Hype Cycle for Cloud Platform Services, 2023

Published 27 July 2023 - ID G00791094 - 153 min read

By Analyst(s): Yefim Natis, David Smith, Nicholas Carter

Initiatives: I&O Platforms

This Hype Cycle helps IT leaders understand the pace and viability of maturing and emerging cloud platform technologies, services and trends. While traditional PaaS capabilities are reaching maturity, a new wave of cloud platform innovations crowd the Peak of Inflated Expectations.

Strategic Planning Assumption

By 2025, all leading cloud service providers will include support of integrated platform and business capabilities, supporting the customers' use of applications as platforms.

Analysis

What You Need to Know

This document was revised on 22 August 2023. The document you are viewing is the corrected version. For more information, see the Corrections *page on gartner.com.*

Cloud platform services have evolved well beyond the traditional domain of platform as a service (PaaS). The division of cloud architecture into the layers of infrastructure as a service (IaaS), PaaS and SaaS has blurred, as technologies have expanded across these artificial boundaries. Customers may utilize any of cloud services as platform capabilities, since they target their cloud investments to outcomes (not technology categories). Most engage some capabilities from at least two, if not all three, of the traditional layers of cloud architecture. IaaS and more technically oriented PaaS have formed a continuum that Gartner defines as cloud infrastructure and platform services (CIPS). SaaS is modularizing and is increasingly used, via its APIs and platform services, as a library of business capabilities. Modular SaaS architecture is thus becoming an integral part of engineering platforms for business technologists and fusion teams. [1] The growing adoption of composable modularity of SaaS has combined with composability tools in advanced industry cloud platforms. [2] In the context of the increasing technological competence of business organizations, a category of "applications as platforms" is emerging as an impactful business technology architecture trend. [3]

Worldwide spending on all public cloud services is forecast to grow 21.7% (21.3% in constant currency) in 2023. Organizations continue investing in the cloud with modernization efforts and cloud-native development driving a five-year compound annual growth rate (CAGR) of 19.5% (18.7% in constant currency).

Cloud platform services is a massive and growing market. The five-year CAGR — 2021 through 2027 — of spend on traditional PaaS capabilities alone is more than double that of on-premises spend across equivalent software markets. Through 2027, the market is projected to more than triple in size, from $90 billion in 2021 to more than $299 billion in 2027. [4]

The Hype Cycle

The 2023 Hype Cycle for Cloud Platform Services covers platform innovations across cloud categories, expanding beyond the traditional PaaS functionality. Infrastructure services like container management and the traditional platform capabilities like communication or Internet of Things (IoT) PaaS are joined by application services like API-centric SaaS, industry cloud platforms and SaaS as a platform.

As the graphic of the Hype Cycle illustrates, the innovation in cloud platform services is unrelenting. New business models and technology breakthroughs drive the need for support and reinforcement at the platform level, and the distributed cloud, driven by its reach and scale, remains the stage of most modern technology innovation.

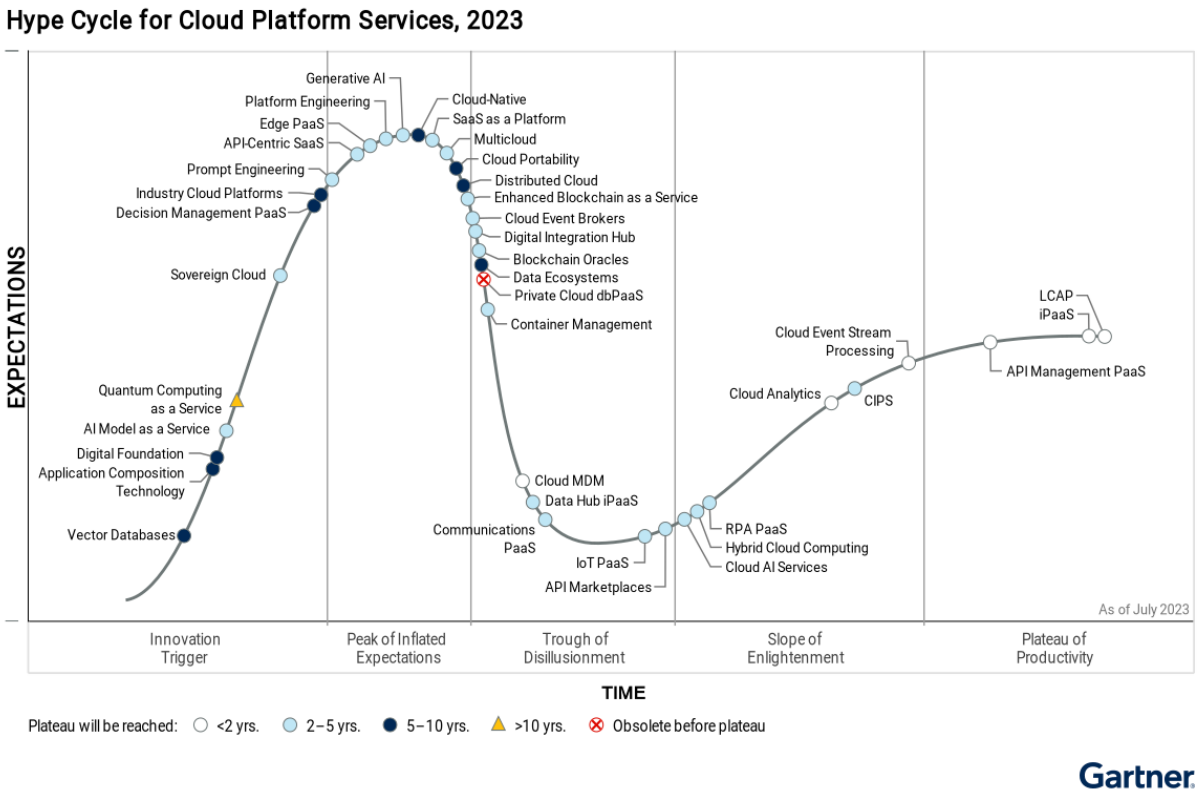
Distributed cloud. The increasing maturity of the core cloud platform services continues to strengthen the more pragmatic use cases of cloud at the place and scale of need. Distributed cloud, sovereign cloud, edge PaaS, multicloud and hybrid cloud innovations remain center stage, evolving from the early adopters toward the mainstream.

Business platforms. Many cloud platform innovations are reflecting the change in the way that SaaS is built, delivered and consumed by its business users. In the cloud, the boundaries between the application and platform capabilities (between SaaS and PaaS) are disappearing to support the shift of business users from consumers to co-creators of their technology experiences. API-centric SaaS, industry cloud platforms, SaaS as a platform and low-code application platforms (LCAP) are all reflections of the emerging fusion of applications and platforms into business empowerment services. Application composition technology and API marketplaces are the necessary enabling technologies that remain the subject of new development and experimentation.

Leading edge. Generative AI is affecting the roadmaps of many products and has spawned a number of new cloud platform technology investments, including the generative AI large language models (LLMs), vector databases and prompt engineering platforms. These new technologies are at the Innovation Trigger stage, but are likely to move fast toward the outsize industry expectations and the associated hype. Generative AI is firmly at the Peak of Inflated Expectations and heading toward the inevitable Trough of Disillusionment. While superseded now by the attention to AI breakthroughs, the blockchain work continues as well with enhanced blockchain as a service and blockchain oracles still in the mix of innovations, which is creating excitement and high expectations.

Maturity. Meanwhile, technologies like robotic process automation (RPA), PaaS, LCAP and IoT PaaS, which were subjects of some intense hype in prior years, are reaching maturity and have become a safe choice for mainstream adoption. IT leaders and planners must balance adoption of the safe and well-proven technologies on the right side of the Hype Cycle and the more risky, but also more differentiating, technologies on the left. Few would benefit from concentrating only on the safety or the leading edge.

Figure 1: Hype Cycle for Cloud Platform Services, 2023



The Priority Matrix

The 2023 state of the cloud platform services is energetic, but overexcited. History tells us that transformational changes are relatively few, and most new ideas end up with useful but moderate impact on the mainstream business operations and outcomes. Transformative changes that evolve to maturity fast are even more rare. However, the impact of generative AI seems to be on the trajectory, considering the rapid pace of its early adoption by industry megavendors and leading business innovators. Industry cloud platforms consolidate some strategic trends in applications. With time, industry cloud technologies will also benefit from the advances in generative user experiences.

Some technologies, now approaching maturity, have had a high degree of positive business impact. Integration platform as a service (iPaaS) has become a trusted model for application integration. LCAPs have bridged the divide between business and technology innovation and laid the foundation for the new business-IT partnership. Cloud analytics, including cloud event stream processing, has become a common practice for most mainstream organizations, assisting the business with more contextual situation awareness for competitive decision making.

IT leaders assessing new business technology opportunities should use caution — some hyped innovations may produce moderate or minimal business impacts. Still, awareness of key trends and leading-edge business technologies, even if surrounded by excessive hype, is required. Some of the hype, including the reception of generative AI, may turn out to be justified, and preparedness for transformative changes is essential for a growing and competitive business.

Table 1: Priority Matrix for Cloud Platform Services, 2023

(Enlarged table in Appendix)

Benefit ↓	Years to Mainstream Adoption: Less Than 2 Years ↓	2 - 5 Years ↓	5 - 10 Years ↓	More Than 10 Years ↓
Transformational		Blockchain Oracles Generative AI Platform Engineering	Industry Cloud Platforms	
High	Cloud Analytics Cloud MDM iPaaS LCAP	AI Model as a Service API-Centric SaaS CIPS Cloud AI Services Cloud Event Brokers Communications PaaS Container Management Data Hub iPaaS Digital Integration Hub Enhanced Blockchain as a Service Hybrid Cloud Computing IoT PaaS Multicloud Prompt Engineering	Cloud-Native Cloud Portability Data Ecosystems Digital Foundation Distributed Cloud Vector Databases	
Moderate	API Management PaaS Cloud Event Stream Processing	API Marketplaces Edge PaaS RPA PaaS SaaS as a Platform Sovereign Cloud	Application Composition Technology Decision Management PaaS	Quantum Computing as a Service
Low				

Source: Gartner (July 2023)

On the Rise

Vector Databases

Analysis By: Arun Chandrasekaran, Radu Miclaus

Benefit Rating: High

Market Penetration: Less than 1% of target audience

Maturity: Embryonic

Definition:

Vector databases store numerical representations of data. In such databases, each point is represented by a vector with a fixed number of dimensions, which can be compared via mathematical operations, such as distance measures. Vector databases are commonly used in machine learning (ML) solutions, where vectors represent data features/attributes, such as text embeddings. Storing these vectors in a database enables users to search for similar data points with low latency.

Why This Is Important

Vector databases serve such use cases as similarity search and product recommendation. Rapid innovation in generative AI and adoption of AI foundation models have spawned interest in vector databases. When customers adopt generative AI models, vector databases store the embeddings that result from the model training. Storing vector embeddings representing the model training, the database can do a similarity search, which matches a prompt (the question) with specific or similar vector embedding.

Business Impact

Businesses thrive by delivering differentiated customer experience (CX). Generative AI is increasingly embedded in applications to empower the human-machine symbiosis, and organizations need scalable and accelerated ways to build and support these applications long term. Vector databases are an important back-end service that allows businesses to future-proof and scale their generative-AI-enabled applications. These drive business value through customer engagement and adoption.

Drivers

- **Popularity of vector embeddings:** With the rise of AI foundational models, embeddings have become the cornerstone for semantic search. Hence, they are the working inputs for training large foundation models.

- **Performance and scalability needs:** The applications looking to embed generative methods that use embeddings-based models need back-end services that can respond with low latency to high concurrency requests (prompts) and responses (completions) for generative AI use cases.
- **Service architecture:** Because most applications are built on service-based architectures, vector databases are ideally presented to applications as services that communicate with the interface via APIs.
- **Hybrid implementation of retrieval and generative models:** Vector databases are optimal for both semantic search (retrieval based on vector similarity) and generative inference through foundation models. This hybrid combination of models drives the need for optimized vector databases, because both generative and retrieval are used together for grounding of facts.
- **Developer focus:** Developers of new applications are driving the demand for vector databases by presenting use cases that cannot scale without the ability for embeddings to be stored in an optimized structure for high-throughput production applications.

Obstacles

- Enterprises lack an understanding of what vector databases do and the unique use cases they enable.
- Vector databases are superspecialized databases that may cause challenges around data migration and integration and limited extensibility across use cases.
- Most vector databases are delivered as cloud-managed service — the complexity of deploying, configuring and operating them outside cloud environments requires deep technical skills and know-how.
- Vector databases can be expensive to implement, given the newness of the technology and lack of industry skills to deploy and manage it.
- The vector database market is nascent and populated mostly by startups, which may not have extensive experience working with enterprise clients, as well as unproven product market fit.

User Recommendations

- Determine whether your functional requirements can be satisfied by incumbent vendors that can support the storage and retrieval of vector embeddings — you may not always need a purpose-built vector database.
- Prioritize developer experience, ecosystem integration, use case fit, reliability and performance as important selection criteria, and validate them thoroughly via a POC process.
- Select managed, cloud-based vector databases as deployment modes, unless you have stringent requirements and deep technical skills for an on-premises, self-managed deployment mode.
- Conduct internal training and education on the appropriate use cases for vector databases, how to leverage their true potential, and effective ways to optimize their deployment and maximize their value.

Sample Vendors

Couchbase; Croma; Elastic; Google; Pinecone Systems; Qdrant; Redis; Weaviate; Zilliz

Gartner Recommended Reading

Innovation Insight for Artificial Intelligence Foundation Models

Quick Answer: What Is GPT-4?

Executive Pulse: AI Investment Gets a Boost From ChatGPT Hype

How Large Language Models and Knowledge Graphs Can Transform Enterprise Search

Application Composition Technology

Analysis By: Paul Vincent, Yefim Natis

Benefit Rating: Moderate

Market Penetration: 5% to 20% of target audience

Maturity: Embryonic

Definition:

Application composition technology provides a catalog for business application components and the means for IT and business technologists to assemble applications, processes and experiences. Optionally this may support the importing or building of new components, a prepopulated catalog, and governance. Cataloged business application components may be UX, data, process, business logic or other software assets.

Why This Is Important

Programmatic access to IT services supporting business functions is now ubiquitous through API and event-driven support. Businesses can buy or build their IT services, then extend these services with application composition technologies including those targeting rapid automation or application delivery with business technologists as well as software engineers. This composition approach enables more support for differing end-user persona and use cases.

Business Impact

If organizations can increase diligence around creating services and components, curating them in catalogs for developer access, and the ability to compose more complex applications through reuse, the economics of application development will improve along with the agility of businesses. Application composition assisted by tooling also encourages democratization of application delivery and encourages exploitation of commodity services like API-enabled SaaS.

Drivers

Application composition technology deployments are emerging to address all of these:

- API ecosystems, API and service marketplaces and catalogs, and API-driven SaaS services are emerging and evolving — especially SaaS vendor marketplaces. Wider and easier availability of these services will reduce the cost of reuse as they become easier to consume by composers and developers using third party as well as SaaS platforms' own composition tools.
- There is increased awareness of practices and technologies to create new scalable, multitenant, cloud-native services and software components. Even low-code vendors are providing the ability for business technologists to create API-driven processes for B2B consumption.

- The complexity limits inherent in some democratized low-code automation developments can be partly solved by delegating processes and logic to external services and components. As low-code advances further to sophisticated enterprise use cases, they rely more on parameterized, high-performance, external and/or vendor-provided technology and business services.
- CIOs are keen to maximize the ROI on their high cost investments in SaaS and exploit these services and APIs via new digital business applications. This implies composing new applications on top of these procured SaaS services and their marketplace ecosystems, via their APIs and possibly the SaaS vendors' own composition technologies.
- The skill challenges faced by IT organizations are forcing a pragmatic approach to application development. Expensive, skilled specialist developers can be redirected to the creation of back-end services, which can then be exploited by less-specialized developers using low-code integration, workflow and user interface development tools.

Obstacles

- Beyond REST API technology there is little appetite for standardized composition mechanisms in enterprises, meaning composition metadata in application platforms is proprietary and mostly for vendor monocultures. Interoperability remains fragile and liable to spoiling tactics from megavendors.
- Similarly, application service catalogs remain vendor-specific and generally private to vendor ecosystems. Application composition approaches will only thrive once platform vendors can inherit service catalogs from other sources such as mainstream providers like Amazon Web Services (AWS), Salesforce and SAP.
- Low-code application development tooling continues to evolve, resulting in continued lowering of the cost of development of medium-complexity enterprise applications. This enables organizations to build systems without costly external business services. Thus, application composition technology may just evolve into a set of low-code application platforms (LCAPs) without the need to address cross-vendor catalogs.

User Recommendations

- Identify the most repeated functionality across applications and package/catalog these for reuse and composition, ensuring best performance and functionality for all users. Track service and component assets as a matter of course to avoid duplication and also measure ROI.
- Prioritize application platforms that allow curation of an extensible catalog of relevant and useful business services and connectors to encourage composability. Consider vendors with a heterogenous approach to creating and applying components to minimize the effects of vendor lock-in.
- Seek collaboration with business technologists and citizen developers and separate the roles of “creators” from “composers.”
- Consider application composition technologies for digital business initiatives involving multiple services where innovation, experimentation and time to market require maximize reuse.
- Build up the competence, skills and technology of API, integration and event management as a precursor to the management of composable packaged capabilities.

Sample Vendors

Automation Anywhere; Betty Blocks; emach.ai; Entando; Novulo; Olympe.io; Oracle; SAP; Spryker

Gartner Recommended Reading

How to Design Enterprise Applications That Are Composable by Default

Essential Patterns for Data-, Event- and Application-Centric Integration and Composition

A Technical Guide to Composable Application Architecture

Predicts 2023: Composable Applications Accelerate Business Innovation

Innovation Insight for Application Composition Technology

Digital Foundation

Analysis By: David Smith, David Cearley

Benefit Rating: High

Market Penetration: 1% to 5% of target audience

Maturity: Emerging

Definition:

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

Why This Is Important

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

Business Impact

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

Drivers

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

Obstacles

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

User Recommendations

- Build digital foundations from broad, composable sets of technologies and services that can create and manage digital solutions, services and experiences. Composable thinking is a key part of the mindset.
- Focus not only on the technology, but also on the principles, processes and governance required to enable the business to deliver value to customers.
- Use proven technology capabilities and services, including cloud computing and other shared capabilities.
- Don’t expect to buy a foundation. A foundation should utilize mostly off-the-shelf capabilities or use co-created and shared capabilities. Some assembly will be required in order to meet your needs.
- Build a roadmap and establish achievement of a digital foundation as an aspirational goal. Isolate legacy applications and grow the digital foundation to support new and updated digital solutions.

Gartner Recommended Reading

Use Collaborative Decision Making to Drive Mixed Portfolios’ Digital Outcomes

Infographic: The Evolving Digital Foundation

How to Improve CRM by Architecting Your Customer Technology Platform

Infographic: Why Digital Ecosystems Are the Future of Business

AI Model as a Service

Analysis By: Rajesh Kandaswamy

Benefit Rating: High

Market Penetration: 20% to 50% of target audience

Maturity: Early mainstream

Definition:

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

Why This Is Important

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

Business Impact

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

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

Drivers

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

- The growing capabilities of foundational models are expanding their potential across a variety of horizontal and vertical applications for companies of all sizes and kinds. This leads to a proliferation of models from many sources, including the cloud providers themselves. This proliferation makes it hard for customers to invest — an opportunity that cloud providers are only eager to capture.

- The use of GenAI in technology products and business is poised to grow, but the costs of utilizing the models are likely to become a serious concern for most businesses. Cloud economies of scale promise users optimized efficiencies and will increase their use.
- Beyond foundational models, AI tooling and nonfoundational models continue to grow. Cloud providers are investing in these complementary services to capture as much AI revenue as possible through these investments.
- There has been significant investment in startups that are leveraging AI capabilities in their own technical and business offerings, directly or through partnerships. These startups typically do not possess the capital or resources to manage their own infrastructure and rely on cloud providers for models and other AI services.
- Enterprises have been shifting large chunks of their IT to cloud providers over the past few years and are operationally ready to use cloud-based services in many aspects of their IT. Such enterprises have a cloud-first approach for most technology investments and will apply the same for their AI investments.
- Automated agents driven by generative AI can propel growth as enterprises start to leverage other models as part of their processes and workflows.

Obstacles

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

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

User Recommendations

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

Sample Vendors

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

Gartner Recommended Reading

Quick Answer: What is GPT-4?

Magic Quadrant for Cloud AI Developer Services

Quantum Computing as a Service

Analysis By: Chirag Dekate, Mark Horvath, Matthew Brisse

Benefit Rating: Moderate

Market Penetration: Less than 1% of target audience

Maturity: Embryonic

Definition:

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

Why This Is Important

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

Business Impact

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

Drivers

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

Obstacles

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

User Recommendations

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

Sample Vendors

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

Gartner Recommended Reading

Cool Vendors in Quantum Computing

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

Preparing for the Quantum World With Crypto-Agility

Sovereign Cloud

Analysis By: Rene Buest, Gregor Petri, Neville Cannon

Benefit Rating: Moderate

Market Penetration: 5% to 20% of target audience

Maturity: Adolescent

Definition:

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

Why This Is Important

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

Business Impact

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

Drivers

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

Obstacles

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

User Recommendations

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

Sample Vendors

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

Gartner Recommended Reading

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

What We Are Hearing About Cross-Border Data Transfers

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

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

Three Critical Use Cases for Privacy-Enhancing Computation Techniques

Decision Management PaaS

Analysis By: Paul Vincent

Benefit Rating: Moderate

Market Penetration: 5% to 20% of target audience

Maturity: Adolescent

Definition:

Decision management platform as a service (dmPaaS) refers to a cloud platform that manages the design, maintenance, governance and deployment of structured executable decisions. Structured decisions are explicit decision models that represent the business rules needed to make a decision.

Why This Is Important

Decision automation enables business consistency and operational excellence. Decision management is often related to high-performance and/or high-value business processes and the automation of business-critical operational decisions in finance and insurance. Increasingly, these cloud offerings are being considered for hyperautomation initiatives.

Business Impact

- dmPaaS in the cloud can improve business performance for automated or assisted decision making, while exploiting the cloud benefits of scale; performance, including real-time, event-driven; and cost predictability.
- It provides frameworks for supporting artificial intelligence/machine learning (AI/ML) initiatives that improve specific operational decisions.
- dmPaaS extends hyperautomation to the decision automation realm.

Drivers

- dmPaaS is driven by cloud-native best practices that support cloud decision services as part of a Mesh App and Service Architecture (MASA) approach. This reduces the cost of business logic development in API-driven services called from mini, micro or monolith services as request-reply or event-driven architectures.
- Enterprises need better-organized, optimized and auditable approaches for decision automation, as they increase straight-through processing of business systems, with an increasing amount of legislation and business competitiveness.
- Cloud-based AI and ML services are increasing the focus on better decision making. However, AI and ML benefit from a structured approach by the management of the executable decisions to which they apply: exploiting predictive models, analytics, ML-driven score models, optimization mechanisms and other advanced decision capabilities.
- Modernized business rule management supports better explicit decision making, and decision engineering, as first-class artifacts and practices are deployed as decision services by business technologists.
- The separation of decision analysis (e.g., business decision modeling using tools such as Signavio and Trisotech) and decision execution (for example, in Camunda and Flowable) exploit standards (for example, the Decision Model and Notification [DMN]) and encourage adoption by larger, more-mature enterprises.

Obstacles

- Decision management in the cloud remains an underexplored area between business governance and software engineering. The concept of decision engineering is unknown to many software engineering and business leaders outside the financial sector.
- It continues to have low levels of support for decision modeling standards, such as DMN. Major AI cloud providers, including Amazon Web Services (AWS) and Google, and major decision management vendors, do not support industry standardization. Instead, they prefer to focus on proprietary innovation with vendor lock-in.
- Decision management continues to be sidelined by legacy practices, such as error-prone, manually coded business logic for decision-rich services, despite the risks and costs of such errors. This results from a poor understanding of the benefits of decision management, and perceived lack of value, particularly due to the historical high cost of, and the skill sets required for, proprietary decision management solutions.

User Recommendations

- Choose dmPaaS for ease of consumption of cloud services to deliver better decision visibility, design complexity, rigor and structure to AI and ML initiatives, or transparent governance and risk management for automated decisions.
- Evaluate dmPaaS as an API-enabled upgrade to the native business-rule capabilities embedded in business process automation or low-code application platforms, especially when supporting decision specialists in their application development fusion teams.
- Assess wider decision strategies and how dmPaaS can support digital business operations and democratized hyperautomation delivery, while exploiting cloud or hybrid decision service solutions.
- Beware of vendor lock-in. Mainstream vendors have limited or incomplete support for decision model standards (e.g., DMN and Predictive Model Markup Language [PMML]). Portability of decision logic across mainstream vendors remains rare.

Sample Vendors

Camunda; Experian; FICO; Flowable; IBM; InRule; Progress; Sapiens; Signavio; Trisotech

Gartner Recommended Reading

Tool: Vendor Identification for Decision Management Suite Providers

How to Choose Your Best-Fit Decision Management Suite Vendor

Should Your Project Use a Decision Management Suite?

The Future of Data and Analytics: Reengineering Business Decisions, 2025

Identify and Evaluate Your Next Low-Code Development Technologies

Industry Cloud Platforms

Analysis By: Gregor Petri

Benefit Rating: Transformational

Market Penetration: 20% to 50% of target audience

Maturity: Adolescent

Definition:

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

Why This Is Important

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

Business Impact

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

Drivers

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

Obstacles

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

User Recommendations

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

Sample Vendors

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

Gartner Recommended Reading

Top Strategic Technology Trends for 2023: Industry Cloud Platforms

Presentation: Industry Cloud Platform Adoption by Vertical Industry

Analyzing Industry Cloud Offerings From CIPS Providers

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

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

At the Peak

Prompt Engineering

Analysis By: Frances Karamouzis, Afraz Jaffri, Jim Hare, Arun Chandrasekaran, Van Baker

Benefit Rating: High

Market Penetration: 1% to 5% of target audience

Maturity: Emerging

Definition:

Prompt engineering is the discipline of providing inputs, in the form of text or images, to generative AI models to specify and confine the set of responses the model can produce. The inputs prompt a set that produces a desired outcome without updating the actual weights of the model (as done with fine-tuning). Prompt engineering is also referred to as "in-context learning," where examples are provided to further guide the model.

Why This Is Important

Prompt engineering is the linchpin to business alignment for desired outcomes. Prompt engineering is important because large language models (LLMs) and generative AI models in general are extremely sensitive to nuances and small variations in input. A slight tweak can change an incorrect answer to one that is usable as an output. Each model has its own sensitivity level, and the discipline of prompt engineering is to uncover the sensitivity through iterative testing and evaluation.

Business Impact

Prompt engineering has the following business impacts:

- **Performance:** It helps improve model performance and reduce hallucinations.
- **Business alignment:** It allows subject data scientists, subject matter experts and software engineers to steer foundation models, which are general-purpose in nature, to align to the business, domain and industry.
- **Efficiency and effectiveness:** Alternative options, such as building a model from scratch or fine-tuning, can be much more complex, drive longer time to market and be more expensive.

Drivers

- **Balance and efficiency:** The fundamental driver for prompt engineering is it allows organizations to strike a balance between consuming an "as is" offering versus pursuing a more expensive and time-consuming approach of fine-tuning. Generative AI models, and in particular LLMs, are pretrained, so the data that enterprises want to use with these models cannot be added to the training set. Instead, prompts can be used to feed content to the model with an instruction to carry out a function.
- **Process or task-specific customizations or new use cases:** The insertion of context and patterns that a model uses to influence the output generated allows for customizations for a particular enterprise or domain, or regulatory items. Prompts are created to help improve the quality for different use cases — such as domain-specific question answering, summarization, categorization, and so on — with or without the need for fine-tuning a model, which can be expensive or impractical. This would also apply to creating and designing new use cases that utilize the model's capability for image and text generation.
- **Validation and verification:** It is important to test, understand and document the limits and weaknesses of the models to ensure a reduced risk of hallucination and unwanted outputs.

Obstacles

- **Embryonic nature of the discipline:** Prompt engineering processes and roles are either unknown or enterprises have a low level of understanding and experience. Gartner webinar polling data (over 2,500 responses; see Executive Pulse: AI Investment Gets a Boost From ChatGPT Hype) revealed that approximately 60% of respondents self-reported that they had not heard of prompt engineering. And 90% of those same respondents revealed that their organization did not currently have prompt engineers.
- **Role alignment:** Data scientists are critical to understanding the capabilities and limits of models, and to determine whether to pursue a purely prompt-based or fine-tuning-based approach (or combination of approaches) for customization. The ultimate goal is to use machine learning itself to generate the best prompts and achieve automated prompt optimization. This is in contrast to an end user of an LLM who concentrates on prompt design to manually alter prompts to give better responses.

- **Lack of business alignment:** There is often a lack of consensus on prompt engineering's business approach, as well as agreed-upon standards, methodology and approaches. This has led to fierce debates on the value of prompt engineering and how to establish governance.

- **Risk:** Beyond the early stages of awareness and understanding, the biggest obstacle may be that prompt engineering is focused on verification, validation, improvement and refinement; however, it's not without risk. Prompt engineering is not the panacea to all of the challenges. It helps to manage risk, not remove it completely. Errors may still occur, and potential liability is at stake.

User Recommendations

- Rapidly build awareness and understanding of prompt engineering in order to quickly start the journey of shape-shifting the appropriate prompt engineering discipline and teams.

- Build critical skills across a number of different team members that will synergistically contribute critical elements. For example, there are important roles for data scientists, business users, domain experts, software engineers and citizen developers.

- Communicate and cascade the message that prompt engineering is not foolproof. Rigor and diligence need to permeate and work across all the enterprise teams to ensure successful solutions.

Sample Vendors

FlowGPT; HoneyHive; LangChain; PromptBase; Prompt Flow; PromptLayer

Gartner Recommended Reading

Quick Answer: How Will Prompt Engineering Impact the Work of Data Scientists?

Quick Answer: What Impact Will Generative AI Have on Search?

Accelerate Adoption of Generative AI by Offering an FMOps- or a Domain-Specific Partner Ecosystem

Glossary of Terms for Generative AI and Large Language Models

API-Centric SaaS

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

Benefit Rating: High

Market Penetration: 1% to 5% of target audience

Maturity: Emerging

Definition:

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

Why This Is Important

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

Business Impact

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

Drivers

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

Obstacles

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

User Recommendations

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

Sample Vendors

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

Gartner Recommended Reading

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

How to Successfully Implement API-First Integration

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

Banking Product Leader Insight: Think Beyond APIs to Address Composability

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

Edge PaaS

Analysis By: David Wright

Benefit Rating: Moderate

Market Penetration: 1% to 5% of target audience

Maturity: Emerging

Definition:

Edge PaaS is a type of cloud-oriented application platform that is purpose-built for capacity-constrained environments at the network edge. Edge PaaS platforms are designed for highly distributed, latency-sensitive and edge-aware use cases, such as interactive gaming, network function virtualization, local device management and real-time analytics.

Why This Is Important

Edge PaaS makes distributed "edge cloud applications" viable for the enterprise by offering:

- Cloud-native runtime environments (e.g., containers or WebAssembly) for easy app deployment at the network edge.
- Optimized performance on edge hardware, enabling compute-intensive processing and analytics to execute closer to edge data sources.
- Cloud-based administration combined with node-level fault tolerance, giving IT ops teams centralized control with built-in distributed resilience.

Business Impact

Edge PaaS will bring the benefits of the cloud to IT environments outside the data center:

- Enterprises can implement applications at the network edge using standard cloud DevOps tools and practices.
- SaaS providers can deploy multitenant edge services that meet enterprise use cases in areas such as AR/VR, device control and real-time analytics.
- Corporate web developers can "escape the browser" by running low-latency edge JavaScript functions at their CDN providers' network points of presence.

Drivers

- The rise of 5G networking and distributed cloud computing is driving demand for "edge cloud applications" that are built with cloud-native tools but deployed at the network edge.
- Edge infrastructure products now support a wider variety of application use cases. Gartner's current analysis of the Edge Infrastructure market forecasts high growth in areas such as distributed business processing, device control, personal monitors and immersive experiences (see Forecast Analysis: Edge Hardware Infrastructure, Worldwide).
- Edge-ready hardware acceleration. Resource-efficient chipsets and function accelerator cards (FACs) are being packaged into edge servers, making it more economical to do processor-intensive work like encryption and AI inferencing at the edge.
- Kubernetes is moving to the edge. Cloud Native Computing Foundation (CNCF)-compatible initiatives, such as MicroK8S, K3S and KubeEdge will introduce the edge to a vast number of container-oriented developers and platform engineers.
- Serverless edge Functions as a Service (Edge FaaS). The ability to compile and execute JavaScript web functions at a CDN PoP (rather than in the browser or on a remote server) will drive the development of mobile and web-based user interfaces optimized for low-latency use cases.
- Distributed cloud computing. Public cloud providers are offering distributed infrastructure solutions capable of running edge cloud applications in a customer data center using locally deployed cloud servers managed by the provider.
- The gradual emergence of Web3. The metaverse, distributed ledgers and the Interplanetary File System (IPFS) are spurring development of a next generation of globally decentralized, peer-to-peer e-commerce and financial systems.
- The rise of Edge AI. Gartner's 2022 Emerging Tech Impact Radar: Edge AI identifies a range of AI-enabled edge technologies that are poised to broadly disrupt markets in the coming decade.

Obstacles

- Lack of coherent edge strategies in the enterprise. Many organizations have been slow to recognize the value of deploying services at the network edge and have yet to formulate a plan.
- Incomplete toolsets. While many standard DevOps tools can be used with Edge PaaS platforms, they lack integrated support for their “edge-native” capabilities.
- Insufficient flexibility. Many Edge PaaS offerings today do not include out-of-the box connectors to legacy enterprise systems or public cloud services, making integration more difficult.
- Potential fracturing of the Kubernetes standard at the edge.
- Standard CNCF-compliant K8S is not optimal for use cases where worker machines in a cluster are dispersed across an unreliable network, as many control plane services needed by nodes require continuous access to a cluster server. While several CNCF incubation projects exist to address this, a market leading alternative has not yet emerged.

User Recommendations

- Work with business stakeholders in your organization to examine their unmet needs at the network edge, and map Edge PaaS platform capabilities to these unmet needs.
- Distinguish opportunities for container-based Edge PaaS applications from lightweight “edge web” needs that would be best addressed by serverless Edge FaaS, and from IoT use cases best addressed by IoT-specific platforms.
- Many Edge PaaS offerings in the market are immature today. Select potential Edge PaaS or FaaS vendors for pilot projects based on a balanced mix of current capabilities and committed roadmap features.
- When building early Edge PaaS and FaaS prototypes, evaluate how well they integrate with your existing DevOps teams, tools and processes to ensure compatibility with your broader cloud strategy.
- Formulate an initial edge strategy around emerging disruptive technologies like Web3 and Edge AI by first determining how these innovations are likely to affect your customers and your market.

Sample Vendors

Akamai; Avassa; Cloudflare; EDJX; F5; Fastly; Section; StarlingX; Taubyte

Platform Engineering

Analysis By: Bill Blosen, Paul Delory

Benefit Rating: Transformational

Market Penetration: 20% to 50% of target audience

Maturity: Adolescent

Definition:

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

Why This Is Important

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

Business Impact

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

Drivers

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

Obstacles

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

User Recommendations

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

Gartner Recommended Reading

How to Start and Scale Your Platform Engineering Team

Guidance Framework for Implementing Cloud Platform Operations

Adopt Platform Engineering to Improve the Developer Experience

Innovation Insight for Internal Developer Portals

Generative AI

Analysis By: Svetlana Sicular, Brian Burke

Benefit Rating: Transformational

Market Penetration: 1% to 5% of target audience

Maturity: Adolescent

Definition:

Generative AI technologies can generate new derived versions of content, strategies, designs and methods by learning from large repositories of original source content. Generative AI has profound business impacts, including on content discovery, creation, authenticity and regulations; automation of human work; and customer and employee experiences.

Why This Is Important

Generative AI exploration is accelerating, thanks to the popularity of Stable Diffusion, Midjourney, ChatGPT and large language models. End-user organizations in most industries aggressively experiment with generative AI. Technology vendors form generative AI groups to prioritize delivery of generative-AI-enabled applications and tools. Numerous startups have emerged in 2023 to innovate with generative AI, and we expect this to grow. Some governments are evaluating the impacts of generative AI and preparing to introduce regulations.

Business Impact

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

Drivers

- The hype around generative AI is accelerating. Currently, ChatGPT is the most hyped technology. It relies on generative foundation models, also called “transformers.”
- New foundation models and their new versions, sizes and capabilities are rapidly coming to market. Transformers keep making an impact on language, images, molecular design and computer code generation. They can combine concepts, attributes and styles, creating original images, video and art from a text description or translating audio to different voices and languages.
- Generative adversarial networks, variational autoencoders, autoregressive models and zero-/one-/few-shot learning have been rapidly improving generative modeling while reducing the need for training data.
- Machine learning (ML) and natural language processing platforms are adding generative AI capabilities for reusability of generative models, making them accessible to AI teams.
- Industry applications of generative AI are growing. In healthcare, generative AI creates medical images that depict disease development. In consumer goods, it generates catalogs. In e-commerce, it helps customers “try on” makeup and outfits. In manufacturing, quality inspection uses synthetic data. In semiconductors, generative AI accelerates chip design. Life sciences companies apply generative AI to speed up drug development. Generative AI helps innovate product development through digital twins. It helps create new materials targeting specific properties to optimize catalysts, agrochemicals, fragrances and flavors.
- Generative AI reaches creative work in marketing, design, music, architecture and content. Content creation and improvement in text, images, video and sound enable personalized copywriting, noise cancellation and visual effects in videoconferencing.
- Synthetic data draws enterprises’ attention by helping to augment scarce data, mitigate bias or preserve data privacy. It boosts the accuracy of brain tumor surgery.
- Generative AI will disrupt software coding. Combined with development automation techniques, it can automate up to 30% of the programmers’ work.

Obstacles

- Democratization of generative AI uncovers new ethical and societal concerns. Government regulations may hinder generative AI research. Governments are currently soliciting input on AI safety measures.
- Hallucinations, factual errors, bias, a black-box nature and inexperience with a full AI life cycle preclude the use of generative AI for critical use cases.
- Reproducing generative AI results and finding references for information produced by general-purpose LLMs will be challenging in the near term.
- Low awareness of generative AI among security professionals causes incidents that could undermine generative AI adoption.
- Some vendors will use generative AI terminology to sell subpar “generative AI” solutions.
- Generative AI can be used for many nefarious purposes. Full and accurate detection of generated content, such as deepfakes, will remain challenging or impossible.
- The compute resources for training large, general-purpose foundation models are heavy and not affordable to most enterprises.
- Sustainability concerns about high energy consumption for training generative models are rising.

User Recommendations

- Identify initial use cases where you can improve your solutions with generative AI by relying on purchased capabilities or partnering with specialists. Consult vendor roadmaps to avoid developing similar solutions in-house.
- Pilot ML-powered coding assistants, with an eye toward fast rollouts, to maximize developer productivity.
- Use synthetic data to accelerate the development cycle and lessen regulatory concerns.
- Quantify the advantages and limitations of generative AI. Supply generative AI guidelines, as it requires skills, funds and caution. Weigh technical capabilities with ethical factors. Beware of subpar offerings that exploit the current hype.
- Mitigate generative AI risks by working with legal, security and fraud experts. Technical, institutional and political interventions will be necessary to fight AI's adversarial impacts. Start with data security guidelines.
- Optimize the cost and efficiency of AI solutions by employing composite AI approaches to combine generative AI with other AI techniques.

Sample Vendors

Adobe; Amazon; Anthropic; Google; Grammarly; Hugging Face; Huma.AI; Microsoft; OpenAI; Schrödinger

Gartner Recommended Reading

Innovation Insight for Generative AI

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

Emerging Tech: Venture Capital Growth Insights for Generative AI

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

ChatGPT Research Highlights

Cloud-Native

Analysis By: David Smith, Michael Warrilow

Benefit Rating: High

Market Penetration: 5% to 20% of target audience

Maturity: Adolescent

Definition:

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

Why This Is Important

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

Business Impact

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

Drivers

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

Obstacles

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

User Recommendations

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

Gartner Recommended Reading

The Cloud Strategy Cookbook, 2023

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

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

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

SaaS as a Platform

Analysis By: Paul Vincent, Yefim Natis

Benefit Rating: Moderate

Market Penetration: 20% to 50% of target audience

Maturity: Early mainstream

Definition:

SaaS as a platform is the application platform underpinning a vendor's SaaS made available for other custom application development. It provides a catalog of the underlying SaaS technologies and business services to a (typically low-code) development environment supporting the development of both SaaS extensions and new applications, while reusing the underlying SaaS platform services for business logic, data, process, user experience and integration.

Why This Is Important

SaaS as a platform offerings are already disrupting the IT business. Salesforce is now the largest vendor in application infrastructure and middleware services. Microsoft, Oracle, SAP and ServiceNow also provide popular SaaS offerings with PaaS extension and development capabilities to their customers. These vendors are often considered by their customers as strategic service providers but are rapidly evolving into strategic technology platforms too. Adoption continues to expand.

Business Impact

SaaS megavendors consume increasing proportions of IT budgets. Business audiences are attracted to SaaS's commoditized services with predictable pricing, but also need differentiating custom applications and extensions. SaaS as a platform enables customers to exploit SaaS vendors' development tooling and composable services to meet their own application goals.

Drivers

- SaaS usage continues to expand, yet the need to differentiate these services to custom audiences and practices remains consistent. SaaS as a platform usage continues to grow to support the need for more customization of more instances of SaaS (and indeed across growing lists of SaaS vendors).
- Consumption of API services to provide composite/composed applications continues to grow. SaaS APIs enable applications to deliver more and extended use cases through access to shared SaaS data and services.
- The low-code application platform (LCAP), business process automation (BPA) and integration platform as a service (iPaaS) capabilities embedded in most core SaaS as a platform offerings also represent high-growth digital platform development technologies. SaaS as a platform provides easy access to these technologies for customer IT and business technologist developers, evidenced by Salesforce's success in becoming the largest application and infrastructure middleware services vendor.
- SaaS platform vendors continue to innovate in order to compete, extending further and faster the scope of their platform offerings. This, in turn, feeds interest and adoption from IT and business technologist teams.
- Hype is low but rising, as SaaS as a platform adoption evolves from departmental to strategic and enterprise use cases, and SaaS vendors increase their marketing to CIOs and IT leaders.

Obstacles

- Lock-in to strategic SaaS partners increases when usage extends to custom extensions and other applications. This results in a customer dilemma: should you invest more in your strategic SaaS partners, or diversify platform investments with best-of-breed solutions using intervendor integration?
- The technology components of SaaS vendors can be incomplete or immature compared with specialist competition. Examples are support for complex architectures, business process standards, common component catalogs, real-time event handling, decision management, or governed citizen development. SaaS-based LCAPs can also inherit a lack of B2C support and cloud portability from their SaaS parentage.
- Overall vendor charges for SaaS and associated platforms are rising as usage increases, typically based on growth of end-user-named user seats, and pricing will continue to be a major inhibitor on wider adoption from departmental to strategic and enterprise use cases.

User Recommendations

- Explore the implications of SaaS as a platform versus a best-of-breed approach for applications and platforms. Overlapping platforms will add cost, skill dilution and overlapping capabilities, but will also reduce exposure and risk to any specific vendor.
- Ensure any SaaS as a platform selection meets your requirements for platforms for custom application platform development. These could include composability to create new and custom applications via APIs, strong integration capabilities, user experience, developer experience, and complex business logic.
- Beware of legacy SaaS vendors that rely on multiple underlying platforms and architectures with disparate API styles and limited integrations, which increase operational management costs and complexity of reuse.
- Beware of high costs associated with menu-driven pricing for these platforms, and mitigate those costs through contract best practices.

Sample Vendors

Creatio; Microsoft; Oracle; Pegasystems; Salesforce; SAP; ServiceNow; Zoho

Gartner Recommended Reading

How to Navigate the Application Platforms Market Including Cloud-Native, Low-Code and SaaS

Identify and Evaluate Your Next Low-Code Development Technologies

How to Govern and Fund Platforms in a Product World

How to Establish Effective SaaS Governance

Multicloud

Analysis By: David Smith

Benefit Rating: High

Market Penetration: 20% to 50% of target audience

Maturity: Early mainstream

Definition:

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

Why This Is Important

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

Business Impact

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

Drivers

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

Obstacles

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

User Recommendations

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

Gartner Recommended Reading

The Cloud Strategy Cookbook, 2023

A Multicloud Strategy Is Complex and Costly, but Improves Flexibility

A CTO's Guide to Multicloud Computing

Cloud Portability

Analysis By: Lydia Leong, David Smith

Benefit Rating: High

Market Penetration: 1% to 5% of target audience

Maturity: Emerging

Definition:

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

Why This Is Important

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

Business Impact

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

Drivers

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

Obstacles

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

User Recommendations

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

Gartner Recommended Reading

Infographic: Mitigate Cloud Risks With Realistic Exit Planning

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

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

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

How to Manage Concentration Risk in Public Cloud Services

Distributed Cloud

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

Benefit Rating: High

Market Penetration: 1% to 5% of target audience

Maturity: Emerging

Definition:

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

Why This Is Important

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

Business Impact

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

Drivers

- Historically, location has not been relevant to cloud computing definitions. In fact, the variations on cloud (e.g., public, private, hybrid) exist because location can vary.
- Distributed cloud supports both tethered and untethered operations of cloud services from the cloud provider, “distributed” out to specific and varied physical locations. This enables an important characteristic of distributed cloud operation — low-latency compute where the compute operations for the cloud services are closer to those that need the capabilities. This can deliver major improvements in performance and reduce the risk of global network-related outages.
- Data sovereignty and other regulatory issues may require services be delivered from locations beyond the data centers of the public cloud service provider.
- Perceived and real security and privacy concerns with off-premises applications and infrastructure drive some consumers to prefer on-premises solutions.
- Latency needs of IoT/edge applications require services to be located close to the edge.
- Distributed cloud is still a single-cloud provider, and the managed cloud assets are still part of the cloud provider’s portfolio.
- Disconnected operations can be supported with distributed services that can operate independently.

Obstacles

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

User Recommendations

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

Sample Vendors

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

Gartner Recommended Reading

The Cloud Strategy Cookbook, 2023

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

Distributed Cloud: Does the Hype Live Up to Reality?

Enhanced Blockchain as a Service

Analysis By: Adrian Leow, Avivah Litan

Benefit Rating: High

Market Penetration: 5% to 20% of target audience

Maturity: Adolescent

Definition:

Enhanced blockchain as a service (eBaaS) is a cloud-based service that provides an application layer for developers to build, deploy and host their blockchain-based applications, functions and smart contracts. eBaaS abstracts the underlying blockchain infrastructure, enabling developers to focus on the application layer without having to worry about the underlying blockchain technology. It enables developers to deploy their applications across multiple clouds and blockchains.

Why This Is Important

eBaaS simplifies blockchain adoption for enterprises and accelerates deployments by providing a secure and scalable environment for app development and deployment. It offers standardized APIs and tools for easier development and integrates with existing enterprise systems. With robust monitoring and management capabilities, eBaaS ensures optimal operation, supports multiple blockchain platforms and provides greater flexibility and choice for developers and enterprises.

Business Impact

eBaaS unlocks new opportunities for businesses to leverage blockchain technology by reducing entry barriers, accelerating time to market and fostering business-related innovation. With the ability to integrate with existing enterprise systems, eBaaS can improve efficiency, security and transparency across industries, helping businesses stay competitive in a rapidly evolving digital landscape. eBaaS removes a good portion of the technical obstacles for businesses looking to take advantage of the benefits of blockchain technology.

Drivers

- eBaaS enables enterprises to focus on their applications without worrying about which blockchain and cloud to use, which smart contract development platform to use and how to connect legacy data to new Web3 applications.
- The benefits of unique blockchain features, such as smart contracts and immutable distributed ledgers, are much more accessible to enterprises when they use eBaaS, as they are shielded from most of the complexities of blockchain infrastructure and protocols.
- Low-code development environments enable developers to deploy applications quickly and avoid costly professional services.
- Some eBaaS vendors have off-the-shelf application templates that accelerate deployment for specific use cases.
- Blockchain application maintenance and performance monitoring are simplified through eBaaS services.
- Some eBaaS vendors support the relatively straightforward migration of applications from permissioned blockchains to public blockchains that support security through decentralization.

Obstacles

- Most eBaaS vendors support limited options for back-end blockchains, so users may be tied to the vendor's blockchains of choice. The historical lack of industry standardization across different blockchain platforms can make it challenging for businesses to select the best fit for their use cases, whether they use eBaaS or not.
- eBaaS vendors may lag in the adoption of public chain innovations unless their architectures are modular and open to allow easy integration of new capabilities.
- For a successful blockchain deployment, enterprises must participate equally with other organizations using their applications. eBaaS services may find it difficult to satisfy the requirements of multiple organizations.
- Enterprises participating in eBaaS (or any non-eBaaS application) must agree on data exchange formats, governance and permissions before deploying their applications.
- Enterprises are often confused on how they can use and benefit from blockchain technology in the first place. eBaaS can help guide them toward worthwhile business scenarios, but does not guarantee alignment.

User Recommendations

- Evaluate and use eBaaS services to accelerate the deployment of your applications, once you have agreed on business and process terms and addressed funding and governance issues with your ecosystem partners.
- Select an eBaaS provider that targets your use case or industry and has experience with it.
- Select an eBaaS provider that supports permissioned blockchain today, and which either can or is planning to support public blockchains in the future. This will ensure that your organization benefits from blockchain decentralization and fast-moving innovations in the public blockchain arena.

Sample Vendors

Alchemy; Fujitsu; IBM; Kaleido; NTT DATA; Oracle; SettleMint; Sky Republic; Vendia

Sliding into the Trough

Cloud Event Brokers

Analysis By: Yefim Natis, Gary Olliffe, Max van den Berk, Keith Guttridge

Benefit Rating: High

Market Penetration: 5% to 20% of target audience

Maturity: Adolescent

Definition:

Event brokers play the role of an intermediary in event-driven architecture (EDA), administering the event topics, registering event publishers and subscribers, and facilitating the capture and distribution of event notifications. Cloud event brokers expand these capabilities with cloud-native characteristics of elastic scalability and multitenancy. They are optimized to support cloud-based applications and operations.

Why This Is Important

Most of the outcomes of digital business transformation depend, in part, on an organization's continuous awareness of relevant business events and its ability to respond in real time. Event brokers facilitate the monitoring and distribution of event notifications to application services for an automated response, dashboards for human action, or data stores for further analysis. Cloud event brokers bring these capabilities to cloud-based applications and operations.

Business Impact

Organizations aware of their ecosystem events are better prepared to manage unexpected interruptions and capitalize on opportunities in business moments. They are equipped to broadcast notable events for simultaneous, multitargeted responses. Event brokers enable organizations' versatility in monitoring multiple sources of events and communicating to many responders in parallel with strong scalability, integrity and resilience. Cloud event brokers extend this functionality to on-cloud operations.

Drivers

- Increased demand for real-time insights drives organizations to manage event streaming and stream analytics, leading, in turn, to event brokers for governance and coordination of event traffic.

- Increased adoption of Apache Kafka, by both businesses and leading technology vendors, promotes organizational awareness of the benefits and opportunities that event-driven application design brings.
- The migration of business applications to the cloud demands new platforms and communication infrastructure, driving many organizations to evaluate and adopt event broker services, paired with integration and API management offerings.
- The availability of multiple vendors' cloud event broker services, based on open-source standards such as Advanced Message Queuing Protocol (AMQP), Apache Pulsar, Apache Kafka, MQTT and NATS, provides competitive and differentiated options in event broker services for a better-tuned fit to customers' use cases.
- Increased maturity of cloud event brokers supports more advanced capabilities in performance, data and process management, and optimization of event-driven applications.
- Most leading SaaS offerings support some event processing, increasing awareness of the benefits and opportunities of event-driven application design in a large number of mainstream business and government organizations.
- The increasing popularity of digital integration hubs and data virtualization approaches deliver near-real-time data to applications by consuming event streams, instead of direct database lookups.
- Growing demand for advanced use patterns of "change cloud-data capture" drive new adoption of cloud event brokers.

Obstacles

- Cloud event broker offerings become too expensive as more proprietary features are added to help differentiate from the competition.
- Event broker functionality, embedded in some platform and application services, fragments control of event streaming across the organization, while delaying a systematic investment in event brokering.
- Some software engineering teams use webhook and WebSocket tools to set up event notifications, delaying the full many-to-many experience of EDA that's implemented via an event broker technology.

- Lack of universally supported standards for protocols or APIs for EDA implementation increases the costs and complexity of managing a large event-driven application infrastructure.
- Some organizations choose the request/reply communication where EDA would be a more effective model, just due to the unmitigated familiarity bias and lack of required skills.

User Recommendations

- Mix up the complementary strengths of the request/reply service-oriented architecture (SOA) and EDA, and encourage new projects to consider the combined use of both, as appropriate.
- Pilot experimental projects using cloud event brokers to gain insight and skills for upcoming, more advanced projects. Even a basic pub/submiddleware service is sufficient as a precursor for a full-featured event broker.
- Give preference to cloud event broker vendors demonstrating the understanding of the full life cycle of event processing functionality and responsibility.
- Plan for coordinated use of an event broker and an event stream processing (ESP) platform (also called stream analytics platform). The technologies are different but contribute to a shared business outcome.

Sample Vendors

Amazon Web Services (AWS); Confluent; Google; IBM; Microsoft; Solace; TIBCO Software

Gartner Recommended Reading

Using Event-Driven Integration With Enterprise Applications

Essential Patterns for Event-Driven and Streaming Architectures

Apply Event-Driven Approaches to Modernize IoT Data Integration

How to Identify Your Event-Driven Architecture Use Cases to Select the Best-Fit Event Broker

Maturity Model for Event-Driven Architecture

Digital Integration Hub

Analysis By: Andrew Humphreys

Benefit Rating: High

Market Penetration: 5% to 20% of target audience

Maturity: Adolescent

Definition:

A digital integration hub (DIH) provides low-latency, high-throughput API- and event-based data access by aggregating and replicating multiple application sources into a data management layer that synchronizes with the applications via event-driven patterns. A DIH enables scalable, 24/7 data access, reduces workloads on the systems of record (SORs) and improves business agility. Organizations can reap additional value by using a DIH in analytics, data integration and composition scenarios.

Why This Is Important

Digital initiatives leverage APIs and, increasingly, events to unlock core business applications data and business logic. Traditional integration architectures can face severe performance, scalability and availability issues stemming from excessive workload generated in the SORs from new digital requests. A DIH is an increasingly popular alternative architecture to overcome these issues and delivers additional benefits that improve business agility.

Business Impact

DIHs can have the following business impacts:

- Provide digital application clients with a richer and more responsive experience than is achieved by accessing applications directly.
- Reduce application workload and limit the fees paid to SaaS providers for API consumption.
- Help enable 24/7 operations. The hub is not dependent on application availability.
- Improve business agility by decoupling the API layer from the applications.
- Maintain an up-to-date picture of fast-changing data used for analytics, notification services and data integration.

Drivers

- DIHs can help organizations lower operational costs by reducing the high workload generated in SORs by digital application requests.
- By enabling data to be stored locally, DIHs reduce the API-limit fees paid to SaaS providers, as well as the number of calls that need to be made to SaaS providers.
- DIHs deliver a more responsive and data-rich user experience than traditional integration architectures, leading to improved customer satisfaction.
- Using a DIH can accelerate the transition to composable business, digital and API economy by maintaining an up-to-date picture of fast-changing data needed to support a comprehensive set of APIs and events.
- DIHs give digital teams greater flexibility in building new offerings, by enabling them to access any data in the hub without needing to involve the SOR teams.
- Many vendors are introducing packaged DIH offerings, at times focused on specific use cases.

Obstacles

- Assembling and managing the varied set of DIH building blocks (API gateways, application platforms, integration platforms, event brokers, data management and metadata management tools) is very complex.
- Dealing with an architecture that is not well-known in the industry implies a scarcity of know-how, experience and skills, leading to high costs.
- Keeping the data management layer in sync with the systems of record by leveraging event-based integration tools (for example, change data capture) is challenging.
- Addressing the data governance issues derived from the creation of yet another copy or data structure outside the systems-of-record data is cumbersome.

User Recommendations

- Adopt a DIH-enabled architecture to provide a rich, responsive omnichannel experience for large audiences (hundreds of thousands of users or greater) and to reduce the cost associated with sustaining the digital initiative generated workload hitting the SORs.
- Enable API "pull" and event "push" services to access data scattered across multiple back-end systems.
- Decouple digital services from SORs to enable more flexible and composable business applications.
- Maintain an up-to-date "single source of truth" for fast-changing data, which can be used to provide additional services (for example, custom analytics or search) or analyzed in real time to detect "business moments."
- Embed DIH initiatives into the overall data hub strategy for governance and integration to avoid ending up with yet another data silo.

Sample Vendors

Cinchy; GigaSpaces, IBM; Informatica; Mia-Platform; Microsoft; Oracle; SAP; Sesam; Software AG

Gartner Recommended Reading

How Software Engineering Leaders Can Overcome the 4 Key Packaged Business Application Challenges

Essential Patterns for Data-, Event- and Application-Centric Integration and Composition

Blockchain Oracles

Analysis By: Avivah Litan

Benefit Rating: Transformational

Market Penetration: 1% to 5% of target audience

Maturity: Adolescent

Definition:

Blockchain oracles are services that support the secure interaction of smart contracts with systems external to a blockchain (or distributed ledger). Oracles enable smart contracts to automate transactions based on real-world data and external events. They enable, for example, the fetching of commodity prices and notifications of changing interest rates.

Why This Is Important

Oracles enable blockchain smart contracts to interact with real-world events, thus integrating blockchain networks to the external world. As distributed ledgers and smart contracts do not have a mechanism to interact with, or process data from, the external world, decentralized oracle networks can help build a trusted mechanism to utilize blockchains in business processes. They contrast with APIs, which represent centralized points of potential compromise.

Business Impact

The availability of reliable and effective blockchain oracles will play a crucial role in the integration of smart contracts with real-world assets and events. Once participants in a blockchain network recognize and trust the authority of an oracle, blockchain business processes can operate without supervision of data inputs and outputs to/from smart contracts. This unsupervised operation can deliver on the promise of blockchain efficiency and trust minimization.

Drivers

- **Recognition of smart contracts' value in reducing transaction costs:** Smart contracts, by reducing human recording, authentication and reconciliation of business transactions, enable a significant increase in the granularity and efficiency of transactions. Blockchain oracles provide necessary controls, triggers and off-chain computation for automating trustless smart contract operations.
- **Decentralized oracle systems provide trusted transaction triggers in support of processing:** Decentralized oracle networks provide security for interfaces between off-chain events and on-chain smart contracts. Multiple nodes participate in validating and processing inputs to smart contracts and the outputs they generate.
- **Needed security:** Decentralized oracle interfaces to blockchain smart contracts are needed in light of severe hacks against centralized interfaces. Sufficient decentralization will mitigate security risks posed by illegitimate data interfacing with smart contracts.
- **Trustworthy data:** Decentralized oracle systems reach consensus on the data that is used to feed smart contracts. This mitigates the risk of incorrect data coming from a single source into a smart contract, which could have damaging consequences. For example, an incorrect exchange rate provided by a single "bad" source could force collateralized loan liquidations that would not occur if the correct exchange rate had been fed into the lending smart contract via a decentralized oracle network.
- **Increasing adoption of decentralized applications (dapps),** which will drive growth in blockchain oracles as they play a critical role in enabling dapp interactions between external systems and smart contracts.

Obstacles

- Integrating existing systems of record and data with blockchain oracles is challenging. Enterprises have internal systems of record, typically built on robust, but dated, systems and software. These systems must be updated to ensure they can interface with blockchain oracle systems.
- Blockchain oracles are dependent on the information sources they draw from. If these data sources are compromised and the network is not sufficiently decentralized, oracles could report falsified transaction events. These falsified events could, for example, cause payments to pass without delivery of products or services.
- If a smart contract commits an incorrect transaction to a blockchain, it is difficult to adjust the records. In cases not involving anonymous parties, the transaction can be reversed by an appended counter transaction.
- There are very few decentralized blockchain oracle network protocols. The market needs more competition to ensure long-term viability for blockchain participants.

User Recommendations

- **Identify integration needs between systems of record with blockchain oracles.** For enterprises, integration of oracles with existing systems of record is critical for a successful blockchain implementation. This integration is not a trivial exercise, so develop plans to identify contact points and define changes where necessary.
- **Build appropriate integrity and reliability mechanisms into blockchain oracle systems.** Unless sufficiently decentralized, blockchain oracles are at risk from lost or corrupted input signals that can trigger fraudulent transactions or suppress legitimate transactions. Choose oracles by carefully selecting and validating their state of decentralization and security.
- **Create processes to adjudicate and reverse incorrect transactions by creating new ones.** When a participant comes to you reporting a failed transaction, you need a predetermined agreement for handling this issue.

Sample Vendors

Band Protocol; Chainlink; DIA; XYO Network

Gartner Recommended Reading

Managing the Risks of Enterprise Blockchain Smart Contracts

How to Mitigate Web3 Blockchain Risks and Security Threats

Guidance for Blockchain Solution Adoption

Data Ecosystems

Analysis By: Adam Ronthal, Robert Thanaraj, Aaron Rosenbaum

Benefit Rating: High

Market Penetration: 1% to 5% of target audience

Maturity: Emerging

Definition:

Data ecosystems provide a cohesive data management environment that supports all data and analytics workloads. They have a common governance and metadata management framework and unified access management. They integrate augmented data management capabilities with a set of services accessible by the business user. They provide streamlined delivery and comprehensive functionality that is straightforward to deploy, optimize and maintain. Third-party vendors also participate in data ecosystems.

Why This Is Important

Data and analytics leaders report that the cloud experience today requires a significant integration effort to ensure that components work well together. Cloud service providers (CSPs) and independent software vendors (ISVs) are responding with more refined data ecosystems as the market moves from “some assembly required” to a “packaged platform experience.”

Business Impact

Data ecosystems unify data management and associated use cases with streamlined delivery and easy integration within a holistic management framework. They address key data management disciplines, such as data integration, data quality, data sharing, governance, metadata and observability — via augmentation — and provide the basis for operational and analytics capabilities. They are delivered both by a combination of native CSP offerings, as well as ISV components that provide enhanced capabilities.

Drivers

- Data and analytics architectures are under significant stress on two fronts: hybrid and multicloud deployment environments, and the diversity of data persistence models required to meet the increasing demands of data and analytics.
- Cloud practitioners need to rationalize data silos, which span multiple deployment environments and frequently require different and potentially conflicting operating models.
- Enterprises are looking to unify the way they engage with different data models, platforms and use cases to improve efficiency and time to value for data-driven initiatives.
- Data ecosystems serve as a unifying approach to resolve these pressure points. Built on a common foundation of governance, metadata and emerging data fabric design, they enable new practices like DataOps, FinOps and PlatformOps. They will become self-optimizing and self-tuning, and support financial governance efforts through cost optimization.
- Data ecosystems promise improved productivity and ROI based on the value of not having to do explicit data and application integration as they are based on a common set of services.

Obstacles

- While data ecosystems have a vision of unifying data management with common governance, security and metadata, significant work is still needed to make this a reality. Gaps exist in data integration, data quality, metadata and governance, which need to be addressed either through native CSP offerings or partnerships with ISVs to fully realize the vision of the cloud data ecosystem.
- When combining native CSP offerings with third-party ISV offerings, end users may find that additional effort is required to integrate these components. This undermines the core concept of a unified, holistic data ecosystem, though the end result leveraging CSP and ISV offerings is likely to be more capable.
- While CSPs are working with third-party ISVs to provide open ecosystems, their initial focus remains on ensuring that their own components are working well within their own cloud and addressing the basic needs of their customers.
- For a cloud ecosystem (that encompasses a CSP and ISVs) to function well, CSP and ISV components must have standard interfaces that exchange metadata bidirectionally. It is essential that participating ISVs and CSP agree on common metadata sharing standards. This remains slow to emerge in the market today.

User Recommendations

- Assess the maturity of these ecosystems and the degree to which they deliver on the promise of a unified environment.
- Assess points of integration between various components (data persistence, use cases, data integration, observability, governance and metadata capabilities) to determine how cohesive the resulting ecosystem is. A less cohesive ecosystem will require significantly more integration time and effort.
- Ensure that the data ecosystem has a well-articulated path to production for a full data life cycle (from discovery to production-optimized delivery).
- Define what CSPs need to deliver as part of the solution and what capabilities to obtain from third-party ISVs; expect to spend more time on integration efforts when combining CSP and ISV offerings.

Sample Vendors

Amazon Web Services (AWS); Cloudera; Databricks; Google Cloud Platform; IBM; Microsoft; Oracle; SAP

Gartner Recommended Reading

The Impacts of Data Ecosystems: A Cloud Architectural Perspective

Innovation Insight: Data Ecosystems Will Reshape the Data Management Market

Strategic Roadmap for Migrating Data Management to the Cloud

Private Cloud dbPaaS

Analysis By: Adam Ronthal

Benefit Rating: Moderate

Market Penetration: 5% to 20% of target audience

Maturity: Early mainstream

Definition:

Private cloud database platform as a service (dbPaaS) offerings bring self-service and scalability to a private cloud infrastructure, which minimizes external exposure. Private cloud dbPaaS offerings should provide similar benefits to their public cloud counterparts — a database management system or a data store engineered as a scalable, elastic, multitenant service, ideally with subscription of chargeback pricing models.

Why This Is Important

Private cloud dbPaaS offerings primarily appeal to enterprises that cannot embrace public cloud offerings (usually due to regulatory or governance concerns), but want cloud-like service delivery. These offerings may leverage the existing container-based infrastructure common to many private cloud offerings. They may also be self-contained products in an appliance form factor or extensions of existing cloud service provider offerings.

Business Impact

Private cloud dbPaaS offerings promise a marketplace-like experience for a range of DBMS offerings: commercial and open-source, relational and nonrelational. Many are still maturing to offer services that go beyond self-provisioned developer environments to true production-class environments. These offer high availability, elastic scalability and solid performance. A number of offerings from established vendors like IBM (Cloud Pak for Data System) and Oracle (Cloud@Customer) are available.

Drivers

- Private cloud dbPaaS will primarily appeal to enterprises that are not yet ready or are unable to embrace public cloud alternatives. Public cloud inhibitors (and private cloud dbPaaS drivers) may include: (1) regulatory, governance or security requirements, or the need to operate in an “air gapped” environment; (2) significant on-premises centers of gravity that are not yet able to move to public cloud; (3) concerns with network connectivity, latency or performance issues in a hybrid cloud environment; (4) data sovereignty requirements that cannot be met by public cloud data centers; and (5) compatibility concerns with on-premises environments associated with native public cloud offerings.

- Initial offerings in the space have come from traditional vendors with a strong on-premises presence. But cloud service providers are now engaging with private cloud dbPaaS offerings as well, almost always based on a container strategy that reaches into on-premises data centers.

Obstacles

- “Public cloud first” strategies remain a primary focus for vendors and end users alike, which may inhibit breadth of offerings as vendors increasingly focus on public cloud approaches.
- Many public cloud inhibitors are transient in nature, and may be addressed as organizational culture adapts to the cloud, regional availability improves and regulatory best practices associated with operating in the cloud improve. Private cloud dbPaaS may be a transitional stage for many adopters.
- Adoption will be limited to established enterprises with existing on-premises data center footprints, or via trusted partners or systems integrator.
- The requirement to leverage governance, security and operational updates by connecting to a control plane raises questions about how truly separated some private cloud offerings are.

User Recommendations

Data and analytics leaders considering private cloud DbPaas must consider:

- **Breadth of DBMS services offered:** Not all offerings support a full range of database types and features.

- **Storage model:** Container-based services require a scalable and persistent data storage tier for effective usage.
- **Pricing model:** Flexibility of pricing models is beneficial to accommodate both capital expenditure (capex) and operating expenditure (opex) approaches. Cloud pricing models (including private cloud offerings) can be complex.
- **Production capabilities:** High-availability and disaster-recovery capabilities must meet your requirements.
- **Disconnected operations:** Many of these offerings have a cloud-based management control plane. They must meet requirements for disconnected operations, if connectivity to the cloud is unreliable.
- **Path to the public cloud:** Private cloud dbPass will be a transitional technology. Continuity from on-premises operations into the public cloud is an important consideration.

Sample Vendors

Aiven; Alibaba Cloud; Amazon Web Services; IBM; Microsoft; Nutanix; Oracle; Pure Storage (Portworx); Robin; VMware

Gartner Recommended Reading

Market Guide for Container Management

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

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

Differentiate Hosted Private Cloud Offerings Using These 7 Dimensions

Magic Quadrant for Cloud Database Management Systems

Container Management

Analysis By: Dennis Smith, Michael Warrilow

Benefit Rating: High

Market Penetration: 20% to 50% of target audience

Maturity: Early mainstream

Definition:

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

Why This Is Important

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

Business Impact

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

Drivers

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

Obstacles

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

User Recommendations

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

Sample Vendors

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

Gartner Recommended Reading

Market Guide for Container Management

Cloud MDM

Analysis By: Sally Parker, Helen Grimster

Benefit Rating: High

Market Penetration: 20% to 50% of target audience

Maturity: Early mainstream

Definition:

Cloud master data management (MDM) solutions ensure the uniformity, accuracy, stewardship, governance, semantic consistency and accountability of the enterprise's official shared master data assets. Available in the cloud across a spectrum of resource delivery models, these range from single-tenant shared nothing (IaaS) to multitenant shared something (PaaS) to multitenant shared everything (SaaS).

Why This Is Important

Trusted master data is a foundational requirement of digital business. As organizations' applications have transitioned to the cloud as part of their digital transformation programs, so too has their associated data. This has subsequently shifted the center of gravity in favor of cloud-based MDM solutions.

Business Impact

The MDM market was relatively late in transitioning from on-premises to cloud-based solutions. Cloud-based MDM solutions have lowered the barrier to entry for MDM with the provision of access to subscription-based licensing models, deployment flexibility and improved time to value. They have effectively enabled support for the MDM best practice of "think big, start small, deliver incremental business value." As the center of gravity shifts to the cloud, it opens up greater opportunities for data sharing and syndication ecosystems.

Drivers

- **Stated vendor direction is cloud:** Vendors will ultimately pull the market to the cloud with product roadmap priorities supporting cloud-based platforms to streamline their own product management overheads.
- **Gravitational pull of their application and data ecosystems:** MDM creates a single source of truth for master data across the enterprise's heterogeneous application landscape. As the center of gravity for these applications and their data shifts to the cloud, MDM logically follows.
- **Acceptance of cloud for master data:** MDM has been slow to follow the broader software solutions market in transitioning to cloud deployment models. Vendors that previously delayed offering cloud-based solutions are responding to demand from end-user organizations now ready to embrace cloud for their most critical data — their master data.
- **Lower barrier to entry:** Cloud has lowered the barrier to entry for MDM programs, permitting expansion into a previously untapped and broader client base. SaaS also reduces some MDM skills pain points.
- **Increased availability of cloud-based offerings:** Few MDM vendors were cloud native from the outset. Through the end of 2022, MDM software solution vendors continued and completed their transitions to subscription and cloud-based solutions.
- **Scalability:** To handle compute intensive requirements such as ML/AI for matching.
- **Delivery of incremental business value:** Facilitates the best practice of a more granular and business outcome drive approach to MDM.

Obstacles

- **Migration complexity:** Not all MDM solutions are cloud-native. Some solutions rearchitected for cloud may lack functional parity in the near term as the products mature and may require a lift-and-shift migration from on-premises to cloud requiring external support services.
- **Installed base:** Although vendors are motivated to transition existing clients to cloud, clients will do this at their own pace and over time — in the absence of a hard trigger.
- **Governance:** As master data is heavily shared, a need for real-time integration into associated data sources and processes arises. Organizations in transition to cloud must optimize the business processes and more complex governance of a hybrid ecosystem.
- **Complexity of navigating the vendor landscape:** SaaS alleviates some MDM skills challenges. PaaS/IaaS offer greater configuration flexibility. Licensing constructs vary with MDM spend for some counting toward clients' cloud provider committed annual spend.
- **Best practices persist:** Cloud does not alleviate the business challenges related to being successful with MDM.

User Recommendations

- Take a "*think big, start small, deliver incremental value*" approach to MDM by leveraging cloud as the enabler for business value.
- Conduct a thorough review of current governance practices as a precursor to cloud readiness. Governance complexity increases in a hybrid ecosystem.
- Map and actively track the center of data gravity within your organization for each master data domain to identify and plan for prospective transition points for the cloud.
- Review and document integration complexity to provide a manageable integration scenario that does not negate any benefits of cloud-based MDM.
- Evaluate gaps in capability between candidate vendors' cloud-based and on-premises MDM solutions to determine when and whether a migration between the cloud and on-premises environments is viable.
- Cost should not be the driver for adoption of cloud MDM. Without appropriate capacity planning and cost modeling, cloud services may prove more expensive on a total cost of ownership (TCO) basis. Due diligence is required around capacity planning and TCO modeling.

Sample Vendors

Ataccama; CluedIn; Informatica; Profisee; Reltio; Semarchy

Gartner Recommended Reading

3 Essentials for Starting and Supporting Master Data Management

Create a Master Data Roadmap With Gartner's MDM Maturity Model

Data and Analytics Essentials: Master Data Management — Presentation Materials

Data Hub iPaaS

Analysis By: Keith Guttridge

Benefit Rating: High

Market Penetration: 1% to 5% of target audience

Maturity: Adolescent

Definition:

Data hub integration platform as a service (iPaaS) supports integration between applications and system endpoints via a centralized intermediary store that persists the (often-normalized) data before delivery to the destination. This is different from the pass-through architecture that most established iPaaS offerings utilize today and provides additional information management capabilities at the data store level.

Why This Is Important

The data landscape for most organizations is fragmenting across on-premises applications and data stores, and an ever-increasing number of SaaS applications and cloud data stores. This is causing complex integration and governance challenges. Some iPaaS vendors include data stores to create standardized data models that are not tightly coupled to vendor data models and expose APIs to accelerate development of newer services.

Business Impact

Data hub iPaaS provides these benefits:

- Simplification of integrating applications and data sources.
- Improved data management and data governance.
- Improved resilience of the production environment with record/replay capabilities for integration errors and system availability.
- Simplified access to the centralized data model via APIs, instead of connecting directly to application APIs.
- Real time analytics to gain business insight, and potentially enable real-time and contextual decision support.

Drivers

- Organizations looking to improve data management across on-premises and cloud-based applications and data sources.
- Organizations looking to build a customer engagement hub.
- Organizations looking to build a digital integration hub.
- Organizations looking to reduce the number of vendors providing integration and data management technologies.
- iPaaS vendors converging various integration technologies.

Obstacles

- Regional compliance policies for data stores.
- Industry compliance policies for data stores.
- Organizations' compliance policies for data stores.
- Preference for best-of-breed integration and data management technology.
- Organizational structure impeding unified approach to integration and data management.
- Vendor landscape made up mostly of small startups with only a handful of large vendors providing this service.

User Recommendations

- Acknowledge that this is a relatively new market. The few vendors that provide this capability often do so for relatively niche use cases. It may take several years before data hub iPaaS becomes general-purpose enough for most clients. Since the data is stored within the data hub iPaaS, this brings with it extra challenges, such as security, resilience and compliance, that regular iPaaS vendors do not have to worry about.
- Combine offerings from several technology categories and vendors (such as iPaaS plus data store plus analytics), if the current offerings in the data hub iPaaS market are not suitable for your needs. Once established, however, the combination of iPaaS, data management, real-time analytics and machine learning has the potential to significantly disrupt how organizations integrate their application and data portfolios, as well as their B2B partners.

Sample Vendors

Cinchy; Domo; IBM, Informatica; K2view; OVHcloud (ForePaaS); SAP; Sesam

Gartner Recommended Reading

Magic Quadrant for Integration Platform as a Service, Worldwide

Magic Quadrant for Data Integration Tools

Communications PaaS

Analysis By: Daniel O'Connell

Benefit Rating: High

Market Penetration: 20% to 50% of target audience

Maturity: Early mainstream

Definition:

Communications platform as a service (CPaaS) is a cloud-based middleware on which organizations can develop, run and distribute communications software. The platform offers APIs that simplify the integration of communication modules — including SMS, voice, messaging apps, email, social and video — into applications, services and business processes, complemented with development tools and documentation. A CPaaS vendor may assemble multiple CPaaS modules into richer solutions, such as e-commerce.

Why This Is Important

CPaaS is important because it easily enables organizations to integrate communications into workflows via developer-friendly software APIs. Even organizations with modest IT skills have developers that can deploy SMS, voice and two-factor authentication (2FA) for basic workflows like notifications and appointment reminders. Digital natives and large enterprises have robust developer teams that can build more complex workflows with features such as email, video, payments, web chat and WhatsApp.

Business Impact

CPaaS plays a prominent role in enterprise IT with the influx of developers joining the IT workforce. A developer ecosystem of APIs, software development kits (SDKs) and documentation provides a low-cost toolset for improving operational efficiency and customer experience. CPaaS vendors now offer visual builders so noncoding business analysts can build simple workflows. Most organizations start by deploying CPaaS for a single business unit (BU) use case, from which it is quickly adopted across other BUs.

Drivers

- CPaaS is highly correlated with the 2023 API economy. Many organizations now have a bigger developer workforce compared to 2018. Megavendors like Amazon, Cisco and Microsoft now have a CPaaS play. The companies' entrance certifies the importance of CPaaS, placing CPaaS on the radar screen of IT leadership.
- CPaaS vendors continue to build out their platforms with an expanded set of modules such as video, WhatsApp, security, authentication, email and payments. This, in turn, enables organizations to build more complex workflows, yielding higher CPaaS revenue for vendors, such as e-commerce, telehealth and insurance claims processing.
- A few CPaaS vendors are building out advanced capabilities in bots, AI, customer data platforms (CDPs) and campaign management. Many of these implementations focus on customer experience.
- CPaaS vendors are building systems integrator (SI) partnerships focused on complex vertical use cases. This provides a scaling opportunity as the SIs have strong CIO relationships for building advanced IT workflows.
- Visual builders continue to be rolled out in the market to allow the participation of noncoding business analysts. This expands the total available market (TAM) to users building simple workflows or making modifications to existing workflows.

Obstacles

- The CPaaS market struggles with brand awareness. Many IT decision makers are not sure which CPaaS providers are best-suited to align with.
- Developer talent constrains CPaaS growth. While organizations add developers to their workforce, their schedules may be booked for other projects.
- The CPaaS landscape is complicated as new vendors enter the market and with others repositioning their product offerings. In addition, CPaaS vendors are expanding their capabilities into CDP, contact center as a service (CCaaS) and campaign management.
- 2022 through 2023 economic uncertainty has forced vendors to focus on profitability rather than growth. This has led to industry layoffs, reduced risk taking and a focus on core competencies further hindering CPaaS adoption.
- CPaaS market adoption is strong with mature offerings like SMS, 2FA, and number anonymization. But they are commodities and have poor margins. CPaaS vendors need greater adoption into the newer capabilities — WhatsApp, video, and conversations — in order to restore their financial health.

User Recommendations

- Proceed first with simple solutions centered on SMS, application-to-person (A2P), 2FA, phone number anonymization and voice if you have modest IT skills.
- Explore the advanced communications modalities such as the messaging apps (e.g., WhatsApp), video, email, payments and e-commerce if you have stronger IT skills. CPaaS is now viable for organizations of all shapes and sizes.
- Adopt CPaaS across the entire business. CPaaS often starts in a single BU, before expanding to others — such as HR, operations and supply chain — to achieve maximum benefits.
- Expand the organization's developer workforce to fully leverage CPaaS for competitive edge. IT core competency is not a luxury, but a necessity for survival.
- Hire SIs or boutique CPaaS development firms for initial projects. Have your IT team learn from the third-party CPaaS firm so you can build your own core competency.

Sample Vendors

Bandwidth; CM.com; Infobip; MessageBird; Route Mobile; Sinch; Twilio; Vonage

Gartner Recommended Reading

Market Guide for Communications Platform as a Service

Quick Answer: Why You Need to Deliver on WhatsApp Business Platform for Rich Business Conversations

How to Evaluate and Select CPaaS Providers to Operationalize Customer Experience

Emerging Tech: Turnkey Yet Customizable Solutions Are Transforming Communication

IoT PaaS

Analysis By: Alfonso Velosa, Eric Goodness

Benefit Rating: High

Market Penetration: 5% to 20% of target audience

Maturity: Adolescent

Definition:

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

Why This Is Important

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

Business Impact

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

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

Drivers

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

Obstacles

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

User Recommendations

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

Sample Vendors

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

Gartner Recommended Reading

Magic Quadrant for Global Industrial IoT Platforms

Technology Opportunity Prism: Internet of Things

Competitive Landscape: IoT Service Providers

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

Connected IoT Brokers for Autocomposing: An Insurance Trend for 2030

API Marketplaces

Analysis By: Andrew Humphreys

Benefit Rating: Moderate

Market Penetration: 5% to 20% of target audience

Maturity: Adolescent

Definition:

An API marketplace is a platform to share APIs. Consumers, mainly developers, use API marketplaces to discover APIs and, in some cases, may purchase access to them. They can be either public commercial marketplaces with APIs from multiple providers, public with APIs from a single provider, or private marketplaces for promoting an organization's internal APIs.

Why This Is Important

API marketplaces enable organizations to publicize their APIs. Marketplaces are usually associated with external marketplaces, which share APIs with a community of developers and enable partners to implement solutions using the APIs. However, as most APIs are meant for consumption by teams within an organization, marketplaces are more frequently internal. They make it easier to find APIs internally, helping with wider sharing of capabilities between different business units and product and development teams.

Business Impact

API marketplaces increase developer visibility and consumer mind share, drive API usage, and, by extension, increase business impact. API consumers can use marketplaces to simplify finding and comparing different APIs when they are looking for specific functionality but have not selected exactly which API to use. There is typically a cost involved with listing in a public API marketplace, but the benefits include exposure to a larger number of API consumers and access to features to enable monetisation.

Drivers

- The number of APIs within an organization is climbing, driving the need for developers to more easily discover which APIs and services are available.
- Composable business, including composable commerce, relies on the use of API marketplaces to share APIs and packaged business capabilities.
- Increased use of low-code platforms, integration platforms, robotic process automation (RPA) and analytics tooling enables more citizen development using APIs that may be sourced from API marketplaces.

Obstacles

- Public API marketplaces that provide a public directory of APIs from multiple providers have had disappointing results, as developers are more likely to go directly to API providers to sign up for APIs. This has resulted in API marketplaces in the Trough of Disillusionment. However, internal API marketplaces have had more success, since they enable developers to share APIs across multiple teams.
- API portals provided as part of API management platforms are typically basic in nature, resulting in significant customization work to create a customer-oriented API marketplace based on such an API portal.
- New open-source platforms, such as Backstage from Spotify, are driving the creation of internal API catalogs as part of larger developer hubs. If your developers are collaborating on solutions around their APIs already, then a simple catalog may be sufficient and a full marketplace is probably overkill.

User Recommendations

API providers:

- Create an internal API marketplace, focused on the needs of software engineers to share APIs across the organization.
- Examine billing terms to understand what the cost of using the marketplace is when considering commercial API marketplaces.
- When considering a commercial API marketplace, examine listing fees and value to your organization before committing.

API consumers:

- Ensure that you use APIs from trusted marketplaces and trusted API providers, examining usage agreements, licensing and billing terms carefully.
- Investigate whether consuming an API directly from the API provider offers better pricing or usage terms than consuming the API via a marketplace.

Sample Vendors

Achieve Internet; Bump; Postman; Pronovix; Readme; Smartbear (Swagger); Spotify (Backstage); Stoplight

Gartner Recommended Reading

Innovation Insight for Internal Developer Portals

Reference Model for API Management Solutions

Climbing the Slope

Cloud AI Services

Analysis By: Van Baker, Bern Elliot

Benefit Rating: High

Market Penetration: 20% to 50% of target audience

Maturity: Early mainstream

Definition:

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

Why This Is Important

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

Business Impact

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

Drivers

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

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

Obstacles

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

User Recommendations

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

Sample Vendors

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

Gartner Recommended Reading

Critical Capabilities for Cloud AI Developer Services

Magic Quadrant for Cloud AI Developer Services

Hybrid Cloud Computing

Analysis By: David Smith, Milind Govekar

Benefit Rating: High

Market Penetration: 20% to 50% of target audience

Maturity: Early mainstream

Definition:

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

Why This Is Important

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

Business Impact

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

Drivers

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

Obstacles

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

User Recommendations

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

Gartner Recommended Reading

Top Strategic Technology Trends for 2021: Distributed Cloud

‘Distributed Cloud’ Fixes What ‘Hybrid Cloud’ Breaks

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

Leverage Platform Engineering to Scale DevOps Platforms Into Hybrid Cloud

RPA PaaS

Analysis By: Arthur Villa, Saikat Ray, Fabrizio Biscotti, Cathy Tornbohm

Benefit Rating: Moderate

Market Penetration: 20% to 50% of target audience

Maturity: Early mainstream

Definition:

Robotic process automation (RPA) platform as a service (PaaS) is a set of public-cloud-hosted services to create and execute RPA scripts against customer applications. It consists of RPA development environments, orchestrators and a performance dashboard. The orchestrator interacts with lightweight agents installed in the client environment that execute bots. Unlike RPA SaaS options that offer a ready-to-use service, RPA PaaS is intended for developers to create a new business function/service.

Why This Is Important

Currently, the majority of RPA implementations have been delivered on-premises, but RPA PaaS adoption will increase significantly over the next few years. This trend will enable greater citizen development, scalability and AI integration.

Business Impact

Like many software markets that begin with an on-premises focus and shift to the cloud, RPA PaaS will reduce barriers to RPA adoption, improve scalability and enable tighter integration with cloud-native services. RPA PaaS allows organizations to realize the value of on-premises RPA with improved process automation and the reduction of manual, repetitive tasks. RPA PaaS further distributes development access, improves governance and enables customers to access new features and capabilities.

Drivers

- One of the main factors driving the adoption of RPA PaaS is demand from business buyers. Through RPA PaaS, business buyers can quickly realize value from RPA without the need to ask busy IT departments to provision servers for RPA orchestration and analytics features. Demand for RPA is often driven by business buyers and RPA PaaS eliminates a major barrier to business buyer adoption.
- For several large software providers that entered the RPA market in recent years, cloud orchestration is the only option offered, driving RPA PaaS adoption. The majority of legacy RPA providers now also offer cloud-delivered access to their development environments and orchestration tools. These legacy providers have created pricing and sales strategies encouraging customers to migrate to cloud delivery, further driving RPA PaaS adoption.
- Organizations interested in the benefits of hyperautomation often require integrating RPA with process automation tech, like intelligent document processing (IDP), business process automation (BPA), integration platform as a service (iPaaS) and low-code application platforms (LCAPs).

- RPA PaaS improves integration with complementary process automation technology by centralizing the development platform in the cloud instead of requiring the installation of add-ons and version control challenges of managing multiple development environments distributed across an organization's developer workstations.
- Another factor driving adoption is decreasing buyer concerns about the security risks created by RPA PaaS. Early RPA buyers had concerns about new security risk vectors since RPA tools often had privileged access to an organization's core applications. As a result of those concerns, the RPA vendors mitigate those risks by developing strong security features including data encryption, credential vault integration, robust role-based access control (RBAC) capabilities, and bot accountability through log integrity and identity-centric approaches.

Obstacles

- Regulated industries may be unable to adopt RPA PaaS due to regulatory and compliance challenges. For example, banking and insurance industries are the greatest adopters of RPA, but many customers may be unable to use RPA PaaS due to data privacy regulations like GDPR.
- Buyers frequently express concerns about RPA tool access to personal information and the transfer of confidential data outside of an organization's firewalls. Years of RPA adoption with limited security events have mitigated buyers' security concerns; however, a highly publicized breach enabled by RPA would affect buyers' willingness to place orchestration and development in the cloud.
- Customers that encounter connectivity issues (e.g., due to latency or incorrect network configurations) between on-premises bots and cloud orchestration have shared this negative experience with other potential buyers, dissuading them from adopting RPA PaaS capabilities.

User Recommendations

- Examine cloud in the form of RPA PaaS to overcome the hurdle of how to implement, maintain and update on-premises orchestrators to manage the bots. Because RPA demand will continue to grow as organizations seek to rapidly digitize and automate legacy processes. RPA PaaS can reduce the total cost of ownership (TCO) of RPA by reducing the maintenance burden of upgrades, patches and so on.
- Pilot a simple use case with a clearly defined ROI achievable within four to eight weeks. Focus on key business outcomes, such as productivity gain, higher accuracy and client satisfaction. When evaluating an RPA PaaS offering, be prepared to define and operationalize your cloud governance model jointly with business and IT to identify applicable processes, acknowledge technical debt and evaluate the performance of the application portfolio.

Sample Vendors

Automation Anywhere; Microsoft; NICE; Pegasystems; SS&C Technologies (Blue Prism); UiPath; WorkFusion

Gartner Recommended Reading

Beyond RPA: Build Your Hyperautomation Technology Portfolio

Magic Quadrant for Robotic Process Automation

Quick Answer: How to Choose the Right Use Cases for Robotic Process Automation

Secure Robotic Process Automation Initiatives With These 4 Essentials

Cloud Analytics

Analysis By: Julian Sun, Fay Fei, Jamie O'Brien

Benefit Rating: High

Market Penetration: More than 50% of target audience

Maturity: Mature mainstream

Definition:

Cloud analytics delivers analytics capabilities as a service. It often comprises database, data integration and analytics tools. As cloud deployments continue, the ability to connect to both cloud-based and on-premises data sources in a hybrid model is increasingly important. Cloud-native architecture and multicloud deployments are also becoming popular in order to cater to the cloud ecosystem.

Why This Is Important

Adoption of cloud analytics is growing, with most analytics deployments originating in the cloud. The majority of respondents to the 2022 Gartner State of Data and Analytics Cloud Adoption Survey say they are using or plan to use the cloud for analytics and data science. Cloud capability among analytics and BI vendors is also expanding, with emerging capabilities coming from cloud-first. The cloud is an ideal place to build modular analytics capabilities that enable greater agility and reuse of existing investments in support of composable business.

Business Impact

A cloud-enabled, composable platform can innovate by assembling modular analytics capabilities on demand. More advanced analytics can complement key components of the analytics infrastructure in the cloud. The high computational power needed to process tasks such as ML and advanced analytics can be more easily accessed and scaled in the cloud. Business users can pilot cloud-first augmented analytics within a sandbox provisioned by the cloud. Cloud deployment offers faster time to value and more targeted analytics for specific business areas.

Drivers

- To better leverage scalability and elasticity from the cloud, many platforms have rearchitected themselves to be cloud-native.
- To bring more flexibility for organizations that are already using multicloud, vendors are adding more deployment options and management capabilities. These additions enable portability through microservices architectures that are readily supported via containerization across multiple clouds.
- Startups continue to join the analytics market with cloud-first or cloud-only solutions, which are complementary to established platforms.
- The range of capabilities is growing too. Reporting and data visualization were already commodified capabilities. Customers can now also subscribe to self-service data preparation; augmented data discovery; predictive modeling; other advanced capabilities, such as ML or streaming analytics; and even data/context broker services from several vendors.
- The growing cloud DBMS market naturally supports and expands the cloud analytics market as companies embrace the cloud for managing their data.

Obstacles

- Security is a top concern for organizations moving to the cloud. Organizations need to plan how they will integrate their growing cloud analytics deployments with additional data sources, provide access to more advanced (potentially open-source) analytics tools, and embed analytics in business processes. Such planning becomes even more challenging across multiple cloud and on-premises ecosystems.
- Organizations' adoption of the cloud is closely tied to data gravity. Data gravity refers to data's attractive force: As data accumulates and the need for customization, integration and access grows, data has greater propensity to "pull" data services, applications and other data/metadata to where it resides. Thus, smaller organizations with data originating in the cloud have higher adoption rates than larger organizations with data predominantly in on-premises legacy solutions.
- Even as cloud analytics becomes more predominant and mature, organizations with deployment and governance challenges face growth obstacles.

User Recommendations

- Establish a measured approach to move to the cloud incrementally — rather than simply "lifting and shifting" — as cloud analytics becomes a dominant option in most scenarios in the analytics space.
- Include innovative cloud analytics solutions in your portfolio, renovating on-premises components or complementing your on-premises platform, to gain competitive advantage through analytics and BI. Completely disregarding cloud analytics solutions means risk for many organizations, as most vendors don't focus their R&D efforts on legacy products.
- Be aware of extra costs and the total cost of ownership (TCO) as you adopt new capabilities and offerings within your vendor's cloud stack. Although cloud analytics solutions do not require significant upfront investment like on-premises solutions do, the former will likely be more expensive to license over four or more years. Also be aware of the performance downgrade in the cloud — benchmark the platform, and carefully plan the data integration approach.

Sample Vendors

Alibaba Cloud; Amazon Web Services; Databricks; Domo; Google; Microsoft; Oracle; Qlik; Sigma Computing; ThoughtSpot

Gartner Recommended Reading

Adopt Cloud Analytics to Drive Innovation

Use Cloud to Compose Analytics, BI and Data Science Capabilities for Reusability and Resilience

Magic Quadrant for Analytics and Business Intelligence Platforms

Critical Capabilities for Analytics and Business Intelligence Platforms

CIPS

Analysis By: Sid Nag

Benefit Rating: High

Market Penetration: More than 50% of target audience

Maturity: Mature mainstream

Definition:

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

Why This Is Important

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

Business Impact

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

Drivers

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

Obstacles

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

User Recommendations

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

Sample Vendors

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

Gartner Recommended Reading

Magic Quadrant for Cloud Infrastructure and Platform Services

Critical Capabilities for Cloud Infrastructure and Platform Services

What Buyers Want From CIPS Providers

Risk and Opportunity Index: Cloud Infrastructure and Platform Services

Extending the CIPS Business to New Markets and Opportunities

Cloud Event Stream Processing

Analysis By: W. Roy Schulte, Pieter den Hamer

Benefit Rating: Moderate

Market Penetration: 20% to 50% of target audience

Maturity: Early mainstream

Definition:

Cloud event stream processing (ESP) services are platform as a service (PaaS) offerings that perform computation on streaming event data in near real time. An event stream is a sequence of event objects arranged in some order (typically by time). Cloud ESP services enable situation awareness and low-latency responses to threats and opportunities. They can also be used to transform and load data streams into files or databases for use in subsequent applications.

Why This Is Important

ESP's data-in-motion architecture is a radical alternative to conventional data-at-rest architecture. ESP derives information from streaming data in near real time as it arrives. As data and applications migrate to the cloud, organizations are using more cloud ESP services. Providers developed cloud ESP services quickly by leveraging years of research and development for on-premises ESP platforms. Cloud ESP services mirror the position of on-premises ESP on the Hype Cycle.

Business Impact

Compared with on-premises ESP, cloud ESP services sometimes reduce the cost of stream analytics and stream data ingestion, particularly if workloads vary greatly over time. Serverless deployment styles, using on-demand processing models, ease management burdens. Cloud ESP brings the generic benefits of cloud: elastic scalability, self-provisioning, continuous patching and versioning, and offloading most of the work required for high availability and disaster recovery.

Drivers

- Organizations have access to ever-increasing amounts of low-cost streaming data from sensors, machines, smartphones, websites, transactional applications, social computing platforms, news and weather feeds, and business partners. Many new AI and other analytical applications need this streaming data to satisfy business requirements for real-time intelligence. When the data comes from distributed sources, including mobile devices, cloud-based ESP is often appropriate.
- The wide use of Apache Kafka and similar streaming messaging systems is reducing the cost of ingesting, storing and using streaming data.
- Cloud ESP services have become widely available. All of the hyperscale cloud providers and numerous smaller companies offer ESP services as part of their PaaS portfolios. Multiple vendors have bundled cloud ESP services into their IoT platforms.
- ESP technology is an essential part of unified real-time platforms that process both data in motion and data at rest. Some unified real-time PaaS offerings embed open-source ESP platform products, while others get their ESP capabilities from custom internal code.
- Cloud ESP services put ESP within reach of businesses with limited technical staff and budget, and replace capital expenditure with operating expenditure.
- The rapid increase in cloud data warehouses is driving demand for cloud ESP services that facilitate rapid ingestion and loading of data from streaming sources and change data capture (CDC) sources. Conventional data engineering pipelines that take hours or days to prepare new, real-time data for data warehouses are increasingly unacceptable. Therefore, a growing number of data engineering pipelines are being reimplemented as real-time data flows in cloud ESP services. These cloud ESP services filter, aggregate, enrich, and perform pattern detection and other transformations on streaming data as it arrives.

Obstacles

- ESP technology is overkill for most applications that process low or moderate volumes of streaming data, or that do not require fast response times. Conventional data-at-rest architectures that store the data in databases before acting on it suffice when volume and latency requirements are less demanding.
- Cloud ESP services will not replace on-premises ESP platforms for applications that require very low-latency computation with on-premises (including edge) data. Edge deployments in factories, field locations and moving vehicles need on-premises ESP to minimize latency and to provide better availability where network connections are slow or unreliable.
- Rearchitecting existing business processes to take advantage of cloud ESP requires significant redesign and investment.
- Until recently, ESP required low-level programming in Java, Scala or proprietary languages. A new generation of SQL and low-code development tools is improving developer productivity for some cloud ESP services.

User Recommendations

- Implement ESP through cloud ESP services or on-premises ESP platforms when conventional data-at-rest architectures cannot process high-volume event streams fast enough to meet business requirements.
- Use a SaaS offering with embedded ESP logic if a product that targets your specific business requirements is available. Verify that the embedded logic can be customized to your specific use cases and SLAs.
- Develop your own ESP-enabled applications on cloud ESP services when appropriate SaaS offerings are not available.
- Choose cloud ESP services over on-premises ESP platforms when most or all of the data sources or targets are in the cloud. Also use cloud ESP services if networks are reliable, workloads are bursty, and ultralow latency is not a requirement.
- Choose a unified real-time platform with embedded ESP capabilities over cloud ESP services if the application needs to use both data at rest and data in motion.

Sample Vendors

Aiven; Amazon Web Services; Axual; Confluent; Decodable; Google; IBM; Microsoft; SAS; TIBCO Software

Gartner Recommended Reading

Market Guide for Event Stream Processing

5 Essential Practices for Real-Time Analytics

Market Share Analysis: Event Stream Processing Platforms, Worldwide, 2021

Entering the Plateau

API Management PaaS

Analysis By: Mark O'Neill

Benefit Rating: Moderate

Market Penetration: 20% to 50% of target audience

Maturity: Mature mainstream

Definition:

API management PaaS (APIM PaaS) takes an on-demand approach to the delivery of API management by providing an alternative to the installation and management of API management software. APIM PaaS manages API access via provider-hosted API gateway services, and it may also include an API developer portal. In some cases, an on-premises API gateway option will be provided. APIM PaaS may be optimized to be used with PaaS services from the same vendor, such as function PaaS (fPaaS).

Why This Is Important

APIM PaaS takes full advantage of cloud benefits, such as autoscaling, resiliency and robust security. It also allows some vendors to offer per-API-call pricing. APIM PaaS may include the ability to deploy on-premises API gateways, to enable hybrid API management architecture with APIs on-premises and to offer cloud-based API management.

Business Impact

APIM PaaS allows costs to scale with the business value of APIs, reducing the impact of a large outlay as an API program grows. It enables APIs to be managed effectively when API traffic is unpredictable and potentially very large. APIM PaaS also brings business benefits when an APIM PaaS offering is provided as part of the PaaS platforms already in use by an organization, through unified procurement and billing.

Drivers

- APIM PaaS is driven by migration to and adoption of cloud platforms.
- An increasing number of available APIs and a growing volume of API interactions drive organizations to APIM PaaS for its high-scale quality of service, including throughput, high availability, and global access.

- SaaS adoption is also a driver, as organizations wish to use API management without needing to operate and maintain an API management software.
- Serverless computing, including fPaaS (also known as function as a service or FaaS), can act as a major driver for APIM PaaS. This is because fPaaS offerings can make use of API management on their associated cloud platforms. In some cases, they can automatically populate API gateways with endpoints so that fPaaS functions can be called via REST APIs.
- Since many organizations build APIs in the cloud, APIM PaaS is also increasingly used in hybrid and multicloud scenarios.
- Automation is also a driver for APIM PaaS. This is because APIM PaaS itself includes documented APIs in the API management platform. These APIs are used for tasks such as registering APIs. APIM PaaS typically can automatically register APIs built on the same platform.

Obstacles

- APIM PaaS tends to focus on runtime (API gateway) capabilities, with limited support for an API developer portal or other aspects of API management beyond API gateways.
- Network latency concerns impact the uptake of APIM PaaS for managing on-premises APIs. Even in a hybrid scenario, any requirement for the remote gateway to connect to the core platform introduces latency concerns.
- Data residency concerns, such as a storage of API payloads that may contain private information, are also an obstacle to the uptake of APIM PaaS for managing on-premises APIs.
- APIM PaaS can result in higher-than-expected costs as API traffic grows.
- APIM PaaS solutions from cloud hyperscalers are generally tied to their larger PaaS platforms, and are not portable for their use on other PaaS platforms.

User Recommendations

- Investigate the use of APIM PaaS to provide a cost-effective means of providing API management. If some or all of your APIs are on-premises, then investigate a hybrid option.
- For organizations migrating to the cloud, investigate hybrid APIM PaaS options.

- Compare the pricing of APIM PaaS vendors, since not all provide consumption-based pricing.
- Include API PaaS as part of your API strategy, since it can accelerate the time to market of your mission-critical digital initiatives.

Sample Vendors

Alibaba Cloud; Amazon Web Services; Google (Apigee); Huawei, IBM; Microsoft; Oracle; VMware

Gartner Recommended Reading

Magic Quadrant for Full Life Cycle API Management

Reference Model for API Management Solutions

Infographic: Decision Point for API and Service Implementation Architecture

iPaaS

Analysis By: Keith Guttridge

Benefit Rating: High

Market Penetration: More than 50% of target audience

Maturity: Mature mainstream

Definition:

Gartner defines integration platform as a service (iPaaS) as a vendor-managed cloud service that enables end users to implement integrations between a variety of applications, services and data sources. iPaaS platforms support at least one of the three main uses of integration technology — data consistency, multistep processes and composite services. iPaaS provides adapters to simplify configuration connectivity. It provides a low-code workflow environment and supports hybrid deployment models.

Why This Is Important

The shift to the cloud along with digital and automation initiatives is boosting the iPaaS market, making it the biggest segment of the integration platform technology market (valued at $6.5B in 2022). Its functional breadth makes it the natural alternative to classic integration software such as enterprise service bus (ESB), and extract, transform and load (ETL) software for large organizations.

Business Impact

By addressing integration needs in a rapid and cost-effective way, iPaaS enables organizations to improve efficiency, increase business agility and introduce innovation faster. iPaaS adoption helps organizations achieve these goals cost-effectively, efficiently and with less specialized personas than are needed for classic integration software. Also, iPaaS makes these benefits accessible to midsize and small organizations that cannot afford the costs of a classic platform.

Drivers

- iPaaS is frequently adopted by organizations that seek to modernize their integration capabilities. This way, they can support digital transformation and cloud adoption (especially SaaS), and enable new roles and personas to be part of the integration delivery.
- There has been relative success for iPaaS targeting particular industries, SaaS ecosystems, business processes or geographies. This is due to its appeal to time-, skill- and resource-constrained organizations.
- The main goal of iPaaS providers now is to maximize opportunities to upsell and cross-sell to their vast installed base. Therefore, they are evolving their offerings into enterprise-class suites that address a wide range of hybrid, multicloud scenarios. Hence, large and global organizations now position iPaaS as a strategic option to complement — but increasingly also to replace — classic integration platform software. This drives deeper commitment among enterprises.
- A growing number of SaaS providers “embed” in their applications their own iPaaS, or one from a third party, which they typically extend with a rich portfolio of packaged integration processes (PIPs). This makes embedded iPaaS offerings attractive to organizations that need to quickly integrate a SaaS application.
- Providers will keep investing to improve developers’ productivity, reduce time to value and shorten the learning curve. The goal is to further expand their potential audience to include business users. Hence, providers’ R&D efforts might focus on using AI, machine learning and natural language processing. These capabilities can assist development and operation, enrich PIP portfolios, and enable continuous integration or delivery (CI/CD) and DevOps to entice professional developers.

Obstacles

- The market's extreme fragmentation (more than 175 providers) and diversity of capabilities makes it hard for user organizations to select the best-fit iPaaS for their needs. This could generate a proliferation of diverse, stand-alone and embedded iPaaS offerings. It could also risk fragmenting service providers' investments in skills building.
- The top five iPaaS providers command about 56% of the market, and only 10 providers have more than a 2% share. The vast majority of providers are smaller vendors with limited brand awareness. This may discourage risk-averse organizations from making strategic investments in smaller vendors or less familiar brands.
- Key obstacles are the API rhetoric of seamless "plug-and-play" integration, and confusion among less technically savvy users about the differences between iPaaS and API management platforms. There is also a growing trend for code-based integration encouraged by the cost of integration platforms and serverless PaaS functions.

User Recommendations

Adopt iPaaS when looking for:

- An integration platform for midsize organizations moving to the cloud and for "greenfield" integration initiatives.
- A strategic complement to traditional integration platforms — increasingly in the context of integration strategies based around hybrid integration capability framework— in order to empower a collaborative, democratized approach to integration.
- An enabler of self-service integration for business technologists, such as SaaS administrators or citizen integrators.
- A platform to support well-defined, tactical integration projects with low budgets, severe time constraints, and informally defined and incrementally formulated requirements.
- A potential replacement for classic integration platforms that are obsolete or cannot support changing requirements.

Sample Vendors

Boomi; Jitterbit; Microsoft; Oracle; Salesforce (MuleSoft); SAP; SnapLogic; TIBCO Software; Tray.io; Workato

Gartner Recommended Reading

Magic Quadrant for Integration Platform as a Service

Critical Capabilities for Integration Platform as a Service

Choose the Best Integration Tool for Your Needs Based on the Three Basic Patterns of Integration

How to Select the Right Mix of Integration Technologies

LCAP

Analysis By: Paul Vincent, Yefim Natis, Oleksandr Matvitskyy

Benefit Rating: High

Market Penetration: More than 50% of target audience

Maturity: Mature mainstream

Definition:

A low-code application platform (LCAP) supports visual programming abstractions, such as model-driven and metadata-based application development. Developers use LCAP to create user interfaces, design data schemas, and implement business logic with simplified tooling and their catalog of packaged capabilities.

Why This Is Important

LCAPs are one of the most popular types of low-code tools. They are used as a general purpose platform for web and mobile application development with high productivity while requiring fewer specialized developer skills. They can facilitate access to new, rich or advanced technology sets for developer personas ranging from enterprise software engineering experts to citizen developers. Over 300 vendors support a wide variety of business use cases and industry specializations.

Business Impact

Accelerating application delivery while reducing developer effort and increasing self-service capabilities are high priorities for business IT. Businesses adopt LCAP to deliver more automation and reduce their application backlogs as well as enable democratized application development. Mostly cloud-based, LCAP vendors are investing in the development of new capabilities to increase their use-case coverage and justify their platform subscription costs.

Drivers

- The demand to automate and digitize business through applications continues to outstrip conventional software development capacity. This is despite the rise of SaaS usage for standard business services — indeed, the latter has resulted in higher demands for application development for differentiating SaaS extensions. This means a large part of the LCAP market is for SaaS vendors' LCAP.
- LCAPs have evolved from rapid application development, business process technologies, mobile application development and SaaS platforms in order to increase productivity for common application types. They have evolved common capabilities around user interface, database, business logic definition, process orchestration and integration of ubiquitous REST API services.
- LCAPs continue to evolve to meet customer needs. Their low-code paradigms have been extended to include ever more functionality such as integration, artificial intelligence/machine learning (AI/ML) services and multiexperience support. LCAPs increasingly overlap with the business process automation market for workflow use cases, and the multiexperience development platform (MXDP) market for user-interface-driven use cases.
- Vendors are focusing more on cloud-native scalability to support larger B2C deployments, and deeper governance and collaboration tooling to support citizen/business and IT development fusion team structures. Support for composing applications from multiple API and service types enables LCAPs to cover an increasingly large set of enterprise application requirements, with some enterprises starting to choose them as a strategic application platform.
- Professional developers have started to realize the opportunities of low-code development, creating new markets for LCAP beyond business developers and citizen developers. This is driving the evolution of new low-code development accelerators for pro-code developers, further widening the LCAP market.

Obstacles

- Current LCAP market share is heavily biased toward some very large SaaS providers and a few successful independent vendors. However, enterprises often have multiple LCAP offerings to support different developers and use cases, often at increased total cost of ownership.
- LCAPs that have been implemented by the main SaaS platform vendors could diminish the opportunities for many small LCAP vendors. Vendors have been known to deprecate features or change low-code offerings, requiring additional maintenance by developers.
- LCAP trades productivity for vendor lock-in (of both applications and developer skills). Lock-in reduces customer flexibility and can increase costs.
- The wider developer audiences encouraged by LCAP increase the requirements for governance and guardrails, which are at best still only partially delivered by vendors.
- Licensing models vary across vendors and often change, and may not scale for new use cases. This can lead to vendor disillusionment.

User Recommendations

- Formulate LCAP usage strategies to address vendor selection against different use cases, developer persona, licensing costs, and technology requirements. Most organizations will support several LCAPs in their portfolios across SaaS, multiexperience and business process automation platforms.
- Review applications for the long-term effects of vendor lock-in, the lack of portability or standards, and ease of technical debt accumulation. Vendor relationships (and contracts) need to be considered strategic and long-term.
- Analyze LCAP subscriptions against their productivity and time-to-market benefits: minimize end-user pricing if desiring B2C rollouts.
- Ensure developers are governed according to their needs. Different developers with different skill sets will vary in their successful adoption of different LCAPs.
- The large number of LCAP vendors implies possible future market instability, although to date, there have been only a few LCAP retirements.

Sample Vendors

Alibaba Cloud; Huawei Technologies; Kintone; Mendix; Microsoft; Newgen Software; OutSystems; Retool; ServiceNow; Unqork

Gartner Recommended Reading

Magic Quadrant for Enterprise Low-Code Application Platforms

Critical Capabilities for Enterprise Low-Code Application Platforms

Identify and Evaluate Your Next Low-Code Development Technologies

How to Navigate the Application Platforms Market Including Cloud-Native, Low-Code and SaaS

Drive High-Value LCAP Outcomes Through Strategic Partnerships With Service Providers

Appendixes

See the previous Hype Cycle: Hype Cycle for Cloud Platform Services, 2022

Hype Cycle Phases, Benefit Ratings and Maturity Levels

Table 2: Hype Cycle Phases

(Enlarged table in Appendix)

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

Source: Gartner (July 2023)

Table 3: Benefit Ratings

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

Source: Gartner (July 2023)

Table 4: Maturity Levels

(Enlarged table in Appendix)

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

Source: Gartner (July 2023)

Acronym Key and Glossary Terms

AD	application development
AI	artificial intelligence
AIM	application integration and middleware
aPaaS	application platform as a service
BPM	business process management
BPMS	business process management suite
CAGR	compound annual growth rate
ESB	enterprise service bus
FLAPIM	full life cycle API management
HIP	hybrid integration platform
hpaPaaS	high-productivity aPaaS
ICC	integration competency center
IMC	in-memory computing
IMDG	in-memory data grid
IoT	Internet of Things
iPaaS	integration platform as a service
LCAP	low-code application platform
LOB	line of business
ML	machine learning
MASA	mesh app and service architecture
MFT	managed file transfer
OSS	open-source software
PaaS	platform as a service
RPA	robotic process automation
SMB	small or midsize business

TPM	transaction processing monitor

Evidence

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

[2] Top Strategic Technology Trends for 2023: Industry Cloud Platforms

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

[4] Forecast: Public Cloud Services, Worldwide, 2021-2027, 2Q23 Update

Document Revision History

Hype Cycle for Cloud Platform Services, 2022 - 19 July 2022

Hype Cycle for Platform as a Service, 2021 - 10 August 2021

Hype Cycle for Platform as a Service, 2020 - 12 August 2020

Hype Cycle for Platform as a Service, 2019 - 29 July 2019

Hype Cycle for Platform as a Service, 2018 - 9 August 2018

Hype Cycle for Platform as a Service, 2017 - 19 July 2017

Hype Cycle for Platform as a Service, 2016 - 12 July 2016

Hype Cycle for Platform as a Service, 2015 - 23 July 2015

Recommended by the Authors

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

Understanding Gartner's Hype Cycles

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

Hype Cycle for Cloud Computing, 2023

Hype Cycle for Application Architecture and Integration, 2023

Hype Cycle for Infrastructure Strategy, 2023

Top 4 Trends at the Peak of the Cloud Platform Services Hype Cycle

Table 1: Priority Matrix for Cloud Platform Services, 2023

Benefit ↓	*Years to Mainstream Adoption*			
	Less Than 2 Years ↓	***2 - 5 Years*** ↓	***5 - 10 Years*** ↓	***More Than 10 Years*** ↓
Transformational		Blockchain Oracles Generative AI Platform Engineering	Industry Cloud Platforms	
High	Cloud Analytics Cloud MDM iPaaS LCAP	AI Model as a Service API-Centric SaaS CIPS Cloud AI Services Cloud Event Brokers Communications PaaS Container Management Data Hub iPaaS Digital Integration Hub Enhanced Blockchain as a Service Hybrid Cloud Computing IoT PaaS Multicloud Prompt Engineering	Cloud-Native Cloud Portability Data Ecosystems Digital Foundation Distributed Cloud Vector Databases	

Benefit ↓	*Years to Mainstream Adoption*			
	Less Than 2 Years ↓	***2 - 5 Years*** ↓	***5 - 10 Years*** ↓	***More Than 10 Years*** ↓
Moderate	API Management PaaS Cloud Event Stream Processing	API Marketplaces Edge PaaS RPA PaaS SaaS as a Platform Sovereign Cloud	Application Composition Technology Decision Management PaaS	Quantum Computing as a Service
Low				

Source: Gartner (July 2023)

Table 2: Hype Cycle Phases

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

Phase ↓	Definition ↓

Source: Gartner (July 2023)

Table 3: Benefit Ratings

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

Source: Gartner (July 2023)

Table 4: Maturity Levels

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

Source: Gartner (July 2023)