Hype Cycle for Cloud Computing, 2021

Published 14 July 2021 - ID G00747400 - 123 min read

By Analyst(s): David Smith, Ed Anderson

Initiatives: Cloud and Edge Infrastructure

Cloud computing is a mainstream computing model that has been validated as a dependable foundation for delivering IT capabilities. This Hype Cycle outlines the key cloud technologies in use today and the innovations that are emerging to support future needs.

Analysis

What You Need to Know

Although cloud computing is not a new topic, industry hype surrounding cloud computing persists. Cloud computing lived up to its potential when tested by the sudden shift in technology and business imperatives driven by the COVID-19 crisis. The unexpected surge in demand for cloud services pushed hyperscale cloud offerings to their limits. In most cases, the cloud services performed well in maintaining service delivery, even in the midst of unprecedented shifts in demand.

The question of whether or not the cloud services model is viable has been answered. It is.

New models leverage cloud capabilities and make use of container and serverless technologies, as organizations seek the pure benefits of cloud by using cloud-native architecture, applications and operating models. Increasing adoption of low-code platform technologies vindicates the vote of confidence for cloud computing by the mainstream business and technology organizations.

Interest in cloud capabilities is moving from central, public cloud data centers toward diverse and distributed environments. Cloud computing and edge computing are two strong partners in delivering the combined power of centralized computing through public cloud services and dispersed computing to the customer data centers and at the edge. New distributed cloud models, delivered through a collection of offerings from the cloud providers — for example, Amazon Web Services (AWS) Outposts, Google Anthos, IBM Satellite, Microsoft Azure Stack Hub and Oracle Cloud@Customer — further extend public cloud capabilities to virtually any location.

Terminology such as hybrid, multicloud, cloud-native, edge computing and distributed cloud describes cloud computing trends and hyped topics, and more terms are likely to follow. CIOs, CTOs and enterprise architects should use this Hype Cycle to understand the continuing evolution of cloud computing and to guide their strategic planning process in pursuit of the full benefits of cloud.

The Hype Cycle

The shift from traditional IT models to cloud computing continues, as demonstrated by the seemingly endless growth of the leading cloud providers. Cloud computing is regularly used as the foundation for business-critical solutions leveraging the flexible and innovative services delivered through cloud models.

As cloud adoption expands across the business organization, the services are separating into two categories:

- The infrastructure as a service (IaaS) + platform as a service (PaaS) offerings target technology providers, advanced IT projects and startups prioritizing designs of unique or highly tuned technology patterns
- Software as a service (SaaS) + PaaS services target business units building up from their SaaS experiences.

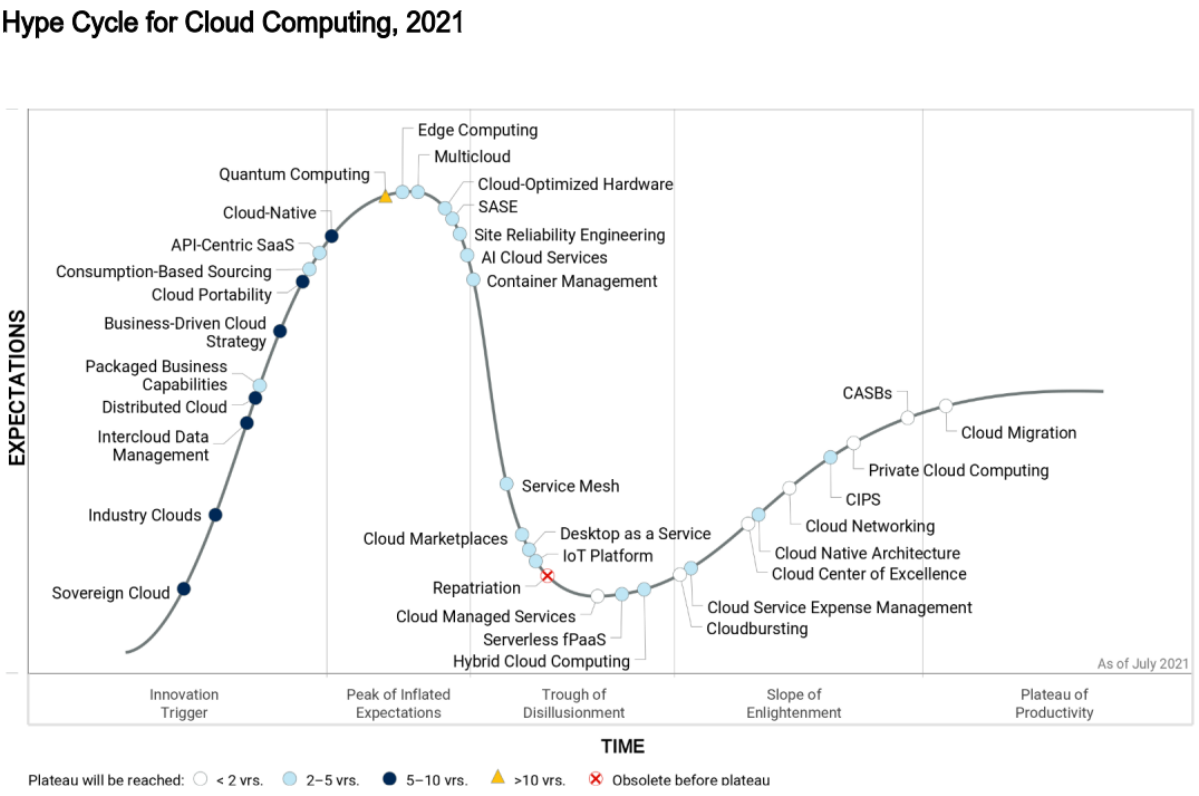
Some major cloud service providers specialize (e.g., AWS for IaaS+PaaS and Salesforce for SaaS+PaaS), while others invest in a platform continuum across both models (Microsoft).

Cloud-native should not be confused with one use of the term cloud native architecture, which appears later in the Hype Cycle. They are near the top of the peak, and both suffer from inconsistent uses that increase confusion and hype. In contrast, hybrid scenarios, including hybrid cloud and hybrid IT (the combination of cloud and noncloud environments), are climbing out of the Trough of Disillusionment and entering the Slope of Enlightenment, illustrating the increasing maturity of hybrid solutions. As mainstream cloud service offerings move into the Plateau of Productivity, cloud foundations enable and stimulate a next wave of innovations. Cloud-based solutions, including the Internet of Things (IoT), artificial intelligence (AI), quantum computing and edge computing, are emerging as best-in-breed approaches to solving a new class of problems, using the power of cloud computing models.

Several new innovation profiles were added this year to represent new cloud-enabled composability scenarios and new types of cloud-related computing approaches found in on-premises systems delivered and billed on a consumption basis. In addition, new hardware innovations, primarily driven by public cloud platform providers, are optimizing hardware platforms for the delivery of optimized and efficient cloud applications. More innovation in cloud-native application architecture and operations is likely as organizations seek the full benefits of cloud. In fact, cloud-native approaches will be a key factor in the success of new technology initiatives as hype associated with cloud continues at its current, elevated status.

In the interest of simplification, several broad category innovation profiles (e.g., SaaS, PaaS, etc.) have been removed.

Figure 1: Hype Cycle for Cloud Computing, 2021



Source: Gartner (July 2021)

Downloadable graphic: Hype Cycle for Cloud Computing, 2021

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 move through the Hype Cycle at an accelerated rate. The transformational nature of cloud-based solutions spawns an ever-increasing collection of innovations as illustrated by the innovation profiles featured on the Innovation Trigger in the Hype Cycle.

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

The relative impact of cloud and cloud-related technologies is high and is often transformational. Organizations building on a cloud foundation will embrace transformational change more quickly and effectively than organizations bound to traditional IT environments.

Table 1: Priority Matrix for Cloud Computing, 2021

(Enlarged table in Appendix)

Benefit	Years to Mainstream Adoption			
↓	Less Than 2 Years ↓	2 - 5 Years ↓	5 - 10 Years ↓	More Than 10 Years ↓
Transformational	CASBs	Edge Computing SASE Site Reliability Engineering	Business-Driven Cloud Strategy Industry Clouds Sovereign Cloud	
High	Cloud Center of Excellence Cloud Managed Services Cloud Migration Cloud Networking	AI Cloud Services API-Centric SaaS CIPS Cloud-Optimized Hardware Cloud Service Expense Management Container Management Desktop as a Service Hybrid Cloud Computing IoT Platform Multicloud Packaged Business Capabilities	Cloud-Native Cloud Portability Distributed Cloud Intercloud Data Management	Quantum Computing
Moderate	Cloudbursting Private Cloud Computing	Cloud Marketplaces Cloud Native Architecture Consumption-Based Sourcing Service Mesh		
Low		Serverless fPaaS		

Source: Gartner (July 2021)

On the Rise

Sovereign Cloud

Analysis By: Rene Buest, Tiny Haynes, Neville Cannon, Gregor Petri

Benefit Rating: Transformational

Market Penetration: 1% to 5% of target audience

Maturity: Emerging

Definition:

Sovereign cloud is the provision of cloud services within a single geography meeting data residency and legislative requirements. Sovereign cloud helps ensure that data remains free from external jurisdiction control and provides protection from foreign legislatively enforced access. Countries engage a sovereign cloud to achieve digital and data sovereignty to provide rules and legal requirements to apply data protection controls, residency requirements, protectionism and intelligence gathering.

Why This Is Important

The importance of digital sovereignty has risen in step with growing discord within global economics, protecting intellectual property, expanding privacy legislation and the desire to be more self-sufficient due to the dominance of a small number of large Chinese and American technology and service providers. The public sector recognizes the value of the digital economy and seeks to develop infrastructure and ecosystems capable of delivering a digital citizen experience while maintaining autonomy.

Business Impact

Legislative mandates can and are being applied to limit the ability to use multinational vendors' services. This clearly impacts current investments and potential sales growth. National vendors may view the changing legislative landscape as a catalyst for further investment and growth. As a result, end users could find themselves in a regulated/fragmented market without access to the software and services that they need to support their ongoing digital business initiatives and drive innovations.

Drivers

- Digitization initiatives need secure and reliable access to data sources and the ability to contextualize and aggregate data from a large number of internal and external data sources. Platform businesses like Alibaba Group, Alphabet, Amazon or Tencent are beneficiaries of this.
- Platform businesses are heavily dominated by U.S. and Chinese companies due in part to their individual market sizes and the common language and currency. Other countries, especially in Europe, lag behind and have already lost national or regional control in core technology areas, such as e-commerce, microprocessors, 5G infrastructure and smartphones, which are necessary to build and run platform businesses. The same applies to the foundational technologies to drive digital transformation and build digital business models or a platform business.
- The market for digital and cloud technology and services is dominated by the U.S. and Asian technology and service providers. As a result, European companies mainly have to access non-European services and technology to build and run digital business models. Hence, data is being stored within non-European cloud and digital service providers, which creates political uneasiness.
- As digital services become increasingly important and system-relevant, companies and regional trade bodies worry about retaining control over their data to stay compliant with local regulations.
- Some more regulated industries and governments are particularly concerned by the U.S. and Chinese legal frameworks that might allow government access under specific circumstances to customers' data.
- In addition, dependence on non-local providers of cloud infrastructure and platform services also comes with economic concerns, such as providers not paying appropriate taxes on transactions conducted within a country or region.

Obstacles

- The range of services and capabilities of hyperscale cloud providers far exceeds pure virtualized infrastructure. Considerable technical obstacles exist if sovereign clouds are expected to deliver the maturity and level of scalability and functionality of hyperscale competitors.
- Too few skilled engineers exist to replicate the design capabilities across multiple countries, simultaneously. With lower levels of skills being available, security and operational maturity will be compromised, potentially leading to greater security and failure risks.

- To date, no non-U.S. or Chinese provider shows the capabilities to compete against one of the hyperscale providers. The market dynamics make it almost impossible for a national vendor to become a strong competitor. Local vendors typically invest significantly less than global players in new infrastructure and find themselves essentially playing catch-up.
- Due to lack of competition, companies will choose to build on technology platforms owned by providers from outside their country.

User Recommendations

- Subject proposals for sovereign cloud to the same level of risk assessment that current cloud computing providers are subjected to. Do not assume that the sovereign cloud conveys any additional security measures in itself.
- Make explicit decisions about your organization's digital sovereignty and track the cloud climate change. Base your plans on the assumption that changes in global cloud climate will potentially disrupt your business.
- Explore evaluating local 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.

Gartner Recommended Reading

Market Trends: Europe Aims to Achieve Digital Sovereignty With GAIA-X

Tech Providers 2025: Strategic Responses to Disruption From Geopolitics and World Events

Tech Providers 2025: Strategic Impacts to the Competitive Landscape

Industry Clouds

Analysis By: Gregor Petri

Benefit Rating: Transformational

Market Penetration: Less than 1% of target audience

Maturity: Embryonic

Definition:

Industry clouds leverage underlying cloud services to offer business and technical capabilities that are specifically relevant to an identified vertical industry. Industry clouds aim to address the unique and evolving functional, regulatory or technical requirements and use cases of the addressed vertical industry and offer a whole product.

Why This Is Important

Cloud adoption has been growing at record pace, but total cloud spend currently makes up only about 10% of global enterprise IT spend. The majority of organizations use one or more cloud services in a variety of areas. But broader adoption within enterprises will require more vertical-targeted whole-product solutions that follow defined industry scenarios and process models, rather than technology-oriented solutions that enterprises have to largely configure and integrate themselves.

Business Impact

Industry clouds will have a lasting impact on cloud customers, blurring the lines between established cloud services such as infrastructure as a service (IaaS), platform as a service (PaaS) and SaaS, and also between cloud providers and cloud SIs and MSPs. The market is taking a classic page from Geoffrey Moore's "Crossing the Chasm," by creating whole-product offerings that cater directly to the established needs of vertical industry enterprises.

Drivers

- Industry clouds have the potential to cover the full functional breadth required for a specific industry, from top (with a suite of applications) to bottom (middleware and infrastructure). So far, organizations bought cloud services as either specific point solutions, in the form of somewhat siloed top to bottom SaaS applications, or as an alternative for a horizontal layer of their technology stack, in the form of IaaS or PaaS services.
- Today, industry clouds are largely being initiated and created by large cloud and software providers, although we see some industry enterprises considering creating an industry cloud as the basis for a manufacturing, food or pharma ecosystem.
- Enterprises can gain value from industry clouds through shared best practices and thought leadership offered through vertically specialized go-to-market and implementation teams; compliance of the infrastructure platform with industry-specific regulations, such as HIPAA or FedRAMP; analytical capabilities to mine the data from their existing applications; industry-specific functionality applications and collections of industry-specific functional building blocks available in industry cloud marketplaces; choosing a combination of the above as a (pre-)composed industry solution.
- Industry clouds create value for enterprises by bringing traditionally separate purchased solutions together in a preintegrated solution. This can simplify the sourcing, implementation and integration process, but enterprises need to realize that providers will initially build out their offering from the infrastructure, application or analytics solutions they historically offered. Leaders in this space are expected to leverage composable cloud and edge approaches to create more holistic and comprehensive industry offerings, which enterprises will be able to recompose to meet unique or special requirements.

Obstacles

- Industry clouds are at risk to follow the same path as community clouds, such as dedicated government clouds. These provided a walled garden where providers added specific vertical functionality, which often led to breaking the compatibility and upgradability with the cloud it was derived from. This left enterprises on a long-term unsupported or unsupportable fork or copy of the cloud of their choosing.
- Initially, industry clouds will cater to the needs of individual enterprises, but to reach their full potential, they will evolve into something best described as ecosystem clouds. Enterprises can leverage these ecosystems by participating in shared (business) processes, such as shared procurement, shared distribution, shared payment procession, and maybe even shared R&D and innovation. This is a step beyond the initial sharing of infrastructure and technology that clouds traditionally brought to the table.

User Recommendations

- Assess the industry-specific features promoted by cloud providers, and distinguish between real technology or functionality offerings versus marketing messages.
- Take into account that vendors choose different paths to add industry value: some focus on compliance with industry regulations, some on adding analytical capabilities to existing or partner-sourced functionality, others invest in building or acquiring specific functional building blocks. Longer term, we expect composability to play a significant role in creating more comprehensive and adaptable industry clouds.
- Select industry cloud offerings that satisfy both your IT/cloud requirements and your vertical industry end-user or line-of-business requirements
- Establish communications with current application and infrastructure providers about which industry cloud ecosystems they plan to support or envision to become part of in the future.

Sample Vendors

Amazon Web Services; Google; IBM; Infor; Microsoft; Oracle

Gartner Recommended Reading

7 Elements for Creating a Pragmatic Enterprise Cloud Strategy

Future of Applications: Delivering the Composable Enterprise

Predicts 2021: Navigating Through the Changes for Vertical Industries

Gartner's Vertical Strategy Framework: Your Roadmap for Successful Industry Go-to-Market Strategies

Tech Providers 2025: Strategic Impacts to the Competitive Landscape

Intercloud Data Management

Analysis By: Adam Ronthal, Donald Feinberg

Benefit Rating: High

Market Penetration: 1% to 5% of target audience

Maturity: Emerging

Definition:

Intercloud distributed data is the process of actively managing data in multiple cloud providers as part of a cohesive application and data management strategy. It relies 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, 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 (i.e., multicloud managed service). As the center of gravity of data shifts to the cloud (or more likely to multiple clouds) data and analytics leaders will seek out means to unite that data in a logical and cohesive consumption tier.

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, there are several drivers for adoption, including:

- Regulatory requirements: Some industries are starting to mandate the use of multiple clouds for resilience and availability reasons.
- 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.

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.

Obstacles

- Intercloud data management typically involves investment in highly specialized technologies from relatively niche vendors. This must be weighed against existing strategic relationships with large cloud service providers, which generally do not support intercloud capabilities broadly at this time. Because the supporting vendors tend to be specialized, most applications will need to be custom-built and written with specific intercloud vendor capabilities in mind. This adds risk because switching costs may be high and technically difficult.
- Any intercloud data management application will also be subject to the laws of physics. They 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 integrated in storage and/or logically to achieve maximum business value.

User Recommendations

Weigh trade-offs in optimization and flexibility in your design decisions. There are three primary deployment options for intercloud data management:

- **Object store:** If data is distributed at the cloud object store (COS) layer, there is not a single service — native or third-party — that cannot read and write from the local COS. This enables diversity of choice in selecting a data management offering for the last-mile delivery.
- **DBMS:** Also called "distributed SQL." The database management system distributes and manages data in a geodistributed cluster. Nodes can reside in multiple clouds and/or on-premises. This approach can support local, low-latency read/write apps with global read capabilities, and data sovereignty enforcements.
- **Application:** This can be thought of as an after-the-fact solution. Data may already be in multiple clouds and require a means of bringing it together. This approach essentially defers data integration to the point of consumption.

Sample Vendors

Cockroach Labs; DataStax; Denodo; HPE; MariaDB; MongoDB; NuoDB; WANdisco; Yugabyte

Gartner Recommended Reading

The Future of Cloud Data Management Is Multicloud

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

Understanding Cloud Data Management Architectures: Hybrid Cloud, Multicloud and Intercloud

Data and Analytics Essentials: Cloud

Distributed Cloud

Analysis By: David Smith, Daryl Plummer, Milind Govekar

Benefit Rating: High

Market Penetration: 1% to 5% of target audience

Maturity: Emerging

Definition:

"Distributed cloud" refers to the distribution of public cloud services to different physical locations, while operation, governance, updates and evolution of the services are the responsibility of the originating public cloud provider.

Why This Is Important

Distributed cloud enables organizations to use cloud computing wherever needed with consistency while the cloud service provider retains the burden and responsibility of managing the technology, implementation and evolution of the capabilities. It gives organizations the flexibility to support use cases that will benefit from the power of cloud services. Organizations can use distributed cloud to reimagine use cases where cloud computing is not currently feasible at the location desired.

Business Impact

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

Drivers

Distributed cloud computing is a style of cloud computing where the location of the cloud services is a critical component of the model. 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. While many people may claim that private cloud or hybrid cloud requires on-premises computing, this is a misconception. Private and hybrid cloud do not require that the private components are in any specific location. With the advent of distributed cloud, location formally enters the definition of a style of cloud services.

Drivers include:

- In hyperscale public cloud implementations, the public cloud is the center of the universe. There has been distribution of cloud services through worldwide regions in public cloud practically since its inception. The major hyperscale cloud providers have different geographic regions around the world, and all are centrally controlled and managed, and provided by the public cloud provider.

- Distributed cloud supports tethered and untethered operation of like-for-like cloud services from the public cloud “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 as well as reduce the risk of global network-related outages.
- Data sovereignty and other regulatory issues.
- Perceived and real security and privacy concerns with off-premises applications and infrastructure.
- Latency needs of IoT/edge applications.
- Disconnected operations.

Obstacles

Customers can’t abandon existing technologies in favor of complete and immediate migration to the public cloud. Sunk costs, latency requirements, regulatory and data residency requirements, and even the need for integration with noncloud, on-premises systems hold them back. Some obstacles include:

- Different approaches to distributed cloud have different value propositions (portability approach, software approach, appliance approach).
- Distributed services are a relatively small subset of the providers centralized services today and will take time to expand. They will never reach 100% parity.
- Distributed cloud running in your data center will inherently have limits to scale and elasticity that 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

- Adopt distributed cloud by overcoming the fear of a single franchise controlling the public cloud and on-premises cloud estates.
- Use the distributed cloud model to prepare for the next generation of cloud computing by targeting location-dependent use cases (such as low latency, tethered scale and data residency) that are enhanced by using a distributed cloud model.
- 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.

Sample Vendors

Amazon Web Services; Google; IBM; Microsoft

Gartner Recommended Reading

Top Strategic Technology Trends for 2021: Distributed Cloud

'Distributed Cloud' Fixes What 'Hybrid Cloud' Breaks

The Cloud Strategy Cookbook, 2021

Packaged Business Capabilities

Analysis By: Yefim Natis

Benefit Rating: High

Market Penetration: 1% to 5% of target audience

Maturity: Emerging

Definition:

Packaged business capabilities (PBCs) are encapsulated software components that represent a well-defined business capability, recognizable as such by a business user and packaged for programmatic access. The definition does not specify the size, functional scope or internal architecture of the implementation, but PBCs are only as useful as they are modular, discoverable, autonomous and ready for composition (integration).

Why This Is Important

As the pandemic disruption forced organizations to seek increased business adaptability, many turned to the model of the composable business. PBCs are a foundational element of the composable application architecture. They act as the building blocks for composition and recomposition of application experiences. When combined with the democratized application composition tools, PBCs empower fast, safe and efficient application and business innovation by the business-IT fusion development teams.

Business Impact

- Adoption of PBCs improves the ability of organizations to involve business professionals in design of application experiences and to make changes to applications by way of recomposition instead of new coding.
- Composable applications, using PBC architecture, equip organizations to innovate faster, safer and smarter, which in turn delivers business resilience, efficiency and adaptability.

Drivers

- Increasing pace of business change, demanding faster, safer and more efficient application innovation.
- Increasing participation of business professionals in software engineering, requiring more business-oriented expression in software modeling, replacing or augmenting the traditional programmatic orientation.
- Increasing democratization of platform technologies, bringing more business professionals to application design work.
- Increasing orientation of vendor applications (SaaS) to API-first and API-only (“headless”) design, leading organizations toward composition and integration instead of the basic customizations of vendor applications.
- Increasing sophistication of agile development practices and product-style application delivery demands more advanced modularity, autonomy, orchestration and discovery for application capabilities.

Obstacles

- Lack of clarity in understanding the fundamentals of composable application architecture, which leads to false starts or “composability-washing” initiatives that do not deliver the expected results.
- Lack of democratized composition tools, which leaves too much of the attempted composition initiatives with technology professionals, limiting the direct business professional participation. This in turn generates designs that are less reflective of the nuance of the required business change and compromise the delivery pace and quality of the outcomes.
- Lack of experience operating fusion teams, which reduces their effectiveness and compromises both technology and business aspects of the products.
- Cultural resistance to change, fear of the shifting business priorities and common familiarity bias — all form obstacles to rapid adoption of architecture of composability and the PBCs.

User Recommendations

- Prioritize expertise in API management, event brokering, integration, business-IT collaboration and democratized tooling to achieve preparedness for composable business applications experience.
- Reject any new monolithic solutions proposed by vendors or in-house developers, and plan to renovate or replace the old ones to enable their participation in composition.
- Accelerate product-style delivery of application capabilities, using agile and DevOps techniques over traditional methods.
- Prioritize democratized tools in support of development, integration (composition) and governance of composed application experiences.
- Give preference to vendor offerings that deliver API-first and API-only (headless) application services.
- Transform the IT organization to the role of a partner and strategic guide to business units, trusted to deliver efficient, safe and fast services to help advance organizations’ business objectives.

Gartner Recommended Reading

Innovation Insight for Composable Modularity of Packaged Business Capabilities

Strategic Architecture Roadmap for Composable Enterprise Applications (Presentation)

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

Kick-Start Your Composable Business Journey With 2 Key Strategies

How to Design Enterprise Applications That Are Composable by Default

Business-Driven Cloud Strategy

Analysis By: David Smith, Lydia Leong

Benefit Rating: Transformational

Market Penetration: 5% to 20% of target audience

Maturity: Adolescent

Definition:

A cloud strategy is a concise viewpoint on the role of cloud computing in an organization. Exit strategy is an important part of a cloud strategy. A true cloud strategy is a business-driven cloud strategy. It is different from other efforts often referred to as strategy that in many cases are better described as cloud adoption or migration plans.

Why This Is Important

Every business requires a cloud strategy, regardless of where it is in its cloud journey. Organizations that lack a cloud strategy don't achieve as much from their use of cloud computing as organizations that have one. Devising a cloud strategy leads to many secondary questions about principles and prioritization. Effective 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. Such questions are becoming more important as they lead to others about principles and prioritization. Cloud strategies need to be aligned with other strategies (for example, data center, security and architecture). Your organization's cloud strategy should enable cloud adoption/migration (tactical implementation plans) that takes the strategy and transforms it into specific actions and digital transformation initiatives.

Drivers

- Organizations are looking for more coherent approaches to cloud usage. They anticipate both the benefits and potential downsides of cloud use, attempting to exploit the benefits while minimizing the downsides. An organization is not fully prepared to meet its needs if it doesn't have a common strategy to draw on and to consult for priorities, with the other subdisciplines of IT.
- Although cloud has been around for over 12 years, there are still disputes among various constituencies about what cloud means. This is exacerbated by vendors that have their own agenda. Business-driven cloud strategy helps to reach a common understanding and language.
- There exists a need to clarify the target business benefits and challenges to address, and establish a repeatable process for evaluating where and how to adopt cloud computing for particular use cases.
- Cloud repatriation (moving workloads back from a public cloud) is rare, but it's vital to have an exit strategy that describes dependencies and choices, even though you may never use it. Several regulators (mainly in the EU and focused on financial services) now mandate a cloud exit strategy, which is a big part of business-driven cloud strategy.

Obstacles

- Need to align cloud strategy with existing strategies (such as those for security, data center, and development and architecture).
- Cloud myths or common mistakes like 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.
- On the organizational front, 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 preclude plans for a business-driven cloud strategy.

User Recommendations

- Maximize the benefits from your use of cloud computing by creating a cloud strategy. Make it a living document that provides a concise view on the role of cloud computing in your organization.
- Align your cloud strategy with other strategic plans (for example, those for data center, security and architecture). Do so by communicating and negotiating with all stakeholders.
- Plan for your cloud strategy to be the launching point for all subsequent cloud activities, such as architecture, assessment, migration and operations. Achieve this by keeping those activities in mind when devising your cloud strategy.
- Safeguard your organization from any potential problems, if you subsequently withdraw from the cloud, by including an exit strategy that describes the dependencies and choices involved in cloud computing. Focus your overall cloud strategy and your exit strategy on answering “what” and “why” questions. Cover the answers to “how” questions in a more detailed cloud adoption and/or exit plan.

Gartner Recommended Reading

The Cloud Strategy Cookbook, 2021

A CIO’s Guide to Strategy Development

Align Your Cloud Strategy With the Organizational Strategic Plans

Top 10 Tips for Avoiding the Most Common Mistakes in Cloud Strategies

Innovation Insight for 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 IaaS, PaaS, or SaaS) is important for reducing both vendor lock-in as well as 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, especially when coupled with the myth of container portability, as offering a negotiation advantage when contracting with cloud providers.

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, increases complexity and increases costs.

Drivers

- **Flexibility:** Organizations seek cloud portability in order to be able to potentially replace one provider with another provider (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.

- **Addressing specific risks:** Organizations are shifting away from envisioning an immediate "disaster recovery" cloud exit, in favor of gradual exit over the period of two years or more. 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.

- **Scenario-based plans:** 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 the period of two years or more. Scenario-based cloud exit planning allows organizations to address cloud risks in a more specific fashion. 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 and add unnecessary cost — and create a different, often riskier point of lock-in.

Obstacles

- SaaS applications and SaaS as a platform are normally entirely nonportable. Any exit strategy for SaaS requires completely replacing the application.
- Open source or COTS-based PaaS may be replaced by running that software directly, but the customer must replace the management and automation capabilities delivered by the service. Proprietary PaaS is nonportable and must be replaced by an alternative solution.
- Workloads running in cloud IaaS are portable, but the customer must replace the management and automation capabilities of the service, including revising DevOps and CI/CD toolchains.
- Portability requires 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, its contractual relationships (including relationships with 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, as portability is impractical.
- Treat cloud IaaS and PaaS portability as an application portability challenge, not as an infrastructure lock-in problem.
- Decide how much portability a particular application needs based on business requirements, and then make architectural choices for that application that reflect the portability requirements.
- Focus on scenario-based exit planning and take a risk management approach to cloud portability.
- Find the balance between desired short-term 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: Balance Public Cloud Business Benefits Against Concentration Risk Reduction

A Guidance Framework for Managing Vendor Lock-In Risks in Cloud IaaS

Assessing Kubernetes for Hybrid and Multicloud Application Portability

A Guidance Framework for Architecting Portable Cloud and Multicloud Applications

Consumption-Based Sourcing

Analysis By: Jeff Vogel, Philip Dawson

Benefit Rating: Moderate

Market Penetration: 5% to 20% of target audience

Maturity: Adolescent

Definition:

A consumption-based sourcing model strategy for on-premises data center storage and compute infrastructure is an acquisition, deployment and support model that includes a cloud-like consumption and platform services model with a variable payment tied to measured use.

Why This Is Important

Server and storage vendors have launched or rebranded consumption-based offerings during the past two years, to provide cloud-inspired or alternatives to the public cloud. Program examples include Dell Technologies' Apex, HPE's GreenLake, Pure Storage's Pure as-a-Service and Cisco Plus. These programs align infrastructure costs with resource use, and to shift infrastructure spending from capital expenditures (capex) to cloudlike operating expenditure (opex).

Business Impact

A consumption-based sourcing strategy is a cloud-like managed service offering that:

- May shift maintenance and support costs' responsibility to vendors investing in artificial intelligence (AI) operations that automate operations
- Will preserve cash by avoiding upfront capex in exchange for strategic priorities
- Will shift IT and finance resource budget cycles to a services-based delivery model

- Will provide a more flexible and agile IT operations environment aligned with business demands

Drivers

Many 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 permanence of the cloud
- The need for a more flexible and agile IT operating model
- The massive growth in enterprise data
- The need for an application-aware services delivery model
- Preferences of 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 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

Obstacles

Consumption-based sourcing may:

- Be more expensive than a purchase.
- Be organizationally challenging to implement during long, protracted periods.
- Be unsuitable for IT operations that don't require flexibility to accommodate uncertain growth and variability in forecast demand.
- Require minimum-usage commitment levels, and do not scale down to zero — that is, minimum commitments require that users pay for most of the deployed infrastructure, regardless of what is actually consumed.

- Require three-to-five-year contracts with mandatory 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 metrics.
- Conflict with established industry accounting standards and operational norms.
- Conflict with consumption-based hardware, making software licensing terms impractical.

User Recommendations

By implementing a consumption-based strategy, IT leaders should:

- Adopt a cloud operating model as a platform strategy to shift to an IT-as-a-services-capable organization.
- Organize a joint team approach to include I&O, vendor management and finance to establish a strategic sourcing strategy and implement.
- Rightsize and align IT I&O to business demand.
- 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.
- Align and link consumption-based costs to specific usage at agreed on elastic utilization levels, 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 (HPE); IBM; Lenovo; NetApp

Gartner Recommended Reading

Enterprise Storage as a Service Is Transforming IT Operating Models

Enterprise Storage Is Growing Consumption-Based Pricing Revenue

Create Consumption-Based On-Premises Infrastructure Bundles for Midsize Enterprises

Key Considerations for CSOs Moving to a Consumption-Based Subscription Model

API-Centric SaaS

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

Benefit Rating: High

Market Penetration: 1% to 5% of target audience

Maturity: Emerging

Definition:

API-centric (“headless”) SaaS is a cloud application service that is offered primarily or entirely for programmatic access via APIs or event subscriptions. Some user experience may be provided, but the strategic intent for API-centric SaaS is to be used as a business capability, programmatically packaged for application composition, development and integration.

Why This Is Important

Increasingly, businesses seek a greater role in defining their digital experiences. API-centric SaaS enables customers to define their differentiated application experiences instead of being bound to the user experience packaged by the vendor. API-centric SaaS allows capabilities such as payments or mapping to be used within larger applications. Although as an application (SaaS), the API-centric SaaS is packaged for use in development platforms (PaaS), blurring the boundaries of SaaS and PaaS.

Business Impact

Business organizations that are equipped to capitalize on API-centric SaaS can create new application experiences for their employees and customers through composition of those prepackaged business capabilities, giving them an opportunity for faster, safer and more efficient business innovation. API-centric SaaS is also a base for fast-paced innovation by startups and other software vendors, contributing to the increasing pace of digital business change.

Drivers

- The patterns of software innovation have embraced integration and composition to a degree that application vendors that want to be included must offer their services, at least optionally, as libraries of API-enabled functional capabilities.
- The technology and skills for integration, including management of APIs and event streams, are widespread, allowing some leading businesses to package their business capabilities behind APIs for customers' and partners' customized use. API products serve that purpose.
- The demands for the depth of customization of applications in business organizations have evolved to the point that SaaS vendors must allow rearrangement of their business functionality by the customers. API SaaS serves that purpose.
- Increasing use of digital twin architecture in IoT application design serves as the early experience of utilizing API-packaged business capabilities, preparing organizations' skills and technologies for the use of the similarly designed and encapsulated API-centric SaaS.
- Many older applications are used as if "headless" by businesses accessing them via API and foregoing the vendor-defined user experiences. This prepares organizations' skills and technologies to include the API-centric SaaS capabilities into their application development experience.
- Business application design has become significantly partitioned into the back-end functionality with its APIs and the front-end multiexperience user interface, each side created 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, when compared to the conventional SaaS offerings. There are not enough dedicated development tools or trained software engineering teams to assure the right balance of cost and value of adopting it.
- API-centric SaaS offers business capabilities, typically delivered as a collection of multiple APIs, but most leading marketplace technologies are designed to manage only the individual APIs, making the discovery and governance of such packaged business capabilities more difficult.
- Minimal or fully 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, inhibiting the wide-spread adoption and API-centric SaaS.

User Recommendations

- Give preference to SaaS offerings that expose more of their business capabilities as API and/or event streams.
- Plan for gradual shift of development to composition and integration of API-centric packaged business capabilities.
- Ensure clean separation of the back-end business logic and the front-end user experience in all applications, to maximize future benefits of the composable application experiences.
- Avoid vendor applications that lock your organization into their user experience technology.
- Give preference to low-code and pro-code PaaS offerings that are well-equipped for access to external API and event marketplaces.
- Practice use of API marketplaces in preparation to greater adoption of API-centric SaaS and API products
- Watch for opportunities to experiment by offering some of your business capabilities as priced API products.

Sample Vendors

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

Gartner Recommended Reading

Create API Portals That Drive API Adoption Among Internal and External Developer Communities

Kick-Start Your Composable Business Journey With 2 Key Strategies

Gartner's API Strategy Maturity Model

Top Trends in Cross-Industry Open APIs for Product Leaders at Banking TSPs for 2021

Choose the Right API Monetization and Pricing Model

At the Peak

Cloud-Native

Analysis By: David Smith

Benefit Rating: High

Market Penetration: 5% to 20% of target audience

Maturity: Adolescent

Definition:

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

Why This Is Important

Cloud-native is a very popular term. Depending on what its meaning is, it can be described taking full advantage of cloud capabilities of a cloud provider, or using approaches pioneered in the cloud to deliver benefits wherever needed via specific technologies such as containers.

Business Impact

Cloud-native is a popular and hyped concept today because many organizations using cloud have not fully realized the benefits they expected from cloud. For example, if a traditional, noncloud application is migrated to cloud using a lift-and-shift approach, the application is unlikely to leverage cloud characteristics and deliver the full benefits of cloud. If an application is rewritten to take advantage of cloud capabilities, then it is more likely to deliver the expected cloud outcomes.

Drivers

- The primary driver for cloud-native is a desire to “get the most out of cloud.” As cloud itself means different things to different constituencies, it is not surprising that this can mean different things. What drives people to one or another of these approaches varies.

- Cloud-native, to deliver cloud characteristics, has potential to enable maximum leverage of the cloud technologies and benefits. Note that the two most common meanings in use are quite contradictory. One (CSP-native) is all about using native features and, therefore, locking yourself into the provider. The other (container-native) is all about containers, which don't guarantee portability but are directionally consistent with the goal. There is also another usage — cloud-native architecture — which refers to principles such as LIFESPAR and 12-Factor Apps.
- Cloud-native is a concept that is not binary. Rather, it can be expressed in degrees. The more something aligns with core cloud characteristics, the more we consider it to be cloud-native and the more cloud-native outcomes the thing will produce.
- 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 so. The more it aligns with relevant principles (aka cloud principles), the more cloud-native it is.
- Organizations should make informed decisions regarding the extent to which they invest in cloud-native for traditional workloads and processes. The investment required to move existing applications to become "cloud-native" may not be worth the expense in some cases.

Obstacles

- Cloud-native is particularly 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 what portability can actually be achieved, and at what cost. Otherwise it couldn't be ensured that these features are used "with your eyes open," and that you are aware that 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 further explanation is required in that case.
- When using the term cloud-native, it is, therefore, imperative that there is clarification of which meaning is being used. Sorting out the definitions and being clear about goals are key to leveraging cloud-native.

User Recommendations

- Focus on the outcomes you want from using cloud rather than focusing purely on the definition of cloud-native. The more your use cases align with the core cloud attributes, the more likely you are to recognize the full benefits of using 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 may be.
- Ensure that the supporting tools, processes and operations support the cloud characteristics when building or acquiring cloud-native applications or services. The value of cloud-native applications can be subverted when supporting elements are not cloud-native in their approach.
- Embrace services design to bring you closer to cloud-native outcomes. This can include the use of containers, microservices architecture, serverless design, functions and many PaaS services. Use of these technologies should, however, be a means, not a goal.

Gartner Recommended Reading

The Cloud Strategy Cookbook, 2021

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

Quantum Computing

Analysis By: Martin Reynolds, Matthew Brisse, Chirag Dekate

Benefit Rating: High

Market Penetration: Less than 1% of target audience

Maturity: Embryonic

Definition:

Quantum computing is a type of nonclassical computing that operates on the quantum state of subatomic particles. The particles represent information as elements denoted as quantum bits (qubits). A qubit can represent all possible values of its two dimensions (superposition) until read. Qubits can be linked with other qubits, a property known as entanglement. Quantum algorithms manipulate linked qubits in their entangled state, a process that addresses problems with vast combinatorial complexity.

Why This Is Important

Quantum computing will not displace conventional computers, but will disrupt areas such as organic chemistry (batteries), materials science and cryptography (security) where it will deliver results very hard to achieve with traditional methods. Quantum computing will also advance machine learning and optimization solutions in speed and/or quality.

Business Impact

For businesses where molecular structures such as drug discovery or materials science are important, quantum computers will enable designers to create tailored molecules with valuable and precise effects. These businesses are at risk of long-term disruption by patents developed using quantum computers. Quantum computers will also enable more efficient competitors in industries such as transportation and logistics where optimization can reduce costs and improve performance.

Drivers

- Significant investment by governments, major corporations and startups.
- Proven small-scale results from recently developed quantum systems.
- Demonstrations of foundational quantum technology using electrons, ions and photons.
- Frequent publication of novel algorithms based on quantum computing technology.
- Error correction algorithms and methods are under development.
- In-chamber control electronics are emerging to help scale qubit counts.
- Emerging control software that can abstract quantum hardware from quantum algorithms.

Obstacles

- Current demonstrated qubit technology is too noisy for error correction.
- External control electronics require too many signal lines to scale.
- High-value algorithms may have impractical readout times.
- Quantum computing may lose funding because compelling results remain far off.
- Lack of enterprise readiness to understand and utilize quantum computing.
- Lack of standardization across programming, middleware and assembler levels.
- Fragmented vendor landscape inhibiting customer adoption and engagement.

User Recommendations

- **Hire quantum capable individuals into other roles now:** When quantum computing becomes relevant to your organization, even a few quantum-capable employees will make a material difference to your organization's ability to make use of this tricky technology.
- **Plan for speed-critical quantum optimization projects for 2025 through 2028:** The first applications for quantum systems will be those where speed of optimization is more important than accuracy. For example, truck loading, where any solution has to be delivered before the truck is dispatched.
- **Plan for performance-critical quantum optimization for 2028 through 2030:** By 2028, optimization should be both faster than and superior to conventional technologies. If you use optimization, plan for quantum systems to deliver superior business results.
- **Plan for chemistry and materials science innovations from 2032 onwards:** By 2032, our model shows that quantum computers could deliver otherwise impossible innovations in materials science and biochemistry.

Sample Vendors

1QBit; D-Wave; Honeywell; IBM; IonQ; Microsoft; PsiQuantum; QC Ware; Rigetti Computing; Zapata Computing

Gartner Recommended Reading

Strategy Guide to Navigating the Quantum Computing Hype

Cool Vendors in Quantum Computing

Quantum Computing Planning for Technology General Managers

Edge Computing

Analysis By: Bob Gill, Philip Dawson

Benefit Rating: Transformational

Market Penetration: 20% to 50% of target audience

Maturity: Adolescent

Definition:

Edge computing describes a distributed computing topology in which data storage and processing are placed in an optimal location 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 unacceptable latency and bandwidth requirements, given the massive increase in edge-located data. The edge-computing topology enables the specifics of the Internet of Things (IoT), digital business and distributed IT solutions, as a foundational element of next-generation applications.

Business Impact

Edge computing improves efficiency and cost control through processing close to the edge (e.g., better automation and quality control), and more business opportunities and growth (e.g., customer experience and new real-time business interactions). Early implementations have succeeded in enterprises that rely on operational systems and data outside core IT, such as the retail and industrial sectors.

Drivers

Drivers to the adoption and implementation of edge computing include:

- Growth in cloud adoption has exposed the disadvantages of extreme centralization. Latency, bandwidth requirements, the need for autonomy, and data sovereignty or location requirements may be optimized by placing workloads closer to the edge and data produced at the edge, rather than centralizing in a hyperscale data center.
- Data growth from interactive applications and systems may not be economically funneled into the cloud.
- Applications featuring customer engagement and analysis favor local processing for speed and autonomy.
- IoT use cases are expanding from the industrial sector to other verticals, driving a move toward a hierarchical and distributed model.

Obstacles

- Extreme diversity of devices and application types amplify complexity issues
- Widespread application of the topology and explicit application and networking architectures are not yet common outside vertical applications, such as retail and manufacturing
- Lack of understanding of benefits/use cases
- Lack of standards
- Although the physical infrastructure for edge is maturing rapidly, the overall management and orchestration challenges of distributed applications are beyond vendor-supplied, component management offerings. The tasks of managing, securing, maintaining and updating the physical infrastructure, software and data requires considerable development before management and orchestration can be considered mature.

User Recommendations

- Create and follow an enterprise edge strategy by focusing first on business benefit and holistic systems, not solely pointing to technical solutions or products.
- Establish a modular, extensible edge approach through the use of emerging edge frameworks and architectures, which allow for the mixing and matching of technologies based on enterprise direction, not simply “what comes with the vendor solution.”
- Accelerate time-to-benefit and derisk technical decisions through the use of vertically aligned system integrators (SIs) and independent software vendors (ISVs) that demonstrate an understanding of and ability to implement and manage the full orchestration stack from top to bottom.
- Evaluate the deployment of “edge as a service” options, which promise to deliver “business-outcome-based solutions” that adhere to specific SLAs, while shifting deployment, complexity and obsolescence risk to the provider.

Gartner Recommended Reading

2021 Strategic Roadmap for Edge Computing

Cool Vendors in Edge Computing, 2021

Multicloud

Analysis By: David Smith

Benefit Rating: High

Market Penetration: 5% to 20% of target audience

Maturity: Adolescent

Definition:

Multicloud refers to the deliberate use of cloud services from multiple public cloud providers for the same general class of IT solutions or workloads. 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.

Business Impact

Multicloud provides agility and potential of cost optimization. It can also provide a basis to lower cloud provider lock-in and increase workload migration opportunities.

Drivers

- Enterprises typically start with one provider but, over time, they become concerned about lock-in. So, the first use of multicloud is often procurement-based to encourage competition.
- As multiple cloud providers are in use, the need to manage and govern those services becomes important.
- Eventually, some enterprises get to multicloud architectures. These rely on architectural principles and portability solutions, and can potentially enable even cloudbursting and other dynamic placement efforts. Most deliberate multicloud strategies are designed to take advantage of differentiated capabilities from multiple cloud providers while applications run in a single cloud provider. Some applications may have a multicloud architecture themselves.
- The hype around multicloud (at the Peak of Inflated Expectations) is driving adoption as providers often use this industry buzz term to justify why their offering 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 sometimes thought of as providing a solution to business continuity and disaster recovery scenarios. It is often discussed as part of an exit strategy and planning, as well as regulatory requirements involving managing risk concentration. While there is a connection, multicloud does not by itself solve any of these issues. Rather, an overall architectural approach (which may use multicloud) is key to solutions in these domains. In fact, in some ways, multicloud complicates and makes these solutions more challenging.
- More time, effort and cost are often required to secure and manage multiple providers.
- The scope of leveraging more advanced services from each provider is reduced.

User Recommendations

- Establish security, management, governance guidelines and standards to manage cloud service sprawl and increasing cost, and develop decision criteria to decide placement of services.
- Focus on coordination and strategy across the enterprise to identify the types of services needed to deliver the benefits of a cloud environment.
- Be prepared to incur additional expenses on training and skilled engineers.
- 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, 2021

Technology Insight for Multicloud Networking

Comparing Cloud Workload Placement Strategies

A Guidance Framework for Managing Vendor Lock-In Risks in Cloud IaaS

A Multicloud Strategy Is Complex and Costly, but Improves Flexibility

Cloud-Optimized Hardware

Analysis By: Alan Priestley

Benefit Rating: High

Market Penetration: 1% to 5% of target audience

Maturity: Emerging

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 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 the optimization of the data center infrastructure to ensure ease of deployment and operation. Power, cost and consumption also have a material impact on operational costs. Cloud service providers are working with equipment supply chain to source custom hardware optimized for deployment within their specific infrastructure. This includes servers, networking and storage equipment and, most recently, optimized chip designs.

Business Impact

Cloud-optimized hardware can enable enterprise architects planning large-scale infrastructure deployments:

- Deliver efficiencies not possible when using standard OEM equipment.
- Lower the cost of data center operations.
- Deploy infrastructure targeted to specific workloads.

Drivers

- Hyperscale cloud service providers' requirements for increased operational efficiency and lower component costs in their data center infrastructure.
- Optimized rack and server designs to reduce power and simplify physical installation; for example, the Open Compute Project (OCP) Foundation driven by Facebook.
- The need for network switches and interfaces that support and isolate multiple workloads and users on servers hosting multiple virtual instances.
- Storage controllers tuned to extend flash device life cycles in solid-state drives (SSDs).
- The use of custom chips to establish a hardware-based root of trust to guarantee the integrity of the hardware and firmware; for example, Amazon's Nitro-based services.
- Function accelerators that support offload of storage and network processing from the CPU.
- 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, AWS Graviton processors and Google's Tensor Processing Units (TPUs) for AI workloads.
- Initiatives like OCP Foundation, encouraging original design manufacturers (ODMs), such as Wistron and Inventec, to produce cloud-optimized servers for a wider range of hosting providers, value-added colocation providers, telecom providers, SaaS ISVs and large enterprises to leverage these developments.

Obstacles

- While it is possible for smaller organizations to work with original design manufacturers (ODMs) to access this hardware, ODMs have a very different business model. Many ODMs have minimum delivery quantities of tens of thousands of standard units.
- Although it is possible to specify precisely what the system configurations will be (CPU, memory, storage, network, etc.), 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.

User Recommendations

- Be aware that cloud-optimized hardware is designed for deployment at scale within hyperscale data centers.
- Acknowledge that customized hardware solutions often use “nonstandard” components, impacting 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.
- Track the use of cloud-optimized hardware by cloud service providers. For workloads based on interpreted languages (such as Java or Python, where an Arm-based interpreter is available), evaluate the use of cloud instances based on proprietary CPU designs, such as AWS Arm-based Graviton2 processor.

Sample Vendors

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

Gartner Recommended Reading

Top 10 Strategic Technology Trends for 2020: Distributed Cloud

Evolve Your Infrastructure and Operations Organization to Remain Relevant in the Cloud Era

How to Bring the Public Cloud On-Premises With AWS Outposts, Azure Stack and Google Anthos

Market Trends: Emerging Opportunities for Semiconductor Vendors at the Hyperscale Cloud Service Providers

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

SASE

Analysis By: Joe Skorupa, Neil MacDonald

Benefit Rating: Transformational

Market Penetration: 5% to 20% of target audience

Maturity: Adolescent

Definition:

Secure access service edge (SASE) delivers multiple converged network and security as a service capabilities, such as SD-WAN, SWG, CASB, NGFW and zero trust network access (ZTNA). SASE supports branch office, remote worker and on-premises general internet security use cases. SASE is 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, resilience and security. SASE also dramatically simplifies the delivery and operation of critical network and network security services mainly via a cloud-delivered model. SASE can reduce the number of vendors required for secure access from four to six today to one to two over the next several years.

Business Impact

SASE enables:

- New digital business use cases (such as digital ecosystem and mobile 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

- SASE is driven by enterprise digital business transformation: the adoption of cloud-based services by distributed and mobile workforces, edge computing and business continuity plans that must include flexible, anywhere, anytime, secure remote access and use of the internet and cloud services.
- The need to flexibly support digital business transformation efforts with a zero trust security architecture while keeping complexity manageable is a significant factor for the adoption of SASE, primarily delivered as a cloud-based service (see The Future of Network Security Is in the Cloud). The COVID-19 pandemic accelerated these trends.
- For IT, SASE can reduce the deployment time for new users, locations, applications and devices as well as reduce attack surface and shorten remediation times by as much as 95%.
- Network security models based on data center perimeter security are ill-suited to address the dynamic needs of a modern digital business and its distributed digital workforce. This is forcing a transformation of the legacy perimeter into a set of cloud-based, converged capabilities created when and where an enterprise needs them — that is, a dynamically created, policy-based secure access service edge, or SASE.

Obstacles

- **Organizational silos, existing investments and skills gaps:** A full SASE implementation requires a coordinated and cohesive approach across network security and networking teams.
- **Organizational bias and regulatory requirements that drive continued 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, Russia and the Middle East, where vendors may have limited cloud presence.
- **SASE security services maturity:** For the next several years, SASE capabilities will vary widely. Sensitive-data visibility and control is often a high-priority capability, but it is difficult for most SASE vendors to address. Your preferred vendor may lack the capabilities you require but two-vendor partnerships can be a viable approach.

User Recommendations

- Involve the CISO and network architect when evaluating offerings and roadmaps from incumbent and emerging vendors to ensure an integrated approach.
- Leverage WAN, firewall, VPN refresh or SD-WAN to update network and network security architectures.
- Strive for not more than two vendors for all core services to minimize complexity and improve performance.
- Identify required capabilities for networking and security, including latency, throughput, geographic coverage and endpoint types to develop evaluation criteria.
- Focus on vendors that include CASB if DLP is a priority. They have the most experience in this area.
- Combine branch office and remote access in a single implementation to ensure consistent policies and minimize the numbers of vendors required. Deploy ZTNA to augment or replace legacy VPN to limit investment in legacy technology.
- Consolidate vendors to cut complexity and cost as contracts renew.
- Leverage branch office transformation and MPLS offload to adopt SASE for security services.

Sample Vendors

Cato Networks; Fortinet; Palo Alto Networks; Versa Networks; VMware; Zscaler

Gartner Recommended Reading

2021 Strategic Roadmap for SASE Convergence

Market Guide for Zero Trust Network Access

The Future of Network Security Is in the Cloud

Magic Quadrant for WAN Edge Infrastructure

Magic Quadrant for Cloud Access Security Brokers

Market Guide for Zero Trust Network Access

Market Trends: How to Win as WAN Edge and Security Converge Into the Secure Access Service Edge

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. 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 DevOps 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 reliability of existing products or platforms as well as to create new products or platforms that warrant the investment in reliability.

Drivers

- SRE is intended to help manage the risks of rapid change, through the use of SLOs, “error budgets,” monitoring, automated rollback of changes and organizational learning.
- SRE teams are often involved in code review, looking for problems that commonly lead to operational issues (for instance, an application that does not do log cleanup and therefore may run out of storage).
- They also ensure that the application comes with appropriate monitoring and resilience mechanisms, and that the application meets SRE-approved standards or guidelines set to achieve negotiated SLOs.
- SRE practices are being adopted by organizations that need to deliver digital business products reliably. These practices require a culture that supports learning and improvement, highly skilled automation practices (and usually DevOps), and usage of infrastructure-as-code capabilities (which usually requires a cloud platform).
- SRE also uses automation to reduce manual processes, and leverages resilient system engineering principles and an agile development process that employs continuous integration/continuous deployment (CI/CD).

Obstacles

- Among the biggest roadblocks toward getting started with SRE is the conceptual foundation necessary to obtain buy-in from both management and agile or DevOps teams. SRE is a significant investment, and organizations need to organize implementation based on the mission and on the adequate budget.
- Implementing SRE without an experienced site reliability engineer requires considerable learning and improvement in order to be effective. By nature, SRE embraces risk, and this conceptual mindset is difficult for many traditional employees to completely understand.
- SRE calls for blameless analysis of understanding how teams can learn from a problem, rather than assigning blame. This means understanding what made the problem happen and how to mitigate, and publishing the results.

User Recommendations

- Engage an executive sponsor to support the initiative.
- Demonstrate sufficient value to improve credibility and support; have an acceptable level of risk.
- Build an initial SRE team with a collaborative engineering mindset and a desire to learn and improve, and automate tasks to reduce repetitive manual work.
- Define clear SLOs and service-level indicators (SLIs) to be monitored and reported against.
- Instill collaborative working between site reliability engineers and developers to help them learn how to design and build their product to meet SLOs.
- Drive an iterative approach to start and evolve SRE practices.

Sample Vendors

BigPanda; Blameless; Datadog; New Relic; OpsRamp; PagerDuty; Splunk

AI Cloud Services

Analysis By: Van Baker, Bern Elliot

Benefit Rating: High

Market Penetration: 5% to 20% of target audience

Maturity: Adolescent

Definition:

Artificial intelligence cloud services provide AI model building tools, APIs and associated middleware that enable the building/training, deployment and consumption of machine learning models running on prebuilt infrastructure as cloud services. These services include automated machine learning, vision and language services.

Why This Is Important

The use and sophistication of AI cloud services continues to increase, with vendors competing to become the platform of choice for developers and citizen data scientists. Applications will increasingly use AI cloud services in language, vision and machine learning to help automate and accelerate achievement of business objectives. However, developers struggle with these offerings as current offerings are limited, but improvements in the service portfolios will drive continued adoption.

Business Impact

The impact of AI will extend to the applications that enable business, allowing developers and data scientists to enhance the functionality of these applications. With the incorporation of forecasts, next best actions and other capabilities, including automation of many workflows that are currently handled manually, these AI cloud services will enable advanced applications that improve business performance.

Drivers

- The explosion of data from both internal and third-party sources enable insight that has previously been unavailable to the business.
- Increasing need for human to machine interactions that are based on conversational capabilities.
- A mandate for businesses to automate processes to improve accuracy, improve responsiveness and reduce costs.
- Improving performance of both AI and machine learning models.
- Reduced need for large quantities of data to train models.
- Ease of accessibility for developers and citizen data scientists to AI and machine learning services due to the availability of API callable cloud hosted services will expand the use of AI.
- Multiple use cases spanning language, vision and automated machine learning services.
- Use of automated machine learning to tailor the off-the-shelf services to more precisely address the specific needs of the business.
- A wide range of AI cloud services from both hyperscaler cloud providers as well as specialized providers in the market.
- Business transformation efforts driving the development of applications with enhanced capabilities to improve business operations and workforce productivity.
- Increasing deployments of sensor networks in IoT-based solutions that facilitate data use to drive model development and facilitate proactive response to changes in the data rather than reactive response.
- The emerging AI model marketplaces should help developers adopt those techniques through AI cloud services.

Obstacles

- Lack of understanding by developers and citizen data sciences about these services and how they can be applied to specific business use cases.
- Pricing models for AI cloud services that make it challenging for businesses to determine the costs associated with use of these services.
- Lack of guidance for solutions that utilize multiple services to address specific use cases for developers and citizen data scientists.
- Lack of understanding about how to use automated machine learning to supplement and enhance the standard language and vision services.
- Minimal marketplaces for pre-built machine learning models that could be used by developers and citizen data scientists.
- Serious lack of ModelOps capabilities that contribute to challenges in integration of AI into applications.

User Recommendations

- Choose AI cloud services over building custom models to address a broader range of use cases and for quicker deployment and built in scalability.
- Improve the chances of success of your AI strategy by experimenting with different AI techniques and AI cloud services providers, using the exact same dataset and then selecting one that best addresses your requirements. Consider using an A/B testing approach.
- Use AI cloud services to build less complex models, giving the enterprise the benefit of more productive AI while freeing up your data science assets for higher priority projects.
- Use features like automated algorithm selection, data set cleansing and preparation, and feature engineering for project elements and leverage existing expertise on operating cloud services. This will assist technical professional teams with little to no data science expertise.

Sample Vendors

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

Gartner Recommended Reading

Critical Capabilities for Cloud AI Developer Services

Magic Quadrant for Cloud AI Developer Services

Sliding into the Trough

Container Management

Analysis By: Dennis Smith

Benefit Rating: High

Market Penetration: 20% to 50% of target audience

Maturity: Early mainstream

Definition:

To manage containers at scale, container management provides capabilities such as container runtimes, container orchestration and scheduling, and resource management. Container management software brokers the communication between the continuous integration/continuous deployment (CI/CD) pipeline and the infrastructure via APIs, and aids in the life cycle management of containers.

Why This Is Important

Container runtimes simplify use of container functionality and enable integration with DevOps tooling and workflows. Productivity and/or agility benefits of containers include accelerating and simplifying the application life cycle, enabling workload portability between different environments and improving resource utilization efficiency. Container management makes it easier to achieve scalability and production readiness. It also optimizes the environment to meet business SLAs.

Business Impact

Gartner surveys and client interactions show that the demand for containers continues to rise. This is due to application developers' and DevOps teams' preference for container runtimes, which have introduced container packaging formats. Developers have quickly 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

- Container runtimes, frameworks and other management software provide capabilities such as packaging, placement and deployment, and fault tolerance (for example, clusters of nodes running the application).

- The emergence of de facto standards (for example, Kubernetes) and offerings from the public cloud providers have simplified deploying containers at scale. Many vendors enable management capabilities across hybrid cloud or multicloud environments by providing an abstraction layer across on-premises and public clouds. 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.).
- Data analytics use cases have emerged over the past few years, as have operational control planes that enable the management of container nodes and clusters.
- All major public cloud service providers now offer on-premises container solutions. Independent software vendors (ISVs) are starting to package their software for container management systems.
- Some enterprises have scaled sophisticated deployments, and many more have recently commenced container deployments or are planning to. This is expected to increase as enterprises restart application modernization projects postpandemic.

Obstacles

- 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 their 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.
- More abstracted, serverless offerings may enable enterprises to forgo container management. Among these services are Knative, AWS Lambda and Fargate, Azure Functions, and Google's Cloud Run. These services embed container management in a manner that is transparent to the user.
- 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 PaaS providers' service offerings by experimenting with process and workflow changes that accommodate the incorporation of containers.
- Avoid open-source deployments unless the organization has ample in-house expertise to support.

Sample Vendors

Amazon Web Services (AWS); Google; IBM; Microsoft; Mirantis; SUSE (Rancher Labs); Red Hat; VMware

Gartner Recommended Reading

Market Guide for Container Management

Service Mesh

Analysis By: Anne Thomas

Benefit Rating: Moderate

Market Penetration: 1% to 5% of target audience

Maturity: Adolescent

Definition:

A service mesh is a distributed computing middleware that optimizes communications between application services within managed container systems. It provides lightweight mediation for service-to-service communications, and supports functions such as authentication, authorization, encryption, service discovery, request routing, load balancing, self-healing recovery and service instrumentation.

Why This Is Important

A service mesh is lightweight middleware for managing and monitoring service-to-service (east-west) communications, especially among microservices running in ephemeral managed container systems, such as Kubernetes. It provides visibility into service interactions, enabling proactive operations and faster diagnostics. It automates complex communication concerns, thereby improving developer productivity and ensuring that certain standards and policies are enforced consistently across applications.

Business Impact

- A service mesh helps ensure resilient and secure request-response communication between services deployed in Kubernetes and other managed container systems.
- Service mesh middleware is one of many management technologies that provide software infrastructure for distributed applications deployed in managed container systems.
- This type of middleware, along with other management and security middleware, helps provide a stable environment that supports “Day 2” operations of containerized workloads.

Drivers

- Service mesh adoption is closely aligned with microservices architectures and managed container systems like Kubernetes. Service mesh supports needed functionality in ephemeral environments, such as service discovery and mutual Transport Layer Security between services.
- As microservice deployments scale and grow more complex, DevOps teams need better ways to track operations, anticipate problems and trace errors. Service mesh automatically instruments the services and feeds logs to visualization dashboards.
- A service mesh implements the various communication stability patterns (including retries, circuit breakers and bulkheads) that enable applications to be more self-healing.
- Many managed container systems now include a service mesh, inspiring DevOps teams to use it. The hyperscale cloud vendors provide a service mesh that is also integrated with their other cloud-native services.
- Independent vendors, such as Buoyant, HashiCorp and Kong provide service meshes that support multiple environments.

Obstacles

- Service mesh technology is immature and complex, and most development teams don't need it. It can be useful when deploying microservices in Kubernetes, but it's never required.
- Users are confused by the overlap in functionality among service meshes, ingress controllers, API gateways and other API proxies. Management and interoperability among these technologies hasn't yet been addressed by the vendor community.
- Many people associate service mesh exclusively with Istio, even though it isn't the most mature product in the market and has a reputation for complexity.
- Independent service mesh solutions face challenges from the availability of platform-integrated service meshes from the major cloud and platform providers.

User Recommendations

- Delay adoption of service mesh until your teams start building applications that will get value from a mesh, such as applications deployed in managed container systems with a large number of service-to-service (east-west) interactions.
- Favor the service meshes that come integrated with your managed container system unless you have a requirement to support a federated model.
- Reduce cultural issues and turf wars by assigning service mesh ownership to a cross-functional PlatformOps team that solicits input and collaborates with networking, security and development teams.
- Accelerate knowledge transfer and consistent application of security policies by collaborating with I&O and security teams that manage existing API gateways and application delivery controllers.

Sample Vendors

Amazon Web Services; Buoyant; Decipher Technology Studios; Envoy; F5; Google; HashiCorp; Istio; Kong; Microsoft; Red Hat; Solo.io; Tetrate; VMware

Gartner Recommended Reading

How a Service Mesh Fits Into Your API Mediation Strategy

Assessing Service Mesh for Use in Microservices Architectures

Cloud Marketplaces

Analysis By: Ed Anderson

Benefit Rating: Moderate

Market Penetration: 20% to 50% of target audience

Maturity: Early mainstream

Definition:

Cloud marketplaces are online storefronts through which customers can find and subscribe to cloud service offerings, including IaaS, PaaS and SaaS. Cloud-related IT service offerings, such as consulting or migration services, are now appearing in some marketplaces. Cloud marketplaces are a type of brokerage service, where the marketplace aggregates disparate cloud service offerings (often curated) and presents them as a collection of cloud service offerings.

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 their 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. Cloud marketplaces may be public or private.

Business Impact

As the frequency and volume of cloud service discovery and procurement occurs through cloud marketplaces, they become increasingly important as a technology and architectural control point. Cloud marketplace providers motivate customers to increase their use of cloud marketplaces through discounts, additional services, solution validations and ease of use. CSPs aspire to lead this trend, while third-party cloud marketplaces (such as technology distributors) cater to multicloud requirements.

Drivers

- **CSPs use cloud marketplaces** to highlight their own cloud service offerings and their ISV and IT service partners' offerings. CSPs also use marketplaces to strengthen their ecosystem of solutions, which in turn validates their respective cloud platform. Cloud marketplaces may be disassociated from any specific cloud platform or cloud service and offered by independent cloud service resellers or distributors such as AppDirect and Ingram Micro's CloudBlue. Many traditional software and hardware distributors provide cloud marketplaces for their customers.
- **Cloud service subscribers use cloud marketplaces** to simplify the process of finding, purchasing, deploying and using cloud services. Cloud marketplaces can also be used to isolate and deliver only validated or sanctioned services, which can complement cloud governance policies enforcing the use of only approved cloud services. Cloud service aggregation features, such as unified billing, software license and entitlement management are often included.
- **Third-party cloud marketplaces** will be most popular with smaller organizations that benefit from the value-added capabilities of cloud marketplace providers. Larger organizations will also make use of cloud marketplaces associated with their chosen cloud platform(s).
- **Internal cloud marketplaces** may be established by organizations with sophisticated IT capabilities to facilitate usage of internal and external cloud services, and to provide management and control over the organization's use of cloud services.

Obstacles

- Cloud marketplaces offered by CSPs typically include cloud solutions for the marketplace provider's platform. Using CSP marketplaces may require organizations to support multiple marketplaces when they have multicloud requirements.
- Cloud marketplaces typically charge marketplace solution vendors a commission on marketplace sales, which could impact the pricing of marketplace offerings. For this reason, many third-party marketplace offerings are also available directly from the provider or through the CSP marketplace with BYOL provisions.
- 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.

User Recommendations

Organizations should use marketplaces to expedite the discovery and delivery of solutions that meet both their business requirements and technology preferences.

- Use cloud marketplaces to find solutions that are 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 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).
- 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 including machine learning (ML) modules, prepackaged containers and validated datasets.

Sample Vendors

Alibaba Cloud; Amazon Web Services; CDW; Cloud28+; Google; IBM; Ingram Micro; Microsoft; Salesforce

Gartner Recommended Reading

Navigating Online Marketplaces: Executive-Level Guide

Emerging Technologies: Digital Solutions to Enable Ecosystems and Marketplaces

Create Enterprise Marketplaces to Accelerate Digital Business

How to Derive Value From APIs Using API Marketplaces

11 Imperatives When Building an Enterprise Marketplace

Desktop as a Service

Analysis By: Stuart Downes, Mark Margevicius, Tony Harvey

Benefit Rating: High

Market Penetration: 20% to 50% of target audience

Maturity: Early mainstream

Definition:

Desktop as a service (DaaS) solutions provide a virtualized desktop experience to workers, entirely from a remote hosted location such as the public cloud. DaaS eliminates the need for businesses to purchase the physical infrastructure associated with desktop virtualization, instead functioning through subscription- and usage-based payment structures. DaaS includes provisioning, patching and maintenance of the management plane and resources to host workloads.

Why This Is Important

DaaS provides secure remote access to applications and desktops using a network connection. No data resides on the endpoint, offering a solution that can increase security, redundancy, bursting and performance for remote workers. DaaS offers scalable services, allowing clients to appropriately size and consume their environments hour by hour, day by day, and month by month; however, not all have such granular billing options.

Business Impact

The pandemic highlighted the requirement to deliver secure desktop and application experiences to users in any location:

- Revenues grew by 98% in 2020 compared to 2019 as clients adopted DaaS to secure work-from-home experiences.
- DaaS enables business continuity and anywhere operations for home-based and hybrid home-office operations.
- DaaS enhances security for BYOPC use cases reducing risks for businesses.

Drivers

- Improve security and compliance — centralize data, minimize exposure on notebook PCs.
- Support remote work with no data residing on the endpoint.
- Endpoint computing models that allow any device models and bring your own PC (BYOPC) endpoints.
- Simplified end-user computing operations that are performed in the cloud, especially evident with a highly distributed workforce.
- Business continuity, the main driver for DaaS growth in 2020.
- On-demand desktops.
- A financial model that allows scaling of cloud resources and an opex model.
- Scaling to meet the needs of short-term employees, such as seasonal workers.
- Securely extend services to external contractors, including IT developers and business process outsourcing (BPO).
- Enabling rapid access to systems during mergers, acquisitions and divestitures.
- Rich graphics use cases like engineering, games development, fashion and geographic information systems (GIS) benefit from GPU-enabled workstation class virtual desktops and applications.
- Where supply of devices, or high attrition rates, makes the provision of a physical device difficult.
- Eliminates the need for complex virtual desktop infrastructure (VDI) implementations.
- Enables business expansion to new regions without the need to deploy data centers.

Obstacles

- Cost — In many cases, direct infrastructure cost comparisons show higher IT costs for DaaS compared to VDI or desktop PCs, usually the business case turns positive only when business and users costs are included.
- Multimedia streaming, web meetings and video call performance in DaaS are not equivalent to that of a physical endpoint.
- Applications and data that are landlocked in on-premises data centers that have, or are perceived to have, network-related performance issues when the desktop or application is hosted in the cloud.
- Some DaaS solutions require complex configuration, which, although simpler than VDI, can in some cases require careful configuration and selection of appropriate storage services to ensure a performant DaaS experience.

User Recommendations

DaaS will continue to mature and increase in adoption through 2025. DaaS is yet to move through the Trough of Disillusionment onto the Slope of Enlightenment. Clients should:

- Understand the three DaaS market segments and select a vendor from the appropriate segment: (1) Client-defined DaaS: Clients configure their DaaS experience and manage their workloads. This requires less skill than VDI; *2) Vendor-defined DaaS: The vendor configures the DaaS experience and clients manage their workloads; and (3) Managed DaaS: The vendor configures the DaaS experience and the vendor manages the workloads.
- Choose a DaaS vendor whose services best align to your requirements; even within each segment, there are differences between the services vendors offer.
- Optimize multimedia streaming, web meetings and video calls.
- Select a DaaS vendor that offers billing granularity that you require; some are granular hour by hour, others day by day or month by month.

Sample Vendors

Amazon Web Services; Citrix; Evolve IP; Microsoft; SACA; SimpleCloud; VMware; Workspot

Gartner Recommended Reading

Market Guide for Desktop as a Service

Microsoft's Restrictions for Licensing Windows and Office 365 for VDI/DaaS on AWS and Other Hyperscale Clouds Require Attention

Physical, Virtual and Cloud Desktops: Is a Hybrid Approach Inevitable?

How to Build a Successful Business Case for Desktop Virtualization

IoT Platform

Analysis By: Alfonso Velosa, Eric Goodness

Benefit Rating: High

Market Penetration: 5% to 20% of target audience

Maturity: Adolescent

Definition:

An Internet of things (IoT) platform enables the connection and capture of data from IoT-enabled assets or endpoints to develop, deploy, and manage business solutions that improve operations such as monitoring remote assets or optimizing maintenance. Capabilities include device management, integration data management, analytics, application enablement and management, and security. It may be delivered as edge or on-premises software, or cloud IoT platform as a service, or a hybrid combination.

Why This Is Important

Enterprises continue adding IoT capabilities to assets and products, seeking benefits such as cost optimization, process optimization, improved customer experience, and new opportunities such as product as a service. The sophistication, scale and business value of these interactions call for specialized technology resources, most often implemented as an IoT platform. While all verticals are deploying IoT, spend is highest in asset intensive industries such as manufacturing or oil and gas.

Business Impact

IoT platforms are usually required to implement IoT-enabled assets in order to make better business decisions from the data and information generated by connected products. Goals include:

- Differentiated smart products
- Cost optimization strategies centered on improved maintenance
- Process improvement by using assets at their best state
- Opportunities to sell new services and data products

Drivers

- Proliferation of IoT projects since IoT is widely proven across many industries to improve business outcomes — see Survey Analysis: Focus on Practical Outcomes for IoT Projects.
- IoT platforms are fit-for-purpose PaaS and on-premises software offerings that specifically help software teams to accelerate and improve the quality of IoT products while consolidating and structuring the data.
- Enterprises leverage their IoT assets to drive differentiation, lower costs, improve processes and enhance worker safety.
- Technology providers are driving marketing and sales efforts to engage their customers with IoT platforms. In parallel they invested in improved ecosystems and channel partners to make it easier for companies developing IoT enabled solutions to achieve business value.
- In parallel, technology providers continue to invest in their IoT platform technology to ensure they can deliver business solutions at scale for their customers.

Obstacles

- IoT platforms require extensive customization 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 platforms to achieve business outcomes.
- Many enterprises operate in siloed fashions, adopting different IoT platforms for each use case, limiting their ability to scale, and adding complexity.
- Projects that use IoT platforms drive greater volumes of data, complicating existing processes and overwhelming employees and other stakeholders. They often lack training or process changes to absorb this new data — leading existing systems and people to reject the output of the IoT platform.
- IoT technical complexity, security and integration challenges remain barriers to scale at enterprises.
- Technology providers have yet to develop a clear value proposition and sales strategy that helps their customers leverage their platforms on scaled up levels.

User Recommendations

- Start with smaller IoT projects that help the business unit and IT organization acquire implementation lessons, identify IoT platform strengths and weaknesses, and verify alignment to business and finance KPI requirements.
- Identify the range of IoT projects for your enterprise, and segment them by their focus (internal vs. external), complexity and business objectives. Use these insights to establish a distributed deployment and a platform of platforms architecture.
- Use a skills gap for IoT projects and IoT platforms to build a plan to improve the IT organization's capabilities such as integration or digital twin model development.
- Prioritize vendors you already work with for their IoT platform. Evaluate candidate vendors on their fit-to-your-business objectives and technology. Key evaluation criteria include: proofs of value projects (for tech and business), the ability to drive operational-scale deployments, vertical market expertise and a partner ecosystem.

Sample Vendors

Alibaba Cloud; Amazon Web Services; AVEVA; ClearBlade; COVACSIS Technologies; Detechtion Technologies; Knowledge Lens; Microsoft; Siemens

Gartner Recommended Reading

Magic Quadrant for Industrial IoT Platforms

Critical Capabilities for Industrial IoT Platforms

Use the IoT Platform Solution Reference Model to Help Design Your End-to-End IoT Business Solutions

Use 4 Building Blocks for Successful Digital Twin Design

Repatriation

Analysis By: Ed Anderson, Lydia Leong

Benefit Rating: Low

Market Penetration: 20% to 50% 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 that are not designed for a cloud environment often fail to deliver the full benefits of a cloud migration such as lower costs, better performance and reliability, enhanced security, innovation and other benefits. One approach to remediate these shortfalls is to repatriate the application back to its original environment. While rare, repatriation has been hyped by noncloud providers to undermine cloud models and to reinforce the benefits of traditional environments.

Business Impact

While repatriation is one approach to addressing the shortcomings of a cloud migration, it is typically not the most efficient approach. Most organizations desire the benefits of cloud and, therefore, pursue optimization strategies rather than move applications back to a noncloud environment. Additionally, many cloud migrations are associated with data center divestiture programs, which means the previous environment may no longer exist, which could result in significant expense to instantiate.

Drivers

- The concept of repatriation has achieved significant hype. Multiple entities with a business interest in noncloud technologies have promoted the notion that public cloud services are failing to deliver on their purported benefits, and as a result, customers are returning to on-premises environments. Nevertheless, repatriation hype is declining in the face of the reality that the overwhelming majority of cloud implementations are ultimately successful.
- However, cloud projects may experience temporary failures and setbacks, some of which may result in returning to an on-premises solution 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 IaaS, the most fatal 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 fails to deliver cloud benefits. In such cases, a migration pause may occur while the project is rethought and rescoped. Ungoverned costs or ungoverned self-service can also cause issues that may lead to a slowdown or pause in a shift toward cloud IaaS and PaaS.
- True migration reversals are very rare. The most commonly cited examples are several years old and involve organizations with unique circumstances. While organizations do sometimes rethink their IT direction in a way that may involve less future cloud adoption, few organizations are willing to give up the benefits of the cloud services they already use.

Obstacles

- Repatriation has the potential to create distraction during a time when thoughtful analysis should be devoted to cloud initiatives. Public cloud environments are not well-suited for every application or workload type. Therefore, employ a thorough process of planning and due diligence, using a business-driven cloud strategy, before moving any workload to the cloud to avoid the need for repatriation in the future.
- The process of repatriation assumes the original environment used to host the application before the cloud migration still exists. This is often not the case, which requires an organization to reestablish the data center environment to rehost the application. This runs counter to the strategy for most organizations, which is to embrace cloud where possible. Therefore, most organizations pursue other remediation strategies rather than repatriation.

User Recommendations

When confronted with the topic of repatriation:

- Assess repatriation anecdotes with skepticism, especially when delivered from entities with a bias toward preserving traditional data center businesses. While repatriation may be mentioned in industry discussions, actual repatriation is rare.
- Establish a thorough and complete cloud strategy, including a process for assessing the viability of each application migration initiative. Identify the expected benefits of moving an application to the cloud before the migration actually occurs, and then purposefully pursue those benefits. Proactive planning prevents the eventual need to repatriate an application.
- Employ best practices in tooling and operations to optimize workloads that have already migrated to the cloud. Ongoing optimization should be part of ongoing cloud management practices. Consider cloud-native application design, architecture and operations to get the greatest benefits from public cloud.

Sample Vendors

Amazon Web Services; Cisco; Dell; Google; Hewlett Packard Enterprise; Microsoft; Oracle

Gartner Recommended Reading

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

The Cloud Strategy Cookbook, 2021

Top 10 Tips for Avoiding the Most Common Mistakes in Cloud Strategies

Decision Point for Migrating Your Data Center to Public Cloud IaaS and PaaS

Market Trends: Public Cloud Repatriation Remains the Exception, Not the Rule

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 are IT service offerings that provide for the day-to-day management of, and operational responsibility for, cloud service environments. A select set of professional services are typically offered and highly coordinated with the managed services to assist with cloud strategy, workload migration, solutions architecture and ongoing transformation efforts. Cloud service brokerage is often delivered as a cloud managed service.

Why This Is Important

Cloud adoption is a critical strategic objective for most organizations, but most need help in achieving it. Cloud managed services providers (MSPs) deliver a mix of professional and managed services that guide their customers through typical cloud adoption patterns and into an ongoing, evolving operational state. This improves the organization's degree of success with a faster time to value.

Business Impact

- The primary benefit of engaging a cloud MSP 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 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 MSPs provide 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 hybrid cloud and multicloud have substantially emphasized the need for professional services expertise.

Obstacles

- Cloud MSPs have varying levels of capability based on the clouds they support, specific use cases, 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 MSP for an organization's purposes.
- Becoming heavily dependent on an MSP can make it difficult to leave them 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), 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 the provider's expertise and resources in 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; Bespin Global; Capgemini; Cognizant; Deloitte; Logicworks; Rackspace Technology; Smartronix; TCS; Wipro

Gartner Recommended Reading

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

Critical Capabilities for Public Cloud Infrastructure Professional and Managed Services, Worldwide

Public Cloud IT Transformation Services to Replace Cloud MSP in Magic Quadrant Coverage

Serverless fPaaS

Analysis By: Anne Thomas

Benefit Rating: Low

Market Penetration: 5% to 20% of target audience

Maturity: Early mainstream

Definition:

Function platform as a service (fPaaS), also known as function as a service (FaaS), is a serverless execution platform for event-triggered application components known as functions. Like all serverless platforms, fPaaS enables you to run code without provisioning or managing the underlying system or application infrastructure. fPaaS pricing models allow users to pay in microincrements only for actual usage, rather than preprovisioned resources required to support projected peak loads.

Why This Is Important

fPaaS can deliver significant savings for certain types of workloads via its consumption-based micropricing model. The programming model also enables software engineers to rapidly deploy and configure new functions with little or no assistance from operations teams.

Business Impact

- An fPaaS can offer significant cost savings and virtually unlimited scalability for applications with highly variable capacity requirements.
- Serverless technologies like fPaaS abstract and commoditize infrastructure technologies, increasing developer and operator productivity, enabling organizations to respond rapidly to digital business moments and deliver new applications and features faster.

Drivers

fPaaS is gaining momentum because of:

- Potential cost savings — The micropricing model charges for small increments of compute time per invocation, which can be advantageous for small, spiky workloads. The model is less favorable for large, consistent workloads.
- Rapid solution delivery — The serverless model reduces the amount of work developers and operations teams need to do to build, deploy and configure solutions.
- Integration with hyperscale xPaaS — The hyperscale vendors make it easy to use their fPaaS with their other cloud-native xPaaS offerings.
- Broad use-case support — fPaas can support a broad spectrum of application use cases, from basic websites to complex analytical processes.
- Edge computing efficiencies — Deploying functions at the edge enables processing close to the source.
- Embedding within other xPaaS — Some xPaaS vendors embed an fPaaS in their platforms to host code components, such as rules and workflow routines. Examples include the InRule decision management platform and the Zoho low-code application platform. These systems hide the complexity of the fPaaS programming and operating model.

Obstacles

fPaaS is facing challenges because:

- Cost savings don't always materialize — fPaaS pricing isn't favorable for applications with consistently high invocation rates. Also, fPaaS-based applications often require other xPaaS, such as API management, data management and notifications.
- Latency — fPaaS-based applications can suffer from cold-start issues.
- Lock-in — fPaaS-based applications aren't portable across vendor solutions.
- Resource constraints — fPaaS is inappropriate for memory- and compute-intensive workloads.
- Lacking infrastructure — DevOps teams require development frameworks, testing and debugging tools, security services, and management technology. A fledgling ecosystem is emerging to address these requirements, although most ecosystem players focus only on Amazon Web Services (AWS) Lambda, and tooling for other fPaaS offerings is limited.
- Alternative solutions — Many developers prefer to use more general-purpose platforms, such as Kubernetes or low-code application platforms.

User Recommendations

- Minimize vendor lock-in by ensuring that your software engineering teams don't limit their skills and practices to the proprietary features of a single fPaaS.
- Estimate fPaaS costs based on expected invocation rates, memory requirements, execution times and other xPaaS dependencies. Don't presume that fPaaS is always a less expensive option.
- Evaluate whether fPaaS is a good fit for your applications and your teams' development skills. Consider aPaaS or managed container solutions as alternatives.
- Identify use cases where fPaaS offers a strategic benefit from a cost or agility perspective. Consider fPaaS for microservices deployments and for applications with highly variable or unpredictable capacity requirements.
- Avoid using fPaaS for high-memory or compute-intensive workloads. Don't port existing monolithic applications to fPaaS.
- Use fPaaS if it's an integral part of another solution, such as edge computing, decision management or low code.

Sample Vendors

Amazon Web Services; Cloudflare; Google; InRule; Microsoft; Netlify; Red Hat; Vercel; Zoho

Gartner Recommended Reading

A CIO's Guide to Serverless Computing

Security Considerations and Best Practices for Securing Serverless PaaS

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

Hybrid Cloud Computing

Analysis By: David Smith, Milind Govekar

Benefit Rating: High

Market Penetration: 5% to 20% of target audience

Maturity: Adolescent

Definition:

Hybrid cloud computing comprises one or more public and private cloud services that operate as separate entities, but ones that 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 — the cost optimization, agility, flexibility, scalability and elasticity benefits of public cloud, in conjunction with the control, compliance, security and reliability of private cloud. As a result, virtually all enterprises have a desire to augment internal IT systems with external cloud services.

Business Impact

Hybrid cloud computing enables an enterprise to scale beyond its data centers to take advantage of the public cloud's elasticity. Therefore, 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. It also sets the stage for new ways for enterprises to work with suppliers and partners (B2B) as well as customers (B2C).

Drivers

- The key driver for hybrid cloud is a desire to evolve data centers to become more cloud-like and therefore have a private cloud that has cost and other characteristics that are 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" means you have a vendor-provided private cloud offering that is packaged and connected to a public cloud in a tethered way. Azure Stack 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 hybrid cloud provides include service integration, availability/disaster recovery, cross-service security, policy-based workload placement and runtime optimization, and cloud service composition and dynamic execution (e.g., cloudbursting).
- Hybrid cloud computing is different from multicloud computing, which is the deliberate use of cloud services from multiple cloud providers for the same general class of IT service.
- Note that internally run, virtualized environments are often recast as "private clouds," and 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.

Obstacles

- Hybrid cloud computing complements multicloud computing. Although most organizations are integrating applications and services across service boundaries, we believe that few large enterprises have implemented hybrid cloud computing beyond this basic approach — and for relatively few services. Most companies will use some form of hybrid cloud computing during the next two years, but more advanced approaches lack maturity and suffer from significant setup and operational complexity.
- Hybrid cloud is different from hybrid IT, which is 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 TCS, and other service providers and system integrators.

User Recommendations

- When using hybrid cloud computing services, establish security, management, and governance guidelines and standards to coordinate the use of these services with public and private services.
- Approach sophisticated cloudbursting 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.
- Consider cloud management platforms, which implement and enforce policies related to cloud services.
- Consider using hybrid cloud computing, if your organization implements hybrid IT, as the foundation for implementing a multicloud broker role and leveraging hybrid IT services and service providers to complement your own capabilities.

Gartner Recommended Reading

Top Strategic Technology Trends for 2021: Distributed Cloud

‘Distributed Cloud’ Fixes What ‘Hybrid Cloud’ Breaks

Predicts 2021: Building on Cloud Computing as the New Normal

Climbing the Slope

Cloud Service Expense Management

Analysis By: Dennis Smith

Benefit Rating: High

Market Penetration: 5% to 20% of target audience

Maturity: Early mainstream

Definition:

Cloud service expense management (CSEM) is the practice of reviewing, reconciling and optimizing charges for services provided by external cloud service providers (CSPs). CSEM vendors provide tools to answer questions such as, “Did you get what you paid for?,” “Are you paying for the right things?” and “Did you optimize spend?” Managing cloud expenses becomes more important as organizations increase the number of cloud services they use.

Why This Is Important

Adoption of external cloud resources has far outpaced efforts to gain a deeper view of cloud expenses. Even as investment in public cloud resources accelerates, CSEM maturity for many organizations remains low. This gap will widen as cloud computing increases in the future.

Business Impact

- CSEM has the potential to impact business services and processes across all verticals.
- CSEM enables organizations to compare costs, identify waste and create more-efficient spending opportunities.
- Treating public cloud resources in a businesslike manner — optimizing resource unitization and managing spending — will enhance IT’s credibility by delivering business value in terms that business leaders understand.

Drivers

- The rise in public cloud adoption has increased the need for cloud cost management and optimization. A growing number of Gartner client inquiries reveal that dissatisfaction with cloud deployments is often related to an inability to manage costs. Our research indicates that many CSEM vendors have been acquired by cloud vendors that want to add the ability to manage cloud costs to their respective portfolios.
- Vendors are now combining CSEM-related functionality with governance, software and hardware asset management and/or cloud workload optimization. Vendor tooling is focusing on areas including cloud resource discovery, and IaaS, PaaS and SaaS for both on-premises and public cloud.
- Vendors are also looking to incorporate AI/ML to handle the volume and dynamic nature of cloud adoption.
- Most cloud management platforms (CMPs) have added CSEM functionality. CSPs and management service providers (MSPs) are also providing CSEM services and/or tools. Their offerings often include professional resources along with the tool, as CSEM products require resources to deliver value.

Obstacles

- Vendors' inability to evolve their technology to handle the innovation in cloud services. Many vendors either plan to have AI/ML or have introduced initial products that incorporate AI/ML. But many of these deployments are rudimentary and will not support future cloud cost management requirements.
- Competition from ITOM markets, which will continually add CSEM capability to existing toolsets. Hyperscale cloud providers will continue to improve their cost management offerings. For many enterprises, particularly those committed to one provider, these tools will be "good enough."
- Competition from cloud MSPs, which will combine tooling with skilled, often homegrown, cloud personnel.

User Recommendations

- Confirm that you got what you paid for. IT leaders can gain insight into cloud expenses by confirming that they received the appropriate services associated with the expenditure.

- Optimize value. IT leaders must track the services they need against the expenses incurred to determine whether there are more efficient ways of obtaining the same required services.
- Avoid overprovisioning. IT leaders must determine whether they are using the services for which they paid.
- If you require only cost management functionality, look for a tool or service that provides that instead of a multifunction CMP. If you are looking to select or have deployed a CMP tool, look to see what CSEM capabilities it provides.
- Do not accept your service provider's invoice at face value. Even if it's accurate, you probably can identify cost reduction opportunities or invest in more processing power for applications or services that create additional business value.

Sample Vendors

Apptio; CloudCheckr; CloudHealth; Flexera; Turbonomic

Gartner Recommended Reading

Market Guide for Cloud Management Tooling

Market Guide for Container Management

Cloudbursting

Analysis By: Ed Anderson

Benefit Rating: Moderate

Market Penetration: 5% to 20% of target audience

Maturity: Early mainstream

Definition:

Cloudbursting is the use of an alternative set of public or private cloud services as a way to augment service capacity and respond to increases in IT system requirements. Cloudbursting can occur between on-premises IT environments and public or private cloud environments, across cloud providers, or across resource pools of a single provider.

Why This Is Important

Cloudbursting is based on the concept of utilizing cloud services as a way to rapidly add resource capacity when applications or services experience spikes in demand. Cloudbursting is seeing some renewed hype as applications are implemented using container-based models and microservices architectures. The growing preference for hybrid and multicloud strategies increases the perceived value of cloudbursting across systems.

Business Impact

Cloudbursting represents the notion that cloud resources can be allocated, dynamically, when needed, to support expanding application resource requirements. While the promise of cloudbursting is attractive, implementation is challenging. Cloudbursting crosses system boundaries, which creates issues of compatibility in multiple dimensions. The business value of cloudbursting models is enticing, but implementation challenges remain.

Drivers

- Applications requiring additional resources or functional components may be hindered if required to operate only in a single environment. Cloudbursting can provide these additional resources through autoscaling using cloud services, allowing applications to scale across system boundaries. Although cloudbursting includes any scenario where cloud services are allocated on-demand when additional capacity is required, cloudbursting also includes scenarios where less critical resources are moved to a cloud service to free up capacity for critical on-premises workloads.
- Scaling applications across systems has historically been a manually initiated process, however, with the development of more sophisticated cloud management and orchestration tools, cloudbursting can support automated application scaling using cloud services through the use of triggers and service governor technology.
- Application workload placement can be simplified if cloudbursting is an option, particularly when operating in dynamic, continually changing conditions. Determining workload placement at provisioning time is the easiest to implement because services are placed based on system compatibility, available capacity and policy. Supporting dynamic workload movement is harder, may require some downtime and will be less common because moving services between cloud environments and across different providers can be complex.

Obstacles

Barriers to cloudbursting usage include:

- The lack of cross-platform cloud API standards
- Inadequate application instrumentation and management tools
- Networking latency between data centers
- Identity and security limitations
- Configuration management
- Technology license restrictions
- Incompatible application architectures

Other obstacles are as follows:

- Most legacy applications have storage and database architectures that cannot be easily adapted to geographically dispersed data centers.
- Runtime capacity expansion requires applications to be specifically written or adapted to cloudbursting, such as scale-out web architectures or batch-computing jobs that can disperse workloads in parallel across distributed resource pools. If a microservices architecture is established on the target system, runtime expansion is much easier.

User Recommendations

Use cloudbursting in situations where there is notable benefit to offset cloudbursting implementation costs and complexities:

- Select workloads and applications that are conducive to scale-out execution using parallel and distributed processing models and experience periodic spikes in resource requirements.
- Do not assume cloudbursting will become a broadly viable approach for cross-cloud workload portability and expansion, even when hyped technologies such as containers and multicloud management solutions are used.

- Assess which applications will get real value from cloudbursting and whether these applications meet the technical criteria and constraints for implementation.
- When pursuing cloudbursting capabilities, leverage cross-platform abstractions to increase compatibility between systems.
- Carefully evaluate the financial implications of cloudbursting; there may be significant costs associated with cloud resource consumption as well as increased networking costs.

Sample Vendors

Amazon Web Services; Dell; Flexera (RightScale); Google; Hewlett Packard Enterprise; IBM; Microsoft; Nutanix; Red Hat; VMware

Gartner Recommended Reading

'Distributed Cloud' Fixes What 'Hybrid Cloud' Breaks

A Multicloud Strategy Is Complex and Costly, but Improves Flexibility

How to Assess an Application Portfolio for Public Cloud Deployment

How to Automate Server Provisioning and Configuration Management

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.

Why This Is Important

A CCOE is usually the most effective way to drive enterprise-scale cloud adoption that results in positive business outcomes. The CCOE leads organizationwide cloud governance, provides cloud architecture guidance, gathers and disseminates cloud best practices, and, if necessary, guides cloud transformation. Although the CCOE is responsible for developing cloud computing policies, it primarily influences, rather than controls, and depends on other teams to implement and enforce policies.

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 management. 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.
- The organization needs a guide on its cloud journey, as cloud use grows within an 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.

Obstacles

- The CCOE needs the sponsorship and mandate of the CIO or other 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. Do not understaff the CCOE, since this can be a major threat to its influence and success. If the organization has a rapidly growing number of cloud projects, it may be wise to open headcount for CCOE architects well in advance of demand, since there is significant competition for this skill set. However, the CCOE should not be fully outsourced, due to its strategic nature.

User Recommendations

- A CCOE 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

Innovation Insight for the Cloud Center of Excellence

How to Build a Cloud Center of Excellence, Part 1: Designing for Cloud Adoption Success

How to Build a Cloud Center of Excellence, Part 2: Implementing the Foundations for Cloud Adoption Success

The Cloud Architect: Skills Guidance for Modern Technical Professionals

The Future Direction and Evolution of Business-Outcome-Driven Enterprise Architecture

Cloud Native Architecture

Analysis By: Anne Thomas

Benefit Rating: Moderate

Market Penetration: 5% to 20% of target audience

Maturity: Adolescent

Definition:

Cloud native architecture is the set of application architecture principles and design patterns that enables applications to fully utilize the agility, scalability, resiliency, elasticity, on-demand and economies of scale benefits provided by cloud computing. Cloud native applications are architected to be latency-aware, instrumented, failure-aware, event-driven, secure, parallelizable, automated and resource-consumption-aware (LIFESPAR).

Why This Is Important

Many organizations are moving to cloud native architecture as they shift their application workloads to cloud native application platforms. Cloud native principles and patterns enable applications to operate efficiently in a dynamic environment and make the most of cloud benefits. Organizations that simply “lift and shift” legacy applications to cloud native platforms often find that the applications perform poorly, consume excessive resources and aren’t able to fail and recover gracefully.

Business Impact

- Cloud native architecture ensures that applications can take full advantage of a cloud platform's capabilities to deliver agility, scalability and resilience.
- It enables DevOps teams to more effectively use cloud self-service and automation capabilities to support continuous delivery of new features and capabilities.
- It can also improve system performance and business continuity, and it can lower costs by optimizing resource utilization.

Drivers

- Organizations want to make the most of cloud computing to support their digital business initiatives, but they can't fully exploit cloud platform benefits without cloud native architecture.
- Software engineering teams are adopting cloud native architecture to support cloud native DevOps practices, including self-service and automated provisioning, blue/green deployments, and canary deployments. A basic set of rules known as the " twelve-factor app" ensures that applications can support these practices.
- Cloud native architecture includes practices such as application decomposition (following the mesh app and service architecture [MASA] structure), containerization, configuration as code, and stateless services.

Obstacles

- Cloud native architecture adds a level of complexity to applications, and development teams require new skills, new frameworks, and new technology to be successful.
- Without proper education, architects and developers can apply the principles poorly and deliver applications that fail to deliver the expected benefits. This leads to developer frustration in adopting the new patterns and practices.
- Not every cloud-hosted application needs to be fully cloud-native and developers may be confused about when they need to use particular patterns to address their specific application requirements.

User Recommendations

- Use the twelve-factor app rules and the LIFESPAR architecture principles to build cloud native applications.
- Incorporate cloud native design principles in all new applications irrespective of whether you currently plan to deploy them in the cloud. All new applications should be able to safely run on a cloud platform, even if it doesn't fully utilize cloud characteristics.
- Apply cloud native design principles as you modernize legacy applications that you plan to port to a cloud platform to ensure that they can tolerate ephemeral or unreliable infrastructure. Otherwise, they are likely to experience stability and reliability issues.
- Select an application platform that matches your cloud native architecture maturity and priorities. Recognize that low-code platforms enable rapid development of cloud-ready applications, but they won't provide you with the full flexibility to apply LIFESPAR and twelve-factor principles.

Sample Vendors

Amazon Web Services; Google; Microsoft; Red Hat; VMware

Gartner Recommended Reading

How to Help Software Engineering Teams Modernize Their Application Architecture Skills

How to Modernize Your Application to Adopt Cloud-Native Architecture

A Guidance Framework for Modernizing Java EE Applications

Guidance Framework for Modernizing Microsoft .NET Applications

Cloud Networking

Analysis By: Sid Nag

Benefit Rating: High

Market Penetration: 20% to 50% of target audience

Maturity: Early mainstream

Definition:

Cloud networking is a cloud provider service to connect hybrid IT, hybrid cloud and multicloud environments and IaaS networking. Such services provide robust interconnectivity within a cloud provider's network and between external cloud data centers; a customer's on-premises, hosted data centers; and colocation facilities.

Why This Is Important

Cloud networking — which includes cloud interconnect, VPC networking, zero-trust networking and network services — is a major factor in the selection of cloud providers by 75% of Gartner-surveyed enterprises.

Business Impact

A cloud service is only as good as the network it runs on. Cloud networking continues to be a top challenge to the adoption of cloud and has gained attention as a critical selection criteria for cloud services. Cloud interconnect architectures are important for all applications across disaggregated multicloud, hybrid cloud and hybrid IT environments.

Drivers

- Over 76% of organizations have adopted or plan to adopt multicloud environments by the end of 2021, making cloud networking an important technology.
- In cloud networking, traditional network functions and services, including connectivity, security, management and control, are delivered as a service in the cloud.

Obstacles

- Cloud networking can often be confused with traditional WAN networking although WAN networking technologies are used to enable cloud networking.
- Cloud networking is specific to connecting cloud estates across multicloud and hybrid cloud environments.
- Cloud networking today is often done via inelegant methods by leveraging exchange providers as opposed to using native exterior gateway protocols such as BGP.

User Recommendations

As users move applications into the cloud, new cloud networking requirements will likely emerge:

- There will often be a requirement for the application running in a cloud provider's data center to interoperate with other applications that reside in the customer's data centers or in a different cloud provider.
- As multicloud adoption increases, organizations need to be familiar with multiple cloud service provider (CSP) stacks.

All of these scenarios create a need for a robust connectivity model. Organizations must:

- Select cloud networking offerings that include capabilities to select the right connectivity provider.
- Institute a requirement where QoS, latency and availability issues are addressed by their cloud provider, especially for mission-critical applications that are running in the cloud.
- Mandate capabilities that include provisioning of MPLS, VPN and SD-WAN support for complex overlays for connecting their on-premises locations into the cloud provider's data center.

Gartner Recommended Reading

Cool Vendors in Cloud Networking

Market Guide for Cloud Networking Software

CIPS

Analysis By: Sid Nag, Yefim Natis

Benefit Rating: High

Market Penetration: 20% to 50% of target audience

Maturity: Early mainstream

Definition:

Cloud infrastructure and platform services is the business and technology arrangement where IaaS and/or PaaS capabilities are offered as cloud services. Integrated CIPS implies integration of IaaS and PaaS. The degree of integration 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 context, unified security, unified monitoring and unified billing.

Why This Is Important

Cloud infrastructure and platform services (CIPS) represents a continuum, and some offerings in the market have characteristics of both. For example, container management IaaS, like Google Kubernetes Engine (GKE), is often compared to an application PaaS. GKE offers no tooling and abstractions to support developer workflows. GKE does, however, offer greater control and flexibility for configuring the compute resources.

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 perspective. Vendors also benefit from CIPS — those coming from IaaS and those specialized in PaaS increase their customer value proposition and ability to compete when covering the broader set of capabilities.

Drivers

- The appeal for CIPS is not in best-of-breed offerings, but in the unification and integration of platform capabilities across these services enabling broad deployment of workloads ranging from ERP to cloud-native.
- Most customers that use a hyperscale CIPS 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. Hyperscalers deliver PaaS services with a direct dependency on their IaaS services. 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.
- The complexity and level of investment required to offer a full, integrated portfolio of multifunctional PaaS and IaaS services will likely limit the vendor options in this market to a handful of hyperscalers. Some of the 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.

Obstacles

- Public CIPS markets are consolidating around the market leaders.
- Infrastructure as a service (IaaS)-only or platform as a service (PaaS)-only cloud providers will continue to exist, but only as secondary cloud providers compared to CIPS providers.
- This, in turn, could make it a monopoly of a handful of cloud providers, stifle competition, and drive stand-alone cloud providers out of the market.

User Recommendations

CIOs, CTOs, IT leaders and planners must:

- Use CIPS in both cloud-native and legacy migration projects to expand your design and deployment options. In some cases, 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.
- Consider integrated CIPS providers to be long-term application platforms. They should be managed as such, with appropriate attention to potential application portability issues.
- Do not assume that all services of the provider are of the same maturity, functional completeness or quality of service.
- When considering a smaller specialist PaaS provider, give extra credit to those that are multicloud and, therefore, can be colocated with multiple larger suites of CIPS capabilities.

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 sometimes isolation is important (e.g., for security or regulatory reasons). Private and public cloud computing are at opposite ends of the isolation spectrum. As public cloud providers have offered virtual private cloud, dedicated instances and dedicated hosts, the contrast between private and public has become a spectrum of isolation choices.

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 but more customized and less automated are more appropriate.

Drivers

- **Regulatory or privacy reasons:** Data and/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.
- **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 costly and complex.
- Building an initial self-service offering is one thing, but maintaining and adding new, innovative features (similar to public cloud providers) is very difficult for enterprises.
- Private clouds being built to support existing applications require significant nonstandardization, which significantly reduces automation potential.
- Users often have little motivation to move to a more standard model, unless they make fundamental business changes, such as usage-based pricing for services.

User Recommendations

- Evaluate third-party options first, 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.
- Manage the scope of work — start small and expand based on the business case.
- Build expertise in managing multicloud — beyond just the private cloud.

Sample Vendors

Hewlett Packard Enterprise (HPE); IBM; VMware

Gartner Recommended Reading

Rethink Your Internal Private Cloud

Building 'Just Enough' Private Cloud With Virtualization Automation

CASBs

Analysis By: Craig Lawson, Neil MacDonald

Benefit Rating: Transformational

Market Penetration: 20% to 50% of target audience

Maturity: Mature mainstream

Definition:

Cloud access security brokers (CASBs) provide crucial cloud governance controls for visibility, compliance, data security and threat protection by consolidating multiple types of security policy enforcement into one place for SaaS, IaaS and PaaS. Examples include authorization, UEBA, adaptive access control, DLP, device profiling, object encryption, tokenization, logging, alerting and malware removal. Majority of CASB deployments are cloud-based; on-premises deployments are rare.

Why This Is Important

CASBs are critical for organizations to secure usage of business-critical cloud services. The four key areas — visibility, compliance, data security and threat protection — are the primary value propositions for the usage of CASBs.

Business Impact

CASBs enable consistent security policies and governance across cloud services. Unlike traditional security products, CASBs are designed to protect data stored in someone else's systems, are suitable for organizations of all sizes in all industries and can demonstrate cloud usage is well-governed. With continued feature expansion, ongoing convergence with SWG/ZTNA and relative ease of switching providers, favor one-year contract terms over lengthier ones when selecting CASBs.

Drivers

- End-user organizations need to secure use of business-critical, cloud-delivered applications and infrastructure; secure general internet to prevent threats to users, regardless of their location; and improve access to existing services while taking advantage of zero trust concepts. Today, CASB is converging with SWG and ZTNA to deliver this "three-legged stool" concept to support all these use cases.
- With SWG vendors enabling secure use of business-critical, cloud applications and infrastructure, and CASB vendors expanding functionality for general internet security and access to existing services, security leaders are now able to successfully deliver on the above-mentioned three capabilities from an increasing number of vendors providing all three.
- The COVID-19 pandemic has increased focus on two specific use cases that CASB technology directly helps with: the huge shift to remote working and the continuously increasing use of cloud services critical to business.

Obstacles

- Lack of a focus on DLP can lead to frustration with a CASB as organizations fail to build comprehensive policies and manage false-positive rates.
- A subset of controls are offered by CSPs themselves, for example, Office 365's native security features and Salesforce Shield.
- Unclear organizational ownership of SaaS tenancy leads to a CASB implementation that fails to secure the SaaS adequately.
- Product consolidation failure in organizations where multiple CASBs exist through expanded licensing agreements.
- Overlapping CASB functionality from a number of vendors leads to duplication and confusion.

User Recommendations

- Examine vendor capabilities in four functionality areas: visibility, data protection, threat detection and compliance (see Magic Quadrant for Cloud Access Security Brokers for a more detailed analysis of these capabilities).
- Seek support for multiple modes of operation, namely forward proxy, reverse proxy (or RBI) and API for the best support of managed and unmanaged devices and cloud services via CASB.
- Aim to move to a single provider for CASB, SWG and ZTNA as these services are on strong convergence paths.

Sample Vendors

Bitglass; Broadcom (Symantec); Lookout (CipherCloud); McAfee; Microsoft; Netskope; Proofpoint

Gartner Recommended Reading

Magic Quadrant for Cloud Access Security Brokers

Critical Capabilities for Cloud Access Security Brokers

Magic Quadrant for Secure Web Gateways

2021 Strategic Roadmap for SASE Convergence

Market Guide for Zero Trust Network Access

Entering the Plateau

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 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.
- A recent and well-justified emphasis on application modernization as part of migration has further confused 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 MSP, 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.
- Ensure you 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; Bespin Global; Capgemini; Cognizant; Deloitte; Logicworks; Rackspace Technology; Smartronix; Tata Consultancy Services; Wipro

Gartner Recommended Reading

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

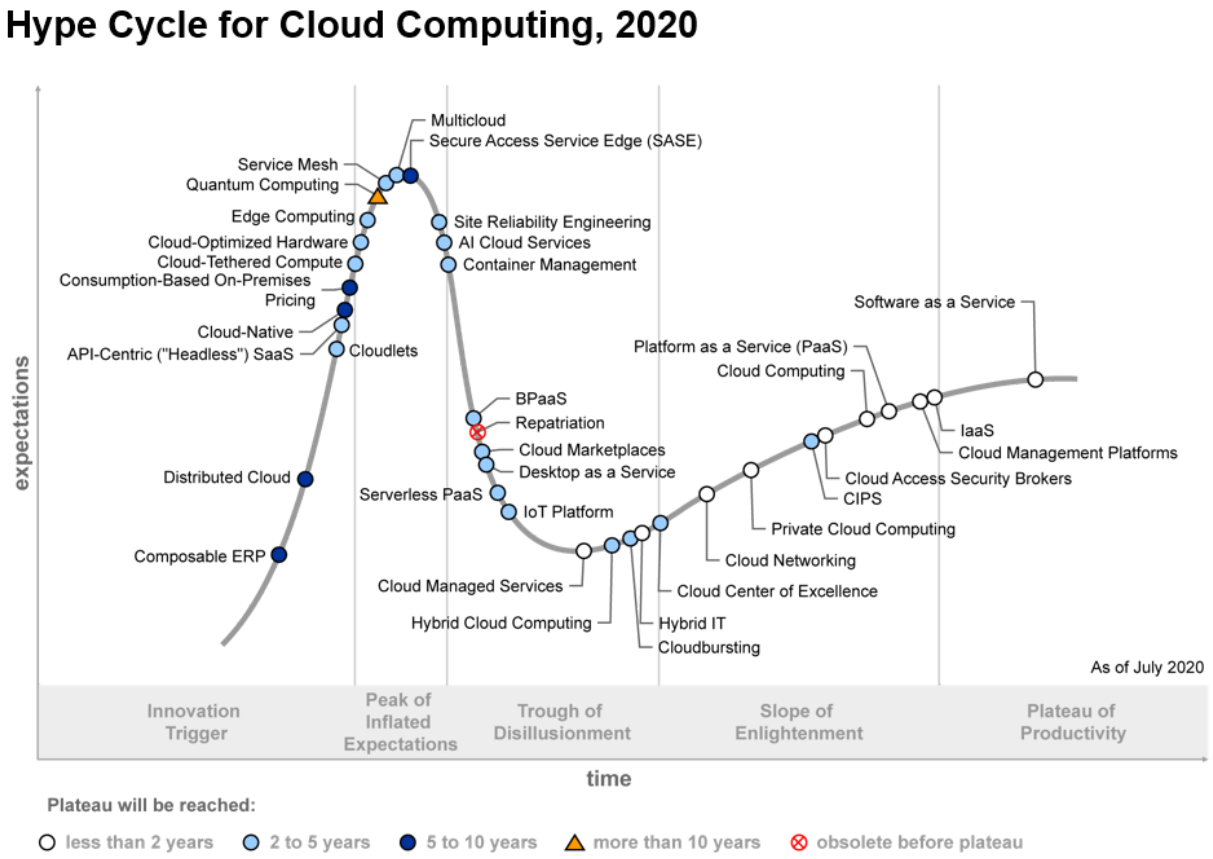
Critical Capabilities for Public Cloud Infrastructure Professional and Managed Services, Worldwide

Cloud Cost Scenario Planning Tool

Ignition Guide to Creating a Migration Plan for Public Cloud

Appendixes

Figure 2: Hype Cycle for Cloud Computing, 2020



Source: Gartner
ID: 450475

Gartner

Source: Gartner (July 2020)

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 2021)

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 2021)

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 2021)

Document Revision History

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.

Cloud and Edge Infrastructure Primer for 2021

The Cloud Strategy Cookbook, 2021

Revisiting the Top 10 Cloud Myths for 2020

Top 10 Trends in PaaS and Platform Innovation, 2020

Devise an Effective Cloud Computing Strategy by Addressing Five Key Areas

© 2021 Gartner, Inc. and/or its affiliates. All rights reserved. Gartner is a registered trademark of Gartner, Inc. and its affiliates. This publication may not be reproduced or distributed in any form without Gartner's prior written permission. It consists of the opinions of Gartner's research organization, which should not be construed as statements of fact. While the information contained in this publication has been obtained from sources believed to be reliable, Gartner disclaims all warranties as to the accuracy, completeness or adequacy of such information. Although Gartner research may address legal and financial issues, Gartner does not provide legal or investment advice and its research should not be construed or used as such. Your access and use of this publication are governed by Gartner's Usage Policy. Gartner prides itself on its reputation for independence and objectivity. Its research is produced independently by its research organization without input or influence from any third party. For further information, see "Guiding Principles on Independence and Objectivity."

Table 1: Priority Matrix for Cloud Computing, 2021

Benefit ↓	*Years to Mainstream Adoption*			
	Less Than 2 Years ↓	***2 - 5 Years*** ↓	***5 - 10 Years*** ↓	***More Than 10 Years*** ↓
Transformational	CASBs	Edge Computing SASE Site Reliability Engineering	Business-Driven Cloud Strategy Industry Clouds Sovereign Cloud	
High	Cloud Center of Excellence Cloud Managed Services Cloud Migration Cloud Networking	AI Cloud Services API-Centric SaaS CIPS Cloud-Optimized Hardware Cloud Service Expense Management Container Management Desktop as a Service Hybrid Cloud Computing IoT Platform Multicloud Packaged Business Capabilities	Cloud-Native Cloud Portability Distributed Cloud Intercloud Data Management	Quantum Computing

Benefit	*Years to Mainstream Adoption*			
↓	***Less Than 2 Years*** ↓	***2 - 5 Years*** ↓	***5 - 10 Years*** ↓	***More Than 10 Years*** ↓
Moderate	Cloudbursting Private Cloud Computing	Cloud Marketplaces Cloud Native Architecture Consumption-Based Sourcing Service Mesh		
Low		Serverless fPaaS		

Source: Gartner (July 2021)

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 2021)

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 2021)

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 2021)