustainability: A Customer Priority and Provider Imperative

FinOps

Analysis By: Lydia Leong

Benefit Rating: High

Market Penetration: 5% to 20% of target audience

Maturity: Adolescent

Definition:

FinOps is the model proposed by the FinOps Foundation to implement cloud financial management (CFM). FinOps applies DevOps principles and practices to cloud financial operations. The FinOps Foundation defines FinOps (which it has trademarked) as "an evolving cloud financial management discipline and cultural practice that enables organizations to get maximum business value by helping engineering, finance, technology and business teams to collaborate on data-driven spending decisions."

Why This Is Important

All organizations with meaningful spend on cloud IaaS or PaaS need to engage in cloud financial management (CFM) so that they are aware of what they are spending and why. Organizations must undertake greater cost governance efforts if they have substantial spending, significant self-service, cost complexity or variability. Further, CFM helps organizations optimize their costs and extract greater value from their spending. FinOps is one possible implementation model for CFM.

Business Impact

FinOps is a partial solution to CFM needs. Cloud economics spans the continuous process of CFM as well as one-off activities, and is focused on maximizing the value of cloud computing to the business, rather than minimizing cloud expenses. For example, business leaders may reasonably make the decision to spend more to deliver a better user experience, or to ignore cost-related technical debt so application teams can focus on delivering more features.

Drivers

- Public cloud costs are of concern to many customers. Many customers experience unexpectedly high cloud costs because they have fulfilled far greater business demand for cloud services than the IT organization had forecast.
- Ungoverned cloud adoption can lead to uncontrolled and careless spending. Organizations without an effective CFM practice cannot properly plan, track or optimize their cloud costs.
- Organizations that do not effectively manage cloud economics cannot assist the business in making thoughtful decisions about the top line versus the bottom line in cloud-enabled digital products and services.
- The FinOps Foundation (FOF), founded in 2019 by several cloud cost management tool vendors, created and popularized the use of the FinOps term. At that time, it defined FinOps as “the practice of bringing financial accountability to the variable spend model of cloud, enabling distributed teams to make business trade-offs between speed, cost and quality.”
- FOF changed the definition of FinOps in November 2021. The new definition is more closely aligned to Gartner’s recommended practices for the management of cloud economics. However, not all entities use the FinOps term in a consistent way.
- In mid-2020, FOF became a project of the Linux Foundation. Its membership has grown over time, and now includes many vendors that sell FinOps tools and services (which FOF will badge as FinOps Certified Platforms), numerous cloud providers, several global systems integrators and some enterprises (mostly financial services institutions). These members promote FinOps concepts to their users, increasing hype.
- FOF has consistently promoted the adoption of a centralized FinOps team, which in turn purchases and uses FinOps tools, leading customers to believe they should adopt this approach.
- FOF launched the FinOps Open Cost Usage Specification (FOCUS) in 2023. It is intended to be an open standard for cloud billing data, and will be broadly useful if widely adopted by cloud providers.

Obstacles

- Not all organizations have a business case for CFM, although almost all organizations that have meaningful cloud adoption need to perform cost management.
- CFM needs to be a cross-functional and cultural practice, not solely the responsibility of a dedicated FinOps team. In addition, many organizations cannot cost-justify having such a team.
- Application teams and the business owners of applications need motivation to optimize their cloud spend. This usually requires coupling showback to incentives (or penalties) or performing chargeback, forcing these entities to be accountable for what they spend.
- Although FinOps tools can often recommend optimizations for cloud infrastructure, they have limited capabilities for PaaS.
- Many organizations do not undertake activities that reduce their cloud spending because these activities do not have an adequate ROI. That is, the spend reduction is insufficient to justify the time and labor necessary to technically implement the optimizations.

User Recommendations

- Establish a cross-functional CFM practice — not just a FinOps team — once you have public cloud IaaS and PaaS adoption that has proceeded beyond the pilot stage.
- Your cloud center of excellence (CCOE) or other cloud governance function should own the CFM practice. The CCOE should set the policies and guidance, but cloud operations teams are typically responsible for implementing CFM tools and assisting application teams with optimizations.
- When custom-developing cloud solutions, design cost-aware architecture. Application design and implementation will be the primary influence on cloud costs.
- When migrating existing applications to the cloud through a lift-and-shift (that is, a rehost) approach, use performance-based rightsizing and accept the smallest recommended size by default. Most organizations tend to oversize virtual machines when they migrate, due to unwarranted concerns about performance and “headroom.”

Sample Vendors

Apptio; Flexera; VMware

Gartner Recommended Reading

Is FinOps the Answer to Cloud Cost Governance?

Managing Cloud Economics: A Cloud Architect's Guide to Productive Relationships With Sourcing Leaders

Effective Cloud Sourcing Strategies Focus on FinOps, Not Cost Reductions

Beyond FinOps: The Gartner Framework for Public Cloud Financial Management

Cloud HPC

Analysis By: Chirag Dekate

Benefit Rating: High

Market Penetration: 5% to 20% of target audience

Maturity: Adolescent

Definition:

Cloud high-performance computing (HPC) services deliver modernized supercomputing environments that accelerate HPC-powered digital innovation, using integrated cloud services, automation, elasticity and AI-infused management. Cloud HPC providers deliver HPC as a service that combines HPC architecture with cloud-native characteristics; elastic scalability, programmable automation and metered usage.

Why This Is Important

Cloud HPC brings the benefits of the cloud to HPC workloads by offering elastic, multitenant HPC infrastructure as a cloud service. Cloud HPC enables enterprises to develop new digital products, scale innovation faster, modernize HPC applications and validate new technologies faster. Further, cloud HPC offers exclusive capabilities including accelerator environments, automated management services, workload-specific continuous cost optimization and early access to HPC technologies.

Business Impact

Cloud HPC services are designed to augment/replace without investing in new capital expenditure (capex), experiment with different configurations in development and test without high upfront investment. This lean approach to HPC will accelerate the modernization of existing HPC apps and allow new apps to be built/tested/refined with less upfront risk and investment.

Drivers

- With enterprise initiatives increasingly leveraging cloud for broader workloads, on-premises HPC infrastructures are transforming into analytics islands, often not directly connected into the broader enterprise architectures. Enterprise IT leaders are actively exploring cloud HPC or a hybrid cloud HPC approach to address these problems and seeking to enable new value creation use cases. This includes simulation-driven digital twins and AI-augmented HPC, and ensures continuity of existing advanced analytics pipelines.
- In most enterprises, HPC systems tend to be oversubscribed with large long-running jobs occupying large chunks of resources. As a result, smaller jobs wait in resource management queues for inordinately long durations. HPC Infrastructure & Operations (I&O) teams are actively seeking to minimize queue wait times for business stakeholders using cloud HPC to offload some of the smaller jobs.
- IT leaders are seeking to modernize HPC I&O practices that have remained relatively unchanged for nearly two-and-a-half decades. In part, leaders are seeking to leverage virtualization techniques including HPC-specific containers to enable portability in some HPC applications.
- Enterprises are evolving IT HPC designs that make “hybrid HPC” possible, moving workloads between on-premises and off-premises environments, opening up the ability to “burst” into more capacity as needed.
- Cloud HPC enables any type of enterprise to rapidly and cost-efficiently explore newer accelerator technologies and validate applicability for their applications.
- Retaining HPC skills is becoming harder due to both “the great resignation” and demand-driven talent attrition, resulting in some enterprises struggling to deliver sustainable HPC services. Some leaders are leveraging cloud HPC to mitigate these impacts.
- Rise of generative AI driving interest in exploring cloud HPC infrastructure capabilities that would otherwise not be accessible on-premises without high upfront investment.

Obstacles

- HPC workloads running in public cloud infrastructure as a service (IaaS) may experience unexpected cost spikes as they begin to integrate with and utilize the other native-cloud services available in that environment, which are billed separately.
- Cloud HPC environments are better-suited to loosely coupled, very parallel workloads. Applications that require tightly coupled systems with high-performance low-latency message passing interconnects will not operate cost-effectively or efficiently in most general-purpose clouds.
- HPC middleware has largely remained unchanged for the last two-and-a-half decades and has not aggressively adopted broader IT practices, including virtualization and containerization, due to the perceived performance impact of virtualization techniques.
- HPC-specific instances are not available from cloud providers in sufficient volumes across the diverse regions, severely limiting enterprise ability to leverage these systems.

User Recommendations

- Devise a hybrid HPC cloud platform that enables your teams to access supercomputing capabilities wherever they are needed (on-premises, cloud and/or edge). Embrace Gartner's platform operations approach to deliver these platforms.
- Modernize HPC skills by upskilling and hiring for HPC talent that can augment cloud HPC in your organization. Seek individuals with cloud competencies in addition to HPC.
- Accelerate disruptive HPC innovation by encouraging rapid prototyping of HPC-specific container and virtualization strategies.

Sample Vendors

Alibaba Group; Amazon; Atos; Google; Hewlett Packard Enterprise; Microsoft; Oracle; OVHcloud; Rescale; UberCloud

Gartner Recommended Reading

Rethink Supercomputing for a Digital Era

Understanding the Opportunity for Arm-Based Servers

Industry Cloud Platforms

Analysis By: Gregor Petri

Benefit Rating: Transformational

Market Penetration: 20% to 50% of target audience

Maturity: Adolescent

Definition:

Industry cloud platforms address industry-relevant business outcomes by combining underlying SaaS, PaaS and IaaS services into a whole product offering with composable capabilities. These typically include an industry data fabric, a library of packaged business capabilities, composition tools and other platform innovations. IT leaders can use the composability of these platforms to gain the adaptability and agility their industries need to respond to accelerating disruption.

Why This Is Important

Cloud, software and service providers are launching industry cloud platforms (ICP) by combining SaaS, PaaS and IaaS offerings with industry-specific functionality and composable capabilities to create more compelling propositions for mainstream customers. Emerging industry cloud platforms are leveraging innovative approaches such as composable packaged business capabilities (PBCs), PBC marketplaces, data grids and fusion teams to accommodate faster change and platform adaptability.

Business Impact

Broader cloud adoption within enterprises requires more whole-product business solutions that enable defined industry scenarios and process models, rather than technology-oriented solutions that enterprises have to largely configure and integrate themselves. ICPs enable enterprises to adopt more holistic cloud strategies that span across established cloud service categories such as SaaS, PaaS and IaaS.

Drivers

- As the complexities of both business and technology continue to increase, enterprises are looking for more outcome-based engagements with their cloud providers. However, such outcomes must be flexible enough to be able to adapt to the changing circumstances.
- To be relevant and be able to resonate with enterprise audiences, such outcomes must be business relevant, specific, measurable and tangible — a goal that is easier achieved when approached in a specific industry context.
- Industry cloud platforms can create value for enterprises by bringing traditionally separately purchased solutions together in a composable and modular way. This simplifies the sourcing, implementation and integration process.
- Currently, industry cloud platforms are being initiated and created by various technology providers. In addition, we see some enterprises considering creating — often in collaboration with a technology provider — a dedicated industry cloud platform as the basis for a more autonomous industry ecosystem.
- Enterprises can gain business value from industry clouds through shared best practices; vertically specialized go-to-market and implementation teams; compliance of the infrastructure platform with industry-specific regulations.
- Value can also be gained through analytical capabilities to integrally mine the data from existing and new applications; industry-specific add-on functionality in front- and back-office enterprise applications; combined with collections of composable building blocks available from industry cloud marketplaces.
- Providers are on a pathway to creating whole-product offerings that cater directly to the established needs of vertical industry enterprises.

Obstacles

- Industry clouds are at risk of following the same path as classic government and community clouds where providers created difficult to support or slightly outdated copies of the original cloud with specific functionality.
- Industry cloud platforms can be overwhelming in terms of the wide breadth of functionality they potentially cover. Customers and providers must therefore be disciplined and not burn precious resources on fixing/replacing things that are not broken.
- Implementing an industry cloud platform must be approached as adding an exoskeleton, bringing new and improved capabilities rather than a vital organ transplant, replacing or repairing functionality that was already present.
- To reach their full potential, industry clouds will need to evolve into something best described as ecosystem clouds. Enterprises can leverage these ecosystems by participating in shared (business) processes, such as procurement, distribution, payment procession, and maybe even R&D and innovation.

User Recommendations

- Target ICPs to complement the existing application portfolio like an exoskeleton by introducing new capabilities that add significant value, rather than as full-scale replacements of largely already existing functionality with more up-to-date technology.
- Start building composability skills by engaging business technologists and fusion teams to create enterprisewide understanding and support for the ICP journey.
- Formulate rules for when to deploy ICP capabilities as a productive platform for optimization and modernization by improving existing processes, and when to actively recompose them for more differentiating transformation and innovation initiatives.

Sample Vendors

Amazon Web Services (AWS); Google; IBM; Infor; Microsoft; Oracle; Salesforce

Gartner Recommended Reading

Top Strategic Technology Trends for 2023: Industry Cloud Platforms

Presentation: Industry Cloud Platform Adoption by Vertical Industry

Analyzing Industry Cloud Offerings From CIPS Providers

Providers of Cloud Managed Services: Use Composable Industry Platforms to Productize Your Offerings

Changes and Emerging Needs Product Managers Must Address in the CIPS Market

At the Peak

Business-Driven Cloud Strategy

Analysis By: David Smith, Lydia Leong

Benefit Rating: Transformational

Market Penetration: 20% to 50% of target audience

Maturity: Early mainstream

Definition:

A cloud strategy is a concise viewpoint on the role of cloud computing in an organization. A true cloud strategy is business-driven, focused on “what” and “why” issues and alignment to business goals. It is different from other efforts, often also referred to as cloud strategy, that in many cases would be better described as cloud adoption or migration plans.

Why This Is Important

Every business needs a cloud strategy, regardless of where it is in its cloud journey. Organizations lacking a cloud strategy don’t achieve as much from cloud computing as those with one. Devising a business-driven cloud strategy leads to secondary questions on product choices, principles and business value prioritization. Effective business-driven cloud strategies align these questions with other strategies and provide a common view of “cloud,” and a launching point for more use-case scenarios.

Business Impact

Many top questions in cloud computing revolve around cloud strategy. But many cloud discussions are typically focused only on technology, not on how cloud capabilities should be assessed against business outcome targets. Business-driven cloud strategies must be aligned with other strategies (e.g., data center, security and architecture). Your organization’s business-driven cloud strategy should enable cloud adoption/migration (e.g., tactical implementation plans), leading to specific actions.

Drivers

- Many organizations lack a cloud strategy. They don’t achieve as much from their use of cloud computing as organizations with a cloud strategy.

- Devising a cloud strategy leads to many secondary questions about principles and prioritization. Effective cloud strategies align these questions with other strategies.
- Cloud computing affects every aspect of organizations' IT and business environments, requiring coordination across multiple domains to ensure successful and safe cloud exploitation.
- Organizations with documented and repeatable processes for evaluating the value and risks using cloud services outperform those with ad hoc processes.
- Organizations that embrace cloud computing best practices across all IT functions reduce the risk and increase the value of exploiting cloud computing.
- Organizations that do not have a high-level cloud computing strategy, driven by their business and IT strategy, will significantly increase their risk of failure and wasted investment.

Obstacles

- Challenges in aligning a business-driven cloud strategy with existing technology and business strategies, such as those for security, data center, and development and architecture
- Cloud myths or common mistakes such as believing a high-level, aspirational statement like "cloud first" is a strategy, or considering implementation plans to be the strategy, or that cloud migration automatically saves money
- A lack of understanding or alignment across the organization regarding the intended business strategy and potential use of technology
- On the organizational front, a lack of senior-level buy-in or backing for the strategy, and failure to assemble a cloud council of interested parties across IT and the business
- Existing cloud adoption/migration plans often precluding plans for a business-driven cloud strategy

User Recommendations

- Maximize the benefits from your use of cloud computing by creating a business-driven cloud strategy. Make it a living document that provides a concise view on the role of cloud computing in your organization and its contribution to business value results.

- Align your cloud strategy with other strategic plans (e.g., those for data center, security and architecture, and business).
- Plan for your cloud strategy to be the launching point for cloud activities like architecture, assessment, migration and operations. Keep those activities in mind when devising your cloud strategy.
- Safeguard your organization from potential problems if you withdraw from the cloud by including an exit strategy describing the dependencies and choices involved in cloud computing. Focus your overall and exit strategies on answering “what” and “why” questions. Cover answers to “how” questions in a more detailed cloud adoption and/or exit plan. Account for potential business impacts.

Gartner Recommended Reading

The Cloud Strategy Cookbook, 2023

Infographic: The Cloud Journey

Approaches to Avoid Common Cloud Strategy Pitfalls

Quick Answer: What’s the Difference Between a Cloud Strategy and a Cloud Adoption Plan?

Cloud Resilience

Analysis By: Lydia Leong

Benefit Rating: High

Market Penetration: 1% to 5% of target audience

Maturity: Emerging

Definition:

Cloud resilience is the application of resilience principles and practices to the delivery and consumption of cloud services. Resilient systems limit the impact of, and recover quickly from, failure. The shared responsibility model of cloud computing applies to cloud resilience responsibilities.

Why This Is Important

Organizations increasingly wonder if their cloud providers are adequately resilient and what customers should do to mitigate associated risks. However, many cloud customers do not choose more resilient, but higher-cost options when purchasing SaaS. They also do not pay adequate attention to the resilient design, implementation or operations of their applications deployed in infrastructure as a service (IaaS) and platform as a service (PaaS).

Business Impact

Cloud outages may create application downtime, which may have a negative impact on application users, business processes or an organization's customers. Lengthy cloud failures may pose a business continuity risk for inadequately prepared cloud customers. Furthermore, many organizations may be impacted by a cloud outage without being a direct customer of that provider, since cloud services are now often used in many solutions delivered to businesses and individuals.

Drivers

- As public cloud providers grow ever larger and more strategic, they become a systemic risk. Thus, governments and industry regulators are growing more concerned about cloud resilience, and may take actions to address perceived risks.
- Cloud providers have become an integral part of driving the global economy, as well as a critical dependency for ordinary people living their daily lives. Cloud-based identity — including social identity from companies such as Apple, Meta and Google — is used for authentication into many consumer and business applications, websites, and other digital products and services.
- As organizations increase their dependence on cloud services for mission-critical business solutions, they become more concerned about the potential impact of major cloud outages.
- Most organizations are adopting a strategic cloud provider that is the focus of their cloud adoption efforts, thus concentrating more of their application portfolios in a single provider. This increases concerns about availability and business continuity risks, and thus cloud resilience.
- Digital products and services often need continuous availability. Therefore, traditional notions of acceptable maintenance downtime, occasional outages and lengthy manually performed disaster recovery efforts are unacceptable for these solutions.
- Cloud providers deliver public assurances of their resilience, but most offer little transparency into how they accomplish that resiliency. Therefore, organizations may place unwarranted trust in marketing promises or service-level agreements.
- It is easier to blame cloud providers for failures than to look inward. Many organizations that purchase cloud IaaS and PaaS do not understand how cloud failures differ from on-premises data center downtime, and hence do not understand how to properly implement cloud resilience.

Obstacles

- Many organizations question whether it is necessary to spend more money on cloud resilience, given that most cloud providers are quite reliable. Customers often do not understand the resilience-related differences between providers.

- Cloud resilience depends on the appropriate physical and logical design of solutions, implementation quality, the quality of change management, and the quality of the design and execution of both proactive and reactive operations processes.
- A smaller cloud provider may be less able to invest in efforts to become more resilient, and is therefore riskier. However, it may be the sole source of a niche solution, forcing customers to consider their risk appetites.
- Cloud services make it easier to rapidly deliver and update applications. However, application release velocity and operational safety are often in conflict, and cloud self-service capabilities can make it harder to govern application resilience.

User Recommendations

- Adopt a holistic view of potential avenues of system failure, addressed both systemically and for individual applications. Many organizations mistakenly focus only on potential infrastructure problems.
- Create tiered architectural standards based on cloud provider resilience capabilities and limitations. Map application criticality classifications to these tiers. Enforce those standards when sourcing or building cloud solutions.
- Apply site reliability engineering (SRE) principles to improve the resilience of fast-changing applications. You do not need to implement an SRE team to usefully adopt SRE principles.
- Use chaos engineering to test complex cloud solutions. The market-leading cloud IaaS and PaaS offer chaos engineering services that simulate a variety of cloud service failures.
- Optimize your risk assessment and triage efforts when evaluating SaaS providers. Ensure SaaS application owners explicitly acknowledge and accept the risks of such providers.

Gartner Recommended Reading

Designing Availability and Resilience for Applications in Public Cloud IaaS and PaaS

9 Principles for Improving Cloud Resilience

Quick Answer: How Should Executive Leaders Plan for Cloud Outages?

How to Establish Effective SaaS Governance

How to Manage Concentration Risk in Public Cloud Services

Consumption-Based Model

Analysis By: Jeff Vogel, Philip Dawson

Benefit Rating: High

Market Penetration: 5% to 20% of target audience

Maturity: Early mainstream

Definition:

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

Why This Is Important

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

Business Impact

A consumption-based sourcing model and services strategy will:

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

Drivers

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

- The need for a more flexible cloud-like operating model for on-premises infrastructure.
- The massive growth of enterprise data that makes capacity planning difficult and upfront purchasing for three to five years of growth expensive and impractical.
- Prolonged procurement lead time increases due to persistent supply shortages.
- The need for an application-aware services delivery model.
- The preference for operating expenditure (opex) to capex with cloud-like benefits, while avoiding risks or costs associated with moving mission-critical workloads to the public cloud.
- The need for a more cost-effective, flexible and efficient sourcing strategy that aligns with business demands.
- The need to augment IT budget priorities to redirect investments to develop cloud-native platform skills that support business growth initiatives.
- The shift from exiting the life cycle management of infrastructure assets in the long term to freeing up IT resources.

Obstacles

A consumption-based sourcing model may:

- Be more expensive than capex financing.
- Be organizationally challenging to implement.
- Be unsuitable for IT operations that have a more stable and predictable growth and variability in forecast demand or lean toward sweating assets.
- Require minimum-usage commitment levels that can't be justified regardless of what is actually consumed.
- Require three- to five-year contracts with vendor-centric services.

- Lack the skills or culture alignment to shift from sourcing products to platform SLA services.
- Not take into account long-term supply chain price fluctuations during the contract period, when declining hardware costs or supply constraints are considered.
- Conflict with financial asset depreciation and amortization schedules or corporate balance sheet objectives.
- Conflict with established industry accounting standards and operational norms.
- Software licensing terms may be incompatible with the use of consumption based hardware.

User Recommendations

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

Sample Vendors

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

Gartner Recommended Reading

Market Guide for Consumption-Based Models for Data Center Infrastructure

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

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

API-Centric SaaS

Analysis By: Yefim Natis, Anne Thomas, Mark O'Neill

Benefit Rating: High

Market Penetration: 1% to 5% of target audience

Maturity: Emerging

Definition:

API-centric SaaS is a cloud application service designed with programmatic request/reply, or event-based, interfaces (APIs) as the primary method of access (instead of the traditional, and now optional, user interfaces). The strategic intent for API-centric SaaS is to contribute a set of business software components, packaged for use by advanced or business technologists as building blocks for composing custom application processes and services for end users.

Why This Is Important

API-centric SaaS serves as a foundation of creative innovation by business and software companies. It exposes modular business software as reusable building blocks in custom application development. Organizations create composite application processes and experiences that are more varied than relying entirely on in-house resources, and better targeted than relying entirely on the SaaS provider. Greater creativity in application engineering translates to business empowerment for faster, safer and more efficient innovation.

Business Impact

Business organizations equipped to use API-centric SaaS create new application experiences for their employees and customers, through composition of the new and prebuilt business software components, some sourced from multiple applications. They gain access to more impactful innovation to be more adaptive to the changing needs of the users and to better respond to competitive opportunities. Procurement of application services gradually becomes better matched to the consumed value.

Drivers

- Modern application design relies on cross-application integration and composition, compelling application vendors to deliver their business functionality, optionally or primarily, equipped for programmatic access.
- The growing popularity of the “headless” SaaS architecture in digital commerce provides the acceleration of creative innovation when using the modular API-first application design, changing the users’ assessment criteria for SaaS to favor the support of composability.
- The technology and skills for integration, including management of APIs, are widespread, promoting increasingly advanced use of programmatic interfaces to business applications.
- The demands for advanced customization of application experience in business organizations have evolved to the point that SaaS vendors must allow rearrangement of their business functionality by their customers. API-centric SaaS serves that purpose.
- Many older applications are increasingly accessed via APIs to include them in the modernization and innovation of organizations’ IT. This prepares organizations’ skills and technologies to include API-centric SaaS capabilities into their software engineering practices.
- Business application design has become significantly partitioned into the back-end functionality with its APIs and the front-end multiexperience, each side using different tools and design expertise. Some business-oriented application vendors find it convenient to concentrate on the back-end data and business logic, and leave the finalized user experience to separate teams, including the customer’s own developers.

Obstacles

- API-centric SaaS is a relatively new phenomenon. Both SaaS vendors and business developers may lack the required skills and tools.
- The best practices for pricing and procurement of API-centric SaaS are not well-developed, delaying adoption or increasing its costs. The pricing of some occasional use of APIs is common (and expensive) and does not match the use practices of API-first application products.
- Using multisourced API-centric components for assembling new application processes and experiences requires some integration work that may not be supported in selected composition tools. This requires advanced software engineering skills and delays adoption of API-centric SaaS by mainstream organizations.
- Reduced or absent user interfaces packaged with an API-centric SaaS assume and require that the customer implement their own differentiated application and user experience. What is a welcome opportunity for innovation for some can be a burden to others, delaying adoption of API-centric SaaS.

User Recommendations

- Build the tools and skills of API management that recognize the added requirements to govern access to imported third-party APIs.
- Give preference to SaaS offerings that expose and price more of their business functionality as APIs and/or event streams.
- Plan for the increasing use of composition and integration of API-centric business software in the design and delivery of application services, processes and experiences.
- Ensure clean API-based separation of the back-end business logic and the front-end user experience in most enterprise applications, to maximize the long-term benefits of adopted API-centric SaaS.
- Give preference to application platform offerings that are well-equipped for managed access to external APIs and event sources.
- Practice use and governance of APIs and event streams in preparation for greater adoption of API-centric SaaS.
- Watch for opportunities to experiment with a new business model by offering some of your business functionality packaged as priced API products or services.

Sample Vendors

Algolia; Alloy; Clearbit; Cloudinary; Lob; MessageBird; Plaid; Strapi; Stripe; Twilio

Gartner Recommended Reading

Accelerate Digital Transformation With an API-Centric Architecture for Enterprise Applications

How to Successfully Implement API-First Integration

Partner With Product Managers to Ensure the Success of API-Based Products

Banking Product Leader Insight: Think Beyond APIs to Address Composability

Quick Answer: What GMs Need to Know About the Composable Future of Applications

Platform Engineering

Analysis By: Bill Blosen, Paul Delory

Benefit Rating: Transformational

Market Penetration: 20% to 50% of target audience

Maturity: Adolescent

Definition:

Platform engineering is the discipline of building and operating self-service developer platforms for software development and delivery. A platform is a layer of tools, automations and information maintained as products by a dedicated platform team, designed to support software developers or other engineers by abstracting underlying complexity. The goal of platform engineering is to optimize the developer experience and accelerate delivery of customer value.

Why This Is Important

Digital enterprises need to respond quickly to customer and internal demands; therefore, flexible, complex distributed software architectures have become popular. Software product teams struggle to focus on features due to this complexity, which results in poor developer experience. Platform engineering provides a self-service, curated set of tools, automations and information driven by developer priorities to accelerate value delivery in line with internal stakeholders, such as security and architecture.

Business Impact

Platform engineering empowers application teams to deliver software value faster. It removes the burden of underlying infrastructure construction and maintenance and increases teams' capacity to dedicate time to customer value and learning. It makes compliance and controls more consistent and simplifies the chaotic explosion of tools used to deliver software. Platform engineering also improves the developer experience, thus reducing employee frustration and attrition.

Drivers

- Scale: As more teams embrace modern software development practices and patterns, economies of scale are created, whereby there is enough value to justify creating a platform capability shared by multiple teams.
- Cognitive load: Adoption of modern, distributed architectural patterns and software delivery practices means that the process of getting software into production involves more tools, subsystems and moving parts than ever before. This places a burden on product teams to build a delivery system in addition to the actual software they are trying to produce.
- Need for increased speed and agility: The speed and agility of software delivery is critical to CIOs. As a result, software organizations are pursuing DevOps which is a tighter collaboration of infrastructure and operations (I&O) and development teams to drive shorter development cycles, faster delivery and increased deployment frequency. This will enable organizations to respond immediately to market changes, handle workload failures better and tap into new market opportunities. Platform engineering can drive this type of cross-team collaboration.
- Emerging platform construction tools: Many organizations have built their own platforms, but to date, these platforms have been homegrown, individual efforts tailored to the unique circumstances of the organizations that build them. Platforms generally have not been transferable to other companies or sometimes even to other teams within the same company. However, a new generation of platform-building tools is emerging to change that.
- Infrastructure modernization: During digital modernization, some forward-looking I&O teams embrace a new platform engineering role as a way to deliver more value, increasing their relevance to the business.

Obstacles

- Lack of skills: Platform engineering requires solid skills in software engineering, product management and modern infrastructure, all of which are in short supply.
- Platform engineering is easily misunderstood: Traditional models of mandated platforms with limited regard for developer experience can easily be relabeled and thus not achieve the true benefits of platform engineering.
- Outdated management/governance models: Many organizations still use request-based provisioning models. Those need to give way to a self-service, declarative model, with the primary focus being the effectiveness of the end users developing and operating solutions using the platform.
- Internal politics: There are many intraorganizational fights that could derail platform engineering. Product teams may resist giving up control of their customized toolchains. There might also be no appetite to improve the developer experience. Enterprises may also refuse to fund platform engineering without a clear ROI.

User Recommendations

- Start small with cloud-native workloads: Begin platform-building efforts with thinnest viable platforms for the infrastructure underneath cloud-native applications such as containers and Kubernetes.
- Embed security into platforms: Enable shift-left security within DevOps pipeline platforms, which will provide a compelling paved road to engineers.
- Don’t expect to buy a complete platform: Any commercially available tool is unlikely to provide the entirety of the platform you need. Thus, the job of the platform team is to integrate the components necessary for the platform to meet your needs.
- Implement a developer portal as part of your platform: An internal developer portal (IDP) serves as the user interface that enables self-service discovery and access to internal developer platform capabilities. Consider Backstage open-source or other commercial tools. Note: “IDP” has multiple meanings in this context, as well as in the industry.

Gartner Recommended Reading

How to Start and Scale Your Platform Engineering Team

Guidance Framework for Implementing Cloud Platform Operations

Adopt Platform Engineering to Improve the Developer Experience

Innovation Insight for Internal Developer Portals

Cloud and Generative AI

Analysis By: Sid Nag

Benefit Rating: Transformational

Market Penetration: 1% to 5% of target audience

Maturity: Adolescent

Definition:

Generative AI technologies can generate new derived versions of content, strategies, designs and methods by learning from large repositories of original source content. GenAI has profound business impacts, including on content discovery, creation, authenticity and regulations; automation of human work; and customer and employee experiences. Public cloud offers scalable infrastructure and is also suited to addressing issues related to cost, sovereignty, sustainability, security and privacy.

Why This Is Important

Generative AI (GenAI) exploration is accelerating and organizations are aggressively experimenting with the innovation. Public cloud providers are driving the GenAI narrative. Cloud technology is best-suited for the delivery of generative AI-enabled applications at scale. However, nontechnology aspects need to be addressed to operationalize GenAI by organizations in order to adopt this innovation for their business needs. These include economics, cost, sovereignty, sustainability, security, privacy, regulation, copyright and ethics issues.

Business Impact

Most technology products and services will incorporate GenAI capabilities in the next 12 months, introducing conversational ways of creating and communicating with technologies, leading to their democratization. GenAI will progress rapidly in industry verticals, scientific discovery and technology commercialization. Sadly, it will also become a security and societal threat when used for nefarious purposes. Responsible AI, trust and security will be necessary for safe exploitation of GenAI.

Drivers

- The hype around GenAI is accelerating. OpenAI's ChatGPT is currently the most hyped application. It is built on top of foundation models (FMs), also called "generative pretrained transformers."
- New FMs and their new versions, sizes and capabilities are rapidly coming to market. FMs keep making an impact on language, images, molecular design and computer code generation.
- GenAI, by its nature, requires a highly scalable platform. The cloud is best-suited to GenAI technology given the scale of compute, storage and networking and other infrastructure demands needed to support large language models (LLMs) and FMs.
- Organizations adopting GenAI will look to the cloud to drive price-to-performance ratios via innovation on silicon in areas such as GPU and DPUs. For example, some hyperscalers (e.g., AWS and Google) have built custom chips for AI workloads.
- Machine learning (ML) and natural language processing platforms available from cloud providers are adding GenAI capabilities for the reusability of generative models, making them accessible to AI teams.
- Industry applications of GenAI are growing, and we will likely see an explosion of domain-specific and use case-specific FMs. In healthcare, GenAI creates medical images that depict disease development. In consumer goods, it generates catalogs. In e-commerce, it helps customers "try on" makeup and outfits. In manufacturing, quality inspection uses synthetic data. In semiconductors, GenAI accelerates chip design. Life sciences companies apply it to speed up drug development. Building vertical GenAI applications on top of industry clouds is ideal.
- GenAI is disrupting software coding. Combined with development automation techniques served up by cloud platform as a service, it can result in automation of up to 70% of the programmers' work. Examples of ML coding assistants are Microsoft 365 Copilot and Amazon CodeWhisperer.

Obstacles

- Democratization of GenAI uncovers new ethical and societal concerns such as copyright and privacy issues.
- Without a “walled garden” approach to building LLMs required for enterprises using virtual private clouds, GenAI will remain the playing ground for consumers and hobbyists.
- Absent country-specific regulatory compliance support as provided by sovereign clouds today, GenAI will face hurdles in certain regions.
- Low awareness of GenAI among security professionals can cause incidents that could undermine its adoption unless cloud-based security platforms are used in its deployment.
- Compute resources for training foundation models are heavy and costly. Most enterprises powering their own data centers will be challenged to train FMs unless highly scalable cloud computing IaaS capabilities are deployed.
- Sustainability and cost concerns about high energy consumption for training generative models are rising. The cloud is best-suited to address these concerns.
- Using generic GPTs requires prompt engineering/fine-tuning and access to enterprise knowledge management sources to give the models context to deliver relevant insights.

User Recommendations

- Identify initial use cases where you can improve your solutions with generative AI by working with your cloud providers, including embedding Gen AI in SaaS and other enterprise applications as well as managed services and custom solutions.
- Leverage the cloud to address scale while attaining cost efficiency delivered via silicon innovation in areas such as GPUs, DPUs and TPUs by your cloud provider.
- Address nontechnology issues such as compliance, privacy, sovereignty and sustainability by building on top of existing cloud offerings such as sovereign clouds, industry clouds and private and distributed clouds.
- Build a solid data foundation by taking advantage of scalable storage offerings from public cloud providers.
- Pilot ML-powered coding assistants using cloud middleware capabilities, with an eye toward fast rollouts, to maximize developer productivity.
- Mitigate generative AI risks by working with legal, security and fraud experts. Use cloud-based techniques such as threat detection, vulnerability management and container security while adhering to your data security guidelines.
- Optimize the cost and efficiency of AI solutions by leveraging the pay-as-you-go cloud principle.

Sample Vendors

Adobe; Amazon; Anthropic; Google; Hugging Face; IBM; Microsoft; OpenAI; OpenShift

Gartner Recommended Reading

Innovation Insight for Generative AI

Emerging Technologies: Cloud and AI Are Top Spending Priorities for Tech Providers

Emerging Tech Roundup: ChatGPT Hype Fuels Urgency for Advancing Conversational AI and Generative AI

Emerging Tech: Generative AI Needs Focus on Accuracy and Veracity to Ensure Widespread B2B Adoption

ChatGPT Research Highlights

SMP

Analysis By: Dan Wilson, Jaswant Kalay, Tom Cipolla, Sid Nag

Benefit Rating: Moderate

Market Penetration: 5% to 20% of target audience

Maturity: Adolescent

Definition:

SaaS management platforms (SMP) help IT discover, manage, automate, operate, optimize and govern organizationwide SaaS use from a centralized console. SMPs also enhance protection of identities and data while using SaaS and SMP's enable SaaS operations — which include the capabilities supporting IT operations, PlatformOps, SecOps and site reliability engineering (SRE). Though SMP vendors focus on operational management or optimization of SaaS, a few have emerged to address both.

Why This Is Important

As SaaS adoption accelerates and managing spend becomes difficult, IT leaders are challenged with discovering and supporting SaaS in accordance with company, market or geographic policy and regulations. The increase in cyberattacks focuses attention on protecting identity and data in SaaS. These trends continue to attract new SMP market entrants, investment and M&A. The SMP market remains fragmented and difficult to navigate, and hyperscale cloud providers approaches differ substantially.

Business Impact

IT leaders can leverage SMP to:

- Improve Saas visibility and manageability
- Reduce or optimize costs
- Improve management of SaaS contracts and renewals, and optimize costs
- Reduce business-acquired SaaS by offering app-store experiences for employees
- Streamline new SaaS onboarding
- Reduce IT overhead with automation
- Improve employee on/offboarding workflows

- Promote collaboration between teams in the SaaS life cycle

Low Gartner client interest keeps hype parked at the peak. Plateau time has been extended, and market penetration and maturity are unchanged from 2022.

Drivers

- SaaS spend continues to grow by 15-20% annually, as organizations maintain an average of over 125 different SaaS applications totaling $1,040 per employee annually.
- IT typically is aware of only a third of those due to decentralized ownership and sourcing.
- SMPs also report that less than half of provisioned subscription-based licenses are regularly used by employees.
- IT teams responsible for discovering, managing, automating, optimizing, protecting and governing SaaS struggle to effectively do this through native SaaS administrator consoles.
- Harmonizing SaaS configurations and employee on/offboarding are also common pain points.
- SMP adoption is higher in small to midsize organizations, as centralized responsibility for SaaS is more common.
- Clients are also struggling to choose between SMP, SaaS security and SAM tools. All three offer SaaS discovery, protection and some optimization capabilities — however, SMPs can do more.
- A lack of common APIs or controls means that SMPs have varying levels of integration and capability to manage and automate SaaS applications.
- Utility continues to improve as new entrants to the SMP and adjacent markets promote unique new capabilities.
- I&O leaders don't have the available tools and capabilities for observability, application monitoring, cost, license or configuration management, and security visibility for this new world of applications. There is no integrated platform approach to these functionalities and tools.
- Broader SaaS operations capabilities to support IT operations, PlatformOps, SecOps and SRE services, as well as continuous integration/continuous delivery (CI/CD) pipelines and other agile approaches.

Obstacles

- Decentralized or shared responsibility within organizations complicates buying decisions.
- Many organizations underestimate SaaS sprawl and do not fully understand how an SMP can help.
- Low maturity organizations generally see SMP as too advanced and have more basic priorities.
- Costs associated with assessing, selecting, implementing and staffing resources to utilize SMP are rarely allocated in budgets.
- Varying breadth and depth of SaaS coverage and integrations. SaaS-heavy organizations often find that SMPs do not cover all of their applications and licensing models.
- Capability overlaps with SAM and SaaS security tools.
- Concerns about the addition of another management tool.
- The SaaS management market is highly fragmented and characterized by wide variability and only partial overlap between tools.
- Gartner client interest remains low compared to other digital workplace tool conversations.
- DevOps, apps, digital workplace and I&O teams generally operate in silos.
- Managing configurations for dozens to hundreds of SaaS apps using their separate admin consoles is untenable.
- Confusion and the business acquiring their own applications, due to a decentralized approach with no clear owner. This results in uncontrolled cost, identity and data security exposure, missing observability and service management processes.

User Recommendations

IT leaders responsible for managing SaaS should:

- Implement an overall SaaS operations strategy and execution plan.
- Avoid overspending by focusing first on discovery.

- Build a business case to fund the SMP by utilizing optimization capability to reduce unnecessary spend on unused and underutilized licenses, and to consolidate similar apps.
- Uncover unsanctioned SaaS by using an SMP with strong discovery capabilities through desktop agents, browser extensions and deep integration with security and finance tools.
- Minimize risk by finding and addressing SaaS that is not integrated with identity and SSO solutions, and documenting discovered SaaS in enterprise architecture tools or CMDBs.
- Choose an SMP that best fits your requirements by reviewing integrations with critical apps to understand if the SMP offers read and write functionality, or is limited to pulling reports.
- Bring together disparate IT teams and processes.

Sample Vendors

Beamy; BetterCloud; LeanIX; Productiv; SailPoint; Snow Software; Torii; Trelica; Zluri; Zylo

Gartner Recommended Reading

Market Guide for SaaS Management Platforms

Market Guide for Software Asset Management Tools

Infographic: Why Are You Wasting Your SaaS Expenditure?

How to Establish Effective SaaS Governance

Cloud-Native

Analysis By: David Smith, Michael Warrilow

Benefit Rating: High

Market Penetration: 5% to 20% of target audience

Maturity: Adolescent

Definition:

Cloud-native refers to something created to optimally leverage or implement cloud characteristics. Those cloud characteristics are part of the original definition of cloud computing, and include capabilities delivered as a service. Cloud computing characteristics also include being scalable and elastic, shared, metered by use, service based, and ubiquitous by means of internet technologies.

Why This Is Important

Cloud-native is a popular term. Depending on its meaning, it can be described as taking full advantage of the cloud capabilities of a cloud provider, or using approaches pioneered in the cloud to deliver benefits wherever needed, via specific technologies such as containers. Cloud-native is not one thing, and there is a battle of ideas.

Business Impact

Cloud-native is a popular, hyped concept that aspires to attain and maximize the benefits of cloud computing; however, the realization of those benefits varies. For example, if a traditional, noncloud application is migrated to the cloud through a lift-and-shift approach, the application is unlikely to fully leverage cloud characteristics and deliver the maximum benefits. An application rewritten to take advantage of cloud capabilities is more likely to deliver the expected cloud outcomes.

Drivers

- The primary driver for cloud-native is the desire to “get the most out of the cloud.” The cloud itself means different things to different constituencies, so it’s not surprising that cloud-native means different things. What drives people to one or another of these approaches varies.
- Cloud-native can optimally leverage cloud technologies and benefits. The two most common meanings in use are contradictory. CSP-native is all about using native features and, therefore, locking yourself into a provider. Container-native focuses on containers, and may evolve into other technologies. This doesn’t guarantee portability, but is directionally consistent with the goal.
- There are multiple aspects to cloud-native, ranging from design to architectural to operational practices. Examples include LIFESPAR and the Twelve-Factor App (i.e., cloud-native application design) and DevOps (cloud-native operations).
- Cloud-native can be viewed on a continuum. It’s not a question of whether something is cloud-native or not; it’s the degree to which it is. The more it aligns with cloud characteristics, the more cloud-native it is.

Obstacles

- Cloud-native is confusing due to its many interpretations. It's especially challenging with respect to hype, because confusion amplifies hype. The biggest obstacle is getting beyond the confusion to focus on desired outcomes.
- It is essential to be realistic about the portability that can be achieved and the cost. Otherwise, these features may not be used "with your eyes open," and you may not be aware you are doing so.
- In cloud strategy efforts, principles are the most important component. Cloud-native and multicloud are often stated as principles in a cloud strategy. These principles can contradict each other, and require further explanation.
- Use of the term "cloud-native" requires clarification of which meaning is being used. This is a function of the hype surrounding cloud-native. Being clear about goals is key to optimally leveraging cloud-native. Assuming that containerizing an application will inherently make it cloud-native is an obstacle. We call this "container-native."

User Recommendations

- Focus on the outcomes you want from using the cloud, rather than focusing purely on the definition of cloud-native. The more your use cases align with core cloud characteristics, the more likely you are to realize the benefits of using the cloud.
- Assess vendor claims about their cloud-native capabilities with skepticism. Vendors use the term "cloud-native" to promote their offerings, regardless of how cloud-native their offerings are.
- Ensure that the supporting tools, processes and operations support cloud characteristics when building or acquiring cloud-native applications or services. The value of cloud-native applications can be subverted when the approaches of the supporting elements are not cloud-native.
- Embrace services designed to bring you closer to cloud-native outcomes. These can include containers, microservices architecture, serverless design, functions and many platform-as-a-service (PaaS) services. However, using these technologies should be a means, not a goal.

Gartner Recommended Reading

The Cloud Strategy Cookbook, 2023

Infographic: Cloud-Native and Multicloud — Buzzwords or Key Principles in Your Cloud Strategy

A CTO's Guide to Cloud-Native: Answering the Top 10 FAQs

Define and Understand New Cloud Terms to Succeed in the New Cloud Era

Intercloud Data Management

Analysis By: Adam Ronthal, Donald Feinberg

Benefit Rating: High

Market Penetration: 1% to 5% of target audience

Maturity: Emerging

Definition:

Intercloud data management is the process of actively managing data in multiple cloud providers as part of a cohesive application and data management strategy. It builds on the foundation of multicloud capabilities, but adds the ability to access and use data across clouds in an operational context. It can be done at the cloud object store (COS), database management system (DBMS) or application tiers.

Why This Is Important

The vast majority of organizations using the public cloud are storing data on more than one cloud. Today, most of that data remains siloed — accessed and managed in the context of a single cloud environment. As data's center of gravity shifts to the cloud (or multiple clouds), data and analytics leaders will seek out means to unite that data in a logical and cohesive consumption tier to improve optimization, efficiency, flexibility and insight.

Business Impact

The ability to access data — regardless of where it is located — is a potentially transformational capability that will serve to break down barriers to access for end users and applications. Intercloud data management is cloud-agnostic and will allow enterprises to access their data in any cloud at any time, and by any means. It will enable globally distributed applications that span cloud providers and geographies, providing resilience and avoiding lock-in to any single cloud service provider.

Drivers

Though actual market penetration remains minimal, several drivers affect adoption, including:

- **Regulatory requirements:** Some industries are starting to mandate the use of multiple clouds for resilience and availability reasons, and some countries have strict data sovereignty laws.
- **Global applications:** Applications that operate on a global basis may require the use of multiple clouds to meet latency and performance requirements.
- **Distributed teams:** Organizations using more than one cloud may need to work with data in more than one cloud and provide continuity in their multicloud environments.
- **Integration:** Distributed data must be integrated (in storage and/or logically) to achieve maximum business value.

Obstacles

- Enterprises seeking to adopt intercloud data management approaches tend to be large, global enterprises with specific application requirements; most small to midsize organizations do not have these requirements. Thus, market penetration is minimal.
- Intercloud data management requires technologies built for distributed data management. Some established cloud service providers — most notably Google Cloud Platform and Microsoft Azure — are entering the space with recent offerings. However, most of the leading data management technologies are not designed to manage data across multiple clouds.
- Intercloud data management requires management of traffic between the clouds. Multiple methods exist, each with different cost, performance and security considerations.
- Any intercloud data management application will also be subject to the laws of physics. Practitioners will need to be aware of both performance implications and trade-offs in consistency and availability, and ensure that their applications are both aware of and designed with these trade-offs in mind.
- As with any data, intercloud data must be governed and integrated in storage and/or logically to achieve maximum business value.

User Recommendations

Weigh trade-offs in optimization and flexibility when making design decisions, as intercloud data management can occur at three different levels:

- **Object store:** Distribute data at the COS layer and replicate between clouds when flexibility and diversity of choice are required. Data science teams, for example, can choose whatever optimization layer they would like for last-mile delivery, as all can read and write to the local COS.
- **DBMS:** Select a DBMS that distributes and manages data in a geodistributed cluster when global operational efficiency and resilience are required. Nodes can reside in multiple clouds and/or on-premises. This approach can not only support local, low-latency read/write applications with global read capabilities, but also enforce data sovereignty requirements.
- **Application:** Use the application tier when data is already in multiple clouds and you require a means of bringing it together. This approach essentially defers data integration to the point of consumption.

Sample Vendors

Cockroach Labs; Couchbase; DataStax; Google; MongoDB; Oracle; PingCAP; Snowflake; WANdisco; Yugabyte

Gartner Recommended Reading

How to Plan for Optimal Multicloud and Intercloud Data Management

What Are the Key Factors to Consider When Choosing a Cloud Data Management Architecture?

6 Best Practices to Create a Cloud Management Services Offering in the World of Multicloud and Hybrid Cloud

Building an Edge Computing Strategy

Distributed Cloud: Does the Hype Live Up to Reality?

Multicloud

Analysis By: David Smith

Benefit Rating: High

Market Penetration: 20% to 50% of target audience

Maturity: Early mainstream

Definition:

Multicloud computing is the use of multiple public cloud providers to provide the same general class of IT solution, workload, application or use case. It is much more common in infrastructure as a service (IaaS) and converged IaaS/platform as a service (PaaS) scenarios than SaaS. While multi-SaaS environments are possible, these would typically be stovepiped situations.

Why This Is Important

Multicloud has the potential to lower the risk of cloud provider lock-in, can provide best-of-breed capabilities for specific use cases and can provide service resilience and migration opportunities, in addition to the core cloud benefits of agility, scalability and elasticity. It also may be used to obtain public cloud services in different geographic locations for global companies.

Business Impact

Multicloud provides agility and can also provide a basis to lower cloud provider lock-in and increase workload migration opportunities. However, multicloud can also create additional complexity and, therefore, cost increases. Also, many organizations find that a multicloud environment is unavoidable for most.

Drivers

- Many organizations end up in a multicloud environment through acquisitions and mergers. Unintended multicloud environments can be rationalized into a purposeful multicloud strategy.
- Enterprises typically start with one provider and focus first on costs, but, over time, become concerned about lock-in. Thus, the first use of multicloud is often based on procurement issues to encourage competition, or as result of a merger and acquisition.
- As multiple cloud providers are in use, the need to manage and govern those services becomes important. Eventually, some enterprises adopt multicloud architectures. This approach relies on architectural principles and portability solutions, and can potentially enable even cloudbursting and other dynamic placement efforts.
- Many deliberate multicloud strategies are designed to take advantage of differentiated capabilities within the same general class (e.g., IaaS) from multiple cloud providers while applications run in a single cloud provider stack. Some applications may have a multicloud architecture themselves.
- The hype around multicloud is driving adoption, as providers often use this industry buzz term to justify why their offerings should be considered when another cloud service already exists.

Obstacles

- Multicloud is often confused with hybrid cloud. The reality is that multicloud and hybrid cloud often coexist in a multi-hybrid cloud environment that spans multiple public cloud providers, as well as between public and private implementations.
- Multicloud is not a practical solution for improving availability and enhancing disaster recovery or business continuity, as these goals are more effectively achieved in other ways within a provider's ecosystem.
- Multicloud environments are complex and often result in cost increases. Effort and cost are more often required to secure and manage multiple cloud environments. Organizations need to invest in the right skills to manage and deal with more complex integration solutions.

User Recommendations

- Ensure your multicloud strategy is coordinated with your overall cloud strategy. When embracing multicloud approaches, account for the tools, skills, processes and other resources to ensure you will achieve the right outcomes.
- Establish security, management, governance guidelines and standards to manage cloud service sprawl and increasing costs, and develop criteria for deciding placement of services.
- Focus on coordination and strategy across the enterprise to identify the types of services needed to deliver the benefits of a multicloud environment. Be prepared to incur additional expenses on training and skill development across roles, including engineers and operators.
- Do not just shift vendor lock-in to a cloud management platform (CMP) and/or a cloud service brokerage (CSB), even though they may enable governance and optimizations in a multicloud environment.

Gartner Recommended Reading

The Cloud Strategy Cookbook, 2023

A Multicloud Strategy Is Complex and Costly, but Improves Flexibility

A CTO's Guide to Multicloud Computing

Cloud Portability

Analysis By: Lydia Leong, David Smith

Benefit Rating: High

Market Penetration: 1% to 5% of target audience

Maturity: Emerging

Definition:

Cloud portability is the ability for a customer to move a cloud-based application or workload from one cloud provider to another.

Why This Is Important

The ability to migrate between cloud providers (whether infrastructure as a service [IaaS], platform as a service [PaaS] or SaaS) is important for reducing both vendor lock-in and the impact of vendor concentration risk. Cloud portability offers customers more freedom and flexibility to alter workload placement based on evolving business or technical needs. It is sometimes perceived as offering a negotiation advantage when contracting with cloud providers, especially when coupled with the myth of container portability.

Business Impact

When a cloud-based application is not portable, the customer is dependent on a single cloud provider for that application. This exposes the customer to a range of vendor-related risks, and potential concerns about the provider's financial viability, service outages, alignment of the provider to the customer's long-term needs and negotiation leverage. However, achieving cloud portability decreases agility, slows cloud implementations and increases complexity and costs.

Drivers

- Flexibility: Organizations seek cloud portability to be able to potentially replace one provider with another (reducing “lock-in”) or, less commonly, to create the possibility of repatriating workloads on-premises. However, lock-in is not caused solely by differentiated capabilities or proprietary API dependencies. Processes, tools, data gravity, the cloud provider’s ecosystem, employee skills and contractual obligations all prevent customers from easily and economically switching between providers.
- “Cloud drift” not “cloud exit:” Organizations are increasingly shifting away from focusing on portability that would enable an immediate “disaster recovery” cloud exit. Rather, they have begun to focus on portability that would allow a gradual exit from a cloud provider over a period of two years or more. This allows the “drift” of workloads from one cloud to another cloud, based on the natural life cycle of each application. This is better aligned to shifts in the organization’s business needs and vendor preferences over time.
- Scenario-based plans: Organizations are building contingency plans that envision specific scenarios and risks that would result in a desire to switch cloud providers. Scenario-based cloud exit planning allows addressing cloud risks in a more specific fashion. Organizations may also be regulatorily required to address cloud provider concentration risks through this sort of planning. While portability needs to be a consideration from the beginning of cloud application architecture and development, shorter exit time frames increase architectural restrictions and portability complexity.
- Container myths: Cloud portability is aggressively hyped by vendors, especially in the context of containers and Kubernetes. However, containers provide limited portability benefits and do not address most of the underlying causes of cloud provider lock-in. Management tools that claim to commoditize the underlying cloud services generally reduce cloud benefits, add unnecessary cost and create a different, often riskier, point of lock-in.

Obstacles

- SaaS applications and platforms are normally entirely nonportable. An exit requires completely replacing the application.
- Open-source or commercial-off-the-shelf-based (COTS-based) PaaS may be replaced by running that software directly. Proprietary PaaS is nonportable and must be replaced by an alternative solution.
- The portability of cloud IaaS workloads may be limited by provider differentiation in infrastructure capabilities. Furthermore, the customer must replace the service's security, management and automation capabilities, including revising DevOps and continuous integration/continuous deployment (CI/CD) toolchains.
- Portability is a program requiring ongoing effort and investment. In addition to the direct technology impacts, customers must consider the impact of cloud provider replacement on the organization's internal skills, contractual relationships (including relationships with independent software vendors [ISVs] whose software they license to run on IaaS), and dependencies upon vendors in the cloud provider's ecosystem.

User Recommendations

- Address SaaS vendor risks through contractual means and integration strategies. SaaS portability is impractical.
- Treat the need for cloud IaaS and PaaS portability as an application portability challenge, not as an infrastructure lock-in problem.
- Balance portability costs and drawbacks against its benefits. Require a business case for larger investments in portability.
- Decide how much portability a particular application needs based on business requirements, and then make architectural choices for that application reflecting the portability requirements.
- Focus on scenario-based exit planning and take a risk management approach to cloud portability.
- Find the balance between desired short- and long-term business outcomes, and the potential future risks of cloud provider dependence. Cloud portability requires compromises between risk reduction and the negative impacts of increased complexity, cost and time to deliver cloud solutions.

Gartner Recommended Reading

Infographic: Mitigate Cloud Risks With Realistic Exit Planning

How to Create a Public Cloud Integrated IaaS and PaaS Exit Plan

Cloud Governance Best Practices: Managing Vendor Lock-In Risks in Public Cloud IaaS and PaaS

Quick Answer: How Should Executive Leaders Respond to Cloud Concentration Risk Concerns?

How to Manage Concentration Risk in Public Cloud Services

Composable Applications

Analysis By: Yefim Natis, Anne Thomas, Paul Vincent

Benefit Rating: High

Market Penetration: 1% to 5% of target audience

Maturity: Emerging

Definition:

Composable applications are built, in part or in whole, as flexible assemblies (compositions) of software components that represent well-defined business capabilities, packaged for programmatic access. The business-centric modularity of composable applications empowers democratized access to technology and business innovation. Composable applications support faster, safe and efficient digital business innovation. Advanced use of composable applications allows cross-application compositions.

Why This Is Important

Composable applications help support resilience, adaptability and growth of business in the context of increasingly frequent challenges, disruptions and opportunities. They support fast-paced business change while protecting the integrity of the outcomes, and bridge application software and business operations by using coarse-grained business-centric software modularity. Organizations that use composable applications maintain customer loyalty by better tracking their changing needs.

Business Impact

The more composable applications there are in the organization's portfolio, the better the organization is prepared to support changing business requirements through digital innovation. In return, greater confidence in the agility of applications promotes faster business thinking. The improved agility of business technology strengthens the ability of an organization to maintain and grow its business, a high value in the modern context of fast innovation, frequent challenges and opportunities.

Drivers

- In the continuously changing business context, demand for business adaptability directs organizations toward technology architecture that supports fast, safe and efficient application change.
- The demand for active participation of business decision makers in the design of their digital experiences promotes the adoption of technology models that are accessible and useful to business experts in addition to, and in cooperation with, technical professionals.
- The need to reduce the costs of redundancy in software capabilities across applications and business units drives organizations to reusable business modularity and from there to composability.
- The increasing number of vendors offering API-centric SaaS (also known as API products or "headless" SaaS) builds up a portfolio of available business-centric packaged application components — promoting their use as building blocks of composable business applications.
- The emerging architecture of micro front ends and superapps advances the principles of composability to the multifunctional user experience, promoting broader adoption of composability in application design.
- Fast-growing competence in mainstream organizations for the management of broad collections of APIs and event streams creates a technology foundation for safe operation of a composable business technology environment.
- The emerging business model of industry cloud, promotes the architecture of modularity and composition inside and across vertical use cases.

Obstacles

- Limited experience of composable thinking and planning in most software engineering organizations complicates composable design efforts and transition plans.
- Limited practice of business-IT collaboration for application design delays the effective composable design that depends on the complementary expert talents in multidisciplinary fusion teams.
- Most legacy applications can participate in composition via their APIs and/or event streams, but their architecture provides only minimal autonomy, delaying the full positive effect of composable architecture.
- Limited development and platform tools dedicated to composable application architecture limit the early success to advanced design teams capable of adapting precursor technologies to new objectives.
- Insufficient mapping of architectural thinking and models between business and technology planners makes digital representation of business functionality less prepared to track real-world business change.

User Recommendations

- Promote modular thinking as the means to great flexibility in business and software innovation.
- Champion API-first business software design, whether or not the application is also packaging the traditional UI capabilities.
- Build competence in API and event stream management as the precursor to managing composable business software modularity.
- Prioritize the formation of business-IT fusion teams to support faster and more effective adaptive change of business applications.
- Use low-code/no-code technologies to facilitate design collaboration of business and technology experts in fusion teams.
- Build an investment case for composability by highlighting how aging digital assets endanger the future success of the business by forming barriers to innovation, competition and customer satisfaction at the pace of market change.
- Gradually modernize (or replace) existing applications toward an architecture of business-centric modularity.

Sample Vendors

Elastic Path Software; Mambu; Novulo; Olympe; Spryker Systems

Gartner Recommended Reading

Becoming Composable: A Gartner Trend Insight Report

Quick Answer: Who's Who in the Life Cycle of Composable Applications?

Case Study: Composable Platform Strategy to Drive Business Agility (Nike)

Predicts 2023: Composable Applications Accelerate Business Innovation

Use Gartner's Reference Model to Deliver Intelligent Composable Business Applications

Distributed Cloud

Analysis By: David Smith, Daryl Plummer, Milind Govekar, David Cearley

Benefit Rating: High

Market Penetration: 1% to 5% of target audience

Maturity: Emerging

Definition:

Distributed cloud refers to the distribution of cloud services to different physical locations, while operation, governance, updates and evolution of the services are the responsibility of the originating cloud provider. Distributed cloud computing is a style of cloud computing where the location of cloud services is a critical component of the model.

Why This Is Important

Distributed cloud enables organizations to use consistent cloud-based services wherever needed, while the cloud service provider retains the responsibility of managing the technology, implementation and evolution of the capabilities. It gives organizations the flexibility to support use cases that will benefit from cloud services, regardless of their dependence on specific locations. Organizations can use distributed cloud to reimagine use cases where cloud computing is not currently feasible.

Business Impact

A major notion of the distributed cloud concept is that the provider is responsible for all aspects of delivery and manages the distributed capabilities “as a service.” This restores cloud value propositions that are broken when customers are responsible for a part of the delivery, as is true in private and some hybrid cloud scenarios. The cloud provider must take responsibility for how the overall system is managed. Otherwise, the value proposition of distributed cloud is compromised.

Drivers

- Historically, location has not been relevant to cloud computing definitions. In fact, the variations on cloud (e.g., public, private, hybrid) exist because location can vary.

- Distributed cloud supports both tethered and untethered operations of cloud services from the cloud provider, “distributed” out to specific and varied physical locations. This enables an important characteristic of distributed cloud operation — low-latency compute where the compute operations for the cloud services are closer to those that need the capabilities. This can deliver major improvements in performance and reduce the risk of global network-related outages.

- Data sovereignty and other regulatory issues may require services be delivered from locations beyond the data centers of the public cloud service provider.
- Perceived and real security and privacy concerns with off-premises applications and infrastructure drive some consumers to prefer on-premises solutions.
- Latency needs of IoT/edge applications require services to be located close to the edge.
- Distributed cloud is still a single-cloud provider, and the managed cloud assets are still part of the cloud provider's portfolio.
- Disconnected operations can be supported with distributed services that can operate independently.

Obstacles

- Customers can't abandon existing technologies in favor of complete and immediate migration to the public cloud, due to sunk costs, latency requirements, regulatory requirements, and the need for integration.
- Different approaches to distributed cloud have different value propositions (e.g., portability, software, appliance). Customers need to maintain visibility back to original goals.
- Distributed services are a relatively small subset of the centralized services, will take time to expand, and will likely never reach 100% parity with public cloud.
- Distributed cloud in your data center will have limits to scale and elasticity, which do not exist with the centralized public cloud. More advanced approaches like distributed cloud embedded in networking or telecom equipment — or delivered as metro area services — are very immature.

User Recommendations

- Overcome the fear of a single franchise controlling the public cloud and on-premises cloud estates, and consider targeted use of distributed cloud.
- Identify scenarios where distributed cloud use-case requirements can be met by evolution of a hybrid cloud model and where the requirements are substantially different. Prefer distributed cloud over building a hybrid cloud. Use the distributed cloud model to prepare for the next generation of cloud computing by targeting location-dependent use cases.
- View vendor claims of the scope of services available and their functional parity with public cloud services skeptically, and demand specific details and data to back up the claims.
- Temper concern about vendor revenue recognition and reporting. As with many capabilities that are thought of as more feature than product, revenue recognition and reporting by vendors are only one indicator of success.

Sample Vendors

Amazon Web Services (AWS); Google; IBM; Microsoft; Oracle

Gartner Recommended Reading

The Cloud Strategy Cookbook, 2023

Comparing On-Premises Public Cloud Appliances: AWS Outposts, Microsoft Azure Stack Hub and Google Distributed Cloud Edge

Distributed Cloud: Does the Hype Live Up to Reality?

Edge Computing

Analysis By: Bob Gill, Philip Dawson

Benefit Rating: Transformational

Market Penetration: More than 50% of target audience

Maturity: Early mainstream

Definition:

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

Why This Is Important

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

Business Impact

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

Drivers

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

Obstacles

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

User Recommendations

IT leaders responsible for cloud and edge infrastructure should:

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

Gartner Recommended Reading

Market Guide for Edge Computing

5 Top Practices of Successful Edge Computing Implementers

Sliding into the Trough

Cloud-Native Application Protection Platforms

Analysis By: Neil MacDonald, Charlie Winckless

Benefit Rating: High

Market Penetration: 1% to 5% of target audience

Maturity: Emerging

Definition:

Cloud-native application protection platforms (CNAPPs) are an integrated set of security and compliance capabilities designed to help secure and protect cloud-native applications across development and production. CNAPPs consolidate a large number of previously siloed capabilities, including container scanning, cloud security posture management, infrastructure as code scanning, cloud infrastructure entitlements management and runtime workload protection.

Why This Is Important

Comprehensively securing cloud-native applications requires the use of multiple tools from multiple vendors that are rarely well-integrated. This lack of integration and automation slows developers down and creates fragmented visibility of risk and friction. CNAPP offerings allow an organization to use a single integrated offering to protect the entire life cycle of a cloud-native application.

Business Impact

Cloud-native application protection platforms consolidate disparate fragmented security testing and protection tools that increase cost and complexity for IT. Using a CNAPP offering will improve developer and security professional efficacy. It will also reduce complexity and costs while maintaining development agility and improving the developer's experience.

Drivers

CNAPPs:

- Reduce the chance of misconfiguration, mistake or mismanagement as cloud-native applications are rapidly developed, released into production and iterated.

- Converge and reduce the number of tools and vendors involved in the continuous integration/continuous delivery (CI/CD) pipeline.
- Reduce the complexity and costs associated with creating secure and compliant cloud-native applications.
- Facilitate the reporting and auditing of cloud security posture/status.
- Improve developer acceptance with security-scanning capabilities that seamlessly integrate into their development pipelines and tooling.
- Place an emphasis on scanning proactively in development and rely less on runtime protection, which is well-suited for container as a service and serverless function environments.

Obstacles

- Cloud workload protection platform (CWPP) vendors that are good at runtime protection aren't necessarily good at integrating into development and vice versa.
- Cloud-native workloads in the form of containers and serverless functions don't require heavyweight runtime protection capabilities.
- There is no single CNAPP offering that does everything. Convergence of capabilities will occur, but will take place over several years.
- Organizations may have siloed purchases of application security testing tooling that is chosen by a different team that manages the runtime protection of workloads. Even at runtime, a separate team may be responsible for web application protection.
- Organizational immaturity in terms of cloud-native application development may inhibit adoption and fragment buying motions.
- Buying centers and influencers are shifting to newer roles such as DevOps architects and cloud security engineering, requiring information security teams to coordinate with these users.

User Recommendations

- Sign contracts of only one to two years because the market for CNAPP is changing rapidly.

- Solicit CWPP vendors to scan containers in development and add cloud security posture management (CSPM) capabilities, including infrastructure-as-code scanning.
- Select integrated offerings with flexible licensing models that allow you to only pay for the capabilities your organization is prepared to use.
- Evaluate the CSPM vendor's ability to add scan of Kubernetes security posture management (KSPM) as well as provide runtime Kubernetes protection capabilities.
- Consolidate open-source software (OSS) vulnerability scanning and software composition analysis through integrations or replacement within a CNAPP offering.
- Scan containers proactively in development for all types of vulnerabilities, not just vulnerable components, including hard-coded secrets, malware and Kubernetes misconfiguration.

Sample Vendors

Aqua Security; Cisco; Lacework; Microsoft; Orca Security; Palo Alto Networks; Rapid7; Sysdig; Trend Micro; Wiz

Gartner Recommended Reading

Market Guide for Cloud-Native Application Protection Platforms

How to Select DevSecOps Tools for Secure Software Delivery

How to Make Cloud More Secure Than Your Own Data Center

Site Reliability Engineering

Analysis By: George Spafford, Daniel Betts

Benefit Rating: Transformational

Market Penetration: 5% to 20% of target audience

Maturity: Adolescent

Definition:

Site reliability engineering (SRE) is a collection of systems and software engineering principles used to design and operate scalable resilient systems. Site reliability engineers work with the customer or product owner to understand operational requirements and define service-level objectives (SLOs). Site reliability engineers work with product or platform teams to design and continuously improve systems that meet defined SLOs.

Why This Is Important

SRE emphasizes the engineering disciplines that lead to resilience; but individual organizations implement SRE in widely varying ways such as a defined role or a set of practices. SRE teams can serve as an operations function, and nearly all such teams have a strong emphasis on blameless root cause analysis. This is to decrease the probability and/or impact of future events and to enable organizational learning, continual improvement and reductions in unplanned work.

Business Impact

The SRE approach to improving reliability and resilience is intended for products and platforms that need to deliver customer value at speed at scale while managing risk. The two primary use cases are to improve the reliability of existing products/platforms or to create new products or platforms that need reliability from the start.

Drivers

- Clients are under pressure to meet customer requirements for reliability while scaling their digital services and are looking for guidance to help them.
- While Google originated what became known as SRE and continued to evolve it, practitioners are developing and sharing new practices as well. Potential practitioners looking for pragmatic guidance to improve the reliability of their systems have a rich body of knowledge they can leverage that works well with agile and DevOps.
- Organizations are adopting highly skilled automation practices (usually DevOps), and usage of infrastructure-as-code capabilities (which usually requires a cloud platform) to deliver digital business products reliably.
- The most common use case based on inquiry calls with clients is to leverage SRE concepts to improve the reliability of existing systems that are not meeting customer requirements for availability, performance or are proving difficult to scale.

Obstacles

- Insufficient internal marketing to understand what agile, DevOps or product teams need or would value and then explaining how the value SRE can deliver will justify the costs and risks incurred. Without marketing its benefits, SRE adoption tends to be less certain or slower. The SRE concept by itself is insufficient — people must continuously believe it is worthwhile.
- Finding SRE candidates who have the right mix of development, operations and people skills is a big challenge for clients. Impacts on initial adoption and scaling efforts as well.
- Rebranding of a traditional operations team without changing to adopt SRE practices, only SRE in name.
- Clients have voiced problems with product owners who overly focus on functional requirements and not nonfunctional requirements thus slowing improvements and support of SRE within the organization.

User Recommendations

- Leverage practices pragmatically based on need. Don't feel that you must implement SRE exactly the way Google does it, learn what works for you.
- Detect an opportunity to begin that is politically friendly, will demonstrate sufficient value and has an acceptable risk profile.
- Start small, focus, learn, improve, and demonstrate value — do not try to change everything at once.
- Work with the customer or product owner to define clear, obtainable SLOs based on their needs.
- Implement monitoring and improve observability to objectively report on actual performance relative to the SLOs.
- Product owners must be accountable for functional and non-functional requirements of their products.
- Instill collaborative working between site reliability engineers, developers and other stakeholders to help them learn how to design, build and evolve their products to meet SLOs.
- Create a community, implement effective organizational learning practices and evolve SRE practices.

Sample Vendors

Atlassian; Blameless; Datadog; Dynatrace; New Relic; OpsRamp; PagerDuty; Splunk

Multicloud Management

Analysis By: Lydia Leong

Benefit Rating: Moderate

Market Penetration: 5% to 20% of target audience

Maturity: Early mainstream

Definition:

Multicloud management encompasses the principles, practices and tools necessary to manage multiple cloud infrastructure as a service (IaaS) or platform as a service (PaaS) providers within the context of a multicloud strategy. There are many possible multicloud management approaches, spanning the spectrum from managing each cloud individually to managing all clouds in a unified way.

Why This Is Important

More than 75% of organizations use multiple public cloud IaaS and PaaS providers. Many of these organizations struggle to coordinate governance and management efforts across these multiple providers. Because these providers are highly differentiated from one another, and are often used by different parts of the business for different use cases, it is almost always impractical and inappropriate to apply a “one size fits all” management approach.

Business Impact

Organizations must implement effective cloud management of their public cloud IaaS and PaaS providers to ensure well-governed, safe and efficient cloud adoption that delivers desired business outcomes. Organizations that do not find the right balance between multicloud management consistency and optimal provider-specific management will suffer unnecessarily high multicloud complexity and cost.

Drivers

- All cloud IaaS and PaaS offerings require management, and an organization's efforts to manage cloud providers become more complicated and expensive with every cloud provider added to their portfolio. This has led to a growing interest in effective multicloud management approaches.
- During the early phase of the cloud management platform (CMP) market, vendors sought to promote and sell "single pane of glass" management tools to multicloud customers. While the commercial success of such tools has been limited, the hype has endured.
- More recently, cloud management tool vendors have successfully focused on providing deep, provider-specific support within focused management offerings. However, most such vendors support at least two cloud providers and market multicloud management capabilities.
- Although the "multicloud management" term is broadly hyped, the term is not used consistently across the market. Multicloud management tools may provide visibility or control of cloud resources across one or more capability domains: provisioning and orchestration; service enablement; monitoring and observability; inventory and classification; cost management and resource optimization; cloud migration, backup and disaster recovery; and identity, security, and compliance.
- Although a cloud provider's first-party management services and tools will often meet an organization's management functionality requirements, third-party tools may provide multicloud aggregation, simplification or additional functions not provided by the provider's native capabilities.

Obstacles

- The most effective multicloud strategies result in the use of multiple cloud providers to better exploit the unique capabilities of each provider. Individuals typically only have deep skills with a single provider. This often leads to a desire to optimize the management of each individual provider, rather than a unified multicloud management approach.
- The increased complexity and breadth of cloud services has made it hard or even impossible for a single CMP to meet all the management needs of an organization.
- Each provider has their own strategy for exposing management functionality for their services and different first-party management capabilities. This makes it costly for customers (and cloud management tool vendors) to engineer multicloud management functionality.
- Public cloud IaaS and PaaS offerings have extensive hooks for automated management, but these capabilities vary per service within each provider's portfolio. Service coverage within cloud management tools varies widely.

User Recommendations

- Decide whether you want a single cloud operations function with a unified multicloud approach or a cloud operations approach for each cloud provider. This may lead to different processes, tools and personnel for each major cloud provider you use. You may unify multiple approaches across points of commonality.
- Minimize the number of cloud management tools in use by balancing provider-specific functionality with the desired degree of consistency across your managed environments. This will allow you to contain complexity, cost and the number of vendors to manage, while mitigating the risks associated with them.
- Try not to select a single CMP to manage multiple cloud providers or hybrid clouds. Do not take a "least common denominator" approach that reduces each cloud provider down to only the commoditized capabilities that can be managed through a unified CMP. Organizations that have tried to do so have often failed.

Sample Vendors

CloudBolt Software; Morpheus Data; Scalr; VMware (CloudHealth)

Gartner Recommended Reading

Comparing Cloud Operations Approaches

Quick Answer: How to Navigate Cloud Management Tooling Selection

A Guidance Framework for Selecting Cloud Management Tools

Cloud-Optimized Hardware

Analysis By: Alan Priestley

Benefit Rating: Moderate

Market Penetration: 1% to 5% of target audience

Maturity: Adolescent

Definition:

Cloud-optimized hardware includes servers, networking, storage and custom silicon, designed specifically for use in cloud operating environments. Optimizations include configuration via software control planes and tuning of hardware functionality for specific software workloads. Cloud-optimized hardware is developed primarily by hyperscale cloud providers to support the execution of cloud services, but is increasingly becoming available to enterprise IT for use in their infrastructure.

Why This Is Important

Efficient delivery of cloud-based services at scale requires optimization of the data center infrastructure to ensure ease of deployment and operation. Cloud service providers (CSPs) are working with equipment supply chains to source hardware optimized for deployment within their specific infrastructure; this includes servers, networking and storage equipment and optimized chip designs. Enterprise customers can benefit by leveraging the experiences gained by the original design manufacturers (ODMs) supplying hardware to CSPs.

Business Impact

Cloud-optimized hardware brings benefits when planning large-scale infrastructure deployments by:

- Delivering efficiencies not possible when using standard OEM equipment.
- Lowering the cost of data center operations.
- Enabling the targeting of infrastructure deployments to specific workloads.

Drivers

- Hyperscale cloud service providers' requirements for increased operational efficiency and lower component costs in their data center infrastructure.
- Initiatives like Open Compute Project (OCP) Foundation, encouraging ODMs — such as Inventec and Wistron — to produce cloud-optimized servers for a wider range of hosting providers, value-added colocation providers, telecom providers, SaaS independent software vendors (ISVs) and large enterprises.
- Major semiconductor vendors offering custom versions of their standard chips, tuned for use in cloud infrastructure.
- Development of custom microprocessors and application-specific integrated circuits (ASICs) for workload acceleration. For example, Amazon Web Services (AWS) Graviton processors and Google's Tensor Processing Units (TPUs) for AI workloads.
- Need for high-performance network interface components that support and isolate multiple workloads and users on servers hosting multiple virtual instances.
- Function accelerators that support the offload of storage and network processing from the CPU.
- The use of custom chips to establish a hardware-based root of trust to guarantee the integrity of the hardware and firmware. For example, AWS Nitro System-based services.

Obstacles

- While it is possible for smaller organizations to work with ODMs to access this hardware, ODMs have a very different business model, often having minimum delivery quantities of tens of thousands of standard units.
- Cloud-optimized systems are usually provided with specific system configurations (CPU, memory, storage and network), however, there is often no flexibility to vary specifications.
- ODMs provide minimal support in terms of software configurations, long-term maintenance contracts or postsales support.
- For many smaller-scale deployments, the cost savings from buying cloud-optimized hardware can be offset by the increase in resources and skill sets necessary to maintain the equipment and associated software stacks.
- Limited desire from ODMs to expand their offerings to nonhyperscale customers.

User Recommendations

- Assess the impact of using customized hardware solutions utilization of “nonstandard” components — when planning deployments at scale using cloud-optimized hardware — by evaluating the impact of these on software configurations and workload optimizations.
- Be prepared to manage an ODM-based supply chain, which is very different from managing traditional vendors and often lacks postsales and long-term maintenance support.
- Evaluate the use of cloud instances based on proprietary CPU designs, such as the Arm-based AWS Graviton for workloads that leverage interpreted languages (such as Java or Python, where an Arm-based interpreter is available), by tracking the use of cloud-optimized hardware by cloud service providers.

Sample Vendors

Amazon Web Services (AWS); Flex; Inspur; Open Compute Project Foundation

Gartner Recommended Reading

Emerging Tech Impact Radar: Compute and Storage

Market Guide for Servers

Market Trends: Arm in the Data Center: Act Now to Develop Plans to Address This Shifting Market

Container Management

Analysis By: Dennis Smith, Michael Warrilow

Benefit Rating: High

Market Penetration: 20% to 50% of target audience

Maturity: Early mainstream

Definition:

Gartner defines container management as offerings that enable the development and operation of containerized workloads. Delivery methods include cloud, managed service and software for containers running on-premises, in the public cloud and/or at the edge. Associated technologies include orchestration and scheduling, service discovery and registration, image registry, routing and networking, service catalog, management user interface, and APIs.

Why This Is Important

Container management automates the provisioning, operation and life cycle management of container images at scale. Centralized governance and security are used to manage container instances and associated resources. Container management supports the requirements of modern applications, including platform engineering, cloud management and continuous integration/continuous delivery (CI/CD) pipelines. Benefits include improved agility, elasticity and access to innovation.

Business Impact

Industry surveys and client interactions show that demand for containers continues to rise. This trend is due to application developers' and DevOps teams' preference for container runtimes, which use container packaging formats. Developers have progressed from leveraging containers on their desktops to needing environments that can run and operate containers at scale, introducing the need for container management.

Drivers

- The adoption of DevOps-based application development processes.
- The rise of cloud-native application architecture based on microservices.
- New system management approaches based on immutable infrastructure, which gives the ability to update systems frequently and reliably maintained in a “last known good state” rather than repeatedly patched.
- Cloud-based services built with replaceable and horizontally scalable components.
- A vibrant open-source ecosystem and competitive vendor market have culminated in a wide range of container management offerings. Many vendors enable management capabilities across hybrid cloud or multicloud environments. Container management software can run on-premises, in public infrastructure as a service (IaaS), or simultaneously in both.
- Container-related edge computing use cases have increased in industries that need to get compute and data closer to the activity (for example, telcos, manufacturing plants, etc.).
- AI/ML use cases have emerged over the past few years, leveraging the scalability capabilities of container orchestration.
- Cluster management tooling that enables the management of container nodes and clusters across different environments is increasingly in demand.
- All major public cloud service providers now offer on-premises container solutions.
- Independent software vendors (ISVs) are increasingly packaging their software for container management systems through container marketplaces.
- Some enterprises have scaled sophisticated deployments, and many more are planning container deployments. This trend is expected to increase as enterprises continue application modernization projects.

Obstacles

- More abstracted, serverless offerings may enable enterprises to forgo container management. These services embed container management in a manner that is transparent to the user.
- Third-party container management software faces huge competition in the container offerings from the public cloud providers, both with public cloud deployments and the extension of software to on-premises environments. These offerings are also challenged by ISVs that choose to craft open-source components with their software during the distribution process.
- Organizations that perform relatively little app development or make limited use of DevOps principles are served by SaaS, ISV and/or traditional application development packaging methods.

User Recommendations

- Determine if your organization is a good candidate for container management software adoption by weighing organizational goals of increased software velocity and immutable infrastructure, and its hybrid cloud requirements, against the effort required to operate third-party container management software.
- Leverage container management capabilities integrated into cloud IaaS and platform as a service (PaaS) providers' service offerings by experimenting with process and workflow changes that accommodate the incorporation of containers.
- Avoid using upstream open source (e.g., Kubernetes) directly unless the organization has adequate in-house expertise to support.

Sample Vendors

Alibaba Cloud; Amazon Web Services; Google; IBM; Microsoft; Mirantis; Red Hat; SUSE; VMware

Gartner Recommended Reading

Market Guide for Container Management

Serverless Infrastructure

Analysis By: Jeffrey Hewitt

Benefit Rating: Transformational

Market Penetration: 20% to 50% of target audience

Maturity: Adolescent

Definition:

Serverless infrastructure is a model of IT service delivery in which the underlying enabling resources are used as an opaque, unlimited, shared pool that is continuously available without advance provisioning and priced in units of the consumed IT service. The runtime environment (that is, the compute, storage, networking and language execution environment) required to execute an application or service is automatically provisioned and operated.

Why This Is Important

Accelerating the development and delivery of software is a core imperative for IT leaders. Not only do serverless technologies enable organizations to build and deliver software faster, but they also offer low operational overheads, resource scaling as needed and an elastic pricing model. Cloud providers and open-source software vendors are all innovating and making serverless products available for a broad set of use cases.

Business Impact

Serverless technologies enable organizations to build cloud-native applications with newer application architectures — such as microservices, which can usher in higher degrees of resiliency, elasticity and agility for digital workloads. Serverless technologies also enable consumption of platform services by developers and business users, with the infrastructure provisioning and life cycle management abstracted away from the consumer.

Drivers

- **Evolution:** In the past few years, the term “serverless infrastructure” has evolved to include much more than function as a service (FaaS) products. Currently, it refers not only to a programming model such as FaaS, but also to an operational model where all provisioning, scaling, monitoring and configuration of compute infrastructure are delegated to the platform. Examples of such architectures include serverless containers and serverless databases.

- **Operational simplicity:** Serverless infrastructure obviates the need for IT departments to perform infrastructure setup, configuration, provisioning and management.

- **"Built-in" scalability:** Infrastructure scaling is automated and elastic.
- **Cost-efficiency:** You only pay for infrastructure resources when they are needed to support requested transactions.
- **Developer productivity and business agility:** Serverless infrastructure abstracts infrastructure architecture and allows developers to focus on writing code and designing applications.

Obstacles

- **Vendor lock-in:** As with most cloud functionality, the leading serverless implementations are proprietary to specific cloud providers. If an application has to move from one cloud platform to another, it will have to be significantly reengineered.
- **Low degree of control:** The managed service model and the runtime virtualization of serverless technologies bestow huge benefits, but at the cost of little or no control over the service. The environment is a "black box" that must be used as it is.
- **Skills gap:** Serverless operations require a major shift in skills and best practices, with much more code and API-oriented service delivery.

User Recommendations

- Ensure cost governance and budget control by evaluating the cost implications of event-driven application architectures and the pricing models of different vendors. Be aware of API gateway, network egress and other costs.
- Revise data classification policies and controls to account for the fact that objects in a content store can now represent code as well as data.
- Rethink IT operations from infrastructure management to application governance, with an emphasis on ensuring that security, monitoring, debugging requirements and application SLAs are being met. In cases where on-premises deployment is merited, IT teams can support FaaS in the role of service provider.

Sample Vendors

Alibaba; Amazon Web Services; Apache Software Foundation; DigitalOcean; Google; IBM; Knative; Microsoft; OpenFaaS; Oracle

Gartner Recommended Reading

A CTO's Guide to Serverless Computing

When to Use Serverless Computing to Optimize Cloud Costs?

Decision Point for Selecting Virtualized Compute: VMs, Containers or Serverless

SASE

Analysis By: Neil MacDonald, Andrew Lerner

Benefit Rating: Transformational

Market Penetration: 5% to 20% of target audience

Maturity: Adolescent

Definition:

Secure access service edge (SASE) delivers converged network and security capabilities, including SD-WAN, SWG, CASB, firewall and zero trust network access (ZTNA). SASE supports branch office, remote worker and on-premises secure access use cases. SASE is primarily delivered as a service and enables zero trust access based on the identity of the device or entity, combined with real-time context and security and compliance policies.

Why This Is Important

SASE is a key enabler of modern digital business transformation, including work from anywhere and the adoption of edge computing and cloud-delivered applications. It increases visibility, agility, performance, resilience and security. SASE also dramatically simplifies the delivery and operation of critical network and security services mainly via a cloud-delivered model. SASE reduces the number of vendors required for secure access to one or two explicitly partnered vendors.

Business Impact

SASE enables:

- Digital business use cases (such as branch office transformation and hybrid workforce enablement) with increased ease of use while reducing costs and complexity via vendor consolidation and dedicated circuit offload.

- Infrastructure and operations and security teams to deliver a rich set of networking and network security services in a consistent and integrated manner to support the needs of digital business transformation, edge computing and work from anywhere.

Drivers

- Digital business transformation including the adoption of cloud-based services by mobile workforces, edge computing and business continuity plans that must include a flexible, anywhere, anytime, secure, identity-based logical perimeter model of SASE.
- The need to flexibly support digital business transformation efforts with a zero trust security architecture while managing complexity is a significant factor for the adoption of SASE, primarily delivered as a cloud-based service.
- For IT, SASE can reduce the deployment time for new users, locations, applications and devices.
- For information security, SASE enables a single way to set policy enforcement consistently across all types of access — internet, web applications and private applications, reducing the attack surface and shortening remediation times.
- Enterprise desire to simplify network and network security deployments via the reduction of policy engines and management consoles.

Obstacles

- **Organizational silos, existing investments and skills gaps:** A full SASE implementation requires a coordinated and cohesive approach across security and networking teams, which is challenging given refresh/renewal cycles, silos and existing staff expertise.
- **Organizational bias and regulatory requirements for on-premises deployment:** Some customers have an aversion to the cloud and want to maintain control.
- **Global coverage:** SASE depends upon cloud delivery, and a vendor's cloud footprint may prevent deployments in certain geographies, such as China, Africa, South America and the Middle East.
- **SASE maturity:** SASE capabilities vary widely. Sensitive data visibility and control is often a high-priority capability, but it is difficult for many SASE vendors to address. While your preferred single vendor may lack the capabilities you require, two-vendor partnerships can be a viable approach.

User Recommendations

- Involve the security architect and network architect when evaluating offerings and roadmaps from the incumbent and emerging vendors to ensure an integrated approach.
- Leverage WAN, firewall, VPN hardware refresh cycles or software-defined WAN (SD-WAN) deployments to update network and network security architectures.
- Explore single-vendor SASE, dual-vendor SASE and managed SASE options when investing, but avoid deploying SASE with more than two vendors, regardless of vendor marketing for all core services to minimize complexity and improve performance.
- Use vendor combinations — when selecting a dual-vendor solution — that have explicit integration including turnkey automation and visibility, and ideally management and data plane integration.
- Combine branch office and remote access in a single implementation to ensure consistent policies and minimize the number of vendors required.
- Leverage branch office transformation and dedicated circuit offload projects to adopt SASE.

Sample Vendors

Cato Networks; Cisco Systems; Cloudflare; Forcepoint; Fortinet; Juniper Networks; Netskope; Palo Alto Networks; Versa Networks; Zscaler

Gartner Recommended Reading

2022 Strategic Roadmap for SASE Convergence

Market Guide for Single-Vendor SASE

The Future of Network Security Is in the Cloud

Magic Quadrant for SD-WAN

Magic Quadrant for Security Service Edge

Repatriation

Analysis By: Ed Anderson, Lydia Leong

Benefit Rating: Low

Market Penetration: 5% to 20% of target audience

Maturity: Legacy

Definition:

In the context of cloud computing, repatriation is the return of an application or data that has been migrated to a public cloud back to its original on-premises environment. Repatriation may occur when an application that has been migrated to a public cloud does not deliver the expected benefits, such as cost savings.

Why This Is Important

Applications migrated to a cloud environment may fail at a technical level or may not deliver the expected benefits such as cost reduction, and improved performance and reliability. One approach to remediate these shortfalls, an approach which has experienced recent hype, is to repatriate the application back to its original environment. While rare, repatriation has been promoted by some vendors to undermine the benefits of cloud and accentuate the benefits of on-premises environments.

Business Impact

While repatriation is one approach to addressing the perceived shortcomings of application performance in the cloud, it is typically not the most efficient approach to remediating issues. Cloud benefits and risks should be clearly identified before engaging in the process of migrating applications to a cloud environment. When the expected benefits are not realized, most organizations are able to address the cloud application shortcomings rather than retreat to the previous environment.

Drivers

- Repatriation hype has gone through waves. Recently, hype has been on the rise, even though actual repatriation remains nearly nonexistent.
- Repatriation has been touted as a means of redress when public cloud services fail to deliver the purported benefits by returning migrated applications back to on-premises environments.
- Cloud projects often experience temporary failures and setbacks, a very small minority of which result in returning applications to their previous on-premises environment for a period of time. Failures in execution are most common in projects that are poorly conceived or inadequately planned. In the specific context of data center migrations to cloud infrastructure as a service (IaaS), the most critical problems are related to the migration strategy, rationale or approach. For instance, a lift-and-shift migration designed to treat the cloud provider as just a virtualization platform is usually costly and often fails to deliver the expected cloud benefits. In such cases, a migration pause may occur while the project is rethought and rescoped.
- Ungoverned consumption can also cause cloud expenses to rise and may lead to a slowdown or pause in the adoption of cloud IaaS and platform as a service (PaaS).
- The most commonly cited examples of true migration reversals typically have unique circumstances and usually involve hosted applications or SaaS.

Obstacles

- Repatriation hype often creates distractions during a time when thoughtful analysis should be devoted to appropriate cloud initiatives. Public cloud environments are not well-suited for every application or workload type.
- The process of repatriation assumes the original environment used to host the application before the cloud migration still exists. This may not be the case, particularly when cloud migration events are accompanied by data center divestitures.
- The process of repatriation assumes that the customer can host the application or otherwise replicate its functionality, including cloud operating characteristics, on-premises. This is almost never the case, especially when attempting to repatriate SaaS.
- There are no mainstream tools to assist with the migration of applications from a cloud to an on-premises environment making repatriation an expensive proposition in most situations.

User Recommendations

- Assess repatriation anecdotes with skepticism, especially when delivered from entities with a bias toward preserving traditional data center businesses. Actual repatriation events are rare, and are becoming rarer as cloud maturity improves.
- Establish a thorough cloud strategy, including a process for assessing the viability of application migration to cloud prior to initiating a cloud migration.
- Manage the economics of cloud services through appropriate governance, cloud financial management and optimization. The business owners of applications, technical teams, finance and sourcing must work together to ensure the business derives maximum value from its cloud usage.
- Ensure your cloud adoption is well-governed in a fashion consistent with your cloud strategy. If cloud adoption outpaces your organization’s skills and ability to deliver, obtain third-party assistance and ensure they are held contractually accountable for project outcomes.

Gartner Recommended Reading

Moving Beyond the Myth of Repatriation: How to Handle Cloud Project Failures

Workload Placement in Hybrid IT — Making Great Decisions About What, Where, When and Why

Is FinOps the Answer to Cloud Cost Governance?

The Cloud Strategy Cookbook, 2023

8 Architectural Principles for Going Cloud-Native

Cloud Marketplaces

Analysis By: Ed Anderson

Benefit Rating: Moderate

Market Penetration: More than 50% of target audience

Maturity: Early mainstream

Definition:

Cloud marketplaces are online storefronts through which customers can find and subscribe to cloud and cloud-related offerings. Resources that can be found and procured through cloud marketplaces include IaaS, PaaS and SaaS services, curated datasets, packaged virtual machine and container images, APIs, and cloud-related IT services. Cloud marketplaces may be public or private, targeting B2B or B2C buyers, and possessing specialized storefronts.

Why This Is Important

Cloud marketplaces are growing in influence as a destination to find and procure cloud service offerings. Cloud service providers (CSPs), distributors, resellers, managed service providers and internal IT organizations build and deliver cloud marketplaces to support their respective constituents. Cloud app stores, cloud service catalogs and portals are also types of cloud marketplaces, but they are typically built for consumption by a closed community of users.

Business Impact

As more cloud services are discovered and procured via cloud marketplaces, they become critical as a technology and architectural control point. Providers motivate customers to use cloud marketplaces through discounts, unified billing, value-add services, solution validations and ease of use. CSPs also often allow marketplace customers to apply marketplace spending to cloud spending commitments, thereby unlocking additional volume discounts.

Drivers

- CSPs use cloud marketplaces to showcase third-party solutions that are available on their cloud platform and to highlight the ecosystem supporting their platform.
- Cloud marketplaces may be dissociated from any specific cloud platform/service and offered by independent cloud service resellers or distributors. Many traditional software and hardware distributors provide cloud marketplaces for their customers, particularly those with multicloud requirements.
- Cloud service subscribers use cloud marketplaces to simplify the process of finding, purchasing, deploying and using cloud services. Cloud marketplaces can be used to isolate and deliver only validated or sanctioned services, which can complement cloud governance policies by enforcing the use of only approved cloud services. Cloud service aggregation features, such as unified billing, software license and entitlement management are typically provided when using a cloud marketplace.
- CSPs often provide financial incentives, such as volume discounts, for cloud spending commitments including both cloud services from the CSP as well as services purchased through the CSP marketplace.
- Cloud marketplaces have expanded to include curated datasets, prepackaged business capabilities (PBCs), prepackaged virtual machine and container images, and other services. Other digital assets are likely to be included in cloud marketplaces in the future.
- Internal cloud marketplaces (sometimes called “private marketplaces”) may be established by organizations to facilitate the use of internal and external cloud services, and to provide management and governance over the organization’s use of cloud services.

Obstacles

- Using CSP marketplaces may require organizations to support multiple marketplaces when they have multicloud requirements.
- There is no common standard for cloud marketplaces, which results in wildly different capabilities and interfaces for each marketplace offering. Organizations will find procuring cloud services from multiple marketplaces to be complex and likely to result in cost inefficiencies.
- Terms related to marketplace purchases may also vary both within and across marketplaces, resulting in inconsistency in asset and entitlement management and governance policies.

User Recommendations

- Use cloud marketplaces to find solutions compatible with specific cloud platforms and validated by the hosting cloud provider.
- Include cloud marketplace spending in your contractual commitments with cloud providers, which may result in committed spending retirement and pricing discounts.
- Use non-CSP cloud marketplaces when looking for a multicloud marketplace solution. These marketplaces are typically offered by technology resellers and distributors, and often include additional IT services such as consulting, implementation, and managed services.
- Leverage the full capabilities of cloud marketplaces including metering, reporting, billing, integration, localization and management services. Seek out marketplace offerings beyond packaged cloud service offerings, such as machine learning (ML) modules, prepackaged containers and validated datasets.

Sample Vendors

Alibaba Cloud; Amazon Web Services (AWS); CDW; Google; IBM; Ingram Micro; Microsoft; Oracle; Salesforce

Gartner Recommended Reading

Incorporate Cloud Marketplaces in Your Digital Delivery Strategy

Tool: IT Sourcing and Procurement Guide to Using Digital Marketplaces

A CTO's Guide to Navigating the Cloud-Native Container Ecosystem

IoT PaaS

Analysis By: Alfonso Velosa, Eric Goodness

Benefit Rating: High

Market Penetration: 5% to 20% of target audience

Maturity: Adolescent

Definition:

An IoT platform as a service (PaaS) solution enables the connection and centralized capture of data or events from IoT-enabled assets. Enterprises use it to improve business operations such as optimizing maintenance of remote assets, or develop new business models, such as product-as-a-service or drive visibility by integrating different IoT platforms. Cloud-based capabilities include device management, integration, data management, analytics, application enablement and management and security.

Why This Is Important

Enterprises keep adding IoT PaaS capabilities to assets to drive business operations cost and process optimization, improve customer experience and build new revenue opportunities, such as product-as-a-service. This effort often requires additional operating information, whether about clients or the supply chain or factors, such as the environment or adjacent equipment. The complexity, scale and business value of this data call for specialized technology, often implemented as an IoT PaaS.

Business Impact

Enterprises use IoT PaaS to improve business decisions by connecting to assets and then to cloud-based applications and contextual data. Goals include:

- New revenue by offering differentiated products and new business models, such as product servitization.
- Process improvement by using assets at their optimal state and starting driving smart connected operations.
- Improving employee safety training and certification plus regulatory and sustainability compliance.

Drivers

- Both asset-intensive (oil and gas or power generation) and asset-light (insurance or medical) industries are increasingly implementing IoT projects to support an abundance of business objectives.
- Enterprise teams need to connect to equipment in order to improve asset health, lower operating costs, improve processes, avoid unplanned downtime and enhance worker safety.
- Business teams need to accelerate time to market and improve the quality of smart products, while consolidating, structuring, and optimizing data and events from the edge into cloud systems.
- Technology providers' marketing and ecosystem partners have increasingly shifted to encourage enterprises to implement IoT-enabled business projects for sustainability or increasing output or lowering operational costs.
- Technology providers continue to improve their IoT cloud services and technology, developer experience, and vertical market templates to ensure they can deliver business solutions at scale and schedule for their customers.

Obstacles

- IoT PaaS requires customization of cloud services such as relational databases, container registries or event hubs to achieve business outcomes for large-scale deployments, driving up cost and schedule.
- Many enterprises approach IoT projects as technology projects, instead of as business projects that use IoT PaaS to achieve business outcomes.
- Many enterprises lack central leadership and operate in a siloed fashion, adopting a different IoT PaaS for each use case, limiting their ability to scale and adding complexity.
- Enterprise leaders often underinvest in culture change processes, employee training, or communicating benefits. This leads local general managers and employees to limit the IoT project deployment scale and to underuse the data produced by the IoT PaaS.
- Small and midsize enterprises lack the teams to implement complex IoT PaaS cloud services.
- Technology providers still over-emphasize technology benefits and under-emphasize business value propositions.

User Recommendations

- Use an IoT PaaS skills gap analysis to map a path to improve the IT organization's capabilities, such as integration or cloud service management or security.
- Start small. Treat initial IoT PaaS projects as business projects to acquire implementation lessons, identify challenges and opportunities, and drive IT/OT alignment.
- Address the lack of a central leadership for IoT projects by developing a multiplatform architecture that addresses differing enterprise needs for IoT PaaS. These include projects with new assets using new protocols compared to complex projects with legacy assets that must connect to a range of legacy protocols.
- Develop a scenario analysis that IT will have to assume IoT PaaS cloud budget and long-term management from the business unit but with a zero sum budget.
- Incorporate a composable architecture to address changes both at the asset level as well as to address platform to platform integration and business solution development.

Sample Vendors

ABB; Alibaba Cloud; Amazon Web Services (AWS); Fujitsu; Microsoft; PTC; Siemens; Toshiba

Gartner Recommended Reading

Magic Quadrant for Global Industrial IoT Platforms

Technology Opportunity Prism: Internet of Things

Competitive Landscape: IoT Service Providers

Now Is the Time to Deliver IoT-Enabled Product Servitization to Manufacturers

Connected IoT Brokers for Autocomposing: An Insurance Trend for 2030

Climbing the Slope

Cloud AI Services

Analysis By: Van Baker, Bern Elliot

Benefit Rating: High

Market Penetration: 20% to 50% of target audience

Maturity: Early mainstream

Definition:

AI cloud services provide AI model building tools, APIs for prebuilt services and associated middleware that enable the building/training, deployment and consumption of machine learning (ML) models running on prebuilt infrastructure as cloud services. These services include vision and language services and automated ML to create new models and customize prebuilt models.

Why This Is Important

The use of AI cloud services continues to increase. Vendors have introduced additional services and solutions with fully integrated MLOps pipelines. The addition of low-code tools has added to ease of use. Applications regularly use AI cloud services in language, vision and automated ML to automate and accelerate business processes. Developers are aware of these offerings, and are using both prebuilt and customized ML models in applications.

Business Impact

The impact of AI extends to the applications that enable business, allowing developers and data scientists to enhance the functionality of these applications. The desire for data-driven decisions in business is driving the incorporation of forecasts and next best actions, including automation of many workflows. AI cloud services enable the embedding of advanced machine learning models in applications that are used to run the day-to-day business operations.

Drivers

- **Opportunities to capitalize on new insights.** The wealth of data from both internal and third-party sources delivers insights that enable data-driven decision intelligence.

- **Support demand for conversational interactions.** The emergence of generative AI and large language models facilitates conversationally enabled applications.
- **To meet business key performance indicators (KPIs).** There is a mandate for businesses to automate processes to improve accuracy, improve responsiveness and reduce costs by deploying both AI and ML models.
- **Reduced barriers of entry.** The ability to do zero-shot learning and model fine-tuning has reduced the need for large quantities of data to train models. Access for developers and citizen data scientists to AI and ML services due to the availability of API callable cloud-hosted services will expand the use of AI.
- **AutoML as an enabler for custom development.** Use of automated ML to customize packaged services to address specific needs of the business is increasing.
- **A wide range of AI cloud services.** AI cloud services from a range of specialized providers in the market, including orchestration layers to streamline deployment of solutions, are available.
- **Emerging AI model marketplaces.** New marketplaces should help developers adopt these techniques through AI cloud services.

Obstacles

- **Lack of understanding** by developers and citizen data scientists about how to adapt these services to specific use cases.
- **Pricing models** for AI cloud services that are usage-based presents a risk for businesses as the costs associated with use of these services can accrue rapidly.
- **Increased need** for packaged solutions that utilize multiple services for developers and citizen data scientists.
- **Lack of marketplaces** for prebuilt ML models that can be adapted for specific enterprise use cases.
- **Continuing need for ModelOps** tools that enable integration of AI and ML models into applications.
- **Lack of skills** for developers to effectively implement these services in a responsible manner.

User Recommendations

- Choose customizable AI cloud services over custom models to address a range of use cases and for quicker deployment and scalability.
- Improve chances of success of your AI strategy by experimenting with AI techniques and cloud services, using the exact same dataset and selecting one that addresses requirements. Consider using an A/B testing approach.
- Use AI cloud services to build less complex models, giving the benefit of more productive AI while freeing up your data science assets for higher-priority projects.
- Empower non-data-scientists with features such as automated algorithm selection, dataset preparation and feature engineering for project elements. Leverage existing expertise on operating cloud services to assist technical professional teams.
- Establish a center of excellence for responsible use of AI that includes all functional areas of the business. This is especially important in light of the advances of generative AI solutions.

Sample Vendors

Alibaba; Amazon Web Services; Baidu; Clarifai; Google; H2O.ai; Huawei; IBM; Microsoft; Tencent

Gartner Recommended Reading

Critical Capabilities for Cloud AI Developer Services

Magic Quadrant for Cloud AI Developer Services

Hybrid Cloud Computing

Analysis By: David Smith, Milind Govekar

Benefit Rating: High

Market Penetration: 20% to 50% of target audience

Maturity: Early mainstream

Definition:

Hybrid cloud computing comprises two or more public and private cloud services that operate as separate entities but are integrated. A hybrid cloud computing service is automated, scalable and elastic. It has self-service interfaces and is delivered as a shared service using internet technologies. Hybrid cloud computing needs integration between the internal and external environments at the data, process, management or security layers.

Why This Is Important

Hybrid cloud theoretically offers enterprises the best of both worlds — cost optimization, agility, flexibility, scalability and elasticity benefits of public cloud, in conjunction with control, compliance, security and reliability of private cloud (assuming their on-premises environments are truly cloud). As a result, virtually all enterprises have the desire to augment internal IT systems with external cloud services. Note that many organizations start with hybrid IT, which lessens the requirement of a true private cloud.

Business Impact

Hybrid cloud computing enables an enterprise to leverage both its data centers as well as the capabilities of the public cloud. It is transformational because changing business requirements drive the optimum use of private and/or public cloud resources. This approach improves the economic model and agility and sets the stage for new ways for enterprises to work with suppliers, partners (B2B) and customers (B2C).

Drivers

- The desire to evolve data centers to become more cloudlike and, therefore, have a private cloud having cost and other characteristics more like a public cloud, while maintaining “in house” infrastructure for key privacy, security, data residency or latency needs.
- As more providers deliver hybrid cloud offerings, they increasingly deliver a packaging of the concept. “Packaged hybrid” (a flavor of distributed cloud) means that you have a vendor-provided private cloud offering that is packaged and connected to a public cloud in a tethered way. Azure Stack HCI from Microsoft is a good example of this packaging, but there is another approach as well. We call these two main approaches “like for like” hybrid and “layered technology” hybrid (spanning different technology bases). Packaged hybrid cloud is a key component of the distributed cloud concept.
- The solutions that the hybrid cloud provides include service integration, availability/disaster recovery, cross-service security, policy-based workload placement and runtime optimization, as well as cloud service composition and dynamic execution (for example, cloud bursting).

Obstacles

- Hybrid cloud computing is different from multicloud computing, which is the use of cloud services from cloud providers for the same general class of IT service.
- Hybrid cloud computing complements multicloud computing. Although most organizations are integrating applications and services across service boundaries, few large enterprises implemented hybrid cloud computing for a few services.
- Hybrid cloud is different from hybrid IT, where IT organizations act as service brokers as part of a broader IT strategy and may use hybrid cloud computing. Hybrid IT can also be enabled by service providers focused on delivering cloud service brokerage, multisourcing, service integration and management capabilities. These services are provided by vendors, such as Accenture, Wipro and Tata Consultancy Services (TCS), and other service providers and systems integrators.

User Recommendations

- Note that internally run, virtualized environments are often recast as “private clouds,” then integrated with a public cloud environment and called a “hybrid cloud.” Hybrid cloud assumes that the internal environment is truly a private cloud. Otherwise, the environment is hybrid IT.
- Establish security, management, and governance guidelines and standards when using hybrid cloud computing services to coordinate the use of these services with public and private services.
- Approach sophisticated cloud bursting and dynamic execution cautiously, because these are the least mature and most problematic hybrid approaches.
- Create guidelines/policies on the appropriate use of the different hybrid cloud models to encourage experimentation and cost savings, and to prevent inappropriately risky implementations.
- Coordinate hybrid cloud services with noncloud applications and infrastructure to support a hybrid IT model.

Gartner Recommended Reading

Top Strategic Technology Trends for 2021: Distributed Cloud

‘Distributed Cloud’ Fixes What ‘Hybrid Cloud’ Breaks

Predicts 2023: The Continuous Rising Tide of Cloud Lifts All Boats

Leverage Platform Engineering to Scale DevOps Platforms Into Hybrid Cloud

Cloud Center of Excellence

Analysis By: Lydia Leong

Benefit Rating: High

Market Penetration: More than 50% of target audience

Maturity: Early mainstream

Definition:

A cloud center of excellence (CCOE) is a centralized enterprise architecture function that leads and governs cloud computing adoption within an organization. Its role is to enable and empower the organization on its cloud journey, not to execute cloud implementation itself.

Why This Is Important

A CCOE is an effective way to drive enterprise-scale cloud adoption that results in positive business outcomes. As an enterprise architecture (EA) function, it leads organizationwide cloud governance and brokerage, and guides cloud transformation. Although the CCOE is responsible for developing cloud computing policies, it primarily influences, rather than controlling. It enables and empowers the teams that are implementing the cloud journey with the guidance they need for successful execution.

Business Impact

The CCOE drives strategic cloud adoption through three core functions — cloud brokerage, cloud governance and cloud transformation. It serves as an internal cloud consulting practice that delivers cloud architectures and recommended solutions. It partners with the sourcing team to provide cloud vendor management, including cloud cost governance. It raises the organization's level of cloud expertise and supports transformation efforts through its leadership of a cloud community of practice.

Drivers

- The impetus to form a CCOE generally comes from the realization that the organization needs to mature or scale its cloud adoption, or to ensure that cloud computing delivers desired business benefits. Specifically, the organization needs to ensure that the “path of least resistance” for cloud use is also the path that is well governed and meets the organization’s security and regulatory compliance requirements.
- An organization needs a guide on its cloud journey, as cloud use grows within the organization and the organization discovers the best practices for cloud usage that are specific to its business needs. The CCOE, governance guidelines and guardrails, and solutions must evolve with the business and its cloud use.
- Many businesses need help with cloud-enabled digital transformation, ideation of new cloud-enabled business innovations and “evangelism” to encourage cloud adoption.
- The creation of a CCOE is strongly encouraged by many cloud providers, as well as cloud managed service providers (MSPs). However, cloud providers often encourage the creation of vendor-specific CCOEs, rather than broad CCOEs that cut across IaaS, PaaS and SaaS lines. MSPs often promote CCOEs because they tend to result in significant managed or professional services revenue.
- Some organizations initially adopt a cloud operating model that uses a consolidated cloud competence center team that blends all functions necessary for cloud implementation. Such organizations often discover that they need to separate cloud governance and other cloud enterprise architecture capabilities from cloud implementation.

Obstacles

- The CCOE needs the sponsorship and mandate of the CIO or another C-level executive to be effective.
- Many organizations make mistakes in setting the structure and mission of the CCOE, resulting in a failure of the CCOE to make the desired business impact. The CCOE is fundamentally a business-outcome-driven enterprise architecture function, not a sourcing or infrastructure and operations function.
- The CCOE is typically small, and its effectiveness depends on educating and influencing others throughout the organization who are actually implementing the use of cloud services.
- Staffing the CCOE is often challenging, even though it is typically small. There is significant labor market competition for this skill set, and it may, therefore, be necessary to open headcount for cloud architects well in advance of demand. Understaffing the CCOE can be a major threat to its influence and success. However, the CCOE should not be fully outsourced, due to its strategic nature.

User Recommendations

- A CCOE is an EA function that should be led by a chief cloud architect and staffed by cloud architects. Some organizations hire an individual contractor with very strong knowledge of their primary cloud services to lead their CCOE, but such an individual needs an enterprise architect partner with very strong knowledge of the business and the necessary cross-functional relationships. In most cases, these are full-time, senior-level, individual-contributor roles.
- The chief cloud architect should also lead a cross-functional cloud computing advisory committee that contains representatives from the business; technical end-user teams (such as the application development teams); infrastructure and operations; and sourcing, security, compliance, risk management and legal teams. This committee is primarily concerned with strategy and policy.
- Neither the committee nor the CCOE should be a cloud implementation function. That role is typically occupied by a cloud operations function.

Gartner Recommended Reading

How to Deploy a Cloud Center of Excellence

Infographic: Gartner's Reference Cloud Operating Model

How to Design a Cloud Operating Model and Build a Cloud Center of Excellence

Solution Path for Implementing a Cloud Center of Excellence

Innovation Insight for the Cloud Center of Excellence

CIPS

Analysis By: Sid Nag

Benefit Rating: High

Market Penetration: More than 50% of target audience

Maturity: Mature mainstream

Definition:

The cloud infrastructure and platform services (CIPS) market is where cloud providers offer infrastructure as a service (IaaS) and platform as a service (PaaS) capabilities in an integrated manner. The degree of integration between IaaS and PaaS may vary, but it includes the use of a single self-service portal and catalog, shared identity and access management, a single integrated low-latency network, unified security, unified monitoring, and unified billing.

Why This Is Important

- Customers are looking for integrated platforms to simplify development, deployment and operations.
- CIPS offerings are the most complete cloud platforms in the industry, thereby driving significant market consolidation.
- Independent software vendors (ISVs), systems integrators (SIs) and management service providers (MSPs) have embraced the leading CIPS platforms, making them the foundation for most organizations' cloud operations.
- Workloads of today are complex, and cloud providers are addressing the problems by offering CIPS platforms.

Business Impact

A well-functioning CIPS will offer enterprises a more natural, flexible and comprehensive cloud computing environment for their workloads, thereby addressing today's IT needs from an application and data perspective.

Drivers

- The appeal for CIPS is not necessarily in best-of-breed offerings, but in the unification and integration of platform capabilities across these services enabling broad support of workloads ranging from ERP to cloud-native.
- Most customers that use a CIPS from a hyperscale provider, such as Amazon Web Services or Microsoft Azure, have adopted a blend of the provider's IaaS and PaaS capabilities. Indeed, the availability of this broad portfolio of services is a key aspect of choosing a strategic cloud platform provider.
- Hyperscale CIPS providers deliver PaaS services that are well integrated with their IaaS services so that both are easily used together. As a customer, whether you are using PaaS services or IaaS services, they are built on a common substrate. The combination of these services means you are making a strategic bet on the cloud provider.

Obstacles

- Public CIPS markets have consolidated around the market leaders which results in limited options and choices for the buyer.
- IaaS-only or PaaS-only cloud providers will continue to exist, but only as secondary cloud providers compared with CIPS providers.
- This, in turn, could make it a market dominated by a handful of cloud providers, which could stifle competition and drive stand-alone cloud providers out of the market.
- The limited set of hyperscale cloud provider options may limit options or create concentration risks for customers.
- The complexity and level of investment required to offer a full, integrated portfolio of multifunctional PaaS and IaaS services have limited the vendor options in this market to a handful of hyperscalers. Some hyperscalers will form ecosystems, enabling smaller PaaS specialists to be included in this market. However, the maturity of this technology will be primarily dependent on the capabilities of the hyperscalers.

User Recommendations

- Use CIPS in cloud-native and legacy migration projects to expand your design and deployment options and reduce complexity. This may involve using capabilities from multiple cloud providers.
- Prioritize consolidating systems on a hyperscaler CIPS offering when you are operating and governing fleets of applications at enterprise scale. This improves your economies of scale, skills and resources through standardization and consistency across your company and industry.
- Treat integrated CIPS providers as long-term strategic technology providers to your organization. Work to optimize costs, limit contractual risk and maintain failover and portability options for business-critical workloads.
- Focus on those services that are multicloud and can be colocated with multiple larger suites of CIPS capabilities as not all services of the providers are of same maturity, functional completeness or quality.
- Invest to maintain an appropriate level of integration across multiple cloud providers for foundational enterprise IT services such as networking, identity and access management, and security.

Sample Vendors

Alibaba Cloud; Amazon Web Services; Google; IBM; Microsoft; Oracle

Gartner Recommended Reading

Magic Quadrant for Cloud Infrastructure and Platform Services

Critical Capabilities for Cloud Infrastructure and Platform Services

What Buyers Want From CIPS Providers

Risk and Opportunity Index: Cloud Infrastructure and Platform Services

Extending the CIPS Business to New Markets and Opportunities

Entering the Plateau

Cloud Managed Services

Analysis By: Craig Lowery

Benefit Rating: High

Market Penetration: More than 50% of target audience

Maturity: Mature mainstream

Definition:

Cloud managed services provide for the day-to-day management of cloud environments. A select set of professional services is typically offered with the managed services to assist with cloud strategy, workload migration, solution architecture and ongoing transformation efforts. Together, cloud professional and managed services are now known as cloud IT services (CITS), though they were simply called “cloud managed services” as characterized initially.

Why This Is Important

Cloud adoption is a critical strategic objective for most organizations, but most need help in achieving it. Because they offer a combination of professional and managed services, cloud managed services providers (MSPs) are now more frequently referred to as cloud IT services (CITS) providers. These professional and managed services help customers progress efficiently through typical cloud adoption patterns and more quickly realize the benefits of an ongoing, evolving operational state.

Business Impact

The primary benefit of cloud managed services is to augment a cloud-adopting organization’s expertise with certified, experienced personnel to provide advice and convey best practices. The secondary benefit is to provide the difficult-to-source tooling and day-to-day management of a highly dynamic operating environment. A long-term benefit of transformative outcomes occurs when organizations work with providers to unlock the disruptive potential possibilities of cloud computing.

Drivers

- As demand for public cloud services has grown steadily, so has the need for professional and managed services to successfully adopt those services.

- Organizations seeking to adopt the public cloud lack the experience, skills, tools and staffing to successfully navigate key milestones: setting strategy, planning, implementation, optimization and ongoing evolution of cloud deployment.
- Strong cloud IT service providers offer cloud capabilities aligned with hyperscale cloud infrastructure and platform services (CIPS) providers, unlocking technology innovations such as artificial intelligence, automation, data services and edge computing.
- The move to more cloud-native solutions and complex deployment scenarios such as distributed cloud, hybrid cloud and multicloud have substantially emphasized the need for professional services expertise.

Obstacles

- Cloud MSPs (CITS providers) have varying levels of capability based on the clouds they support, specific use cases, and the geographies and industries they target, as well as the technologies and personnel roles they use to deliver their services. This can make it difficult to choose the right provider for an organization's purposes.
- Becoming heavily dependent on an MSP can make it difficult to leave it later. MSPs differentiate from each other in how they automate the delivery of their professional and managed services. They often create proprietary tools or "trade secret" integrations of open-source solutions on which a customer can become dependent.
- MSPs face the same challenges as end users in developing and retaining a skilled workforce. Although MSPs are generally better positioned to attract and cultivate those resources, customers may have highly variable service delivery experiences from one interaction to the next or when compared with other customers.

User Recommendations

- Assess providers to ensure the provider has up-to-date expertise and a track record of success.
- Assess providers with capabilities across the adoption spectrum — from initial and ongoing advisory services (design) to implementation services (build) and managed services (run).
- Give first consideration to providers that demonstrate partnership status and accomplishments with the organization's primary public cloud provider.
- Assess providers' expertise and resources by industry, region and country.

- Plan for long-term value by choosing providers that can deliver innovations and support additional use cases and cloud providers as your needs change.
- Assess providers' ability to deliver to the organization's specific hybrid deployment patterns, which range from management of on-premises virtualization farms to distributed container and Kubernetes deployments to distributed public cloud solutions.

Sample Vendors

Accenture; Deloitte; HCL Technologies; Infosys; Tata Consultancy Services; Wipro

Gartner Recommended Reading

Magic Quadrant for Public Cloud IT Transformation Services

Critical Capabilities for Public Cloud IT Transformation Services

Private Cloud Computing

Analysis By: Thomas Bittman

Benefit Rating: Moderate

Market Penetration: More than 50% of target audience

Maturity: Mature mainstream

Definition:

Private cloud computing is a form of cloud computing used by only one organization, or one that ensures an organization is completely isolated from others. As a form of cloud computing, it has full self-service, full automation behind self-service, and usage metering. It does not have to be on-premises, or owned or managed by the enterprise.

Why This Is Important

Cloud services offer many benefits, but isolation may also be required for security or regulatory reasons. Private cloud offers complete isolation, while providing the convenience and ease of use of cloud services. Private and public cloud are at opposite ends of the isolation spectrum. Public cloud providers have offered virtual private cloud, dedicated instances, dedicated hosts and distributed cloud offerings. Thus, there are a variety of isolation choices between public and private cloud.

Business Impact

Organizations that build a private cloud service are emulating public cloud computing providers to acquire similar benefits — mainly agility for new cloud-native applications, and business value and growth. When the goals are IT modernization or efficiency for existing applications, cloud-inspired deployments — that are more customized and less automated — are more appropriate.

Drivers

- **Regulatory or privacy reasons:** Private cloud may be useful where data or applications cannot reside on a public cloud, or need to reside in a specific location or on-premises.
- **Unique cloud service requirements:** Due to specific enterprise requirements (or support of existing applications), public cloud providers may not offer specific capabilities needed by the enterprise.
- **Evolution of virtualization:** Private cloud can be seen as a natural evolution of an existing virtualized environment, virtualizing all infrastructure and providing a service interface.
- **Standardization:** Specific private cloud services can be used to drive users to more standard offerings, further reducing costs and increasing automation.
- **Platform-level services:** Cloud services can provide rapid deployment, agile change, rapid scaling and innovation at the platform level.

Obstacles

- Building a custom private cloud can be very costly and complex. Moreover, most deployments that are called “private cloud” actually do not have cloud characteristics.
- Building an initial self-service offering is one thing, but maintaining and adding new, innovative features is usually untenable for enterprises.
- Private clouds being built to support existing applications require significant nonstandardization, which reduces automation potential.
- Users often have little motivation to move to a more standard model, unless they make fundamental business changes, such as adopting usage-based pricing for services.

User Recommendations

- Rule out public cloud offerings — including forms of distributed cloud — before investigating private cloud.
- Evaluate third-party hosting options and avoid building your own, if possible.
- Choose cloud-inspired technologies rather than true cloud, if IT efficiency or modernization for existing applications is the goal.
- Focus on business and application needs first, and let that determine the cloud service offerings.
- Focus on services that fit the cloud model (standard, high-volume and self-service); those that require agility and horizontal scalability; and usages that might be short-lived.
- Build or buy private cloud services with the potential to interoperate with, integrate with, or migrate to public cloud services in the future. Develop an exit strategy.
- Manage the scope of work. Start small and expand based on the business case.
- Build expertise in managing multicloud by going beyond just the private cloud.

Sample Vendors

HPE; IBM; VMware

Gartner Recommended Reading

Quick Answer: How Do I Obtain Isolated Private Cloud Services?

Differentiate Hosted Private Cloud Offerings Using These 7 Dimensions

Cloud Migration

Analysis By: Craig Lowery

Benefit Rating: High

Market Penetration: More than 50% of target audience

Maturity: Mature mainstream

Definition:

Cloud migration is the process of planning and executing the movement of applications or workloads from on-premises infrastructure to external cloud services, or between different external cloud services. At a minimum, applications are rehosted (moved largely as-is to public cloud infrastructure), but are ideally modernized through refactoring or rewriting, or potentially replaced with software as a service (SaaS).

Why This Is Important

Cloud migration is a necessary step for organizations wishing to maximize the business benefits of their use of public cloud computing services. When applications and data in the organization's private data centers are moved into a public cloud, new opportunities are unlocked for the benefit of the business. A structured, well-informed approach to such a move is necessary to avoid negative impacts such as wasted time and investment, or business disruptions.

Business Impact

- The business is able to access the benefits of public cloud without building all of its cloud application portfolio from scratch.
- Existing applications can be migrated into the public cloud and modified to take some advantage of cloud capabilities, yielding near-term cloud-related benefits.
- Cloud migration enables a business to adopt public cloud with the least disruption by evolving organizational structures, operational capabilities and user experiences over time.

Drivers

- Organizations see benefit in public cloud deployments and seek to move all or a portion of their IT data center deployments there to derive positive business impacts.
- Migration allows an organization to leverage existing deployments in their private data centers rather than building everything anew in the cloud.
- Competitors who built their businesses in the cloud or migrated earlier have cloud-based advantages that the organization must counter as quickly as possible. Conversely, getting to the cloud before competitors can give an organization a competitive cloud-based advantage.

Obstacles

- Most organizations lack the tools, expertise and resources to plan and execute a migration. Although some existing tools and skills can be leveraged, they are usually not sufficient to effect a successful migration.
- Organizations often struggle with the new operational aspects of cloud computing, which can result in technically successful migrations that fail to meet business objectives such as improved agility or more efficient IT spending..
- Emphasis on application modernization as part of migration further confuses the market on best practices and strategies.
- Not everything should be moved to the cloud and most organizations will be in a hybrid deployment for some period of time. There are many approaches to achieving a hybrid deployment and organizations may find it difficult to identify, understand and act on their options.

User Recommendations

- Use an external service provider, such as a cloud IT service provider, to improve the chances of a successful migration. Almost all successful large-scale migrations to public cloud infrastructure and platform services (CIPS) are done in conjunction with a service provider. They provide consulting for strategy and planning, tools, and technical staff to implement the move.
- Set a strategy based on business objectives, provide CSP recommended training (badges) for key personnel, and source migration tools from ISVs or the CSP's native toolset if you are an I&O leader electing to perform your own migration.
- Choose the best approach for modernizing an existing application. Although rehosting without modification might be easiest, it brings far fewer benefits than some degree of modernization or outright replacement. Some rehosting migrations are still done for expediency but expectations of cloud-native benefits have become mainstream.

Sample Vendors

Accenture; Deloitte; HCLTech; Infosys; Tata Consultancy Services; Wipro

Gartner Recommended Reading

Magic Quadrant for Public Cloud IT Transformation Services

Critical Capabilities for Public Cloud IT Transformation Services

How to Choose the Right Approach for Application Modernization and Cloud Migration

Decision Point for Choosing a Cloud Migration Strategy for Your Application

Appendixes

See the previous Hype Cycle: Hype Cycle for Cloud Computing, 2022

Hype Cycle Phases, Benefit Ratings and Maturity Levels

Table 2: Hype Cycle Phases

(Enlarged table in Appendix)

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

Source: Gartner (July 2023)

Table 3: Benefit Ratings

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

Source: Gartner (July 2023)

Table 4: Maturity Levels

(Enlarged table in Appendix)

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

Source: Gartner (July 2023)

Document Revision History

Hype Cycle for Cloud Computing, 2022 - 12 July 2022

Hype Cycle for Cloud Computing, 2021 - 14 July 2021

Hype Cycle for Cloud Computing, 2020 - 31 July 2020

Hype Cycle for Cloud Computing, 2019 - 8 August 2019

Hype Cycle for Cloud Computing, 2018 - 31 July 2018

Hype Cycle for Cloud Computing, 2017 - 1 August 2017

Hype Cycle for Cloud Computing, 2016 - 3 August 2016

Hype Cycle for Cloud Computing, 2015 - 5 August 2015

Hype Cycle for Cloud Computing, 2014 - 24 July 2014

Hype Cycle for Cloud Computing, 2013 - 31 July 2013

Hype Cycle for Cloud Computing, 2012 - 1 August 2012

Hype Cycle for Cloud Computing, 2011 - 27 July 2011

Hype Cycle for Cloud Computing, 2010 - 27 July 2010

Hype Cycle for Cloud Computing, 2009 - 16 July 2009

Recommended by the Authors

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

Understanding Gartner's Hype Cycles

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

Top 4 Trends Are Shaping the Future of Public Cloud

The Cloud Strategy Cookbook, 2023

The Future of Cloud Computing in 2027: From Technology to Business Innovation

Table 1: Priority Matrix for Cloud Computing, 2023

Benefit	*Years to Mainstream Adoption*			
↓	***Less Than 2 Years*** ↓	***2 - 5 Years*** ↓	***5 - 10 Years*** ↓	***More Than 10 Years*** ↓
Transformational	Serverless Infrastructure	Cloud and Generative AI Edge Computing Platform Engineering SASE Site Reliability Engineering	Business-Driven Cloud Strategy Industry Cloud Platforms	
High	Cloud Center of Excellence Cloud Managed Services Cloud Migration	AI Model as a Service API-Centric SaaS CIPS Cloud AI Services Cloud HPC Cloud Resilience Cloud Sustainability Composable Applications Consumption-Based Model Container Management FinOps Hybrid Cloud Computing IoT PaaS Multicloud	Cloud-Native Cloud-Native Application Protection Platforms Cloud Portability Digital Foundation Distributed Cloud Intercloud Data Management	

Benefit	*Years to Mainstream Adoption*			
↓	***Less Than 2 Years*** ↓	***2 - 5 Years*** ↓	***5 - 10 Years*** ↓	***More Than 10 Years*** ↓
Moderate	Private Cloud Computing	Cloud Marketplaces Cloud-Optimized Hardware Multicloud Management Sovereign Cloud	SMP	Quantum Computing as a Service
Low				

Source: Gartner (July 2023)

Table 2: Hype Cycle Phases

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

Phase ↓	Definition ↓

Source: Gartner (July 2023)

Table 3: Benefit Ratings

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

Source: Gartner (July 2023)

Table 4: Maturity Levels

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

Source: Gartner (July 2023)