Gartner

Research

Publication Date: 27 July 2011 **ID Number: G00214915**

Hype Cycle for Cloud Computing, 2011

David Mitchell Smith

While clearly maturing, cloud computing continues to be the most hyped subject in IT today. We look at the different aspects of the topic and where they are on Gartner's Hype Cycle for Cloud Computing, 2011.

TABLE OF CONTENTS

Gartner

LIST OF TABLES

LIST OF FIGURES

What You Need to Know

Cloud computing is maturing, but it continues to be the latest, most hyped concept in IT. As we stated in "Hype Cycle for Cloud Computing, 2010," it is simplistic to look only at the hype around the high-level term. As aspects of the cloud move into mainstream adoption, each needs to be looked at separately. There will be misunderstandings and confusion specific to each aspect of the cloud, not just to the overall term. There will be overestimations and underestimations that will cause users to be confused, have doubts and be disillusioned. There will be misuse and miscommunication among users and vendors, making it a subject that is ripe for Gartner's Hype Cycle. As usual, in cases like this, there is indeed overhype, but also potential benefits. Understanding those benefits requires tearing apart the hype surrounding cloud computing (now just beyond the Peak of Inflated Expectations, and headed for the Trough of Disillusionment), and looking at the many more-granular topics, which are all part of the cloud phenomenon. This follows the pattern observed with other similarly broad labels, such as the Internet and the Web. Cloud is already showing some signs of disillusionment. Salesforce.com's CEO Marc Benioff is beginning to distance himself from the concept. Over-reactions to high-profile outages, such as Amazon, have fed skeptical views. Yet, we continue to see an increased focus on the term and continued abuse of the term through "cloudwashing," increasing the hype.

One of the things that is moving cloud computing toward the Trough of Disillusionment is cloudwashing. Vendors across the board have grabbed the term and use it for marketing purposes. Hosting solutions that have a pay-per-user-per-month pricing model, but without shared elastic capabilities, are being called cloud. As cloud is used to apply to an ever-widening set of external services without regard to sharing, elasticity and the other characteristics that set cloud apart, the term becomes even more cloudy. Companies will buy these cloud services and then be surprised to find that they do not get the agility and cost savings promised by cloud computing. This has pushed cloud toward the Trough.

Our continued research on cloud computing looks at the big picture, as well as the details of cloud computing and the many issues facing enterprises today. With hype comes confusion. While cloud computing is about a very simple idea — consuming and/or delivering services from the cloud — there are many issues regarding the types of cloud computing and the scope of deployment that make the details not nearly so simple. Everyone has a perspective and an opinion; and while some aspects are coming into focus, confusion remains the norm. Many misconceptions exist around potential benefits, pitfalls and, of course, cost savings. Cloud is often part of cost-cutting discussions, even though its ability to cut costs is not a given. There are also many reasons to talk about the capabilities enabled by cloud computing: agility, speed and innovation. These are the potential benefits that can be overlooked if hype fatigue sets in.

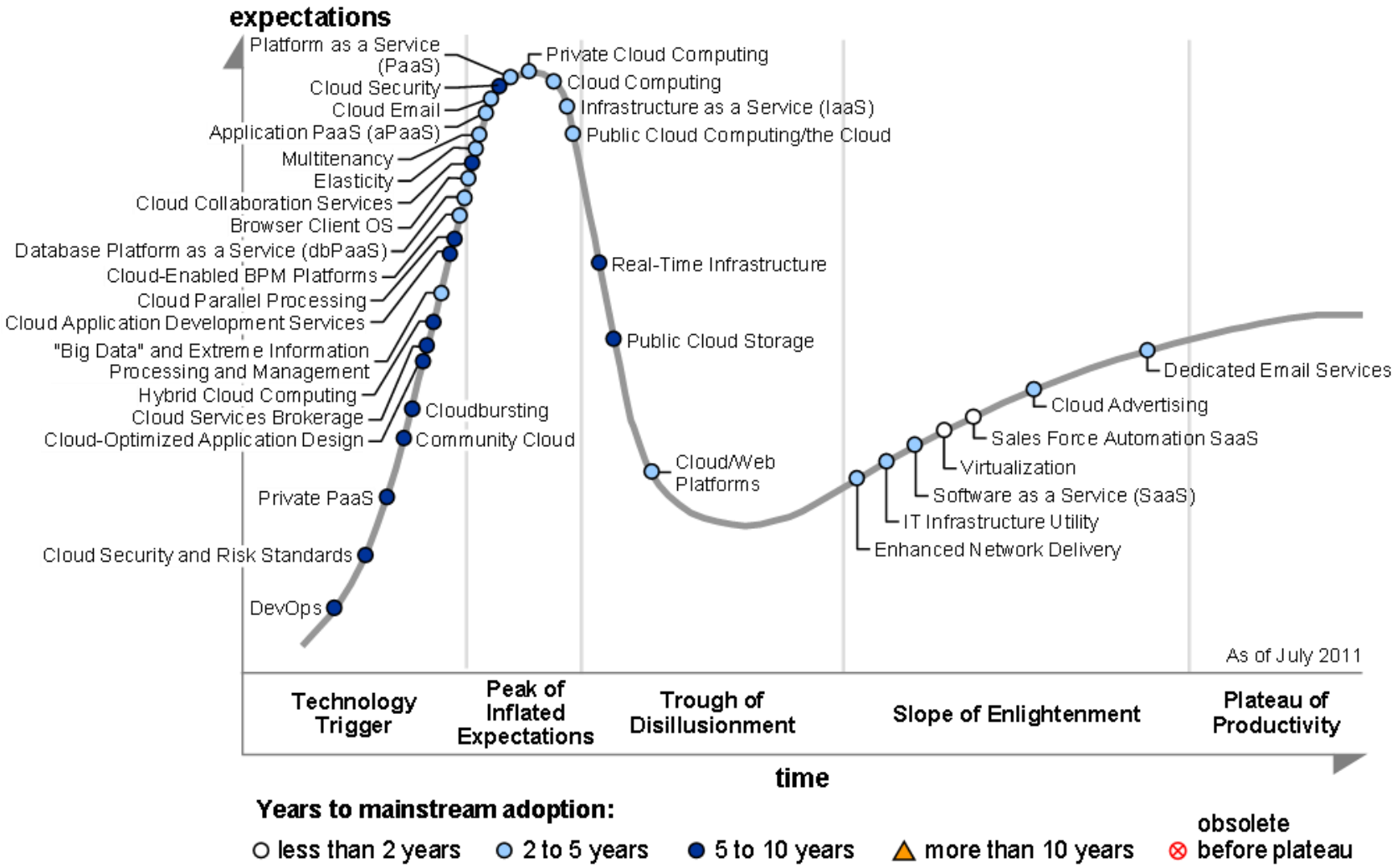
This Hype Cycle for Cloud Computing identifies which aspects of cloud computing are still primarily in the hype stage, which applications/technologies are approaching significant adoption and which ones are reasonably mature. There continues to be a seemingly never-ending supply of new concepts that are pre-Peak, and are attempting to "piggy back" onto the cloud hype. There are many aspects of cloud (including cloud computing itself) clustered around the Peak of Inflated Expectations. Cloud computing has just passed the Peak, although it is nowhere near the Trough as yet. And while the term cloud computing is relatively new, it incorporates derivations of ideas that have been in use for some time. Hosting, software as a service (SaaS) and virtualization are well-established and being used in many ways. The prevalence of inexpensive computing power, inexpensive bandwidth and companies that have developed extensive capabilities in managing large data centers are all relatively new and are all required for the cloud to come to fruition. Some concepts born in the cloud have begun to take on largely hyped lives of their own. Private

cloud computing and platform as a service (PaaS) are becoming more hyped than the term cloud. As cloud computing continues to move beyond the pure hype stage and into mainstream adoption, it is important to dig beyond the main cloud term to the actual ideas and technologies, to dodge the hype and take advantage of the benefits that exist. As always, once the hype dies down, the true value will arrive.

The Hype Cycle

This Hype Cycle covers a broad collection of concepts and technologies associated with cloud computing (see Figure 1). It is the third instance of this Hype Cycle.

Figure 1. Hype Cycle for Cloud Computing, 2011

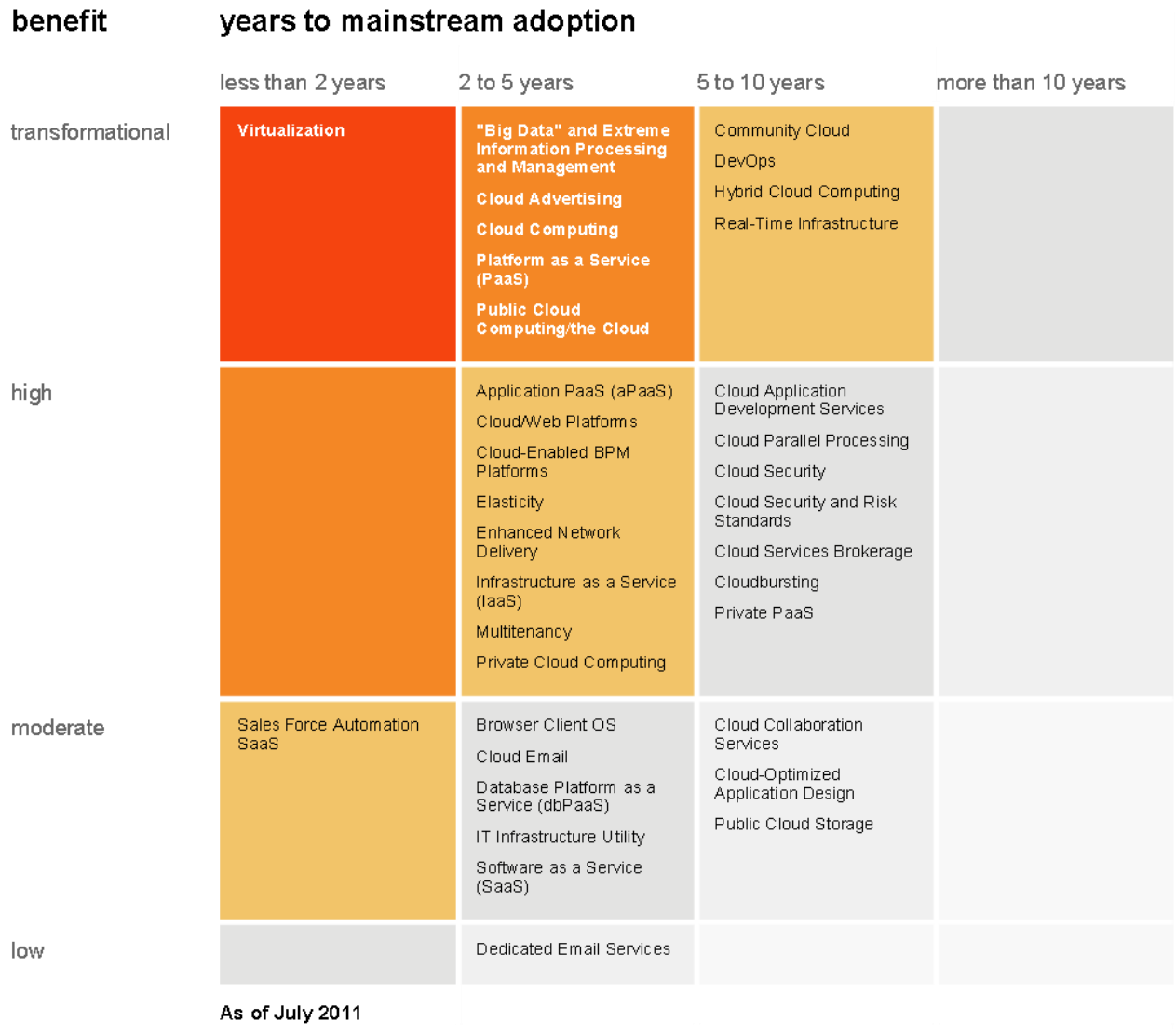


Source: Gartner (July 2011)

The Priority Matrix

Most cloud computing technologies and concepts are more than two years from mainstream adoption, with the exception of certain types of SaaS, such as sales force automation. Many cloud technologies and concepts will see mainstream adoption in two to five years. Some of the more impactful items include application platform as a service (aPaaS), virtualization, elasticity and private cloud computing. Some technologies and concepts (such as cloudbursting/overdraft) will take five to 10 years for mainstream adoption to occur (see Figure 2).

Figure 2. Priority Matrix for Cloud Computing, 2010



Source: Gartner (July 2011)

Off The Hype Cycle

We have renamed several entries:

- Cloud Storage, now Public Cloud Storage
- Cloud Security Concerns, now Cloud Security
- DBMS as Cloud Service, now Database Platform as a Service (dbPaaS)
- Cloud aPAAS, now Application PaaS (aPaaS)
- Cloudbursting/Overdraft, now Cloudbursting
- Cloud Compute Services, now Infrastructure as a Service (IaaS)
- Cloud Testing Tools, mow Cloud Application Development Services
- Cloud App Dev Tools, now Cloud-Optimized Application Design

Cloud Services Integration has been removed from the Hype Cycle.

On the Rise

DevOps

Analysis By: Cameron Haight; Ronni J. Colville; Jim Duggan

Definition: Gartner's working definition of DevOps is "an IT service delivery approach rooted in agile philosophy, with an emphasis on business outcomes, not process orthodoxy. The DevOps philosophy (if not the term itself) was born primarily from the activities of cloud service providers and Web 2.0 adopters as they worked to address scale-out problems due to increasing online service adoption. DevOps is bottom-up-based, with roots in the Agile Manifesto and its guiding principles. Because it doesn't have a concrete set of mandates or standards, or a known framework (e.g., ITIL, CMMI), it is subject to a more liberal interpretation.

In Gartner's view, DevOps has two main focuses. First is the notion of a DevOps "culture," which seeks to establish a trust-based foundation between development and operations teams. In practice, this is often centered on the release management process (i.e., the managed delivery of code into production), as this can be a source of conflict between these two groups often due to differing objectives. The second is the leveraging of the concept of "infrastructure as code," whereby the increasing programmability of today's modern data centers provides IT an ability to be more agile in response to changing business demands. Here, again, the release management process is often targeted through the increasing adoption of automation to improve overall application life cycle management (ALM). Practices like continuous integration and automated regression testing should also be mastered to increase release frequency, while maintaining service levels.

Position and Adoption Speed Justification: While DevOps has its roots in agile methodologies, and, from that perspective, is not totally new, its adoption within traditional enterprises is still very limited, and, hence, the primary reason for our positioning. For IT organizations, the early focus is on streamlining release deployments across the application life cycle from development through to production. Tools are emerging that address building out a consistent application or service model to reduce the risks stemming from customized scripting while improving deployment success due to more-predictable configurations. The adoption of these tools is not usually associated with development or production support staff per se, but rather with groups that "straddle" both development and production, typically requiring higher organizational maturity.

DevOps does not preclude the use of other frameworks or methodologies, such as ITIL, and, in fact, the potential exists to incorporate some of these other best-practice approaches to enhance overall service delivery. Enterprises that are adopting a DevOps approach often begin with one process that can span both development and operations. Release management, while not mature in its adoption, is becoming the pivotal starting point for many DevOps projects.

User Advice: While there is growing hype about DevOps, potential practitioners need to know that the successful adoption or incorporation of this approach is contingent on an organizational philosophy shift — something that is not easy to achieve. Because DevOps is not prescriptive, it will likely result in a variety of different manifestations, making it more difficult to know whether one is "doing" DevOps. However, this lack of a formal process framework should not prevent IT organizations from developing their own repeatable processes that can give them both agility as well as control. Because DevOps is emerging in terms of definition and practice, IT organizations should approach it as a set of guiding principles, not as process dogma. Select a project involving development and operations teams to test the fit of a DevOps-based approach in your enterprise. Often, this is aligned with one application environment. If adopted, consider expanding DevOps to incorporate technical architecture. At a minimum, examine activities along the existing developer-to-operations continuum, and experiment with the adoption of more-agile communications processes and patterns to improve production deployments.

Business Impact: DevOps is focused on improving business outcomes via the adoption of agile methodologies. While agility often equates to speed (and faster time to market), there is a somewhat paradoxical impact, as well as smaller, more-frequent updates to production that can also work to improve overall stability and control, thus reducing risk.

Benefit Rating: Transformational

Market Penetration: Less than 1% of target audience

Maturity: Emerging

Cloud Security and Risk Standards

Analysis By: Jay Heiser

Definition: The buyers of externally provisioned services always need a standardized set of requirements against which they can assess the relative security and continuity risks. This is especially true for cloud-based services, because they insert additional levels of abstraction that make it more difficult for the business to evaluate the risk of using the service. A cloud risk framework consists of a taxonomy or outline of risk-relevant issues. Every standard today includes an evaluation questionnaire, which is understandably the most widely anticipated component of the standard, but current and anticipated frameworks also include discussion papers, evaluation guidelines and cross-references to related security standards. In order to be treated as a standard, a framework must be widely considered by buyers, sellers, auditors and other stakeholders as incorporating the current understanding of what constitutes "best practice" for the security and continuity issues of an externally provisioned service.

Position and Adoption Speed Justification: One of the most commonly expressed buyer frustrations is the lack of a "checklist" of items that need to be addressed when assessing a service provider's relative ability to maintain information confidentiality, integrity, service reliability and the likelihood that a provider can restore data and service after a disaster or data loss. Four different organizations have committed to delivering standardized frameworks for the evaluation of service provider risk, which will likely include some form of questionnaire, along with other supporting tools, background information and assessment practice guidelines.

Versions of two frameworks exist already, including questionnaires that can be downloaded and used at no cost. Shared Assessments is an organization consisting of 60 member companies (primarily in the financial services industry) that publishes the Standardized Information Gathering questionnaire. Colloquially referred to as the SIG, it is intended to be applied to a wide range of service provision scenarios, with the latest version, v.6, significantly updated to take into account the unique risk aspects of cloud computing. The Cloud Security Alliance (CSA) was formed in 2009, and is still on the first version of its "Consensus Assessments Initiative Questionnaire" or "CAIQ." Although both questionnaires arguably can be positioned as representing today's understanding of best practices for service provider risk review, neither is in widespread use.

Two additional cloud security assessment frameworks are expected to release their initial standards at the end of 2011. The U.S. government has a program specifically intended for federal agencies, Federal Risk and Authorization Management Program, usually referred to as FedRAMP. A European initiative, the Common Assurance Maturity Model (CAMM), will be unique in that it is expected to provide an actual rating.

2011 is the initial year in what will be a multiyear effort to reach consensus — first on what risk issues are most relevant to externally provisioned services, and second, how to most efficiently assess a provider's relative ability to maintain a low level of risk. The practicality and completeness of these standards will be tested as they are applied, and it will take at least five years before they can be fully accepted as being useful and sufficient.

User Advice: Instead of creating a unique questionnaire in-house, buyers of commercial cloud services should demand a complete response to one of the published cloud risk questionnaires. Providers that refuse to comply can legitimately be asked why they would not want to meet current best practices for transparency. The SIG and CAIQ are in the form of yes/no questions, but in critical use cases, significantly better risk assessment information can be obtained by replacing the binary "do you" with "how do you."

Organizations using International Organization for Standardization (ISO) 27001 and 27002 should consider incorporating material from the CSA, which recently announced an alliance with ISO. U.S. government agencies will have to wait for FedRAMP, which is not expected to have anything meaningful until 4Q11 at the earliest. However, it will draw from Federal Information Processing Standards (FIPS) publication (PUB) 800-53, which is already mapped to ISO 270001, so the CSA material should be useful in the interim, as well as the background material in the draft National Institute of Standards and Technology (NIST) document "Guidelines on Security and Privacy in Public Cloud Computing," SP 800-144.

Be cautious in accepting vendor claims that their products are fit for purpose because they have undergone a Statement on Auditing Standards (SAS) 70 or Statement on Standards for Attestation Engagements (SSAE) 16. Such a claim is meaningless without a review of the detailed auditor's report to ensure that the scope of the assessment was complete, and that the set of controls evaluated truly was adequate in meeting the organization's business requirements for security, continuity and recoverability.

Business Impact: The inconvenience of current risk assessment processes acts as a significant inhibitor on the cloud computing market, adding unwanted momentum to what should in many ways be a very fluid market. The advantages of extreme specialization are encouraging ever-more-complex chains of provider. If it is difficult for a buyer to evaluate the risks associated with a single provider, it's virtually impossible today to evaluate the risks of a multivendor service. It isn't just buyers that would benefit from a streamlined approach to provider evaluation. Investors and insurance underwriters would also be more supportive of cloud service providers if they had access to a standardized way to assess risk.

A standard for controls and their evaluation is a typical first step toward the creation of third-party evaluation programs. Ultimately, some form of cloud risk certification would be hugely beneficial in reducing the market friction caused by concerns over security and the current lack of transparency.

Benefit Rating: High

Market Penetration: Less than 1% of target audience

Maturity: Embryonic

Recommended Reading: "Cloud Security and Risk Standards"

"Survey Results: Assessment Practices for Cloud, SaaS and Partner Risks"

Private PaaS

Analysis By: Yefim V. Natis

Definition: Platform as a service (PaaS) is a reference to the functionality of application infrastructure software, offered as a service. Application servers, database management systems (DBMSs), integration brokers, portal products, messaging technology and business process management technology are all examples of application infrastructure software (also known as middleware). When offered as services, these become examples of PaaS.

Most of the discussion about PaaS at this stage of cloud computing adoption is focused on the public cloud. Offerings like salesforce.com's Force.com, Microsoft SQL Azure and IBM Cast Iron are typical examples.

The majority of early investments in the private cloud have focused of the technology layer below PaaS: the system infrastructure services (or IaaS). However, as the initial typical private cloud implementations (which might be referred to as "private IaaS") reach maturity, some IT organizations turn their attention to private PaaS. The initial objective of many private cloud projects was to consolidate control over the proliferating networks of virtual machines (VM) and, in the more advanced cases, improve hardware resource utilization through the dynamic allocation of VMs. The objective of private PaaS is to further improve the agility of the enterprise computing infrastructure by expanding the logic of sharing computing resources to higher levels of software stacks; to further improve resource utilization through more-advanced elasticity and multitenancy; and to further reduce costs through these and other advanced technology patterns available with PaaS architecture.

In a private PaaS, multitenancy is applied primarily to multiple applications acting as tenants and competing for shared resources (this is in contrast to the public PaaS, where the tenants are often IT organizations; see "Understanding Tenancy: Salesforce.com Versus Google.com").

It is important to note that deploying a traditional application server, DBMS, ESB or another middleware product over private IaaS most of the time does not constitute a viable private PaaS. The expected benefits can only be achieved by using a cloud-enabled application infrastructure technology — one that extends the resource sharing and optimization up the stack into the application container and application infrastructure. Placing a precloud middleware product over managed virtualization technology most of the time amounts to simply hosting that product on VMs, with minimal or no sharing of any resources between tenants.

Implementation of a private PaaS requires that the application infrastructure technology is inherently cloud-enabled: It must embed support for multitenancy or, at a minimum, be dynamically and elastically horizontally-scalable. In the first case, multitenancy and elasticity are

implemented in the application infrastructure layer (shared-container or shared-everything multitenancy, see “Gartner Reference Architecture for Multitenancy” G00205983). In the second case, the elasticity is implemented in the IaaS layer by dynamically allocating VMs based on established policies, and the application infrastructure expands or contracts to follow the changing underlying resources, without impacting the performance of the running applications.

The application infrastructure products that are enabled for cloud computing are referred to as cloud-enabled application platforms (CEAPs). The same products are also used by providers of public PaaS and by independent software vendors (ISVs) providing cloud application services (software as a service [SaaS]). CEAPs are available from many software vendors (such as Rollbase, Cordys, LongJump and GigaSpaces), often in conjunction with a public PaaS offered by the same vendor (see "Platform as a Service: Definition, Taxonomy and Vendor Landscape, 2011").

Position and Adoption Speed Justification: Most private cloud projects remain focused on managing pools of VMs. More-advanced private cloud projects add elastic allocation of VMs and other resources. Few extend their investment to cloud-enabled application infrastructure products to build true private PaaSs. Private cloud projects are likely to be an important area of growth for CEAP providers. However, in 2011, this process is in its early stages. Lack of standards or industry experience in building CEAP-based environments delay mainstream adoption. However, as more private cloud installations reach maturity in their use of private IaaS, many will turn to private PaaS as the logical next step.

User Advice: Start a private cloud project focusing on private IaaS, managing and optimizing your network of VMs. Run traditional application infrastructure and traditional applications in the first stage of your private cloud to become familiar with the costs and benefits of cloud computing incrementally, avoiding major discontinuities.

Start experimenting with private PaaS by deploying horizontally scalable application infrastructure over elastically managed pools of VMs (shared-hardware PaaS architecture). Some of the traditional application servers and DBMSs are or will be available in the elastic clustering mode. This approach offers greater "cloudiness," while still preserving the familiar programming skills.

Look into native cloud-enabled application infrastructure products to achieve maximum agility and the highest levels of density in utilizing your computing resources. These platforms may require adopting a new model of programming.

Give preference to cloud-enabled application infrastructure products that have one or, even better, multiple public PaaSs using them. This will set you up for future expansion or transition to the public cloud in a minimally intrusive manner.

Before moving to public or private PaaS, make sure that your IT environment is well-versed in SOA administration. Adjusting to using cloud application resources is a lot like building SOA services.

Business Impact: The transition of mainstream enterprise IT from traditional stove-piped data center architectures to private cloud computing is a transformational change. Private PaaS is an important step in this transformation. IT organizations that adopt private PaaS architectures will position themselves to enjoy higher degrees of agility, efficiency and manageability of their information resources. They will also be better prepared to adopt the hybrid combination of public and private cloud, expanding their technology utilization options, as well as increase the scope of the available information processing resources.

Benefit Rating: High

Market Penetration: 1% to 5% of target audience

Maturity: Emerging

Sample Vendors: Appistry; Apprenda; Cordys; Expanz; IBM; Intalio; IS Tools; Joyent; LongJump; Magic Software Enterprises; Pegasystems; Red Hat; Rollbase; Seeburger; VMware; Wolf Frameworks; WSO2

Recommended Reading: "Understanding Tenancy: Salesforce.com Versus Google.com"

"PaaS Road Map: A Continent Emerging"

"The 10 Fundamentals of Building a Private Cloud Service"

"Platform as a Service: Definition, Taxonomy and Vendor Landscape, 2011"

"The Road Map From Virtualization to Cloud Computing"

"Gartner Reference Architecture for Cloud-Enabled Application Platforms"

Community Cloud

Analysis By: David W. Cearley

Definition: Community cloud computing refers to a shared cloud computing service environment that is targeted to a limited set of organizations or individuals (e.g., government agencies). The organizing principle for the community will vary, but the members of the community generally share similar security, privacy, compliance, performance, reliability, availability and scalability requirements.

It important to have a mechanism to review those seeking entry into the community. At a minimum, there must be a strict and narrow admission policy to create trust among the members of the community, as well as runtime authentication to ensure that only members of the community are accessing the community resources. Ideally, other members of the community should have some mechanism to identify other members and, in some cases, have an opportunity to exert some level of influence over the admission of new members to the community.

Unlike the more limited forms of private cloud computing that serve a single organization, community cloud computing is similar to public cloud computing, in that it delivers shared services to multiple organizations. Therefore, a community cloud environment can achieve larger economies of scale, spread costs across the community and further reduce redundancy. However, unlike a public cloud service, a community cloud service serves substantially fewer users, and may require additional technologies and unique policies/practices to address the more extensive privacy, security, compliance, availability and performance needs of the community. As a result, although a community cloud computing approach is likely to be less expensive than a custom, single-enterprise private cloud implementation, it is likely to be more expensive than public cloud computing.

An industry-based community cloud is a cloud environment that is built for and targeted exclusively to companies in a particular industry. This is the most common model for community clouds and is a major theme in the government sector. The greater maturity of government community cloud efforts stems from previous efforts in government to create centralized and consolidated shared-service environments prior to the emergence of cloud computing.

The initial focus for community cloud computing is limiting membership to known and trusted entities (e.g., defense agencies) to address common security needs. These factors have meant that government-focused community clouds were the first to emerge. Finance is the most recent market to see significant community cloud activity with the NYSE announcing a financial services

community cloud. A number of business structures are in place in the financial services and banking industry (e.g., business process networks such as SWIFT) that complement a community cloud approach. One potential concern with an industry-based community cloud is that members may have similar seasonal resource demands (e.g., retail has large demand spikes around Christmas) that limit the ability of the community cloud to elastically satisfy all member needs at peak times.

Although an industry model is the most likely approach to forming community cloud environments, community cloud services can be delivered based on other organizing principles. For example, a membership-based community has controlled membership, but is not limited to a particular industry. Membership-based communities are emerging as alternatives to industry-specific community cloud platforms, because the ability to blend resource demands across industries provides additional flexibility and more closely approximates the leverage that can be obtained with public cloud models. A broad membership-based community is used by both IBM and HP to deliver their cloud infrastructure services. In both cases, potential consumers must approach the provider and establish a business relationship first and then they are admitted to the "community" of "clients." Although the community links, in this case, are minimal, the approach prevents unknown individuals with credit cards (possibly stolen) from accessing the services.

A more extensive membership community would include more rigorous process to review membership applications and ensure that anyone with access to the cloud environment is trusted by other members of the community. Membership in these community cloud environments may require that applicants for membership meet certain requirements with regard to their own security environments, and there could be a voting process by which community members approve or deny a potential applicant.

A community cloud service provider may organize itself in several ways. In the case of government community cloud computing, an agency is often specifically tasked with providing shared cloud computing services to other government agencies. Commercial third parties (e.g., Google, Amazon, Microsoft and HP) will also establish community cloud environments to provide services to community members. In other cases, the members of the community may appoint one or more members, or may establish a jointly owned entity, to implement, run and manage the community cloud environment. In financial services, this may be a network provider or an exchange, as well as software vendors with a large interest in financial services.

Position and Adoption Speed Justification: The leading concerns expressed by those considering the use of public cloud computing services are security, privacy, compliance and regulatory approval. These concerns stem from the fact that public cloud computing services are provided via an open Internet-based model to the general public, and that the workloads and data of all consumers run in a common, shared environment. A community cloud model is one way to begin addressing these concerns, by limiting access and focusing the types of services and performance levels delivered. Although interest in community cloud computing is growing, its full scope and benefits are not yet well-understood.

Although there are more mature examples of government community cloud computing, and it is growing in the banking and financial services sector, its use in communities such as healthcare is more embryonic. We expect growth in the interest and popularity of community cloud computing in other market sectors during the next three years.

User Advice: Community cloud computing models can offer enterprises with significant legal and compliance requirements an approach to using third-party cloud services in a more controlled manner. By limiting access to a cloud service environment to entities that have met some common membership criteria, trust in the cloud environment can theoretically be increased. However, the community cloud is an emerging concept, and users should examine community cloud offerings closely to ensure that they truly meet their needs.

A community cloud provided by a public cloud provider should offer physical isolation (at least separate servers, though separate data centers is even better) between the community cloud and the vendor's public cloud offerings. In addition, community cloud providers with a global audience should offer the ability to keep programs and data isolated in defined political jurisdictions. Users considering a community cloud must examine the process of vetting potential members, as well as the mechanism for isolating the community cloud resource environment of a provider's public cloud environment, if one exists.

Business Impact: Community cloud computing offers a blend of public and private cloud benefits and challenges, and is likely to serve as an intermediary stage between private and public cloud computing. Like the public cloud, the cost of the cloud service environment is spread across multiple organizations, enabling economies of scale, higher utilization rates and reduced cost versus a single-enterprise private cloud. In addition, the community cloud offers a more controlled and limited environment where members are known, which reduces risk, versus an open, public cloud computing environment.

In the future, we expect industry-based community clouds to provide a common set of shared infrastructure services, as well as specialized application, information and business process services of unique value and interest to the target industry. The emergence of community clouds based on a pooling of member resources will provide the basis for a hybrid private-and-community cloud model.

Benefit Rating: Transformational

Market Penetration: 1% to 5% of target audience

Maturity: Emerging

Sample Vendors: BT; Google; IBM; Microsoft; NYSE Euronext; Verizon

Recommended Reading: "Three Areas of Risk and Compliance Management Where the Cloud Can Add Value"

Cloudbursting

Analysis By: Neil MacDonald; Donna Scott; Daryl C. Plummer

Definition: Cloudbursting is the use of an alternative set of public or private cloud-based services as a way to augment and handle peaks in IT system requirements. It is similar to bank account overdrafting, where money is borrowed from one account to cover insufficient funds in another. Cloudbursting can span between on-premises IT systems and the cloud, as well as across multiple cloud providers. It can also be enabled across multiple internal data centers, across multiple external data centers, or between internal and external data centers. It is a specific implementation of elasticity, whereby capacity-bound services can seek and obtain additional capacity through programmatic means from alternative providers. Cloudbursting can be achieved at startup of a specific service, as well as accommodate additional demands while running.

Position and Adoption Speed Justification: For enterprise IT consumers, initial interest will be to provide capacity for peak workloads from on-premises to cloud-based infrastructure as a service (IaaS), or by moving less critical resources to a cloud provider to free up capacity for more critical, on-premises workloads. Initially, this usage scenario will be a manual process, but it will increasingly be automated over the next decade through the use of service governor technology. Standards for the seamless exchange of workloads, security and service-level agreement (SLA) requirements between alternative providers have not yet matured. For this reason, automation of this process will initially be tied to a specific vendor's implementation or will require migration/conversion.

Cloudbursting will not be limited to IaaS, although that is where most organizations will initially adopt this technology. Platform as a service (PaaS) and software as a service (SaaS) overdraft will also become viable over the next decade, built directly into application platforms.

User Advice:

- Consider cloudbursting for workloads with highly variable levels of utilization to address peak periods of usage without having to overprovision on-premises-based IT systems.
- Consider cloudbursting as a means to broker service delivery through many internal and external resources.
- Explore the use of cloudbursting with IaaS, but also evaluate PaaS and SaaS vendors on their ability to provide such capabilities.

Business Impact: Cloudbursting reduces the cost of provisioning IT systems for peak workloads (such as the holiday system or quarterly close process) or peak overall capacity. It enables alternative providers' systems to be used as an overflow for on-premises or cloud-based systems.

Benefit Rating: High

Market Penetration: 1% to 5% of target audience

Maturity: Embryonic

Sample Vendors: Amazon; BMC Software; CA Technologies; DynamicOps; HP; IBM; ManageIQ; ServiceMesh; VMware

Recommended Reading: "Clearing the Confusion About Fabric-Based Infrastructure: A Taxonomy"

"Hype Cycle for Real-Time Infrastructure, 2010"

Cloud-Optimized Application Design

Analysis By: Eric Knipp; Mark Driver; Ray Valdes

Definition: Cloud-optimized solutions are designed to take full advantage of the global-class characteristics of cloud computing platforms. Horizontal scalability, fault tolerance, high performance, efficiency and ease of interoperability are several of the key principles underlying successful cloud-optimized solutions. Cloud-optimized application design embodies the practices and patterns required to support these principles during solution delivery. The individual practices and patterns applied in cloud-optimized application design are not new to architects of distributed computing systems; however, the availability of massive, distributed, multitenant, pay-per-use cloud computing runtime environments makes them more important than ever. Furthermore, the interplay between the key tenets of cloud-optimized solutions necessitates greatly improved developer skills in large-scale, high-volume application design.

Cloud-oriented application design is applicable to new cloud-aware solutions delivered via infrastructure as a service (IaaS), as well as application platform as a service (aPaaS). IaaS offerings — for example, Amazon Elastic Compute Cloud (EC2), Rackspace Cloud Servers, or similar private cloud services constructed with VMware or OpenStack virtualization technologies — lend themselves better to high-control project styles where developers maintain fine-grained decisions over every attribute of the system beyond basic compute and storage virtualization. Architects and system administrators working within IaaS have complete control over component selection, freely choosing from a wide selection of application infrastructures compatible with virtualization software, such as Java- or .NET-based applications and Web servers, commercial

and open-source databases, and myriad programming languages. Developers constructing cloud-optimized applications on IaaS directly manipulate every attribute of their solutions, enabling low-level control of the user experience, reliability, portability and performance. To maximize the potential of the cloud services used in solution delivery, developers and architects should become familiar with event-driven and parallel programming, adding tools such as the actor model to the well-worn model-view-controller pattern used in many Web applications. The additional flexibility available with IaaS is a positive and negative attribute, as many Amazon EC2 customers discovered in April 2011, when a major service disruption in an Amazon Web Services (AWS) data center exposed inadequate cloud application design. Dozens of high-profile websites and applications created by skilled programmers and thought to be cloud-optimized were cloud failures. A handful of high-profile cloud services, including Twilio and Netflix, survived the disruption relatively unscathed, due to greater consideration for the key principles of cloud-optimized application design, and to the fortuitous avoidance of a key point of failure: the Elastic Block Store (EBS). Being aware that failure is inevitable will be a key consideration in the development of cloud-optimized solutions; "design for failure" will become a best practice.

APaaS offers an alternative for enterprises that are unable or unwilling to develop the skills required to construct cloud-optimized solutions from a low level. The canonical aPaaS provides a complete prepackaged development and runtime environment, embodying everything from development tools, to databases, to programming languages, to elastic application and integration middleware. While aPaaS offerings tend to be proprietary, they also abstract some percentage of cloud-optimized application design decisions away from developers, enabling less-skilled programmers to build scalable, reliable cloud solutions. We see here a trade-off: less control, flexibility and portability in exchange for less-challenging development and shorter time to market, compared with IaaS-based solutions. Architectural divergence is the norm across the aPaaS market, with some platforms, such as Amazon Elastic Beanstalk, offering little more than a thin IaaS management layer for IT programmers, while others, such as TrackVia, provide a codeless environment suitable for end users and citizen developers. One major caveat with aPaaS versus IaaS is warranted: With abstraction comes opacity. APaaS providers add value in large part by hiding the complexity of "autoscaling" to provide horizontal scalability, of automatic failover and redundancy to provide fault tolerance, and of managing tenant service levels automatically. If an aPaaS provider fails to adequately consider and implement the key principles of cloud-optimized application design, then enterprise customers will be at risk, and they will have no way to know it, much less remediate it. For these reasons, IT leaders should place a premium on provider transparency when evaluating aPaaS alternatives.

Position and Adoption Speed Justification: While the adoption (or, at least, attempted adoption) of cloud-optimized application design practices and patterns remains primarily in the realm of cutting-edge Web startups, Gartner clients report increasing levels of interest in constructing their own cloud solutions to capture the reduced operational complexity and time-to-market benefits of cloud services. Increased adoption of private cloud computing, which tends to manifest itself primarily in coarse-grained compute and storage virtualization (e.g., IaaS), will provide application development (AD) departments with self-service resources that can only be maximized with cloud-optimized application design. Popular programming frameworks like Spring, Rails and Node.js are evolving to encompass and simplify the use of the principles of cloud-optimized application design. Finally, the aforementioned high-profile failures spawned by the Amazon EC2 outage have made it clear, even to Web startups, that more attention must be paid to those principles.

User Advice: IT leaders considering custom cloud solution delivery should apply the principles of cloud-optimized application design when high scalability, fault tolerance, performance and efficiency are of paramount importance. Low-importance solutions for which some downtime is not a major business challenge may not justify the full gamut of due diligence that cloud-optimized application design implies; in these cases, consider aPaaS, which will provide lower-

cost solution delivery in exchange for degraded portability and increased opacity, and a more constrained design space. High-importance solutions should be constructed on top of either multiple IaaS offerings and fully utilize cloud-optimized application design principles, or aPaaS offerings that can provide abstractions that fully implement those principles.

Business Impact: Failure to consider and address the principles of cloud-optimized application design will burden the enterprise with an unknown risk likely to be realized at an inconvenient time. At the same time, applying the full gamut of those principles to every custom cloud solution will burden the enterprise with unnecessary costs, intensive skill requirements and loss of agility. IT leaders should establish guidelines that define when and where the various practices of cloud-optimized application design should be applied, based on cost, risk and time factors applicable to a given cloud solution delivery initiative.

Benefit Rating: Moderate

Market Penetration: Less than 1% of target audience

Maturity: Emerging

Recommended Reading: "Creating Cloud Solutions: A Decision Framework"

"Symposium 2010 Q&A: Web and Cloud AD Gain Mind Share"

"How to Balance the Business Benefits and IT Costs of SOA"

Cloud Services Brokerage

Analysis By: Benoit J. Lheureux; Daryl C. Plummer

Definition: Cloud services brokerage (CSB) is a form of cloud services intermediation. We define CSB as an IT role and business model in which a company or other entity adds value to one or more (generally public or hybrid, but possibly private) cloud services on behalf of one or more consumers of that service (see "Defining Cloud Services Brokerage: Taking Intermediation to the Next Level"). In an internal private cloud scenario, the internal IT organization takes on the role of broker. A CSB provides technology to implement CSB functionality, people and methodologies to implement and manage CSB-related projects, or a combination of the two.

A CSB does many of the same things that a traditional IT services provider does in a service aggregator role, but also addresses additional complexities, particularly relevant to cloud computing and to achieving specific IT outcomes. CSBs with enduring business models will offer compelling combinations of the capabilities necessary to address these complexities:

Management of the proliferation of solution assets — Traditional consulting or outsourcing projects tend to focus on a solution (for example, implementing an ERP solution, building a custom application, managing the ongoing operations of an ERP solution or providing payroll services). In comparison, a cloud service may be more atomic, such as a credit check cloud service in conjunction with a few other cloud services. A CSB can help intermediate this complexity.

Rapid pace of change — The more service providers involved in providing the parts of a solution on an ongoing basis, the more likely the changes in the solution. The pace of change is likely to be faster than is normally seen in many traditional consulting, system integrator (SI) and outsourcing relationships, and CSBs will need to adopt more agile approaches for dealing with change (see "Eight Tactics for Managing the Tempo of Process Change in BPO and Cloud-Enabled Business Process Services").

IP and compliance risk — Cloud Services Provider A may consume a service from Cloud Services Provider B. The end consumer of Cloud Services Provider A may or may not know about the services from Cloud Services Provider B. A CSB is necessary to indemnify the end consumer from the licensing and intellectual property (IP) risks of "embedded" cloud services. Although traditional service providers supply these capabilities, they tend to do so for solutions involving a smaller number of other service providers.

Security and risk — Using multiple cloud service providers poses challenges for even rudimentary security policies (e.g., single sign-on [SSO]). CSBs can provide enhanced security and compliance through services in the cloud (e.g., SSO capabilities across multiple cloud providers), or even by supporting on-premises solutions at the provider's or the consumer's site. There is an increased interest in recovery and restoration services, as well as a need for increased transparency of provider security practices and certifications.

A CSB offers some combination of the following functional capabilities:

- **Integration brokerage —** integration project implementation and management
- **Governance —** for security and policy compliance of cloud services consumption
- **Community management —** to manage the provisioning of consumers and providers
- **Service enrichment —** such as billing, aggregation, arbitrage, context and analytics
- **Service delivery —** IT services for cloud services and CSB project implementation and management
- **Distributed quality of service (QoS) —** delivering good QoS across a distributed set of cloud services
- **Analytics and operational intelligence** — delivering business activity monitoring
- **Software as a service (SaaS) and custom SaaS —** where appropriate, to support specific business processes

A CSB may or may not have a direct commercial relationship with the providers of services, but will more often have a commercial relationship, particularly when enriching services, for example, by extending a common billing mechanism to consumers across a wide range of services. Although the right combination of cloud-based technology underpinnings is essential to enable CSB offerings, the primary value proposition of CSBs lies in delivering such technology within a larger IT service offering. For example, although, by definition, a CSB may use integration-platform-as-a-service (iPaaS) functionality (see "Integration Platform as a Service: Moving Integration to the Cloud") to support cloud services integration, the more important thing is that a CSB typically offers integration brokerage as part of its overall CSB offering.

A viable CSB provider can potentially make it easier, safer, less expensive and more productive for companies to navigate, integrate, consume and extend well-governed cloud services, particularly when they span multiple, diverse cloud services providers. A viable CSB provider may offer some combination of various value-added business and technical CSB functions, including:

- Reducing the risk of consuming services — for example, via federated security and compliance
- Integrating diverse cloud services, including SaaS — for example, via protocol intermediation and semantic translation, sometimes enabled by iPaaS
- Adding significant value to services — for example, context and analytics

- Providing relevant subject matter expertise — for example, if the CSB involves procurement business process outsourcing (BPO), then it requires expertise in procurement processes
- Providing exclusivity and transparency
- Centralizing cloud services functionality — for example, service aggregation, billing, archival and auditability
- Providing a central point for governance — for example, for U.S. federal government mandates or European Union directives
- Offering various IT services to help consumers with CSB-related project implementations, ranging from managed services to business process utility
- Customizing services to create a new layer of value, and to meet the specific requirements of consumers
- Implementing consistent QoS to service consumers across services originating from, potentially, many discrete cloud offerings.

Position and Adoption Speed Justification: As cloud services proliferate across industries and geographies, CSBs will also proliferate, as cloud services consumers simplify and improve their consumption of cloud services across multiple providers. Although users will consume a large portion of cloud services directly, the diversity and complexity of direct cloud services consumption and associated management issues will drive some users toward CSB providers to simplify and improve the process.

CSB providers are emerging from diverse backgrounds. Examples include:

- Cloud solution aggregators such as Okta (cloud-centric identity management) and Common IT (virtual, distributed browsing) solve problems with safely and effectively accessing and browsing multiple cloud services.
- Microsoft (WIndows Azure DataMarket), StrikeIron and Xignite aggregate business information services from a diverse range of sources and providers, normalizing access and providing value-added services for such data via one interface and billing mechanism.
- Providers such as OxygenCloud aggregate and add value (e.g., content metadata tagging) to cloud storage services.
- Providers such as Cloudmore and Jamcracker assume the CSB role when aggregating cloud services in conjunction with SaaS enablement.
- Providers of integration brokerage such as GCommerce, GXS, Hubspan and Liaison Technologies have made specific investments and are evolving their service offerings to not only aggregate trading partners, but also to aggregate cloud services, and, therefore, are adopting the CSB role. This is a convergent scenario (e.g., whereby orders in a supply chain may be received either from trading partners or directly from SaaS, such as salesforce.com, NetSuite or SugarCRM; see "Cloud Service Brokerages Create a New Role for Integration Service Providers").
- Providers of iPaaS such as Boomi, IBM (Cast Iron), Informatica, Jitterbit and Pervasive Software assume the CSB role when they "wrap" their iPaaS in a service offering to deliver cloud services integration to their customers.

- Providers such as Appirio, Celigo, and Verecloud assume the CSB role as they deliver system integration services for cloud-centric projects. For example, implementing customer applications or composite applications (delivered as SaaS) in conjunction with third-party SaaS functionality, and managing delivery of the combination of functionality, including support, SLAs, etc.

- Cloud-focused, value-added resellers such as LTech assume the CSB role when they combine and add value to various third-party cloud services and deliver the combined as one outcome to their customers.

- BPO providers such as ADP or Workday assume the CSB role when they aggregate and incorporate value-added cloud services aggregation into their business process utility (BPU) offerings (see "Cloud Services Brokerages Will Enable Business Process Utilities to Deliver Off-the-Shelf Processes").

- Providers such as Apigee, Mashery, SOA Software and Vordel take on the role of a CSB when they deliver their fine-grained security and policy management capabilities for use by companies exposing cloud services for consumption by many consumers.

- Traditional IT service providers such as Accenture, Capgemini, CSC, TCS and Wipro are all actively offering different types of advisory and cloud-enabled outsourcing services that selectively qualify as CSBs.

Examples and adoption of CSB are proliferating, which justified moving this technology a bit further up the Technology Trigger portion of the Hype Cycle. Nevertheless, the CSB role is still relatively immature, many offerings do not yet offer the full range of potential CSB value-added capabilities, and CSB is not well-recognized or understood by IT consumers; therefore, we believe it will be at least a few more years before CSB reaches the Peak of Inflated Expectations.

User Advice: Consider leveraging a CSB in cloud services consumption scenarios that involve multiple cloud services providers, a large number of consumers or when the CSB offers sufficient value-add to the services executed between providers and consumers to justify its role as intermediary. In particular, consider using a CSB when:

- The CSB makes it less expensive, easier or faster to navigate, integrate, consume or extend services

- The CSB reduces the risk of consuming services from an operational (IT) or business point of view

- The CSB adds significant value to services (e.g., aggregation or enrichment)

- Centralized services address a community need (e.g., consistent service billing, governance and auditability)

- An IT organization wishes to delegate its cloud exposure to a skilled third party

- When a community can't or won't do its own CSB (i.e., when CSB is a sourcing decision or mandate of, for example, the federal government or European Union)

Beware of CSB provider maturity both at the commercial level (e.g., new IT role, unproven business model) and the technical level (e.g., lack of fault-tolerant architecture or a strategy for dealing with third-party cloud services provider failures).

Business Impact: As cloud services proliferate across industries and geographies, CSBs will correspondingly proliferate as more companies seek help in dealing with multiple cloud providers.

While a large portion of cloud services will be consumed directly, the diversity and complexity of direct cloud services consumption will drive at least some users to CSB providers to simplify and improve the process. By 2015, at least 20% of all cloud services will be intermediated via CSBs, up from less than 5% today (see "Predicts 2011: Application Integration: A Nimble Veteran of IT").

Benefit Rating: High

Market Penetration: 1% to 5% of target audience

Maturity: Adolescent

Sample Vendors: Alcatel-Lucent; Apigee; Appirio; Axeda; Capgemini; Celigo; CommonIT; GCommerce; Gigya; GXS; Hubspan; IBM; Jamcracker; Layer 7; Liaison Technologies; LTech; Microsoft; Okta; Oracle-AmberPoint; Oxygen; StrikeIron; Verecloud; Vordel; Wipro; Xignite

Recommended Reading: "Defining Cloud Services Brokerage: Taking Intermediation to the Next Level"

"Cloud Service Brokerages Create a New Role for Integration Service Providers"

"Cool Vendors in Cloud Services Brokerage, 2011"

"What's Between You and the Cloud?"

"Knitting Clouds Together: How Integration as a Service Enables B2B Integration Outsourcing"

Hybrid Cloud Computing

Analysis By: David W. Cearley; Donna Scott

Definition: Virtually all enterprises have a desire to augment internal IT systems with those of cloud services for various reasons, including for additional capacity, financial optimization and improved service quality. Even enterprises that primarily outsource their runtime operations want to augment their systems with cloud resources. Some IT organizations envision third parties providing an infrastructure to their companies, while managing the operations themselves to meet SLAs. However, hybrid cloud computing does not refer to using internal systems and external cloud-based services in a disconnected or loosely connected fashion. Hybrid cloud computing refers to the combination of external public cloud computing services and internal resources (either a private cloud or traditional infrastructure, operations and applications) in a coordinated fashion to assemble a particular solution. Hybrid cloud computing implies significant integration or coordination between the internal and external environments at the data, process, management or security layers.

Hybrid cloud computing can take a number of forms. These approaches can be used individually or in combination to support a hybrid cloud computing approach:

- Joint security and management — Security and/or management processes and tools are applied to the creation and operation of internal systems and external cloud services.
- Workload/service placement and runtime optimization — Using data center policies to drive placement decisions to resources located internally or externally, as well as balancing resources to meet SLAs (e.g., real-time infrastructure).
- Cloudbursting — Dynamically extending an application or a portion of it from an internal, private cloud platform to an external public cloud service based on the need for additional resources.

- Cloud service composition — Creating a solution with a portion running on internal systems, and another portion delivered from the external cloud environment in which there is ongoing data exchanges and process coordination between the internal and external environments. Mashups are a form of integrated solutions where public cloud-based services are combined with internal application components to create a composite application using Web APIs and data success mechanisms (such as RSS feeds).

- Dynamic cloud execution — The most ambitious form of hybrid cloud computing combines joint security and management, cloudbursting and cloud services compositions. In this model, a solution is defined as a series of services that can run in whole or in part on an internal private cloud platform or on a number of external cloud platforms, in which the actual execution (internal and external) is dynamically determined based on changing technical (e.g., performance), financial (e.g., cost of internal versus external resources) and business (e.g., regulatory requirements) conditions.

Position and Adoption Speed Justification: Most companies will use some form of hybrid cloud computing. Early adopters are already using mashups and joint security and management approaches. Some are building simple integrated solutions and experimenting with cloudbursting. The grid computing world already supports hybrid models executing across internal and external resources, and these are increasingly being applied to cloud computing. More sophisticated, integrated solutions and dynamic execution interest users, but beyond the current state of the art.

User Advice: When using public cloud computing services, establish security, management and governance models to coordinate the use of these external services with internal applications and services. Where public cloud application services or custom applications running on public cloud infrastructures are used, establish guidelines and standards for how these elements will combine with internal systems to form a hybrid environment. Approach sophisticated integrated solutions, cloudbursting and dynamic execution cautiously, because these are the least mature and most problematic hybrid approaches. To encourage experimentation and cost savings, and to prevent inappropriately risky implementations, create guidelines/policies on the appropriate use of the different hybrid cloud models.

Business Impact: Hybrid cloud computing leads the way toward a unified cloud computing model in which there is a single "cloud" that is made up of multiple cloud platforms (internal or external) that can be used, as needed, based on changing business requirements. This ideal approach would offer the best-possible economic model and maximum agility. It also sets the stage for new ways for enterprises to work with suppliers and partners (B2B) and customers (business-to-consumer) as these constituencies also move toward a hybrid cloud computing model. In the meantime, less-ambitious hybrid cloud approaches still allow for cost optimization, flexible application deployment options, and a coordinated use of internal and external resources.

Benefit Rating: Transformational

Market Penetration: Less than 1% of target audience

Maturity: Emerging

Sample Vendors: Enomaly; HP; IBM; Rackspace; Zimory

"Big Data" and Extreme Information Processing and Management

Analysis By: Mark A. Beyer

Definition: According to Gartner research, information quantification includes issues emerging from the volume, the wide variety of information assets, and the velocity of their arrival (which includes both rapid record creation and highly variable rates of data creation, including suddenly lower velocity forcing the issue of scaling "down"). Within each variety of data types, there are significant variations and standards, creating additional complexity as well as complexity in the processing of these various resources. "Big data" is the term adopted by the market to describe extreme information management and processing issues which exceed the capability of traditional information technology along one or multiple dimensions to support the use of the information assets. Throughout 2010 and into 2011, big data has focused primarily on the volume issues of extremely large datasets generated from technology practices such as social media, operational technology, Internet logging and streaming sources. A wide array of hardware and software solutions have emerged to address the partial issue of volume. At this point, big data, or extreme information processing and management, is essentially a practice that presents new business opportunities.

Position and Adoption Speed Justification: The larger context of big data refers to the massive data creation venues in the 21st century. It challenges existing practices of selecting which data to integrate with the concept that **all** information can be integrated, and that technology should be developed to support this. As a new issue with requirements that demand an approach, the expansion of traditional boundaries will occur extremely fast because the many sources of new information assets are increasing geometrically (for example, desktops became notebooks and now tablets; portable data is everywhere and in multiple context formats), which is causing exponential increases in data volumes. Additionally, the information assets include the entire spectrum of the information content continuum, from fully undetermined structure ("unstructured") to fully documented and traditionally accessed structures ("structured"). As a result, organizations will seek to address the full spectrum of extreme information management issues, and seek this as differentiation from their competitors, so they can become leaders in their markets in the next two to five years. Big data is thus a current issue (focused on volume and including the velocity and variety of data, or, together, V3) which highlights a much larger extreme information management topic demanding almost immediate solutions. Vendors are almost universally claiming they have a big data strategy or solution. However, Gartner clients have made it clear that big data must include large volumes processed in streams and batch (not just MapReduce), an extensible services framework which can deploy processing to the data or bring data to the process, and which spans more than one variety of asset type (for example, not just tabular, or just streams or just text). Partial solutions are acceptable but should be evaluated for what they do — not the additional claims. Gartner clients are making a very large number of inquiries into this topic; however, this is only evidence of true hype in the market, and, as a result, big data will hit the Peak of Inflated Expectations sometime in 2012, then suffer a long, slow and painful maturation in the Trough of Disillusionment. Importantly, the different aspects and types of big data have been around for more than a decade — it is only recent market hype around legitimate new techniques and solutions which has created this heightened demand.

User Advice:

- Identify existing business processes which are hampered in their use of information because the volume is too large, the variety is too widespread or the velocity creates processing issues. Then identify business processes which are currently attempting to solve these issues with one-off or manual solutions.

- Review existing information assets which were previously beyond existing analytic or processing capabilities, determine if they have untapped value to the business, and make them a first or pilot target of your big data strategy. In particular, look for information use cases which combine extremely diverse information assets into analysis and data mining solutions.

- Plan on utilizing scalable information management resources, whether this is public cloud, private cloud or resource allocation (commissioning and decommissioning of infrastructure), or some other strategy. Do not forget that this is not just a storage and access issue. Complex, multi-level, highly correlated information processing will demand similar elasticity in compute resources as those in storage/persistence needs.

- Extend the metadata management strategies already in place and recognize that more is needed to enable the documentation of these information assets, their pervasiveness of use, and the fidelity or assurance of the assets, tracking how information assets relate to each other and more.

Business Impact: Big data, and addressing all the extreme aspects of 21st-century information management, permits greater analysis of all available data, detecting even the smallest details of the information corpus — a precursor to effective pattern-based strategies and the new type of applications they enable. Big data has multiple use cases. In the case of complex event processing, queries are complex with many different feeds, and the volume may be high or not high, the velocity will vary from high to low, and so on. Volume analytics using approaches such as MapReduce (the Apache Hadoop project, for example) are valid big data use cases. In addition to MapReduce approaches which access data in external Hadoop Distributed File System (HDFS) files, the business intelligence use case can utilize it in-database (for example, Aster Data and Greenplum), or as a service call managed by the database management system (IBM Big Insights, for example), or externally through third-party software (such as Cloudera or MapR). Enterprises using portals as a business delivery channel have the opportunity already to combine geospatial, demographic, economic and engagement preferences data in analyzing their operations, and/or to leverage this type of data in developing new process models. For example, supply chain situations include location tracking through route and time, which can be combined with business process tracking. Life sciences generate enormous volumes of data in clinical trials, genomic research and environmental analysis as contributing factors to health conditions. Gartner estimates that organizations which have introduced the full spectrum of extreme information management issues to their information management strategies by 2015 will begin to outperform their unprepared competitors within their industry sectors by 20% in every available financial metric.

Benefit Rating: Transformational

Market Penetration: 1% to 5% of target audience

Maturity: Emerging

Sample Vendors: Cloudera; EMC-Greenplum; IBM; MapR; Teradata-Aster Data

Recommended Reading: "'Big Data' Is Only the Beginning of Extreme Information Management"

"Hadoop and MapReduce: Big Data Analytics"

"CEO Advisory: 'Big Data' Equals Big Opportunity"

Cloud Application Development Services

Analysis By: Jim Duggan; Thomas E. Murphy

Definition: Cloud application development (AD) services are tool offerings delivered as a service and used to create custom software applications deployed on an application platform as a service (aPaaS), a cloud-enabled application platform (CEAP) or infrastructure as a service. Target applications can range from simple, situational business process management (BPM) to complex,

mission-critical, line-of-business systems. The distinguishing features of cloud AD services include awareness of, integration with and control of multiple, target cloud runtime environments. This is distinct from specialized AD services included within, which are tightly integrated and optimized for a particular aPaaS.

The development of applications targeting cloud infrastructures as a deployment medium is possible using traditional AD tools; however, these tools fail to deliver seamless awareness, integration and control of the cloud. Some cloud infrastructure vendors are offering enhancements, usually delivered as plug-ins, to popular development tools, and nearly all tool vendors are including some form of cloud-enablement in their product road maps. Cloud AD tools can be categorized along two axes: target audience and runtime environment.

Some applications targeted for aPaaSs or CEAPs may be authored by non-IT staff, requiring a programming environment that is amenable to use by citizen developers. For example, platforms like WaveMaker Cloud Edition, Rollbase, TrackVia and dbFlex provide fourth-generation-language (4GL)-style features in a "what you see is what you get" (WYSIWYG) graphical editing environment. To further simplify the development of end-user-created applications, these programming environments are delivered as software-as-a-service (SaaS)-style offerings, avoiding the need for installation and maintenance of a traditional, local integrated development environment (IDE).

Mission-critical, line-of-business applications generally require more customization than is possible using simple WYSIWYG tools. Advanced customization is supported by the most flexible aPaaS and CEAP platforms, which may use a proprietary or standard programming language, coupled with rich, proprietary application programming interfaces (APIs) or libraries, and a service-oriented-architecture (SOA)-style architecture. For example, Google App Engine (GAE) enables developers to write code in a reduced version of Java using a plug-in for the popular Eclipse IDE, but provides no graphical user interface for development. Force.com also offers a downloadable plug-in for Eclipse, which enables both WYSIWYG editing and customization using the proprietary Apex language.

Applications that will be deployed into IaaS offerings, such as Amazon Elastic Compute Cloud (Amazon EC2) or Rackspace Cloud Servers, require a different approach than those targeted at aPaaS deployment. IaaS enables a high degree of architectural flexibility, but necessitates more-customized approaches to development that require professional programmers to implement successfully. Developers constructing applications for IaaS choose from a variety of middleware components, including application servers, Web servers, and databases, and they need development tools to match them. To that end, companies like CloudOne, which offers on-demand versions of popular IBM Rational products, and services like Cloud9 IDE, which offers a fully browser-based IDE, exist to support them.

A range of different delivery styles are in this category, from tools that are adapted for cloud hosting, to AD tool services specifically designed for cloud platforms, to tools that are installed locally, but leverage the cloud for pieces like Soasta (and others), using EC2 to generate test load.

Position and Adoption Speed Justification: The growth in interest in cloud AD tooling directly tracks the growth in adoption of aPaaS and IaaS offerings as runtime environments for custom software applications. New types of cloud AD tools are a requirement for working with an aPaaS. Developing applications that target IaaS as a deployment medium is possible using traditional AD tools; however, these tools fail to deliver seamless awareness, integration and control of the cloud service. Some IaaS vendors are offering enhancements, usually delivered as plug-ins, to popular development tools, and nearly all tool vendors are including some form of cloud-enablement in their product road maps.

The evolution of continuous delivery and the rise of new cloud platforms will drive new combinations or stratifications in the packaging of development execution and application life cycle management (ALM) functionality. ALM plan/measure/manage/report capabilities may or may not be packaged with an IDE, functional or performance test execution tools, requirements elicitation, design or any other execution tool. When ALM capabilities; development capabilities like design, code and test; and runtime facilities are packaged together, those products will become complete application platforms and move to the aPaas category.

For the present, specialized services for a single AD practice or stage (planning, requirements gathering, analysis, design, coding test and build) will continue to exist. These persist partly because, in development, there are tasks done by individuals with different skills and needs, and partly because when we try to mash everything together, we lose as much to complexity as we gain through unification. Furthermore, sunk investments in "one-stop shop" tooling often constrain the adoption of new tools to isolated tasks. This last effect will cause the adoption of broader offerings to lag behind some of the most useful point solutions.

The market continues to see growth in SaaS offerings and tools that utilize cloud resources to dynamically provision resources. The AD offerings of both IBM and Microsoft are being provisioned by partners. In addition, both of these vendors are moving toward cloud-based models of their own. HP is beginning its third year of offering cloud-sourced load testing.

Involving cloud AD tools provides a variety of benefits:

- Dynamic provisioning and self-service for the creation of development and test labs.
- Broader variety of development platforms provided at lower administrative costs.
- More-flexible provisioning of tool types. Adjust the mix of licenses to suit the stage of development.
- Greater capacity in load tests, together with lower costs, and, for external-facing Web applications, more-realistic tests.
- Testing on virtualized infrastructure can provide richer information to developers in defect reports.

This space is currently dominated by small innovators. As such, these solutions are best for organizations that have spikes in need for tools and labs, or those with very dynamic team structures. Broader implementations will accelerate once offerings from larger vendors like HP, IBM and Google are actively promoted.

User Advice: Users need to create a road map for their cloud development needs to clarify where cloud AD services will be used instead of ALMaaS or aPaaS offerings. Tools used for creating cloud applications — especially for those targeting an aPaaS or a CEAP platform — must reflect the reality that applications designed for the cloud are different from their traditional on-premises counterparts. The spectrum of available cloud AD tools is composed of a wide range of capabilities and varying degrees of user-friendliness. When evaluating the capabilities of a cloud platform, include the toolset used for that platform as part of your evaluation, and do not assume that existing AD tools will be sufficient to build applications for this new medium. For example, citizen developers working outside the IT department will be better-served by a visual development environment, but they will only be able to build somewhat limited applications, as compared with professional programmers who need advanced tooling (such as Eclipse with an AWS or a GAE plug-in) to construct advanced applications.

Finally, recognize that many cloud AD services are proprietary to the cloud system infrastructure, aPaaS or CEAP offering chosen as the deployment medium; once you have invested in custom

applications "in the cloud," there's no clean migration path to another vendor's offering. More-general services are appearing like the Eclipse Foundation Orion, CodeRun's Studio or Code9's IDE.

Business Impact: The best cloud AD tools will be transformational in nature, but they depend on the maturation of aPaaSs, CEAPs and cloud system infrastructures; the most powerful cloud AD approaches are likely several years off. This next generation of cloud AD tooling will enable business users to create custom software applications that once required the assistance of professional programmers to realize. Professional programmers will still be critical of more-complicated applications, but they will be able to deliver them faster, less expensively and without the need for expensive infrastructure.

Benefit Rating: High

Market Penetration: 1% to 5% of target audience

Maturity: Emerging

Sample Vendors: Cloud9; CloudBees; Eclipse Foundation; Google; HP; Microsoft; Neustar; Qrimp; Rollbase; salesforce.com; Soasta; TrackVia

Recommended Reading: "Platform as a Service: Definition, Taxonomy and Vendor Landscape, 2011"

"Test and Quality Management: The First Frontier for Private Cloud"

"Productivity vs. Control: Cloud Application Platforms Must Split to Win"

"Client-Cloud Applications: The Rebirth of Client/Server Architecture"

"Citizen Developers Are Poised to Grow"

"APaaS: A Step to a 'Killer App' for Cloud Computing?"

"On Apples, Oranges, Amazon and Google"

Cloud Parallel Processing

Analysis By: Daniel Sholler; Tom Austin

Definition: Parallel processing techniques are algorithmic and code-structuring methods that enable the parallelization of program functions. These techniques have been automated at the processor level, but the coming availability of large-scale grid systems through the adoption of cloud architectures creates an opportunity to apply these techniques to application system design. Approaches to parallelism at this level are becoming necessary for applications to leverage the massive amounts of data available from the Web, social networks and large-scale systems.

Position and Adoption Speed Justification: While the concepts of parallel processing have been studied for years, the reality has not affected most developers. Parallelism techniques have been used to improve the performance of system software, databases and other specialty programs, but the typical enterprise developer has been shielded from the need to understand how to structure programs for parallelization through increasingly sophisticated middleware and other system components. Most of the progress of parallelization has focused on moving serial workloads onto multicore and multithreaded processes. However, as cloud computing concepts become a reality, the need for some systems to operate in a highly dynamic grid environment will require these techniques to be incorporated into some mainstream programs. This is particularly

true for analytical tasks that incorporate large datasets. The popularity of the Hadoop implementation of the map/reduce technique is a good example of this trend. While vendors will continue to package the middleware and extreme transaction processing components that will simplify parallelization, the application developer can no longer ignore this as a design point for applications that leverage huge sensor networks or Internet-generated data.

Currently, few developers are trained in these tools and techniques, and there will likely be a flurry of interest as large-scale grids (public and private clouds) become readily available. While only a small percentage of applications will require the use of these techniques, these applications will be high-value, Web-facing systems. During the next five years, knowledge and skills will penetrate the market; by 2015, we expect that a sufficient percentage of developers will be able to use these techniques to build the required applications. Systems supporting parallelization techniques exist, but will develop in range and sophistication during this evolution. For most purposes, developers will rely on improvements in automated parallelization techniques, but a small number of applications that leverage huge (usually Web-based) information sources will require these manual or assisted techniques. Generally, the applications will be those in which the cloud infrastructure enables a truly differentiated user experience. The applications are also likely to be linked to contextual computing, where (for example) the relationships among millions of locations must be recalculated every time one changes, or for Pattern-Based Strategy processes, where systems will sift masses of information for relevant patterns.

User Advice: Users should determine the timeline for using cloud- and grid-based computing, and should ensure that, as these infrastructures become available, there are appropriate skills in the developer community to fully utilize their potential.

Business Impact: Parallelization will make economically viable the implementation of many algorithms and workloads not currently feasible. The types of problems, the granularity of the analysis (and simulation), the scope of data and the speed with which it can be accomplished will all be affected. The widespread use of these techniques will enable many organizations to produce truly global-class applications, and business data providers will be the most aggressive in developing systems that use these techniques. In many cases, parallelization can create a qualitative shift in user experience, which results from dramatically speeding certain processes. By enabling decisions in seconds that currently take days and weeks, businesses create new opportunities for managing their processes, and for increasing the accuracy and cycle times of decision making.

Benefit Rating: High

Market Penetration: Less than 1% of target audience

Maturity: Emerging

Recommended Reading: "Top 10 Technology Trends Impacting Information Infrastructure, 2011"

"The Future of Server Platforms"

"Key Issues for Web and Cloud Application Development, 2010"

At the Peak

Cloud-Enabled BPM Platforms

Analysis By: Michele Cantara

Definition: BPM technologies help manage the work of a single organization or multiple organizations. Business processes are the actual work of a single organization or multiple organizations. Business processes include formally defined activities as well as informal work practices. Additionally, business processes may involve human and application activities, and they may be structured or unstructured. A cloud-enabled BPM (CE-BPM) platform is a platform for managing business processes in a private or public cloud.

CE-BPM is often confused with bpmPaaS. bpmPaaS refers to the delivery of BPM technology functionality as a service by a cloud service provider, while CE-BPM refers to a cloud-enabled BPM technology product. CE-BPMs are typically purchased by enterprises to run shared business process service centers in a private cloud. A vendor may use the exact same technology in its bpmPaaS and its CE-BPM. The only difference is in the delivery model. bpmPaaS is delivered as a service. CE-BPM is delivered as a product and then used to provide a public or private cloud service by an ESP or an internal IT organization. ESPs use CE-BPM as the underlying application infrastructure to deliver SaaS and business process utilities (BPUs) in the public cloud, as well as cloud-enabled outsourcing in community clouds. Providers of bpmPaaS may use their own or a third-party CE-BPM platform.

A CE-BPM exhibits cloud-enabled application platform (CEAP) capabilities (see "Gartner Reference Architecture for Cloud-Enabled Application Platforms"). A CE-BPM must include at least one of the following BPM runtime capabilities: flow management, rule management, optimization and simulation, or BAM. It may optionally include a variety of design-time BPM capabilities, such as business process modeling and automated business process discovery.

A CE-BPM may be a lightweight platform, such as Nimbus Control, for mapping strategic objectives to KPIs, modeling processes, guiding workflows and integrating to applications. Most CE-BPMs are full-fledged BPMS products, such as the Appian BPM Suite, Cordys Business Operations Platform or Pegasystems Pega BPM.

Position and Adoption Speed Justification: More than two-thirds of vendors that offer cloud-enabled BPM offer CE-BPM, and the majority of these offer both CE-BPM and bpmPaaS. Cloud-enabled BPM platforms are positioned at pre-peak with a time to plateau of two to five years. The 2010 Gartner BPM end-user adoption survey showed that organizations that were doing BPM included both bpmPaaS and CE-BPM in the top five BPM technology capabilities critical to the success of their BPM efforts. Furthermore, the importance of bpmPaaS/CE-BPM was significantly higher among respondents in Latin America and Asia/Pacific, and these regions are also the fastest-growing regions for BPMS software (see "The State of the BPM Platform Cloud Market, 2011" and "Forecast: Enterprise Software Markets, Worldwide, 2008-2015, 1Q11 Update").

The increase of CE-BPM hype will follow the market's focus on private PaaS. CE-BPM adoption has been spurred by a move to standardize processes across multiple business units and regions and to deliver these as shared business services in a private cloud. CE-BPM helps organizations better understand the business impact of process standardization and where process variation is warranted. The cloud-enabling features in more-advanced CE-BPMs let an IT department manage a standardized process and at the same time treat each region, business unit or partner as a separate tenant with their own process variations. For example, Wesfarmers, an insurance provider and retailer in Australia, built a new process-based solution on a CE-BPM in a private cloud to support both direct online sales and online sales of white-labeled products by its retail partners (www.techworld.com.au/article/348073/wesfarmers_insurance_modernises_private_cloud/). UPS is another example (www.kmworld.com/Articles/Editorial/Feature/2010-Promise-and-Reality-award-winners-72955.aspx).

The most prevalent use of CE-BPM is by service providers, acting as cloud service brokerages (CSBs) that use the platform as a foundation for SaaS, BPU and BPO offerings. End users may

not even be aware of the CE-BPM. They typically view CE-BPM as a service feature that enables more visibility and flexibility in the SaaS BPU or BPO services they consume. Examples of SaaS providers that embed CE-BPM in their offerings include Anacomp, Enkata, ICCM and L@W. ADP is an example of a BPU provider that uses CE-BPM. Some of the outsourcers that embed a CE-BPM within selected application outsourcing or BPO offerings include Capgemini, CrimsonLogic, CSC, DST Systems, Eucon, First Data International, Genpact, i3, KMD, ISCorp, Logica, Maximus, Mercer, MphasiS, NCCW, iGATE Patni, SOS International, Resolve, SunGard, Target Group, Transcom, and Xchanging (see "Emerging Service Analysis: BPO + BPMS = Vendor Agility and Buyer Alignment" and "BPM Delivers Agility to Buyers and Provides Leverageable Service Delivery Models for Service Providers").

User Advice:

- Consider CE-BPM as one of the options to help you standardize processes and support them in a shared-service center.
- Also consider CE-BPM to provide end-to-end visibility into processes that span multiple organizations in a partner/supplier network.
- When evaluating CE-BPM for your own use, scrutinize the multitenancy and metering-by-use capabilities of the platform.
- If you expect to make changes to the business processes you outsource or need to outsource a differentiating process, then ensure that your BPO provider uses a bpmPaaS as a delivery platform for the services you consume, and that your CE-BPM can integrate to it.
- In these cases, use BPO providers that have based their services on a commercially available CE-BPM rather than a homegrown one. Commercial software vendor investment in CE-BPM products is generally 15% to 20% net revenue, and most BPO providers lack the margins to keep pace with that investment. As a result, functionality in a homegrown, noncommercial CE-BPM is less likely to be market leading and is less likely to support the "round-tripping" capabilities needed for rapid or ad hoc process change. Homegrown CE-BPM may also result in more "vendor lock-in," and it may be more difficult to move your process assets to another BPO provider.
- Establish a more transformative sourcing relationship with your outsourcer to deal with ad hoc process changes and to obtain more visibility in your outsourced processes. Use "Eight Tactics for Managing the Tempo of Process Change in BPO and Cloud-Enabled Business Process Services"as a guide.
- Many CE-BPM products have evolved from a single enterprise BPMS product. They were not purpose-built with the multitenancy, elasticity and metering-by-use features of cloud computing in mind, and these features may be limited. These limited cloud capabilities are not likely to matter as much to consumers of BPU, BPO or SaaS, but they will matter to service providers. ESPs and internal data centers that are providing private cloud services need to seek out CE-BPM products with advanced cloud capabilities to ensure appropriate levels of agility and efficiencies.

Business Impact: CE-BPM will have a moderate impact on end-user organizations that use CE-BPM for private cloud shared-service centers. These customers will gain better insight into the work of their organization, cut costs and improve operational resilience. The increase of CE-BPM hype will follow the market's focus on private PaaS.

CE-BPM will have a more transformational impact on outsourcers and on the end users that consume their services. In "Steering Your Business Through the IT Services and Outsourcing Revolution,"Gartner outlined two opposing trends that challenge outsourcers. First, current service providers' offerings are neither agile nor flexible enough to enable buyers' organizations to grow and change in response to triggers in the business environment. Second, service providers are challenged to deliver the requisite business agility that clients need, and, at the same time, sustain reasonable margins in their service businesses. Service buyers are price-sensitive, but they want more agility from service providers. However, ESPs' current practices limit their ability to provide agile solutions profitably. CE-BPM is a foundation for building and optimizing next-generation solutions that deliver agility to buyers and underpin a leveraged versus labor-based business model for ESPs. At present, most outsourcers are focused on using CE-BPM for platform BPO and improving their service margins. Nevertheless, a few visionary BPO providers such as Maximus (www.gartner.com/it/page.jsp?id=1588516) and i3 are actively using the business agility capabilities of CE-BPM to continuously improve their customers' processes, and they are shifting their sourcing relationships from cost-based efficiency deals to higher-value transformation relationships. Gartner worldwide BPM adoption surveys conducted in 2008 and 2010 show a marked shift in the pace of process change from 2008, where the majority of process changes in business process projects took place several times a year, to 2010, where a majority of process changes occurred on an ad hoc basis. This shift will drive more outsourcers to embed CE-BPM in their offerings and evolve into cloud service brokerages.

Benefit Rating: High

Market Penetration: 1% to 5% of target audience

Maturity: Emerging

Sample Vendors: Active Endpoints; Adeptia; Appian; AppPoint; Be Informed; Cordys; Fujitsu; Intalio; Invensys; Nimbus; Pegasystems; Progress (Savvion); Singularity

Recommended Reading: "The State of the BPM Platform Cloud Market, 2011"

"Platform as a Service: Definition, Taxonomy and Vendor Landscape, 2011"

"Cloud-Enabled Outsourcing: Use BPM to Ensure Process Agility in Alternative BPO Service Delivery Models"

Database Platform as a Service (dbPaaS)

Analysis By: Donald Feinberg

Definition: Database platforms as a service (dbPaaS) is engineered to run as a scalable, "elastic" service that is available on a cloud infrastructure. They are available as a one-to-many service, not necessarily relational and offer some degree of self-service. For example, Microsoft's SQL Azure and database.com from salesforce.com are fully relational dbPaaS, while Amazon's SimpleDB and Google's BigTable are not relational and have different persistence models. Provided in a shared, multi-tenant environment with automatic and elastic scaling, dbPaaS is for use in simple through to complex transactions.

Database management systems (DBMSs) and data stores that run on cloud infrastructure, but are not purpose-built as a cloud service are excluded from dbPaaS. Most of the currently available DBMS engines will run on cloud infrastructure (for example, Amazon's EC2), but are not considered dbPaaS as they are not made available as one-to-many, multi-tenant, elastic services, which includes Amazon's RDS, MySQL or Oracle, IBM's DB2, MySQL, Oracle and many others.

Position and Adoption Speed Justification: All the currently available dbPaaS offerings are relatively new. SQL Azure began full production at the beginning of 2010 and still has some size limitations; Microsoft plans to reduce and eventually lift, these restrictions. The only other fully-relational dbPaaS with atomicity, consistency, isolation, durability (ACID) properties is database.com.

There are several non-ACID dbPaaS offerings that allow for eventual consistency; restricting them to less complex and normally single-user, transactions — especially where data is used by only one transaction at a time and locking is not required. non-ACID dbPaaS is becoming more widely used for Web 2.0 development projects, where sharing of data among multiple tenants in the cloud is not a requirement of the application. This lack of offerings in the market is due to the difficulties of creating multi-tenant, automatically elastic database services. Maturity will begin for dbPaaS at a fast pace as more offerings become available.

Currently, dbPaaS is used primarily for development and testing of applications — where database sizes are small and issues of security and sharing with multiple users are not a concern. Recently, we have seen examples of applications using dbPaaS in production applications deployed in the cloud on SQL Azure.

Many Web 2.0 applications may be experimenting with some of these services, but most still rely on non-cloud-based DBMS implementations. One exception is where all the data already exists in the cloud and it is desirable to have the application there, with the data. One big advantage of dbPaaS is that it doesn't use license-based pricing, it uses either elastic pricing (the more you use, the more you pay; the less you use, the less you pay), or fixed subscription pricing (a flat price per user).

The rate of adoption of dbPaaS will depend on its increasing maturity and acceptance of cloud system infrastructure in general and on the maturation of dbPaaS offerings. Adoption will also depend on the usage model and whether the relaxed consistency model can be used by an application. Gartner believes that there will be additional types of dbPaaS available, as true cloud services during the next few years, in line with what Microsoft has done with SQL Azure.

User Advice: Advice for mainstream users in the short term (the next two years):

- Use of dbPaaS should be restricted to development and test systems, single-user systems, or those requiring file storage in the cloud with one writer and multiple readers.
- Consider limited use of dbPaaS for hosting Web-specific content.
- Use caution with dbPaaS, as there are still issues with security and reliability and with the non-relational DBMSs there are issues with concurrent user control.
- Exercise care in systems with high amounts of data transfer; most cloud infrastructure vendors charge for movement of data in and out of the cloud.
- Latency is another data transfer issue — the time available to transfer large amounts of data to the cloud (for example, to support a data warehouse in the cloud) can be restricted. For these reasons, initial usage for development systems (with minimal data transfer) can be beneficial before moving the systems in-house after development.

Business Impact: Initially, dbPaaS will have an impact for software vendors (especially smaller ones) requiring a less-expensive platform for development. As dbPaaS gains maturity (especially in scalability, reliability and security), dbPaaS implementations used for short-term projects (such as small departmental applications and rapid development platforms) will show some marked cost reductions, compared with implementations within IT departments. This cost saving will be primarily based on the ability to set up a dbPaaS environment without the use of expensive IT

personnel. The speed of setup will be a primary driver to rapid deployment of systems — without the usual requirements and planning necessary for IT projects within the IT department. This will also reduce the necessity for IT staff to respond to short-notice and short-duration projects, reducing overall IT costs.

As dbPaaS offerings mature during the next two to five years, it will be possible for an organization to host its entire DBMS infrastructure as dbPaaS, with potential reductions in the cost of servers, storage, DBMS licenses, maintenance and support, storage management and database administration.

Benefit Rating: Moderate

Market Penetration: Less than 1% of target audience

Maturity: Emerging

Sample Vendors: Amazon; Google; Microsoft; salesforce.com

Recommended Reading: "Platform as a Service: Definition, Taxonomy and Vendor Landscape, 2011"

"Global-Class Persistence for Cloud-Based Web Applications"

"Salesforce.com Unveils Potentially Disruptive Cloud DBMS, database.com"

Browser Client OS

Analysis By: Annette Jump; Michael A. Silver

Definition: The increasing focus on Web-based applications is creating interest in a new form of simpler, pared-down operating systems (OSs) targeted at supporting just a browser and connection to the Internet. In browser client OSs, all applications are Web-based, and many traditional PC OS functions are missing. However, they may provide simpler and lower-cost computing, but they will rely on browser applications being available.

Position and Adoption Speed Justification: Ongoing complexity, security and support issues with client OSs, combined with the success of devices with more locked, managed operating systems, such as iOS and Android, have proved the desire for smaller, simpler OSs that can run on those devices. Additionally, the hardware configurations of those devices are significantly lower compared with traditional PCs, and lighter-weight, alternative OSs, such as Chrome OS and Android, may provide a better user experience. Further, Web-based applications continue to become more pervasive (see "Windows Applications Will Be Critical Through the Planning Horizon, but Lose Majority in 2012,"), so a slimmer client OS will at some point be sufficient for many users to run most of the applications they need.

The application ecosystem and application stores will be important in determining if a browser OS will be important or compete successfully with other thinner/lighter models. The development of the browser OS is another symptom of the general trend toward commoditizing the footprint for applications on client devices. Microsoft's ARM-based version of Windows 8, expected in 2H12, would seem to be Microsoft's answer to keep Windows relevant on alternative devices.

Google's Chrome OS represents the first example of a new class of OS designed to hide all the complexity and configurability of the base OS, resulting in a lightweight operating environment. While Chrome OS is based on Linux, it has been reworked to support applications based only on Web standards. Momentum for browser client OSs slowed down in 2010, as Google has been postponing the introduction of Chrome OS and OEMs' excitement and support has weakened. However, in May 2011, Google announced launch plans for two Chromebooks — Web books

running Chrome OS, which are available in selected countries from the middle of June 2011. The devices will be sold by Samsung and Acer, which are eager to explore and test the user interest for those new devices.

Generally, browser client OSs tend to suffer when compared directly with traditional PC OSs because they lack much of the open extensibility, broad device support or the ability to support local rich applications. As such, consumers may be disappointed if the marketing doesn't accurately reflect the appropriate usage models. However, with the emergence of richer Web development environments (for example, Flash, Silverlight and, most notably, HTML5), much-richer applications can be created to satisfy user needs, and within the next three to five years, it is likely that users will adapt to the new paradigms.

User Advice: For enterprise users:

- Continue to focus on browser-based application development.
- Understand the limitations of these OSs and their associated devices, and plan potential use accordingly.

For device vendors:

- Use browser client OSs as an alternative OS to Windows on Web books only. Current browser client OSs are not suitable for the media tablet category because of a lack of multitouch capability. In the medium term, when the touch support will be added, consider it for media tablets, too.
- In the short to medium term, offer devices with browser client OSs only for consumers in mature markets, and be conservative in your forecasts.

Business Impact: The browser client OS will be relevant first in the consumer client market, in combination with a new class of devices, such as Web books and media tablets. The success of a browser client OS as an enterprise platform will likely take much longer because of application compatibility issues. Enterprises are concerned that these OSs are too immature and lack sufficient compatibility with current application requirements. Additionally, it is also unclear how OS vendors may evolve these OSs, and the potential exists for interoperability issues. To succeed, browser client OSs must gain a significant market share in consumer Web books or media tablets before they will even be considered for other PC computing devices. Therefore, it is unlikely that any browser client OS will gain any meaningful market share among enterprise PC users as a PC replacement in the next three to five years. However, if browser client OSs are successful on Web books or media tablets, some enterprise users might bring them into the enterprise as an additional client device when they travel, and they might want to use them to access their corporate email, calendar and contacts.

Benefit Rating: Moderate

Market Penetration: Less than 1% of target audience

Maturity: Emerging

Sample Vendors: Google

Cloud Collaboration Services

Analysis By: Jeffrey Mann

Definition: Many collaboration services are particularly suited for cloud-based deployment. These services include email, instant messaging, document repositories, team work spaces,

discussion forums, wikis, blogs, Web/audio/video conferencing and many types of social software. They are generally well understood and reasonably mature technologies, require less integration and tailoring than many other software categories, and are most valuable when deployed widely.

While some vendors provide narrow offerings that concentrate on one or two of these technologies, the larger trend is to provide a bundled suite of capabilities across the spectrum of collaboration services. Megavendors like Microsoft (with its Office 365 offering), Google (with Google Apps) and IBM (with LotusLive) provide these cloud-based collaboration suites. At the other end of the market, many social software startups are launching specific collaboration services based on cloud deployment models.

Position and Adoption Speed Justification: Collaboration services in some form have been available from the cloud for some time. In fact, Web conferencing was the first technology to gain widespread acceptance using the cloud deployment model. However, reports of sporadic outages and privacy, confidentiality and security concerns with other cloud collaboration services have limited widespread adoption. Gartner believes many of these attitudes are poised to change. Cloud email in the consumer market from services like Yahoo, Hotmail and Gmail has convinced many that the model works, even though business requirements for availability, confidentiality, archiving, security and compliance are higher for enterprise users. While cloud email is usually the primary driver for cloud collaboration services, usage will expand into other areas as the services prove their worth.

The highly anticipated release of Microsoft Office 365 in June 2011 is driving a great deal of attention to cloud collaboration services. It is likely to serve as a bellwether for enterprise cloud collaboration services in general. If it succeeds, the space as a whole will grow quickly. If it fails, it is likely to set back the concept of cloud collaboration services for some time.

User Advice: Unless the cloud model is unappealing or impossible due to organizational culture or compliance, privacy and security concerns, most organizations should be considering some form of cloud collaboration services, especially if they are facing some change in their current collaboration infrastructure due to major upgrades or vendor swaps. Those that are happy with their current infrastructure do not need to abandon what they have to move to the cloud, however. Cloud collaboration services are not suited to all organizations. Interenterprise collaboration opportunities not easily addressed with current internal, on-premises technology should be evaluated as an adjunct to current investments.

Business Impact: While the expected growth in movement to the cloud for collaboration services will have wide implications for the industry as a whole, it will have moderate to low impact on individual organizations. They will experience lower costs and more flexibility, as fewer resources will be needed to provide collaboration services to employees. While these benefits will allow companies to provide more employees with access to these services, they will not suddenly enable large shifts in business practices or new capabilities. The potential benefits are real, however. For those organizations comfortable with the idea of cloud deployment, collaboration services will be a natural contender.

Benefit Rating: Moderate

Market Penetration: 1% to 5% of target audience

Maturity: Adolescent

Sample Vendors: Google; IBM; Microsoft; salesforce.com; Yammer; Zoho

Recommended Reading: "The Cloud E-Mail and Collaboration Services Market"

"Exploit the Differing Business Models of Google and Microsoft for Cloud Office, E-Mail and Collaboration Services"

Elasticity

Analysis By: Daryl C. Plummer

Definition: In the service provider view, cloud service elasticity is the ability to increase or decrease the amount of system capacity (for example, CPU, storage, memory and input/output [I/O] bandwidth) that is available for a given cloud service on demand in an automated fashion. This stops the need to overprovision for peak demand. The degree of automation of elasticity is determined by the service provider. Gartner covers an area called real-time infrastructure (RTI) — that is, the architecture behind the elasticity. It focuses attention on the automated nature of the changes to capacity, but also to services — dynamic scale up/down, in/out based on incoming demand, schedules and other planned events. Manual means of adding capacity are generally discouraged, but are more commonplace in 2011 (as they were in 2010), although writing the automation necessary to scale services is being adopted in an ad hoc way for services that have high variability in demand.

The consumer view does not require the "terminology of elasticity" to be used. Consumers simply want the appropriate resources available to them while they are working. For cloud service consumers, the difference between traditional scalability and elasticity is the "scale down" aspect — the consumer of elastic services does not pay for a fixed portion of overall capacity. The consumer pays for only what is used, up to the limits that were previously agreed on with the provider. Payment, price or cost must also be metered and billed in a way that matches the elasticity of the service.

Position and Adoption Speed Justification: Elasticity is one of the more hyped aspects of cloud computing. It is rising rapidly toward the Peak of Inflated Expectations. There is little understanding of what elasticity truly is in the market, and even less of how it can be achieved consistently and automatically. However, elasticity techniques and products have existed for some time, and all they need to grow to maturity in the application of these technologies and techniques in a more automated fashion. As elasticity is automated, the ability to take advantage of shared pools of resources grows. The real barrier to elasticity is lack of modeling of the service architecture, inclusive of thresholds for service scaling, and methods to do it, which are implemented at runtime.

In 2011, elasticity is a part of almost every cloud conversation, yet its use is still relatively misunderstood. Its use, however, is increasingly delivered as part of the value proposition of notable service offerings in the cloud, but this often means "the perception of limitless capacity," not autoscaling based on demand. Thus, elasticity is still growing in hype, because cloud consumers feel it is one of the critical aspects of cloud computing.

User Advice: Companies seeking to use cloud computing should include elasticity as a critical element of capacity planning and cloud service pricing. Use elasticity to reduce the amount of overall capacity you plan to use. Supplement your capacity using elastic cloud services. Use elasticity in any formula for optimizing cloud computing costs. Treat elasticity as an expected feature of cloud services embedded in the offering. Understand the time frame it takes a cloud provider to be truly elastic (automated, it should take seconds; manual, it may take hours).

Business Impact: Elasticity is an inherent trait of shared pools of resources, and it refers to tailoring system capacity on demand to its use (see "Cloud Services Elasticity Is About Capacity, Not Just Load"). The ability to add capacity and release is necessary for the economics of cloud computing's usage-based models to work. Without this capability, it becomes difficult to enable two key parts of the model:

- **Pay for only what you use:** Consumers of cloud services must be aligned with the amount of capacity they use. A service provider that wishes to deliver a certain quality of service can throttle a system up or down by using scaling policies and elasticity engines liberally.
- **Economies of scale through sharing:** In an environment where capacity is large or underused, it is possible to share that capacity among unrelated users with the goal of reducing costs for all through sharing. When one segment of capacity is left unused, it can be allocated to any user requiring space. Doing this on a fixed-capacity model, or a model in which capacity is always increased, does not allow a service provider to respond as effectively to market conditions or bad economic conditions.

To be sure, no amount of elasticity can prevent a service provider from going out of business. The fact that service providers charge for usage requires them to ensure that they can still make a profit — even if demand/usage is down for long periods of time. As a result, we expect service providers to charge extra for elasticity features, thus enabling them to recoup some of their fixed costs in down periods. Elasticity benefits the consumers and, therefore, is worth any reasonable incremental price they may pay. However, the use of an elastic approach provides more flexible options for determining how much capacity should be purchased, as well as when and how that capacity will be allocated from hour to hour. Finally, because the largest cloud computing providers are able to realize massive economies of scale, the cost to offer a marginal increase in capacity through an elastic pricing model is negligible, when compared with the provider's overall fixed costs. For this reason, providers are likely to encourage high consumption of computing resources by offering low marginal prices.

Benefit Rating: High

Market Penetration: 5% to 20% of target audience

Maturity: Adolescent

Sample Vendors: Amazon; Enomaly; salesforce.com; VMware

Recommended Reading: "Scalability, Elasticity and Multitenancy on the Road to Cloud Services"

"Three Levels of Elasticity for Cloud Computing Expand Provider Options"

Multitenancy

Analysis By: Yefim V. Natis

Definition: Multitenancy is a reference to the mode of operation of software where multiple independent instances of one or multiple applications operate in a shared environment. The instances (tenants) are logically isolated, but physically integrated. The degree of logical isolation must be complete, but the degree of physical integration will vary. The more physical integration, the harder it is to preserve the logical isolation.

The tenants (application instances) can be representations of organizations that obtained access to the multitenant application (this is the scenario of an ISV offering services of an application to multiple customer organizations). The tenants may also be multiple applications competing for shared underlying resources (this is the scenario of a private or public cloud where multiple applications are offered in a common cloud environment).

Multitenancy can be implemented at different layers of the technology stack of a multitenant application service. "Gartner Reference Architecture for Multitenancy" identifies seven different

models of multitenancy, starting with shared-nothing multitenancy through shared-hardware, shared-OS, shared-database, shared-container, shared-everything and custom multitenancy. Generally, the higher the level in the stack where multitenancy is implemented the more elastic the environment, and the greater the density of packing the workload into the available underlying resources. Generally, the lower the level in the stack, the more the environment is backward-compatible with the precloud applications.

Multitenancy may also be nested. This occurs when a provider hosts tenants that are ISVs and have their own tenants (subtenants of the cloud provider). A well-designed multitenancy platform would have provisions to support subtenants as well as tenants.

Position and Adoption Speed Justification: The notion of multitenancy is well-popularized with the software designers and buyers. However, the essential trade-offs between different models of multitenancy often are not understood or appreciated. Lured by the backward-compatibility of the lower levels of multitenancy, many organizations and vendors choose shared-hardware or even shared-nothing offerings. As the major enterprise application infrastructure vendors (IBM, VMware, Microsoft and Red Hat) promote the shared-hardware or shared-OS cloud infrastructure offerings — enterprise users are likely to follow. These offerings allow rapid transition of existing workloads and skills to cloud computing, but will lead to disappointment over time, as the full value of elastic cloud multitenancy requires the use of the advanced higher-level implementations of multitenancy. Shared-hardware platforms cannot match the parallelism and elastic agility of shared-everything platforms. Native cloud applications, especially for social enterprise operations, gaming and global e-commerce — will push the high-end (and typically discontinuous) forms of multitenancy to the peak of hype and inflated expectations and on to the disillusionment and finally — the Plateau of Productivity.

User Advice:

- Shared-hardware, shared-OS and shared-database multitenancy can be compatible with precloud application infrastructure and applications. Choose these offerings to migrate the existing workloads and skills to cloud platforms with reduced costs.
- Invest in some initiatives using shared-container or shared-everything multitenancy to gain experience with the full-function multitenant cloud environment and its software design practices.
- When looking for a PaaS offering, give preference to providers that offer or plan to offer multiple types of multitenancy to cover the different needs of varied current and future projects.
- When looking for a SaaS offering, give preference to providers using the high-end, high-density forms of multitenancy, expecting higher degrees of elasticity, efficiency and, often, productivity as well. These platforms are also immediately available for cloud-specific innovation, because they will support the fine-grained resource management required for optimal parallel multitenant execution.

Although the "higher" forms of multitenancy generally offer a more complete cloud experience, the right choice of architecture for any particular project depends on the circumstances of the project: regulatory constraints, available skills, required backward-compatibility, required levels of productivity, manageability and others. In other words, more cloudiness is not always better.

Business Impact: Competent use of multitenancy reduces the costs of IT operations for providers of cloud services by packing workloads with higher density. It also decreases the levels of overcapacity in advanced private cloud data centers that establish private multitenant PaaS. Multitenant application environments are more agile in their utilization of computing power.

Overall, well-applied multitenancy reduces the cost of computing, and increases the agility with which applications respond to changing levels of demand.

Benefit Rating: High

Market Penetration: 1% to 5% of target audience

Maturity: Adolescent

Sample Vendors: Appistry; CloudBees; Cordys; GigaSpaces Technologies; IS Tools; LongJump; maatG; OrangeScape; Red Hat; Rollbase; salesforce.com; VMware; Wolf

Recommended Reading: "Gartner Reference Architecture for Multitenancy"

"Platform as a Service: Definition, Taxonomy and Vendor Landscape, 2011"

"Scalability, Elasticity and Multitenancy on the Road to Cloud Services"

Application PaaS (aPaaS)

Analysis By: Yefim V. Natis

Definition: Cloud application platform as a service (aPaaS) is a cloud service that offers development and deployment environments for application services. It's an extended application server "in the sky." An aPaaS is a general-purpose platform for building software-as-a-service (SaaS)-style applications. It implements such characteristics of cloud computing as multitenancy, elastic scalability, self-service and use tracking. It frees application designers and developers from having to reinvent this nontrivial “wheel” (see "Gartner Reference Architecture for Multitenancy: Enterprise Computing 'in the Cloud'"). Although many older SaaS applications use embedded, proprietary enabling technology, rather than a general-purpose aPaaS (Workday, Oracle CRM on-demand), this practice is rapidly becoming outdated with all application independent software vendors (ISVs) except some of the largest ones that aim to own the entire software stack for their offerings.

The internal aPaaS technology that facilitates the service is referred to as cloud-enabled application platform (CEAP) technology. Some CEAPs are available as general-purpose products; others are used only as internal technology to power a particular aPaaS, and not offered for sale to third parties. Despite the commonality of the underlying technology stack, the two must not be confused: CEAP is acquired as a product and aPaaS is acquired as a service.

To ensure competitive performance of the resulting application services, aPaaS is typically associated with a collocated database PaaS (dbPaaS). Technically, dbPaaS is not a part of aPaaS, but neither is fully useful without the other, and each is consumed integrated with the other — salesforce.com Force.com with database.com; Microsoft Azure AppFabric with SQL Azure, Google App Engine with BigTable).

The aPaaS technology is part of a larger-scope, comprehensive PaaS that encompasses other middleware-style service offerings, such as dbPaaS, Messaging PaaS, business process management (BPM) PaaS and many others (see "Platform as a Service: Definition, Taxonomy and Vendor Landscape, 2011"). The more application infrastructure services a provider offers, the closer it is to offering a comprehensive PaaS (see "PaaS Road Map: A Continent Emerging").

Position and Adoption Speed Justification: In 2011, more than 40 vendors offer aPaaS (see "Platform as a Service: Definition, Taxonomy and Vendor Landscape, 2011"). They run the gamut of types of multitenancy from shared hardware to shared everything (see "Gartner Reference Architecture for Multitenancy: Enterprise Computing 'in the Cloud'"). Many are targeting the high-productivity market, and some are focusing on the high-control market (see "Productivity vs.

Control: Cloud Application Platforms Must Split to Win"). In recent months, a number of software infrastructure vendors (including Red Hat OpenShift, SAP HANA/River, IBM Workload Deployer, VMware CloudFoundry), some for the first time, have announced new strategic initiatives in aPaaS space, and more are expected in the next six months. Key software vendors have now targeted strategic resources at the aPaaS opportunity.

Meanwhile, ISVs, system integrators (SIs) and enterprise IT organizations are increasing their adoption of aPaaS offerings for creating cloud business applications. With the growing real-world experience will come the inevitable "reality check" and a future stage of disillusionment. However, expectations are high and the industry hype is near its peak.

User Advice:

- Application ISVs — and, especially, smaller startup ISVs — should look at the aPaaS opportunity as a serious long-term, game-changing option. Its low cost of entry, low burden of operations, and, often, high degree of productivity and scale enable less-technical application ISVs to concentrate on their business expertise and leave the IT issues to others.

- Large-enterprise IT organizations using traditional on-premises software may be able to achieve, at a cost, greater degrees of uniqueness, availability, performance and security for their mission-critical applications than that offered by aPaaS providers. However, projects that need to be completed quickly will find that the aPaaS option can't be beat for the fast time to initiate the project and, often, the fast time to deliver value.

- Most users should plan for a hybrid environment and ensure that their application integration infrastructure (including possibly the use of integration PaaS) is prepared to manage both on-premises and aPaaS-based cloud application services.

- IT users should be careful evaluating the long-term total cost of ownership (TCO) of aPaaS-based applications. The pricing strategies for aPaaS are also evolving, and a future change in the pricing model can increase TCO. On the other hand, the increasing competition is likely to drive down costs, at least for some period. As standards emerge, allowing the development and sale of generalized aPaaS, prices will also be driven down. In the next three years, the real costs of aPaaS will fluctuate across providers and use patterns.

- Most aPaaS offerings are new or are being re-engineered. Users should take the relative immaturity of most aPaaS offerings into account in their long-term planning and realize that the stability of established standards, market leaders and best practices for aPaaS is probably two to four years away. (Salesforce.com Force.com is an exception — a mature aPaaS offering that is challenged by its proprietary programming model.)

- In the absence of de jure or de facto standards, users of many aPaaS offerings become locked into their providers' offerings. Users planning the use of an aPaaS should take into account the high cost of migrating their applications from one aPaaS to another (although the probability that such migration will become necessary is relatively low). Open PaaS will offer some modest relief of this problem in the future.

- IT organizations that are not ready to use aPaaS, but anticipate its use in the future, should invest in developing a competent infrastructure environment for service-oriented architecture (SOA). A well-functioning SOA platform can begin to adopt aPaaS and SaaS as a gradual and incremental change. Without an SOA background, the adoption of cloud computing will face higher costs and reduced value.

Business Impact: As a new model for developing business solutions and delivering business value of IT, aPaaS affects the business of enterprise IT, ISVs, SIs and technology vendors. IT organizations transitioning to use PaaS and aPaaS will have to review their organizations, their required IT skills and the model of their interaction with vendors and the business side of the enterprise. The models of calculating and managing costs will change as well. The vendors will have to learn a new kind of business relationship with their customers, and their cost, revenue and margin calculations will change as well. Along with other forms of PaaS and SaaS, aPaaS brings a new model of IT to enterprises and technology vendors across industries and geographies.

Benefit Rating: High

Market Penetration: 1% to 5% of target audience

Maturity: Emerging

Sample Vendors: Appistry; Apprenda; Aptana; Bungee Labs; Caspio; CloudBees; Cordys; DotCloud; Engine Yard; Express Dynamics; ForeSoft; GigaSpaces Technologies; Google; Intuit; IS Tools; Joyent; LongJump; maatG; Micro Focus; Microsoft; MorphLabs; OrangeScape; Qrimp; Rackspace; Red Hat; Relational Networks (LongJump); Rollbase; salesforce.com; Tibco Software; Trackvia; VMware; Wolf; WorkXpress; WSO2; Zoho

Recommended Reading: "CloudFoundry.com: VMware as a PaaS Provider"

"Platform as a Service: Definition, Taxonomy and Vendor Landscape, 2011"

"PaaS Road Map: A Continent Emerging"

"Productivity vs. Control: Cloud Application Platforms Must Split to Win"

"Cool Vendors in Platforms as a Service, 2011"

"Salesforce.com at an Inflection Point"

"Gartner Reference Architecture for Multitenancy: Enterprise Computing 'in the Cloud'"

"Microsoft AppFabric: A Platform for the Cloud Era Is Under Construction"

"Predicts 2011: Platform as a Service: The Architectural Center of the Cloud"

"From OLTP to Cloud TP: The Third Era of Transaction Processing Aims to the Cloud"

"Scalability, Elasticity and Multitenancy on the Road to Cloud Services"

"Who's Who in Application Platforms for Cloud Computing: The Cloud Specialists"

"Who's Who in Application Platforms for Cloud Computing: The Enterprise Generalists"

"On Apples, Oranges, Amazon and Google"

"Cloud Computing Will Cement the Mainstream Role of SOA"

Cloud Email

Analysis By: Tom Austin; Matthew W. Cain

Definition: Cloud email describes a vendor-offered, multitenant, Internet-delivered email service that is scalable and flexible.

Position and Adoption Speed Justification: Email in the cloud continues to be a white-hot topic for IT professionals. Drawn by the lure of rock-bottom pricing and the perception that cloud email is wildly popular, organizations are aggressively investigating the new provisioning model. In early 2008, we predicted that by year-end 2012, 20% of enterprise email users would be employing a cloud-based email service. We were wrong; the uptake of cloud email services has been happening at a slower pace than previously expected. We now think that the percentage of corporate mailboxes using cloud services will be 10% by year-end 2012 — up from 2% to 3% at the end of 2009 and approaching 5% by the end of 2010. There are about a dozen reasons why cloud uptake is slower than we expected and these can be split into four categories: vendor dynamics, economics, service immaturity and customer dynamics.

User Advice: Organizations should approach cloud email services with caution. These services are immature in their ability to provide a rich management and reporting infrastructure, and in some cases cloud email services lack features commonly found with on-premises email systems.

Organizations also need to examine where they are in the email cycle: those that have just gone through a significant version upgrade, or vendor migration, should generally realize more ROI from that change before moving on. Conversely, companies on the brink of changing email vendors (or undertaking a substantial version upgrade) should look more closely at cloud email services.

Companies should also determine internal email costs, to ascertain if a move to cloud email makes economic sense.

Business Impact: Organizations with large populations of users that don't rely heavily on email services — such as retail or manufacturing floor workers, data entry clerks or hospitality personnel — can immediately benefit from cloud email services.

Small to midsize enterprises can achieve significant cost savings, but the more complex the requirements the lower the cost benefits. The larger the enterprise, the more likely that there are more complex requirements — leading to a reduction of economic benefit in many cases (this requirements-based issue is independent of the scale effects which give larger firms greater internal operating efficiencies).

Larger entities with highly customized email environments will probably not be in a position to save on operating costs with cloud email.

Benefit Rating: Moderate

Market Penetration: 1% to 5% of target audience

Maturity: Adolescent

Sample Vendors: Google; IBM; Microsoft

Recommended Reading: "The Cloud E-mail and Collaboration Services Market"

Cloud Security

Analysis By: Jay Heiser; David W. Cearley

Definition: The relative security of the cloud computing style, compared to classic approaches that do not employ a service-delivery and shared-service approach, remains a major concern. User polls consistently indicate that security is the most significant inhibitor to the use of cloud computing. The primary challenge is the lack of best practices, including control standards, for cloud security assessment. The secondary challenge is the relative nontransparency of commercial cloud service providers. The final challenge is the technical one of ensuring that the

various virtualization and multitenancy models are able to simultaneously protect data confidentiality without loss of availability, within the hostile environment represented by today's Internet. While cloud security is a legitimate area of concern, it is also the case that many organizations are erring on the side of caution, avoiding useful and convenient forms of service because of an exaggerated perception of the risks or their security requirements.

Position and Adoption Speed Justification: The same characteristics that make cloud computing an economically desirable and convenient computing model simultaneously exacerbate both the exposure to attack and the difficulty in confirming the relative ability of the technology to resist that attack.

The history of computer science provides a parallel example as to how a significant change in a computing model can provide desirable economic value, while significantly impacting security. The actual practice of computer security, and the first technical security mechanism, the password, was unnecessary before the introduction of remotely accessible multitasking and multiuser operating environments. Computer hacks that were impossible on batch-oriented computing models slowly but surely became a common occurrence during the 1970s. The introduction of multitenanted cloud computing environments, sometimes supporting millions of user accounts within a single multisite Internet-based platform, represents an equally profound change in risk implication.

Private clouds avoid many of the security complexities of cross-company multitenancy, although the virtualization element of private cloud does introduce a layer of infrastructure that would potentially be leverageable across multiple virtual machines if it were compromised. Private clouds, especially as processing migrates out of the enterprise onto infrastructure services, create a new and not-yet completely fulfilled need to manage the status and location of machine images and any associated data. Cloud governance practices and risk control technology are still evolving, but for organizations that are sophisticated enough to build their own private cloud, the vulnerabilities can be managed, and the threat represented by external attackers — and other customers — is exponentially smaller than the threat from internal "tenants." While externally provisioned services are more security-awkward and risk-ambiguous than private clouds, community cloud and virtual private cloud models are useful compromises that still retain most of the flexibility of public clouds.

The greater the value-add on the part of a service provider, the harder it is for a customer to determine the risk exposure and security robustness. The users of infrastructure services are generally relying on their own machine images, and encryption can be easily applied within storage services if desired. In contrast, platform services, and especially software services, place the majority of the security functionality and oversight in the control of the service provider, not the buyer. With the growing exception of strong authentication, most security solutions that are commonly used within the enterprise, such as encryption and data loss prevention, are not available or not deployed within externally provisioned services (public clouds).

The market for cloud-specific security mechanisms is nascent, with the overwhelming majority of solutions aimed at improving the control and security of private clouds, and enterprise-managed virtual machines within infrastructure services. Although many have long offered a variety of backup solutions, only recently have cloud service providers begun offering any significant degree of security technology, such as firewalls, virtual private networks (VPNs), intrusion prevention and vulnerability scanning. This is more often available in infrastructure as a service (IaaS), although SaaS providers are increasingly supporting VPNs, along with extra price options for strong authentication and offline backups. The IaaS market will not reach a level of security maturity until service providers begin reselling the security and governance technologies that are being used experimentally in the enterprise today. Much work has been put into security of the

private cloud, and the pieces are falling into place to allow the enterprise to extend control into the public cloud.

The most significant current development in cloud computing security is the development of cloud-specific risk management frameworks and assessment questionnaires. Recent work from the Cloud Security Alliance (CSA) and Shared Assessments represents two early attempts to establish best practices for cloud security review, including detailed lists of specific technical and process controls that are most relevant to cloud computing. While this work is useful for private clouds, its primary purpose is to simplify the difficulty in service provider assessment. The existence of these frameworks is expected to encourage significantly higher levels of transparency from an industry that has been notoriously tight-lipped. The most ambitious attempt to force higher levels of transparency on the service provider community is expected in late 2011 from the U.S. federal government in the form of the Federal Risk and Authorization Management Program (FedRAMP). This program intends to introduce two innovations: the use of independent third parties to evaluate service providers, and a continuous control mechanism to ensure that changes in the provider's offering or technology do not negatively impact security expectations.

User Advice: There are many appropriate uses of internal and external cloud computing today, although highly regulated or sensitive proprietary information should not be stored or processed in an external public cloud-based service without appropriate visibility into the provider's security controls, technology and processes. Alternatively, the use of encryption and other security mechanisms can ensure the appropriate level of information protection. Ask for design documentation if data protection mechanisms are part of the provider's offering.

Develop a pragmatic and balanced strategy for the safe use of cloud services by your organization, and support it with policies that provide firm guidance on both appropriate and undesirable use cases and processes to enable security professionals, legal staff and procurement specialists to carefully analyze business and legal security requirements.

Look for opportunities to chain security mechanisms in front of cloud-based services. Within the enterprise, secure Web gateways and security appliances for cloud application control can provide some enterprise control over use of a public cloud service, and some cloud-based vendors are providing similar controls without the need to route remote traffic through the enterprise.

Although it is not yet a common practice for vendors to completely fill out a standardized security questionnaire, buyers should demand that they meet this best practice by doing so. They should not be considered as stable or complete at this point, but Gartner does recommend the use of emerging standards from CSA and Shared Assessments as defensible starting points.

Choose your battles carefully and understand your security- and privacy-related needs if you intend to put sensitive data in an external service. In critical use cases, ask for details on provider technology, security models and operating procedures. Ask what security technology they use and how they implement encryption, key management and application security. Although most SaaS vendors are very reluctant to modify their standard master service agreement terms of service, if you are planning on putting sensitive data in an external cloud, try to negotiate security and privacy clauses beyond the standard for data breach notification and vendor responsibility, including service-level agreements, data ownership language and ongoing access to the results of vulnerability assessments and audits. Also, ask for details on data center coverage, and whether you can restrict data to a specific geography. Some cloud-based service providers are providing greater visibility into their security procedures "on an exception basis," usually for larger customers, providing information under nondisclosure agreement beyond that which is provided on their websites.

Business Impact: Organizations that avoid taking advantage of new forms of externally provisioned service will pay an opportunity cost by failing to take advantage of products that will be effectively used by their competitors. However, organizations that use cloud computing without adequately understanding the risks will suffer from security, compliance or continuity failures that can have devastating consequences. A well-planned strategy will enable companies to maximize the benefit, while ensuring that they are not undertaking unacceptably high levels of risk.

Benefit Rating: High

Market Penetration: 1% to 5% of target audience

Maturity: Adolescent

Recommended Reading: "Cloud Security and Risk Standards"

"Key Issues for Securing Public and Private Cloud Computing, 2011"

"Four Risky Issues When Contracting for Cloud Services"

"Three Styles of Securing Public and Private Clouds"

"Securing and Managing Private and Public Cloud Computing"

"From Secure Virtualization to Secure Private Clouds"

" Will Your Data Rain When the Cloud Bursts?"

"Location Matters: A Model for Assessing Information Risk"

"Gartner Survey Highlights Company Burden of Vetting Third-Party Security Controls"

"Gartner's 30 Leading Locations for Offshore Services"

Platform as a Service (PaaS)

Analysis By: Yefim V. Natis

Definition: Platform as a service (PaaS) is a generally accepted reference to the middle layer of the cloud technology stack, which Gartner also refers to, in a more precise way, as application infrastructure services. In traditional contexts and in the cloud context as well, the term "application infrastructure" is often used to mean middleware and vice versa. PaaS is a highly hyped concept that has a partly uncertain meaning.

A comprehensive PaaS suite, usually depicted in cloud diagrams, is a broad collection of application infrastructure services offered by a cloud services provider. Such comprehensive PaaS suites would include technologies of application servers, database management systems (DBMSs), portal products, application and data integration, business process management suites (BPMSs), messaging and many other forms of application infrastructure — all formatted to be offered as a service. However, no providers offer such a comprehensive offering, and we don't expect comprehensive PaaS offerings to be a mature part of the market before 2015 (see "PaaS Road Map: A Continent Emerging").

Today's PaaS market is chaotic. Various middleware technologies are offered as services, with varying degrees of cloud characteristics. The hype surrounding the PaaS concept is focused mainly on such notable suites of application infrastructure services as the cloud application PaaS (aPaaS) or integration PaaS (iPaaS; see "Platform as a Service: Definition, Taxonomy and Vendor Landscape, 2011").

Users should carefully evaluate vendor claims about their support of PaaS, because a "PaaS" offering may, in reality, be a reference to one of many relatively small subsets of the complete list of application infrastructure services. However, for the purposes of measuring the level of hype, we treat PaaS as the overall concept of an application infrastructure platform offered as a service, rather than a specific rendition (special-purpose, suite of functionality or all-inclusive/comprehensive) of PaaS offering.

Position and Adoption Speed Justification: In prior years, much of the attention in cloud computing was focused on software as a service (SaaS). More recently, the emphasis has been on infrastructure as a service (IaaS), especially on the private cloud rendition of IaaS. However, as leading software vendors begin to adjust their long-term strategies to reflect the emerging importance of cloud computing to their customer and prospect bases, they are investing to establish a leadership position in the middle layer of PaaS. Just in April and May of 2011, IBM, VMware, SAP and Red Hat introduced new strategic offerings in this space.

Historically, a prerequisite for leadership of software infrastructure markets was a vendor's ability to take a leading role in establishing the prevailing programming models and architectures for software developers. This role enabled vendors to build ecosystems of partners, and a leading ecosystem amounted to sustainable industry leadership. In cloud computing, analogously, the leadership of the PaaS market will require leadership in the evolution of standards, architectures and best practices in cloud application platforms and application services.

Ahead of its enterprise competitors, Microsoft has made a strategic investment in Windows Azure to advance its .NET application infrastructure to the cloud. VMware relies on several notable middleware acquisitions to elevate itself from a "standard" in the server virtualization market to a standard for PaaS. IBM, Red Hat, SAP and others are investing in bringing familiar Java programming models to cloud computing. Salesforce.com leads the market with a native (and new) cloud platform architecture of Force.com. Microsoft benefits from its early entry into the market and leads among the familiar enterprise computing players. IBM, through its acquisitions or Sterling and Cast Iron, has taken a strong position in iPaaS.

Vendors take different approaches to the database PaaS (dbPaaS) services. Each has implemented its own model, from Google BigTable, to salesforce.com database.com, to Microsoft SQL Azure. This activity has generated high levels of buzz and hype regarding the PaaS concept, while most potential users are still discovering its essentials.

The process of establishing the platform architecture and standards for PaaS is in its early stages, but most of the largest software vendors are already actively engaged. By the end of 2011, Gartner expects all of the enterprise software leaders to have delivered or announced a strategic PaaS offering. We expect the next two years to see a strong increase in innovation and competition in the PaaS space, with the introduction of new services and a move toward comprehensive PaaS suites. This activity is pushing the concept of PaaS toward the Peak of Inflated Expectations and beyond, to the inevitable Trough of Disillusionment.

User Advice: The available PaaS offerings are at varying stages of development in terms of maturity, functional completeness and "cloudiness." Users should select PaaS services for their projects carefully, matching the functionality and service levels of PaaS offerings with their requirements. No provider offers a comprehensive PaaS suite, so users should use services from multiple providers and combine them with applications and platform technologies on-premises to form a hybrid computing environment.

Users that delay the adoption of PaaS to sometime in the future, when the standards, leading providers and best practices are better established, should invest now in building expertise in service-oriented architecture (SOA), including its event-driven form. SOA is a bridge from the

traditional computing in the enterprise data center to the hybrid model of computing, engaging enterprise data center and cloud resources.

When engaged in long-term planning, users should give preference to vendors that are more likely to accumulate a comprehensive PaaS offering over time. Unlike on-premises — where users can take a best-of-breed approach to selecting component technologies from different vendors — in the cloud, the winning scenario will be where many platform requirements of an application are provided out of one data center footprint of one cloud provider.

Business Impact: During the next five years, reliable and functionally rich PaaS offerings from industry-leading providers will alter the business of engineering and delivering software to enterprises and consumers. A mature, functional, always-on, high-productivity PaaS will be the foundation for a wave of innovation in business application services, as independent software vendors turn their engineering efforts to these platforms. The new levels of agility, resource sharing and productivity of software engineering will change the way IT organizations plan and develop software, the kinds of skills they'll require, and how they'll be managed, evaluated and budgeted.

Enterprise IT will refocus on its core differentiated business and become more responsive; however, the costs of IT will not decline. Instead, they'll be rearranged, with more spending going to cloud service providers in the form of subscriptions and more internal IT spending focused on the management and integration of the enterprise's IT resources (internal and cloud-sourced), as well as custom engineering of specialized "family jewels" software.

Benefit Rating: Transformational

Market Penetration: 1% to 5% of target audience

Maturity: Emerging

Sample Vendors: 3Scale Networks; Amazon; Apigee; Appian; Appirio; Appistry; CloudBees; CloudPrime; Compuware; Cordys; Dell Boomi; Engine Yard; Fujitsu America; Google; IBM; Informatica; IS Tools; Jitterbit; Layer 7 Technologies; Mashery; Microsoft; Nimbus; OrangeScape; Pegasystems; Pervasive Software; Red Hat; Relational Networks (LongJump); salesforce.com; Skytap; StormMQ; Tibco Software; VMware; Wolf Frameworks; WSO2; Xerox

Recommended Reading: "Platform as a Service: Definition, Taxonomy and Vendor Landscape, 2011"

"PaaS Road Map: A Continent Emerging"

"Cloud-Based Messaging Services and Technology Are Positioned for Rapid Growth"

"The State of the BPM Platform Cloud Market, 2011"

"Salesforce.com at an Inflection Point"

"Microsoft AppFabric: A Platform for the Cloud Era Is Under Construction"

"Salesforce.com Unveils Potentially Disruptive Cloud DBMS, database.com"

"Data in the Cloud: The Changing Nature of Managing Data Delivery"

"Integration Platform as a Service: Moving Integration to the Cloud"

"Windows Azure AppFabric: A Strategic Core of Microsoft's Cloud Platform"

"Who's Who in Application Platforms for Cloud Computing: The Cloud Specialists"

"Who's Who in Application Platforms for Cloud Computing: The Enterprise Generalists"

"Who's Who in Cloud-Computing/SaaS Integration, Volume 1"

"Who's Who in Cloud-Computing/SaaS Integration, Volume 2"

"Portals in the Cloud Will Take Five Forms"

Private Cloud Computing

Analysis By: Thomas J. Bittman

Definition: Cloud computing is a style of computing in which scalable and elastic IT-enabled capabilities are delivered as a service to customers using Internet technologies. In the broadest terms, private cloud computing is a form of cloud computing that is used by only one organization, or that ensures that an organization is completely isolated from others. This contrasts with public cloud computing, where access to the service is open to any customer willing to pay (unless the service is subsidized, for example, by advertising). For our purposes here, the focus will be on private cloud computing that is internal to an organization — in other words, the customer has control/ownership of the service, and service access is limited to the internal organization. However, three other variants of private cloud computing (not discussed here) are community cloud computing (in which a third-party provider offers services to a limited set of customers), virtual private cloud computing (in which a third-party provider offers the services, but the customer has some control over the implementation, usually in terms of limiting hardware/software sharing), and hybrid cloud computing (in which a service is delivered by a federation of private and public cloud resources).

Organizations building a private cloud service are trying to emulate public cloud computing providers in order to get similar benefits, but within their control and on-premises. In most cases, this is based on a virtualization foundation, but private cloud computing requires more (see "Private Cloud Computing: An Essential Overview"). This includes standardization, automation, self-service tools and service management, metering and chargeback, to name a few. Many of these technologies are still evolving, and early deployments often require custom tools. Regardless, the biggest challenges with private cloud computing tend to be process-related, cultural, political and organizational.

Unlike public cloud providers, which maintain a small number of offered services, enterprises have many complex and interrelated services to deliver. A private cloud computing service can fit within a broader portfolio of services delivered by a real-time infrastructure.

Position and Adoption Speed Justification: Although some of the technologies required for private cloud computing exist, many do not, or are immature. Many early examples of private cloud computing services are focused on development and test provisioning. However, the private cloud has become a marketing buzzword for most of the largest IT vendors, and many new products shipped in 2010 or will be shipped in 2011 to address technology gaps. Since private cloud computing is a natural evolution of the rapidly growing server virtualization trend, no vendor wants to miss the "next big thing." The hype (both external, from vendors, and internal, from directives "to do something about cloud computing") is already tremendous, and it's going to increase during the next year.

Enterprise interest is already high, with 66% of respondents in a recent poll saying they plan to pursue a private cloud computing strategy (at least for a small set of services) by 2014 (see "Private Cloud Computing Ramps Up in 2011").

User Advice: Let service requirements lead your private cloud computing plans, rather than technologies (see "Getting Starting With Private Cloud: Services First").

Create a business case for developing a full private cloud service using public cloud services, or modernizing established architectures.

Consider the long-term road map for your private cloud service (see "The Road Map From Virtualization to Cloud Computing"). Build with the potential to take advantage of hybrid sourcing (using both your private cloud services and public) at some point in the future.

Start slowly with development/test lab provisioning; short-term, low-service-level-agreement computing requests; and simple, non-mission-critical Web services (e.g., self-service requests and dynamic provisioning for Web environments). Pilot a private cloud implementation to gain support for shared services and to build transparency in IT service costing and chargebacks.

Implement change and configuration management processes and tools prior to implementing private cloud services to ensure that you can standardize on the software stacks to be delivered through self-service provisioning, and adequately maintain them.

Business Impact: Most private cloud implementations will evolve from a virtualization foundation. Virtualization reduces capital costs, but private cloud computing will reduce the cost of operations and enable faster service delivery. It's primarily attractive to the business, because it enables agility — self-service ordering of frequently requested services, as well as dynamic provisioning. Test lab provisioning is an early example of a private cloud service that enables testers to improve time-to-market and efficiencies, while labor costs associated with provisioning are reduced.

Private cloud computing also changes the relationship between the business and IT, transforming how IT is consumed. The shift to services (rather than implementation and assets), pay per use and chargeback enables the business to focus on rapidly changing service requirements and consuming IT based on variable costs, while IT can focus on efficient implementation and sourcing (including the potential to leverage public cloud services in the future, without negatively affecting the business).

Benefit Rating: High

Market Penetration: 5% to 20% of target audience

Maturity: Emerging

Sample Vendors: Abiquo; Adaptive Computing; BMC Software; CA Technologies; DynamicOps; Elastra; Eucalyptus; HP; IBM; VMware

Recommended Reading: "Key Issues for Private Cloud Computing, 2011"

"Private Cloud Computing Ramps Up in 2011"

"The Drivers and Challenges of Private Cloud Computing"

"The 10 Fundamentals of Building a Private Cloud Service"

"Private Cloud Computing: An Essential Overview"

"The Architecture of a Private Cloud Service"

Cloud Computing

Analysis By: David Mitchell Smith; Gene Phifer

Definition: Gartner defines cloud computing as a style of computing in which scalable and elastic IT-enabled capabilities are delivered as a service using Internet technologies.

Position and Adoption Speed Justification: Cloud computing remains near the Peak of Inflated Expectations. Although there are signs of fatigue and disillusionment (e.g., Amazon's highly visible failure), cloud computing is showing notable staying power as a major force in IT. Users are changing their buying behaviors, and, although it's unlikely they'll completely abandon on-premises models or soon buy complex, mission-critical processes as services through the cloud, there will be a movement toward consuming services in a more cost-effective way.

As expected of technology near the Peak of Inflated Expectations, there is deafening hype around cloud computing. Every IT vendor has a cloud strategy, although many aren't cloud-centric. Variations, such as private cloud computing and hybrid approaches, compound the hype and demonstrate that one dot on a Hype Cycle cannot adequately represent all that is cloud computing. Cloud computing has moved just past the Peak and is likely to spend some time in the Trough of Disillusionment. Subjects that generate this much hype rarely skip through the Trough quickly.

User Advice: Vendor organizations must begin to focus their cloud strategies around more-specific scenarios, and unify them into high-level messages that encompass the breadth of their offerings. User organizations must demand road maps for the cloud from their vendors. Users should look at specific usage scenarios and workloads, and map their view of the cloud to that of potential providers, and focus more on specifics than on general cloud ideas.

Cloud computing involves many components, and some aspects are immature. Care must be taken to assess maturity and assess the risks of deployment. Tools such as cloud services brokerages can help.

Business Impact: The cloud computing model is changing the way the IT industry looks at user and vendor relationships. As service provisions (a critical aspect of cloud computing) grow, vendors must become, or partner with, service providers to deliver technologies indirectly to users. User organizations will watch portfolios of owned technologies decline as service portfolios grow. The key activity will be to determine which cloud services will be viable, and when.

Benefit Rating: Transformational

Market Penetration: 5% to 20% of target audience

Maturity: Early mainstream

Sample Vendors: Amazon; Google; Microsoft; salesforce.com; VMware

Recommended Reading: "Key Issues for Cloud Computing, 2011"

"The What, Why and When of Cloud Computing"

Infrastructure as a Service (IaaS)

Analysis By: Lydia Leong

Definition: Compute infrastructure services are a type of infrastructure as a service (IaaS) offering. They offer on-demand computing capacity from a service provider. Rather than buying servers and running them within its own data center, a business simply obtains the necessary infrastructure from a service provider in a shared, scalable, "elastic" way, and accesses it via the Internet or a private network.

Position and Adoption Speed Justification: Four main use cases exist for cloud-based compute infrastructure services: Web hosting, high-performance computing, test and development infrastructure, and general production infrastructure.

The most rapidly maturing use case is Web hosting, as it is convergent with the general Web hosting market. Features and capabilities formerly available only on dedicated hardware are now being extended to shared cloud resources.

The use of these services for test and development infrastructure is growing for pilot projects, rapid application development environments and formal lab environments. As test and development-specific features and management tools improve, formal development environments will become more common. Batch-oriented, compute-intensive workloads (such as modeling, simulation, scientific computing and one-time processing needs such as transcoding), are highly cost-effective in the cloud.

However, before cloud computing for general workloads can achieve mainstream adoption, security, risk and compliance issues must be overcome and costs driven down even further.

These services are maturing and being adopted most quickly in the U.S. Although global demand is robust, the growth of the market is slower outside the U.S., due to less competition, less mature offerings, and fragmentation resulting from users' desire to keep data and processing in-country.

User Advice: Cloud providers' capabilities vary significantly, but enterprise-class clouds, with strong service-level agreements backed by financial penalties, high levels of security and solid service and support, have emerged. Businesses can safely adopt these services; the risks are not significantly greater than other outsourced hosting approaches, assuming the cloud service used matches the service-level and security needs of the application.

Businesses should consider pilot projects for test and development, compute capacity augmentation, and Web content and applications. Successful pilots can be expanded into broader production use.

Both public multi-tenant and private single-tenant offerings are available, but the distinction between public and private cloud IaaS is blurring. The most cost-effective clouds are highly standardized, and use a shared capacity pool. Hybrid public-private cloud offerings — enabling "cloud bursting" for on-demand capacity and business continuity — currently exist, but the technology will not be mature until at least 2016.

This market is evolving extremely quickly, so the suitability of these services should be re-evaluated at least once every six months.

Business Impact: Cloud compute infrastructure services will be broadly advantageous to IT organizations. The cost benefits, driven primarily by automation, will be particularly significant for small and midsize businesses. Larger enterprises will benefit primarily from greater flexibility, rather than direct cost reduction.

In the short term, the benefits will be driven primarily by rapid provisioning that requires minimal manual intervention. Over the longer term, more system management tasks will be automated, leading to more efficient infrastructure management.

The metered-by-use attribute of these services will result in more efficient use of capacity. The self-service nature of these services will empower employees outside IT operations, improving developer productivity and making it easier for business buyers to obtain infrastructure.

Benefit Rating: High

Market Penetration: 1% to 5% of target audience

Maturity: Emerging

Sample Vendors: Amazon; AT&T; GoGrid (previously ServePath); Rackspace; Savvis; Terremark; Verizon Business

Public Cloud Computing/the Cloud

Analysis By: Daryl C. Plummer

Definition: Gartner defines public cloud computing as a style in which scalable and elastic IT-enabled capabilities are provided as services to external customers using Internet technologies. Therefore, public cloud computing involves the use of cloud computing technologies to support customers that are external to the provider's organization. Through public consumption of cloud services, economies of scale and sharing of resources will be generated to reduce costs and to increase the choices available to consumers.

Public cloud computing carries with it the concerns that security, data management, trust, control and guarantees of appropriate performance will not be sufficient to support enterprise needs. Enterprises want the value delivered through cloud computing services, but also need to ensure that the concept is ready for delivering services that a company can rely on. However, public cloud computing has proved itself time and again, in the context of the Internet and the Web, from what is commonly referred to as the "consumer perspective." Sites such as Flickr and Facebook, as well as countless business sites delivering services from entertainment to healthcare records, have been in use for some time in the public context.

It's important to distinguish among the cloud, cloud services and cloud computing. The cloud is an abstract concept that refers to a collection of one or more cloud services. The cloud is used euphemistically to indicate that work is being performed somewhere else, with little regard for where or how. Beyond this, the idea of cloud services is more tangible, in that each cloud provider (e.g., Amazon or Workday) delivers cloud services that provide value to their consumers. In the end, cloud computing is a style of computing, whereas cloud services are used to deliver value through this abstract thing that we call the cloud.

Position and Adoption Speed Justification: The public cloud is at (and a little past) the Peak of Inflated Expectations. As enterprises heavily experiment with the concept, they begin serious budgeting efforts for real projects. The evaluation of peer projects that solve actual problems are under way. In addition, cloud providers are advertising their ability to deliver enterprise services and reduce cost. Customers should still be cautious about the claims of most providers, because their models are still unproved for enterprise use. The potential advantages in terms of agility, as well as in time to market to stand up new applications, or set up new users in an existing environment, are worth investigating for cloud computing adoption. Many providers are unprepared to deliver these advantages to the enterprise, but besides reducing the initial investment in assets, agility is certainly something that clients will want to look at closely.

User Advice: User companies should be moving experimental projects to feasibility discussions for serious implementation in 2010. There will continued investment in 2011, with rapid growth in cloud computing through 2012.

Business Impact: The business impact of cloud computing in the public sense can be varied, but the basic opportunity is for businesses to consume services from other companies that free them from the need to provide those services themselves. This can enable to companies to eliminate work that previously might have been done in-house. It can also lead to massive changes in the way money is spent — for example, using operating expenses to fund external services, rather than using capital expenses to fund IT projects.

Benefit Rating: Transformational

Market Penetration: More than 50% of target audience

Maturity: Early mainstream

Sample Vendors: Amazon; Google; Rackspace; salesforce.com

Sliding Into the Trough

Real-Time Infrastructure

Analysis By: Donna Scott

Definition: Real-time infrastructure (RTI) represents a shared IT infrastructure (across customers, business units or applications) in which business policies and SLAs drive dynamic and automatic allocation and optimization of IT resources (that is, services are elastic), so that service levels are predictable and consistent, despite the unpredictable demand for IT services. Where resources are constrained, business policies determine how resources are allocated to meet business goals. RTI may be implemented in private and public cloud architectures, as well as hybrid environments where data center and service policies would drive optimum placement of services and workloads. Moreover, in all these configurations, RTI provides the elasticity functionality, as well as dynamic optimization and tuning, of the runtime environment based on policies and priorities.

Position and Adoption Speed Justification: This technology is immature from the standpoint of architecting and automating an entire data center and its IT services for RTI. However, point solutions have emerged that optimize specific applications or specific environments, such as dynamically optimizing virtual servers (through the use of performance management metrics and virtual server live-migration technologies) and dynamically optimizing Java Platform, Enterprise Edition (Java EE)-based shared application environments. It is also emerging in cloud solutions, initially for optimizing placement of workloads or services upon startup based on pre-established policies. Moreover, enterprises are implementing shared disaster recovery data centers, whereby they dynamically reconfigure test/development environments to look like the production environment for disaster recovery testing and disaster strikes. This type of architecture can typically achieve recovery time objectives in the range of one to four hours after a disaster is declared. Because of the advancement in server virtualization, RTI solutions are making some degree of progress in the market, especially for targeted use cases where enterprises write specific automation, such as to scale a website up/down and in/out. However, there is low market penetration, primarily because of lack of modeling the service (inclusive of runtime policies and triggers for elasticity), lack of standards and lack of strong service governors/policy engines in the market. This leaves customers that desire dynamic optimization to integrate multiple technologies together and orchestrate analytics with actions.

User Advice: Surveys of Gartner clients indicate that the majority of IT organizations view RTI architectures as desirable for gaining agility, reducing costs and attaining higher IT service quality, and that about 20% of organizations have implemented RTI for some portion of their portfolios. Overall progress is slow for internal deployments of RTI architectures because of many impediments, especially the lack of IT management process and technology maturity levels, but also because of organizational and cultural issues.

It is also slow for public cloud services, where applications may have to be written to a specific and proprietary set of technologies to get dynamic elasticity. We see technology as a significant barrier to RTI, specifically in the areas of root cause analysis (required to determine what optimization actions to take), service governors (the runtime execution engine behind RTI analysis and actions), and integrated IT process/tool architectures and standards. However, RTI has taken a step forward in particular focused areas, such as:

- Dynamic provisioning of development/testing/staging and production environments

- Server virtualization and dynamic workload movement
- Reconfiguring capacity during failure or disaster events
- Service-oriented architecture (SOA) and Java EE environments for dynamic scaling of application instances
- Specific and customized automation written for specific use cases, such as scaling up/down or out/in a website that has variable demand

Many IT organizations that have been maturing their IT management processes and using IT process automation tools (aka run book automation) to integrate processes (and tools) together to enable complex, automated actions are moving closer to RTI through these actions. IT organizations desiring RTI should focus on maturing their management processes using ITIL and maturity models (such as Gartner's ITScore for I&O Maturity Model), and their technology architectures (such as through standardization, consolidation and virtualization). They should also build a culture conducive to sharing the infrastructure, and should provide incentives such as through reduced costs for shared infrastructures. Organizations should investigate and consider implementing early RTI solutions in the public or private cloud or across data centers in a hybrid implementation, which can add business value and solve a particular pain point, but should not embark on data-center-wide RTI initiatives.

Business Impact: RTI has three value propositions expressed as business goals:

- Reduced costs achieved by better, more-efficient resource use, and by reduced IT operations management (labor) costs
- Improved service levels achieved by dynamic tuning of IT services
- Increased agility achieved by rapid provisioning of new services or resources, and scaling capacity (up and down) of established services across both internally and externally sourced data centers

Benefit Rating: Transformational

Market Penetration: 5% to 20% of target audience

Maturity: Emerging

Sample Vendors: Adaptive Computing; CA Technologies; IBM Tivoli; NetIQ; Univa; VMware

Recommended Reading: "Provisioning and Configuration Management for Private Cloud Computing and Real-Time Infrastructure"

"Private Cloud Computing Ramps Up in 2011"

"Survey Shows High Interest in RTI, Private Cloud"

"Building Private Clouds With Real-Time Infrastructure Architectures"

"Cloud Services Elasticity Is About Capacity, Not Just Load"

Public Cloud Storage

Analysis By: Michele C. Caminos; Stanley Zaffos

Definition: Cloud storage is a storage utility offering that is defined by the following characteristics: pay-per-use model, software-agnostic, reservationless provisioning and provider-

owned; and usually geographically separated from the servers that are using it. For the purposes of this Hype Cycle, we focus on the enterprise, rather than the consumer community, and exclude software as a service (SaaS) because the storage associated with SaaS is unavailable for other applications.

Position and Adoption Speed Justification: Several variations of cloud storage are currently available in the U.S. market, with other regions just starting to evaluate and service providers coming to market with cloud storage capability. Its evolution is being driven primarily by market demand for low-cost storage alternatives. User willingness to experiment with and integrate cloud storage into existing storage infrastructures seems greater in the U.S. than in other geographic areas for a variety of regulatory, economic and cultural reasons. This has contributed to differences in expectations between users in the U.S. and other geographic areas. In the U.S., larger established vendors have recently entered the market to join emerging cloud storage vendors that created the market. Some have exited the market, and others are competing with different architectural approaches to improve agility, provide differentiated services and lower costs. Hence, the impact of three major cloud storage providers pulling out of the market or selling their cloud storage business in the past 18 months, the failure of multiple emerging cloud storage companies, and a growing awareness of the technology's limitations and cloud storage alternatives have had more of an impact in the U.S. than in Asia/Pacific and Europe. Outside of the U.S., we are seeing just the beginnings of cloud storage take off with a number of local country and regional providers coming to market. Gartner expects these regions to follow the same path as the U.S. market, albeit at a much quicker pace. No doubt, the U.S. headlines are reaching other regions, helping end users approach this delivery model carefully. Gartner does not expect full-scale adoption to occur earlier due to as-yet-unanswered security issues, which could lead to potential legal exposure. Unpredictable monthly costs due to usage variability and the current lack of sufficiently differentiated storage services will also affect the absorption rate of the market.

User Advice: Evaluate cloud storage as a low-cost option for certain applications, such as archiving and backup, because they generally are not classified as mission-critical workloads and they tolerate relatively long latencies better than transactional workloads. Additionally, security requirements may dictate the cost and management expense of data encryption, both in flight and at rest. Some vendors are offering on-premises appliances providing data deduplication, thin provisioning, encryption and further security capability that will assist with security and latency concerns. Due diligence should also include an evaluation of the organization's viability, SLAs, data location, transport costs and ingestion fees, and disaster recovery service options. Considerable investments in time and money will often be required to integrate cloud storage options into current applications and environments.

Business Impact: The cost and agility expectations set by the public cloud storage vendors are forcing in-house IT operations to change their storage infrastructure management procedures and storage infrastructure strategy. User demands for lower costs, more agility and operations that are more autonomic are also influencing vendor R&D investments and cloud services offerings. Those services that have already been influenced by user demands include: backup, versioning, encryption and secure erasure. And in response to cost concerns, vendors are offering a variety of pricing models that allow end users to align their storage costs with their usage rates, with the goal of lowering costs in the short and long term.

Benefit Rating: Moderate

Market Penetration: 1% to 5% of target audience

Maturity: Emerging

Sample Vendors: Amazon; AT&T; Mezeo Software; Microsoft; Nirvanix; Zetta

Recommended Reading: "Search Analytics Trends: Cloud Storage"

"Hybrid Cloud Gateway Appliances Expand Cloud Storage Use Cases"

"Cloud Storage Infrastructure-as-a-Service Providers, North America"

"Clients Must Prepare as Iron Mountain Looks to Exit Digital Business"

"Case Study: Reducing Storage Costs via the Public Cloud"

Cloud/Web Platforms

Analysis By: Gene Phifer; David Mitchell Smith

Definition: "Cloud/Web platforms use Web technologies to provide programmatic access to functionality on the Web, including capabilities enabled not only by technology, but also by community and business aspects. This includes, but is not limited to, storage and computing power. We use the terms "Web platform" and "cloud platform" interchangeably, as well as the merged term "Web/cloud platforms." They have ecosystems similar to traditional platforms, but the concept has emerged as a result of market and technology changes collectively known as "Web 2.0." These platforms will serve as broad, general-purpose platforms, but, more specifically, they will support business flexibility and speed requirements by exploiting new and enhanced forms of application development and delivery. Web platforms reuse many of the capabilities and technologies that have been accessible on websites for more than a decade through browsers by adding programmatic access to the underlying global-class capabilities. Reuse is occurring via Web services, and is being delivered via Web-oriented architecture (WOA) interfaces, such as representational state transfer (REST), plain old XML (POX) and Really Simple Syndication (RSS). In addition to the capabilities of Web 2.0, these platforms provide programmatic access to cloud computing capabilities. The public API phenomenon has taken WOA beyond consumer markets (e.g., Twitter) into enterprise B2B integration.

The cloud/Web platform term is not the same as platform as a service (PaaS). PaaS, according to the National Institute of Standards and Technology (NIST), refers to the middleware layer in cloud architectures. The cloud/Web platform is broader and employs a more accurate use of the term platform (as a relative term, see "NIST and Gartner Cloud Approaches Are More Similar Than Different") and can span all layers of cloud architecture. Our use of the term cloud/Web platform predates the PaaS term and current cloud terminology, but is not inconsistent with it.

Position and Adoption Speed Justification: The use of Web/cloud platforms is happening first in consumer markets. As further adoption of all the cloud service layers increases, use and maturity will evolve. Enterprise use of Web-based capabilities, such as Amazon Simple Storage Service (Amazon S3) and Amazon Elastic Compute Cloud (Amazon EC2), has begun as well.

User Advice: Web platforms and related phenomena have affected consumer markets, but enterprises should evaluate the growing space as an appropriate extension to internal computing capabilities. The use of Web platforms will drive WOA, which enterprises should adopt, where appropriate, along with simple interfaces, such as REST, POX and RSS (wherever possible), to exploit the interoperability, reach and real-time agility of the Internet.

Business Impact: Web platforms can be leveraged as part of business solutions, and will form much of the basis for the next generation of interest in the virtual enterprise. Web platforms can decrease barriers to entry, and can deliver substantial value for small and midsize businesses that cannot afford to build and maintain capabilities and infrastructures. Examples include Amazon Web Services (including S3 and EC2), salesforce.com's Force.com, Google's App Engine and Microsoft Azure Services Platform. Note that the term "Web/cloud platform" is broader than, and includes multiple layers in, cloud computing terminology (e.g., infrastructure as

a service [IaaS], PaaS and software as a service [SaaS]), and the use of the term "platform" is different from the term "PaaS."

Benefit Rating: High

Market Penetration: 1% to 5% of target audience

Maturity: Early mainstream

Sample Vendors: Amazon.com; Google; Microsoft; salesforce.com

Recommended Reading: "Web Platforms Are Coming to an Enterprise Near You"

"NIST and Gartner Cloud Approaches Are More Similar Than Different"

Climbing the Slope

Enhanced Network Delivery

Analysis By: Joe Skorupa

Definition: Enhanced network delivery comprises a combination of network-based WAN optimization services and WAN optimization controller (WOC)-equipment-based deployments. Enhanced network delivery uses a combination of protocol spoofing, route control, compression and caching to provide enterprises with improved application performance across the WAN for cloud-based services.

Position and Adoption Speed Justification: As the cloud computing market matures, consumers and providers are realizing that applications often suffer from performance problems as network latency increases. Additionally, providers also realize that their bandwidth costs will be a significant component of their overhead. As a result, this market is being driven simultaneously by consumers and providers. Many software as a service (SaaS) consumers (including a number of Microsoft Business Productivity Online Standard Suite [BPOS] customers) are demanding that their preferred WOCs be supported by their SaaS provider, primarily to optimize HTTP/HTTPS traffic. SaaS providers are realizing significant increases in customer satisfaction and lower support costs when they bundle network-based optimization (often from Akamai) into their offerings. In order to eliminate the need to deploy WOCs in the SaaS providers' data centers and to provide end-to-end optimized delivery, Akamai and Riverbed have announced a joint development and go-to-market partnership to integrate Riverbed WOCs and Akamai's network. In some cases, SaaS providers are leveraging asymmetrical acceleration for HTTP/HTTPS applications from companies, such as Aptimize, Citrix Systems, F5 Networks and Strangeloop Networks. Cloud infrastructure as a service (IaaS) providers (particularly cloud storage) are adopting WOC-based acceleration to lower bandwidth costs and to improve performance of backup and replication traffic.

Enhanced network delivery comprises a combination of mature technologies, such as application delivery controllers and WAN optimization controllers, and route optimization with emerging technologies to address the rapidly growing need of integration as IaaS and SaaS providers. The rapid adoption of enhanced network delivery is being driven by the great need for immediate solutions and by a mix of products and services that bring immediate value. Over time, all of these approaches will become mainstream.

User Advice: Cloud services consumers should test applications across actual network topologies rather than assuming that they will perform adequately. If performance problems (latency- or bandwidth-related) appear, insist that your cloud provider support the required WAN optimization products or services. Cloud providers that have customer-facing HTTP/HTTPS

applications should consider enhanced network delivery based on WAN optimization services and equipment-based solutions as part of their core portfolios.

Business Impact: Enhanced network delivery can reduce the cost of WAN bandwidth, while delivering significant gains in application performance and customer satisfaction.

Benefit Rating: High

Market Penetration: Less than 1% of target audience

Maturity: Emerging

Sample Vendors: Akamai; Aptimize; AT&T; Blue Coat Systems; BT; Cisco; Citrix Systems; F5; Orange Business Services; Riverbed Technology; Strangeloop Networks; Verizon Business; Virtela

Recommended Reading: "You Can't Do Cloud Computing Without the Right Cloud (Network)"

"Magic Quadrant for WAN Optimization Controllers, 2010"

"Magic Quadrant for Application Delivery Controllers, 2010"

IT Infrastructure Utility

Analysis By: Claudio Da Rold; Frank Ridder; Philip Dawson

Definition: An IT infrastructure utility (IU) is a shared IT infrastructure architecture provided through on-demand services. Pricing is based on service use and proven, ongoing reductions in the fixed baseline (or subscription fees) and unit costs. The IU is open, flexible, predesigned and standardized, as well as virtualized, highly automated, secure and reliable.

The most basic IU style is utility hosting, which has evolved from traditional dedicated hosting via virtualization. The most developed IU offerings are currently built on standard infrastructure blocks (such as virtualized computing, storage, networking), to which elements designed to support a specific application landscape — such as ERP, communication, collaboration or CRM — are added. The client is still in full control of the customized applications, while the service provider controls and manages the operating platform up to a level below the logic of the application. The provider tailors the architecture/performance/price of the service to the application requirements; for example, billing on a per-user or per-SAP Application Performance Standard (SAPS) basis.

Companies such as Amazon.com (with its EC2 and S3 offerings); smaller providers such as GoGrid, Joyent, OpSource and SoftLayer Technologies; and virtual data center hosting companies, all deliver IU services that leverage a cloud computing infrastructure as a service (IaaS) approach. Virtual data center hosting companies enable the implementation of complex virtual architectures in their physical data centers. Traditional outsourcers and small startups, such as ThinkGrid, are also introducing virtualized desktop utility services into the market. Public cloud solutions do not give enough visibility into the structure, architecture, operations and security of the global data centers or computing environments; a fact that causes compliance issues for multiple industries (such as banking, insurance and the public sector). IU solutions must close this transparency gap and enable regulated industries to leverage solutions based on this approach; this will happen through private/hybrid, hosted cloud computing.

Position and Adoption Speed Justification: During 2010, IU services (IUS) continued its impressive growth curve. In a worldwide IUS survey during early 2011, 39% of respondents confirmed that they were already users of IUS, and an additional 22% of the respondents planned adoption within the next 12 months (see "Survey Analysis: End-User Trends in Infrastructure

Utility Services and Infrastructure as a Service"and "User Survey Analysis: Infrastructure Utility Services, 1Q11"). This growth is mainly driven by cost reduction opportunities, quality improvement options and the increased flexibility that IUS bring to an organization. As more vendors offer infrastructure utilities (CSC announced an IU for SAP on 9 May 2011, a few months after having announced a hybrid cloud; see "CSC Introduces the First Hybrid Cloud IaaS for Security and Simplicity"), increasing adoption and competition causes the IU market to further mature. During 2009 to 2011, strong competition between IU for SAP (IU4SAP) solutions brought down the price of this industrialized service very significantly, showing that, in many cases, the next step in industrialization will be low-cost services. Examples of industrialized, one-to-many, low-cost IT services that exist today, include virtualized servers available for $400 or less per server per month, and Infrastructure Utility for SAP, available at €5 (approximately $7) per user per month — the traditional client cost used to be five to 10 times as much (see "Behind the Cloud: The Rise of Industrialized, Low-Cost IT Services" and "Industrialized Low-Cost IT Services Drive Value, By Definition").

Across the challenging economic environment of 2009, and the uncertain climate of 2010/2011, the industrialization of the IT services industry actually accelerated. The evolution from traditional outsourcing delivery models toward cloud computing is driving innovation at an increased pace, and is leading to significant investments at different service layers. Many of these investments are being made in the infrastructure layer, because this is an area where technology is mature, sharing is possible, willingness to outsource is high and knowledge is widely available. Also, the ongoing discussion about cloud delivery models ranging from private to hybrid to public, and the mix of hype and fear, uncertainty and doubt associated with cloud computing, are fueling interest in IU — which is perceived as more secure and reliable than public cloud. (Marketing claims about the robustness of cloud computing are encouraging many customers to ignore vital continuity and recovery practices, leaving them dangerously exposed to service interruptions and data loss, see "Will Your Data Rain When the Cloud Bursts?" and "Managing Availability and Performance Risks in the Cloud: Expect the Unexpected.")

Most service providers have already incorporated, or are currently adding, IU solutions into their portfolios: often rebranding on-demand and utility offerings as hybrid "cloud computing." Finally, an increasing number of telco providers are adding computing to their network, and offer IaaS IUS as an alternative for colocation or hosting (see "Magic Quadrant for Data Center Outsourcing and Utility Services, Europe." "Magic Quadrant for Web Hosting and Hosted Cloud System Infrastructure Services (On Demand)" and "Competitive Landscape: Data Center Outsourcing and Infrastructure Utility Services, Europe"). This creates plenty of choices for buyers, but also adds to the confusion; the boundaries between various IUS and cloud solutions are not always clear (see "Comparing Infrastructure Utility Services and Private Clouds"). That is the main reason why organizations should understand the seven attributes that define IUS (as described in "The Seven Golden Rules for Industrialized IU Services"), creating unique value for organizations of all sizes.

IUS are outcome-focused, ready-to-use and charged on a usage basis. Enterprises can scale their IU use up or down. IUs are also highly virtualized and shared, automated, lean and standardized. While evaluating IU and cloud services, an organization must be aware of the potential lock-in and vendor strategies associated with private cloud implementations, and adopt the relevant emerging practices (as depicted in "Cloud Sourcing Deals Anatomy: From Public to Private, From Services to Technology Lock-In").

From a maturity perspective, we map the advancement of IU against our Infrastructure Utility Maturity Model (IUMM); see "Gartner Introduces the Infrastructure Utility Maturity Model." Leading IU providers have implemented Level 3 maturity (virtualized) and are progressively implementing Level 4, which is all about automation and will close the gap between IU and IaaS. What's stopping many service providers from running full speed into Level 4 is that an increased level of

automation decreases the number of touchpoints with the client, something they currently rely on for upselling and their relationship-improvement efforts. Additionally, before automating IU providers must define their service approach in great detail, and this requires time and offerings maturity.

While some IUs, such as IU4SAP, are quite mature and still see double-digit growth rates, more complex and complete IU architectures will emerge within leading IU providers. Basic IU services (such as virtual server, storage and network) are modular and can be grouped and combined with managed services, standardization and automation to support more complex client requirements that are aligned to a specific application landscape or more broad vertical or segment-specific requirements.

Traditional providers must continue to invest in, and further industrialize, their IT infrastructure service delivery, because new and disruptive approaches — especially those based on cloud computing — and new providers will progressively threaten the existing status of every insourced or outsourced solution. During the next five years, IUS will drive consolidation, and large providers will end up winning the market share battle — growing organically or by acquisition.

Overall, the outsourced services that are delivered through an IU approach (and that fulfill the seven traits mentioned above) have been grouped into a subset of the IT services marketplace — IUS. These services represent the provision of outsourced, industrialized, asset-based IT infrastructure managed services (below the business application functional layer). IUS are defined by service outcomes, technical options and interfaces, and are paid based on resource usage, allocation or number of users served. For this market, Gartner has created a formal market sizing and forecast (see "Forecast: Infrastructure Utility Services, Worldwide, 2009-2013"). The forecast shows that the IUS market was worth $7,101 million in 2009. By 2013 it will grow to $23,501 million, representing a compound annual growth rate of 34%. However, this will represent only 11.8% of the combined infrastructure managed services market in 2013. This clearly underlines the huge potential impact associated with the development of IUS, both on traditional and cloud-based architectures.

User Advice: IUS are infrastructure managed services delivered by an industrialized delivery model. All clients should:

- Gain an awareness and understanding of these new offerings in order to leverage the value for their enterprise.
- Include IUS in the set of service options under evaluation as part of their sourcing strategy and enterprise architecture.
- Investigate critical areas, including pricing mechanisms and demand management, architectural specifications and limits, impact on application software licenses, transition in and out, contract terms and conditions, security, compliance, auditing and risk management.
- Use the Gartner IUMM as a road map to follow the evolution of infrastructure toward the real-time infrastructure concept. This evolution will affect most organizations, regardless of their decision to transform and run their infrastructure internally (insourced delivery) or externally (outsourced delivery or IU).

Organizations delivering their IT infrastructure services in house should:

- Regularly check how IU offerings are advancing in the market. Increasingly, these offerings will become the external benchmark for price, efficiency and flexibility. Examples include a SAP production managed platform (excluding SAP licenses) starting

at under $10 per user per month (PUPM), or a Microsoft Exchange IU service at $5 PUPM. These entry-level prices represent the ongoing trend toward industrialized low-cost services (see "Behind the Cloud: The Rise of Industrialized, Low-Cost IT Services" and "Industrialized Low-Cost IT Services Drive Value, By Definition").

Organizations considering outsourcing deals or utility offerings should:

- Concentrate on pricing units and pricing schema — and on the related tools for service requests, metering, billing, and financial and service reporting — to understand the maturity of offerings. The degree of flexibility must align with client requirements and the maturity of the offerings.
- Request references from other clients using these offerings and pricing units, and exercise due diligence in actively checking those references.
- Ask the provider to carefully describe the processes, automation tools and SLAs underpinning service delivery quality and efficiency, because a focus on unit definition and pricing alone is insufficient to achieve the best value for money.
- Request that providers communicate their service/architecture road map: to give an understanding of how their offerings evolve over time, and to judge the potential for lock-in into their specific architecture. Ask providers how they are moving from traditional to cloud-based IU services during the next few years.
- Understand how their sourcing life cycle (sourcing strategy, vendor selection, contracting and ongoing management) will change when embracing highly standardized solutions.
- Start piloting or using IUs as part of their IT value chain.
- Request proof regarding statements of regulatory compliance and verification of security and location transparency of data stores.
- Verify the impact of software licensing models when moving from dedicated to shared IU-based hosting solutions.

Business Impact: IT IU can:

- Optimize the cost-efficiency and service effectiveness of IT infrastructure.
- Increase flexibility in response to business requirements.
- Deliver an open, predefined and automated platform for innovation.

To benefit, clients must overcome significant cultural, financial and technical issues such as standardization acceptance, independent software vendor pricing strategies, application portability, virtualization and policy-driven management for heterogeneous environments. Continuing economic uncertainty and the further rise of cloud-enabled services solutions will accelerate the evolution toward industrialized low-cost IT services.

Benefit Rating: Moderate

Market Penetration: 5% to 20% of target audience

Maturity: Early mainstream

Sample Vendors: Amazon.com; Atos Origin; AT&T; BT Global Services; Capgemini; CSC; Dell; Fujitsu; HCL Technologies; HP; IBM; Logica; Rackspace; Savvis; Siemens IT Solutions and Services; T-Systems; Terremark; Unisys; Verizon Business

Recommended Reading: "Infrastructure Utility Services: The Business Between Outsourcing and the Cloud"

"Infrastructure Utility for SAP: Comparing Five Leading Offerings"

"Infrastructure Utility for SAP: Comparing Contract Terms and Service Levels"

"Keiper: Adopting an Infrastructure Utility for Flexibility and Efficiency"

"Case Study: Areva Gains IT Flexibility Through an Infrastructure Utility"

"Oxea Shows How Infrastructure Utility Can Deliver Speed and Efficiency"

"Case Study: How IT Utilities Support Rio Tinto's IT Dynamics and Company Moves"

"Case Study: Nampac Adopts the IBM Infrastructure Utility for SAP Applications"

"Comparing Cloud Computing and Infrastructure Utility"

Software as a Service (SaaS)

Analysis By: Robert P. Desisto

Definition: Software as a service (SaaS) is software that is owned, delivered and managed remotely by one or more providers. If the vendor requires user organizations to install software on-premises using their infrastructures, then the application isn't SaaS. SaaS delivery requires a vendor to provide remote, outsourced access to the application, as well as maintenance and upgrade services for it. The infrastructure and IT operations supporting the applications must also be outsourced to the vendor or another provider.

The provider delivers an application based on a single set of common code and data definitions, which are consumed in a one-to-many model by all contracted customers at any time. Customers may be able to extend the data model by using configuration tools supplied by the provider, but without altering the source code. This approach is in contrast with the traditional application-hosting model, in which the provider supports multiple application codes and multiple application versions, or a customized data definition for each customer.

SaaS is purchased on a pay-for-use basis or as a subscription based on usage metrics. Purchasing is based on a subscription (for example, a per-user, per-month fee) or usage basis (for example, allocating a certain number of transactions for a fixed time period). A perpetual license purchase is not considered SaaS.

Position and Adoption Speed Justification: The purpose of the SaaS positioning on the Hype Cycle is to provide an aggregate view of the state of SaaS in the context of cloud computing. There are examples of SaaS, such as isolated tenancy, that would not be considered cloud computing. The SaaS positioning reflects the maturity of SaaS in the context of leveraging the cloud computing infrastructure. Because different software applications using SaaS would be placed on different points on a Hype Cycle, the post-Trough of Disillusionment positioning reflects the fact that SaaS is proven in certain markets, but remains nascent in complex application markets, such as ERP.

User Advice: Companies with complex requirements should not assume they will significantly lower their total cost of ownership or reduce complexity by moving to SaaS. Companies with tight

capital budgets, and those that are IT resource-constrained or want to get something simple deployed quickly, should consider SaaS. Even if one or more of the three elements involved in considering SaaS are not met, a SaaS solution still may be best for a company. As with any product, however, a company should evaluate the functional capabilities of the SaaS offering to meet the company's specific requirements.

Business Impact: SaaS has the effect of lowering expenses for the first two years, because it does not require an upfront capital investment. However, in outlying years, SaaS may become more expensive, because the operating expense does not decrease (see "Fact Checking: The Five Most-Common SaaS Assumptions"). SaaS is also helpful to companies that do not have IT resources to deploy and maintain on-premises software. This is prevalent in small and midsize businesses, but also is applicable to large businesses that may have experienced downsizing in the IT department.

Benefit Rating: Moderate

Market Penetration: 20% to 50% of target audience

Maturity: Early mainstream

Sample Vendors: Concur; Microsoft; NetSuite; Oracle; RightNow; salesforce.com; SuccessFactors; Taleo; Workday

Recommended Reading: "Essential SaaS Overview and 2011 Guide to SaaS Research"

"Key Issues for Software as a Service, 2011"

"SaaS Industry Report Card"

"Four Components Define Software as a Service"

Virtualization

Analysis By: Thomas J. Bittman

Definition: Virtualization is the abstraction of IT resources that masks the physical nature and boundaries of those resources from resource users. An IT resource can be a server, a client, storage, networks, applications or OSs. Essentially, any IT building block can potentially be abstracted from resource users.

Abstraction enables better flexibility in how different parts of an IT stack are delivered, thus enabling better efficiency (through consolidation or variable usage) and mobility (shifting which resources are used behind the abstraction interface), and even alternative sourcing (shifting the service provider behind the abstraction interface, such as in cloud computing). A key to virtualization is being able to effectively describe what is required from the resource in an independent, abstracted and standardized manner. In essence, cloud computing is about abstracting service implementation away from the consumers of the services by using service-based interfaces (i.e., the interface for cloud-computing services is about virtualization — an abstraction interface). To a provider, virtualization creates the flexibility to deliver resources to meet service needs in a very flexible, elastic, rapidly changing manner. The tools that make that happen could be virtual machines, virtual LANs (VLANs), or grid/parallel programming.

Position and Adoption Speed Justification: Virtualization is not simply one technology; rather, it is many technologies that are all evolving at different rates. Virtual machines for servers, for example, were introduced on the mainframe more than 30 years ago. However, virtual machines for x86 architecture servers were introduced in 2001, and, today, are used for about 40% of all x86 architecture workloads. The x86 architecture server virtualization is being adopted quickly,

however, and is expected to be used by three-quarters of all workloads by 2015. Storage virtualization is relatively mature within storage vendor offerings; originally in homogeneous forms only, there are a growing number of heterogeneous offerings. Networking is extremely mature in terms of virtualization. There are many forms of virtualization, and the challenge is choosing which form of virtualization to use. For example, server virtualization tools tend to dictate how storage and networking will be virtualized. Abstracting a resource usually means commoditizing it. Vendors will not promote an abstraction technology that hides their differentiation, and thus will promote their own virtualization technologies. There are many technologies and many competing solutions, and not all the technologies are mature yet.

User Advice: The virtualization trend has caused huge turmoil among vendors, threatening commoditization status and removing the vendor's ability to differentiate and influence buyers. Vendors are focusing on competing for ownership of the virtualization layers, and the management/automation tools that work with those virtualization layers. Be cautious about vendors that promote one technology to "virtualize" everything. Different technologies will be appropriate for different situations. Be cautious about vendors that promote one technology to "manage" everything. In the end, architectures will include many different virtualization layers, with many different management mechanisms. Effective tools will work with those different management mechanisms, rather than replace them. Virtualization is about much more than technology architecture. Virtualization causes cultural and political change. Virtualization projects, therefore, need strong executive support to drive those changes. Finally, virtualization requires processes and management tools to fundamentally change how they work. Processes need to account for speed, agility and granularity, which are very different in virtualized environments. At the highest level, management tools need to shift from managing vertical, tightly integrated silos into managing horizontal resource pools to meet service needs (leading to new architectures, such as fabric-based infrastructures). Virtualization projects that don't have effective management strategies will fail.

Business Impact: Virtualization makes it easier for IT to deliver IT capabilities faster, to have a lower barrier to access IT capabilities, and to deliver exactly the right amount of resources needed — no more and no less. This puts more pressure (and opportunity) on the user to use IT efficiently, and to make good business decisions about the use of IT. As an IT catalyst, virtualization can help a business adjust to changing market trends faster than before, transforming the business and its use of IT. However, a lubricant used badly can also cause a business to "slip and fall," negatively affecting business performance.

Virtualization starts to change the relationship between the enterprise and IT, based more on service levels, and tracked and perhaps paid based on usage (i.e., not simply as a cost center, but as an investment). In many cases, business units will need to adjust to leverage the speed and granularity that virtualization provides (for example, virtual machines can be deployed roughly 30 times faster than physical servers, which have a tendency to double demand for servers). That includes making good decisions about how many resources to ask for, and taking advantage of speed to deploy IT-based solutions much faster to meet the business's needs. Part of an effective virtualization deployment requires shifting to usage-based costing and chargeback to treat this more-fluid IT resource as any other business.

Benefit Rating: Transformational

Market Penetration: 20% to 50% of target audience

Maturity: Early mainstream

Sample Vendors: Cisco; Citrix Systems; EMC; HP; IBM; Microsoft; NetApp; Oracle; Parallels; Red Hat; VMware

Recommended Reading: "Virtual Machines Will Slow in the Enterprise, Grow in the Cloud"

"The Road Map From Virtualization to Cloud Computing"

"Q&A: The Changing Dynamics of Virtual Machine Density"

"Q&A: Six Misconceptions About Server Virtualization"

Sales Force Automation SaaS

Analysis By: Robert P. Desisto

Definition: Although there are many components of a sales force automation (SFA) solution (such as proposal generation, sales configuration and pricing management), the primary focus of software-as-a-service (SaaS) SFA is opportunity management. This is the practice of systemizing how sales channels pursue sales opportunities in the context of preferred philosophies, methodologies and strategies. As opportunities enter the sales cycle in the form of leads and move through a defined sales process, they are tracked and updated. Leads may be developed by an inside sales channel, then distributed to direct salespeople or selling partners to close. Pipeline management capabilities provide a view of sales opportunities by sales stage or potential close date. As required, salespeople can create and submit forecasts from their active opportunities. With visibility into opportunity data, sales management can inspect, coach and mentor sales representatives, while analyzing the opportunity data at each level in the sales hierarchy to predict sales performance.

The underlying customer information layer of opportunity management is contact management. Contact management systems provide internal data — such as names, addresses, interactions and product/revenue values — for the accounts in a territory and the contacts in an account. Salespeople can create and share valuable account/contact intelligence, such as account strategies, corporate goals or strategic objectives, competitive presence, and decision-making authority and disposition. Customer information can be enhanced further by incorporating external or third-party data, as well as customer profiles derived from marketing or customer service (for example, customer value or churn risks).

Position and Adoption Speed Justification: SFA SaaS is more than 11 years old. Gartner estimates there are at least 3 million SFA SaaS users. All market segments, from small to large, have validated its use, and users have a better expectation of the real benefits and costs associated with SFA SaaS. IT departments continue to be more engaged in purchasing decisions, and are becoming more involved in vendor evaluations, placing higher scrutiny on the on-demand provider's data center operations, upgrade practices and SLAs.

In some industry verticals, such as financial services, data privacy and security issues have slowed SaaS adoption somewhat; however, even in those industries where data privacy is an issue, SFA SaaS is continuing to take hold.

User Advice: Sales organizations with complex requirements (such as significant process automation outside of opportunity management and complex real-time integration) should not assume that they will significantly lower their total cost of ownership simply by moving to SaaS. Companies with tight capital budgets, those that are IT-resource-constrained and enterprises that want to get something simple deployed quickly should consider SFA SaaS. Even if one or more of the three elements involved in considering SFA SaaS is not relevant, a SaaS solution still may be best for a company. However, as with any product, the sales organization should evaluate the functional capabilities of SFA SaaS offerings to meet the company's specific requirements.

Business Impact: The primary business effect of SFA SaaS involves the capabilities that manage accounts, contacts, opportunities and sales pipelines. Sales organizations gain greater visibility, sales process formalization and help with bottom-up forecasting.

Benefit Rating: Moderate

Market Penetration: 20% to 50% of target audience

Maturity: Early mainstream

Sample Vendors: 37signals (Highrise); Aplicor; B-kin; CDC Software (Pivotal); Landslide Technologies; Microsoft; NetSuite; Oracle; salesforce.com; Sage North America; Sawfish Software; Soffront Software; SugarCRM; Zoho

Recommended Reading: "Essential SaaS Overview and 2011 Guide to SaaS Research"

"CRM Vendor Market Guide for Software as a Service, 2011"

"A Framework for Evaluating Sales Force Automation Application Functionality"

"Key Issues for CRM Sales Strategies, Processes and Technologies, 2011"

Cloud Advertising

Analysis By: Andrew Frank

Definition: Cloud advertising refers to cloud service business processes that support selection, transaction and delivery of advertising and ad-related data where content and price are determined at the time of end-user access, usually by an auction mechanism that matches bidders with impressions as they become available. Search engine marketing (SEM) and various forms of online display advertising (for example, banners) are the most-developed formats, but the concept is spreading to other channels and platforms, such as mobile devices, online video, addressable television and out-of-home digital signage.

Position and Adoption Speed Justification: Cloud advertising has been developing steadily since its search-based origins in the late 1990s, although its classification as "cloud" began in 2009, when cloud computing unleashed a wave of scalable real-time marketplace capabilities. Search continues to command the largest share of cloud advertising services, although growth in search has begun to slow while other digital formats, notably online display, mobile, social, and video, have been accelerating. The widespread application of these technologies to television advertising, still the largest global ad medium, is yet to come.

Cloud advertising is forecast to be the largest contributor to the growth of cloud services revenue during the next five years and, as such, is a large and slow-moving trend. Although it is considered somewhat mature in the online category, we expect it to spread to other digital media, and it is the focus of major investments all along the advertising, media and communications value chain.

The designation of this category as "business process as a cloud service" acknowledges the significance of its revenue contribution to the development of cloud infrastructure and applications.

A note on market penetration: as in the 2010 Hype Cycles, we've identified cloud advertising as having penetrated more than 50% of the target audience. It is rare that Gartner reports this level of penetration for a technology that has not yet entered the Plateau of Productivity, however we believe it's justified in this case based on the penetration of search engine marketing, a significant

subset of cloud advertising, but just the first in a broad composite of cloud advertising formats to reach maturity.

User Advice: Marketers should hire or partner with specialists to realign advertising practices around cloud-based processes.

Marketers especially need to develop integrated enterprisewide platforms and approaches to measure and optimize marketing activities across channels, based on the capabilities presented by cloud advertising services.

Cloud service vendors should understand the opportunities and costs associated with various cloud advertising models, such as real-time bidding, and evaluate their impact on and applicability to cloud architectures.

Business Impact: Cloud advertising will have a high impact on the economics and accountability of advertising and marketing in general. Advertising is often characterized as largely unaccountable spending, particularly when economic conditions bring cost-cutting pressures to bear. Cloud-based marketing data and exchanges that support highly targeted delivery of rich messages and interactive response will create new opportunities to reduce risk and improve the efficiency and effectiveness of advertising, tying activities more closely to sales and, thus, making it less vulnerable to cuts.

Benefit Rating: Transformational

Market Penetration: More than 50% of target audience

Maturity: Early mainstream

Sample Vendors: AOL; Apple; AppNexus; Baidu; Facebook; Google; Microsoft; OpenX; Yahoo

Recommended Reading: "Forecast: Sizing the Cloud; Understanding the Opportunities in Cloud Services"

"Real-Time Bidding Heralds Growth in Cloud Advertising"

Dedicated Email Services

Analysis By: Matthew W. Cain; Tom Austin

Definition: Dedicated email services (DES) are one of three off-premises ways of provisioning email. The other two are outsourced email services and true multitenant mail (MTM) services, the only type that meets Gartner's definition of cloud-based services.

Although DES do not meet our definition of cloud-based services, they are more cloud-like than outsourced email services. Outsourced email can be delivered in a wide range of ways, with customized terms and conditions — at a cost. DES, by contrast, are highly standardized and typically "fixed price, fixed service" offerings, such as Microsoft Exchange Online Dedicated.

DES do meet Gartner's definition of software as a service (SaaS), but they fail the multitenancy test. They represent the ultimate evolution of the earlier application service provider business model.

True cloud-based, MTM services (such as the Gmail component of Google Apps Premier Edition [GAPE] and Microsoft Exchange Online Standard) reflect a more advanced underlying technology model, offering, at least in theory, greater flexibility and lower cost. Like DES, true MTM offerings are highly standardized and typically come as fixed-price, fixed-service offerings.

With Office 365, Microsoft will limit sales of its dedicated implementation of Exchange Online to organizations with at least 30,000 users. There will therefore be limited uptake of DES due to vendor restrictions. The extra vendor costs for DES, as compared with MTM, are typically passed on to customers.

Position and Adoption Speed Justification: DES has passed the Trough of Disillusionment and is approaching the Plateau of Productivity. This advance reflects a growing understanding of how organizations can work within the constraints of the system, and better vendor support in the form of, for example, application programming interfaces that enable an on-premises trouble ticketing system to take direct operational data from the vendor's platform.

User Advice: Although Gartner believes that MTM services will eventually dominate SaaS email, we also believe that DES will thrive in the long term for more security-conscious customers.

Business Impact: Both DES and MTM services offer enterprises the opportunity to exploit many of the benefits of SaaS and cloud provisioning models, such as fixed pricing, elasticity, avoidance of complexity and "evergreen" technology.

Benefit Rating: Low

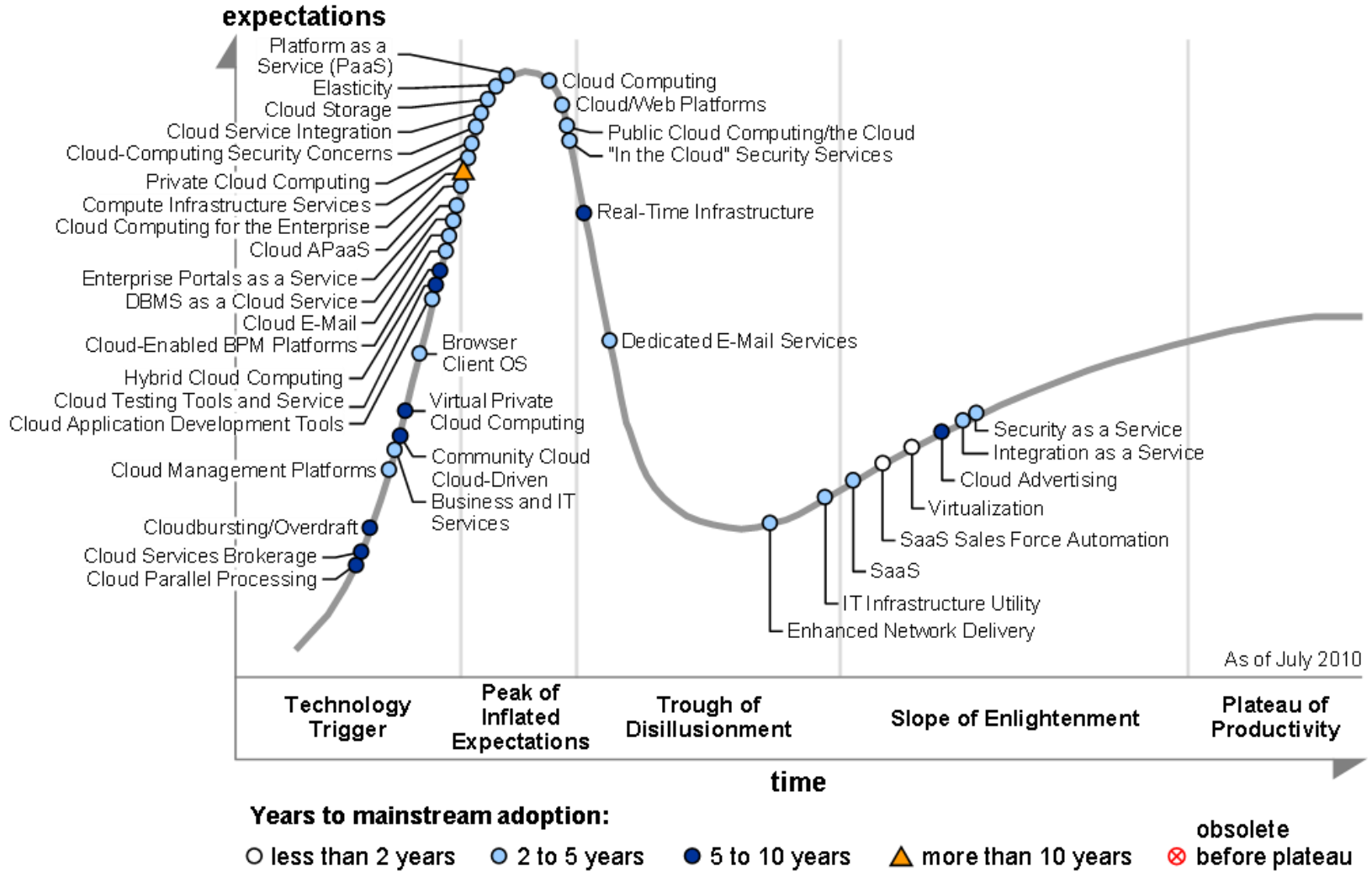
Market Penetration: 1% to 5% of target audience

Maturity: Adolescent

Sample Vendors: Microsoft

Appendixes

Figure 3. Hype Cycle for Cloud Computing, 2010



Gartner

Source: Gartner (July 2010)

Hype Cycle Phases, Benefit Ratings and Maturity Levels

Table 1. Hype Cycle Phases

Phase	Definition
Technology Trigger	A breakthrough, public demonstration, product launch or other event generates significant press 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 technology is pushed to its limits. The only enterprises making money are conference organizers and magazine publishers.
Trough of Disillusionment	Because the technology 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 technology'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 technology 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 technology to reach the Plateau of Productivity.

Source: Gartner (July 2011)

Table 2. 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

Benefit Rating	Definition
Low	Slightly improves processes (for example, improved user experience) that will be difficult to translate into increased revenue or cost savings

Source: Gartner (July 2011)

Table 3. Maturity Levels

Maturity Level	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 2011)

RECOMMENDED READING

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

"Understanding Gartner's Hype Cycles, 2011"

"The What, Why and When of Cloud Computing"

"Key Issues for Cloud Computing, 2011"

This research is part of a set of related research pieces. See "Gartner's Hype Cycle Special Report for 2011" for an overview.

REGIONAL HEADQUARTERS

Corporate Headquarters
56 Top Gallant Road
Stamford, CT 06902-7700
U.S.A.
+1 203 964 0096

European Headquarters
Tamesis
The Glanty
Egham
Surrey, TW20 9AW
UNITED KINGDOM
+44 1784 431611

Asia/Pacific Headquarters
Gartner Australasia Pty. Ltd.
Level 9, 141 Walker Street
North Sydney
New South Wales 2060
AUSTRALIA
+61 2 9459 4600

Japan Headquarters
Gartner Japan Ltd.
Aobadai Hills, 6F
7-7, Aobadai, 4-chome
Meguro-ku, Tokyo 153-0042
JAPAN
+81 3 3481 3670

Latin America Headquarters
Gartner do Brazil
Av. das Nações Unidas, 12551
9° andar—World Trade Center
04578-903—São Paulo SP
BRAZIL
+55 11 3443 1509

Gartner