Gartner

Research

Publication Date: 18 July 2008 **ID Number: G00158969**

Hype Cycle for Real-Time Infrastructure, 2008

Donna Scott, Dave Russell, Ronni J. Colville, Carl Claunch, Rakesh Kumar, John R. Phelps, Philip Dawson, Thomas J. Bittman, Debra Curtis, Will Cappelli, David Williams, Milind Govekar, Kris Brittain, Mark A. Beyer

Real-time infrastructure continues making gains in its longer-term journey, spurred on significantly by the spread of virtualization in the data center and by service-oriented architectures. However, significant technology hype, standardization and process inhibitors still exist.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Gartner

What You Need to Know

Gartner has analyzed and published research on real-time infrastructure (RTI) for more than six years. Although we are closer to seeing RTI realized than we were in 2002, a data-center-wide RTI is still a vision and five to 10 years to reality. However, the past three years have spurred RTI greatly due to new service-oriented architectures (SOAs), as well as data center space and energy issues that are forcing systematic changes in strategy and implementation (with consolidation and virtualization). Server virtualization, in particular, is showcasing the potential for RTI as companies that implemented it for development and test efficiencies turn to production environments. They are consolidating first and then testing virtual machine (VM) movement and policy-based VM movement capabilities (which provide a form of RTI).

Moreover, cost pressures and agility requirements are increasing the focus on standardization and on automation for greater efficiencies and speed. As IT organizations mature toward service alignment, they focus on optimization of service delivery, which has increased the interest in RTI. For these reasons, most IT organizations see RTI as a key part of their next-generation data center vision (see "Survey Shows the RTI Journey Continues"). As a result, despite some large vendors de-emphasizing their RTI messages (for example, IBM On Demand and Sun Microsystems Sun N1), the user momentum to build a strong, agile and efficient IT platform to meet ever-changing business needs will continue to drive RTI during the long term.

The Hype Cycle

RTI has progressed beyond the conception stage and has materialized into early solutions being deployed by visionary enterprises. These solutions consist of infrastructure and/or IT operation management technologies (such as server virtualization, server provisioning and configuration management, and run book automation), which generally have been highly customized to achieve early RTI architectures and their associated benefits. However, the promise of completely dynamic management of IT service levels is far off. Today's early solutions offer partial RTI architectures and are focused on:

- Dynamic (and self-service) provisioning of development and test environments
- Virtualization and workload movement
- Reconfiguring capacity during failure or disaster events
- SOA and Java Platform, Enterprise Edition (Java EE) environments for dynamic scaling of application instances
- Application scaling (solutions, such as those from SAP, built for specific applications)
- Using run book automation technology to integrate multiple management processes and toolsets as a prerequisite for orchestrating processes and tools for RTI
- Repurposing production environments, for example, from daytime online transaction processing (OLTP) to night-time batch optimization
- Emerging software-hosting environments that offer runtime optimization for applications built on their platforms (hosted internally or at service providers/cloud providers)

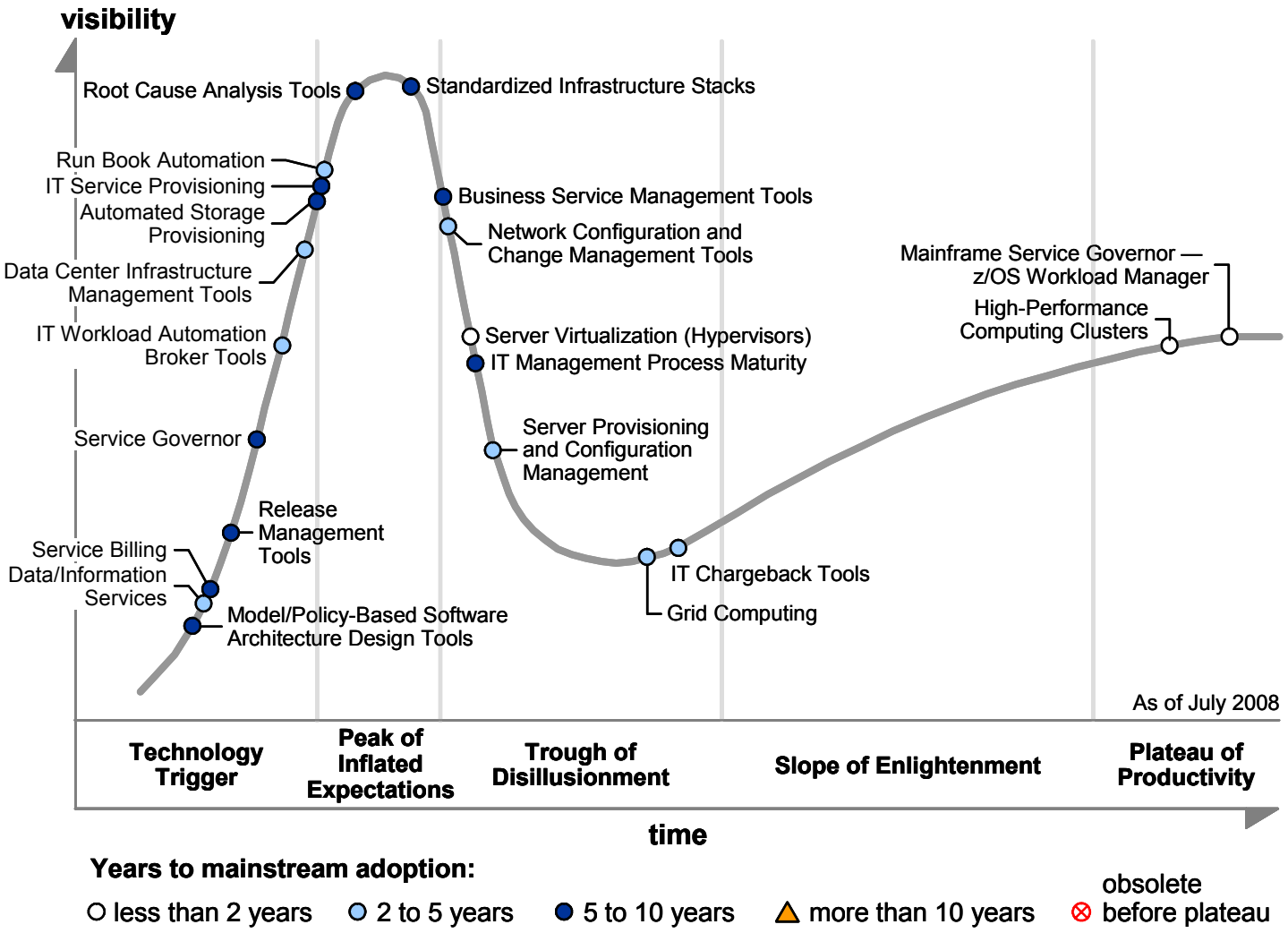
Completing the full objectives of data-center-wide RTI will require a higher-level understanding of IT services, service-level agreements (SLAs), and dynamic capacity and configuration management — which are still many years off. Inhibitors such as these remain:

- Some technologies are embryonic, especially in model-based/policy-based design tools and the service governor
- A lack of IT management process maturity and technology maturity (see "Survey Results: IT Infrastructure and Operations Management Maturity")
- Resistance from enterprises to adopt policies for standardized hardware and software infrastructure
- Organizational cultures associated with dedicated hardware to specific IT services are not conducive to shared use for RTI

Overcoming these challenges will take at least five to 10 years for most enterprises to achieve; thus, businesses should appropriately set their RTI expectations and their plans to get there. With relentless pressures to reduce costs and increase quality of service and agility to respond to changes in business requirements, visionary and early adopter companies that are investing to remove barriers surrounding the harder-to-achieve cultural and process changes will benefit from these investments with earlier realization of deployments of early RTI technologies. At the same time, every technology and platform supplier wants to control its runtime environment and not have others control it. For this reason, solutions that support more-homogeneous environments (for particular platforms, applications and others) will be the norm in the short term to midterm; the longer-term potential for RTI across heterogeneous environments will depend on the ability of vendor partnerships and on standards to take effect, specifically in the area of configuration management (for example, to standardize on how resources are talked to and change state, as well as how IT services are described).

The desires of IT leaders for data-center-wide RTI and its promised benefits are ahead of the technologies and of the IT department's ability to implement them. In addition, architectures are becoming more complex, rather than more standardized, which further threatens the ability to achieve RTI. The RTI Hype Cycle will aid these leaders in identifying important prerequisites in technologies and other people- and process-oriented inhibitors as they chart a course toward RTI.

Figure 1. Hype Cycle for Real-Time Infrastructure, 2008



The Priority Matrix

The Priority Matrix maps a technology's or other prerequisite's time to maturity on a grid for an easy-to-read format that answers these questions:

- How much value will an enterprise get from a technology or other prerequisite to RTI?
- When will the technology be mature enough to get this value?
- In the case of a people/process inhibitor, when will most enterprises surpass the obstacle that inhibits their ability to achieve RTI?

The Hype Cycle graphic is generated from the Benefit Rating and Time to Plateau data.

High people, process and technology maturity levels are needed to dynamically and automatically manage workloads, resources and applications to meet the SLAs that are set for IT or business services. Technology maturity is achieved partially through standardization and virtualization; both appear directly on the RTI Hype Cycle, as does IT management process maturity:

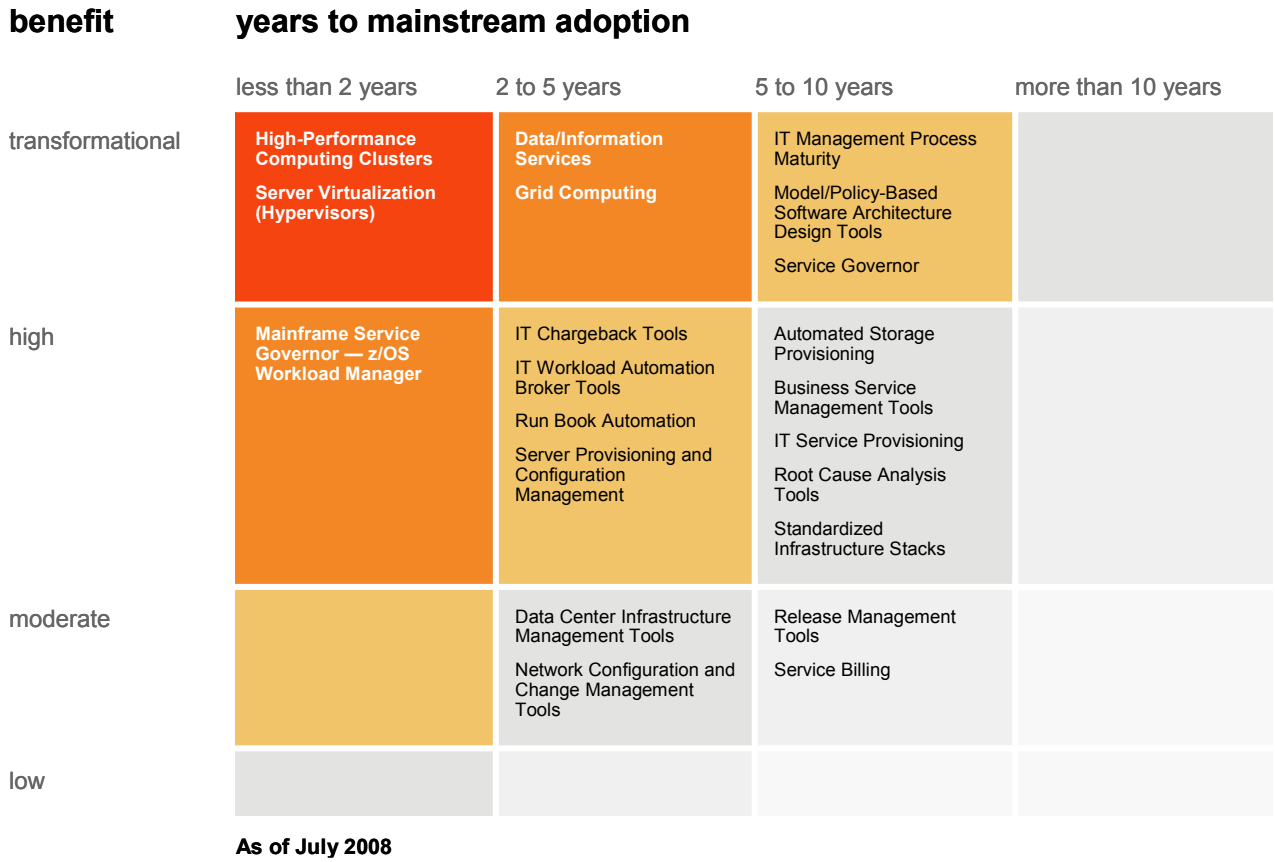
- Standardization is less specific to technologies and is more often a conceptual set of architectural policies and enforcement that enterprises put in place to reduce the complexity of their hardware and software environments. Therefore, it enables a greater degree of automation and agility.
- Virtual server technologies enable abstraction between software and hardware layers. Adoption is common in development/test environments and is growing rapidly in production environments. With server virtualization comes the promise of greater levels of dynamic workload optimization and agility; however, it adds challenges in managing and securing the environment. In addition, because IT services will continue to be deployed on physical and virtual servers, RTI architectures must enable optimization across physical and virtual, not just virtual, levels. Moreover, there are many levels of virtualization in servers, storage, clients, applications, and databases that also must be managed. Nevertheless, virtual server technologies have been an important data point for RTI, to showcase the potential for dynamic optimization and benefits.

The IT management process maturity is positioned midway to the Trough of Disillusionment on the Hype Cycle — with a five- to 10-year time frame to reach maturity. This is because transforming IT to run like a business requires massive cultural, organizational and process changes and takes a minimum of two to three years to achieve, with many processes taking five years or more. Gartner's Strategic Planning Assumption states that, by year-end 2012, fewer than 14% of IT infrastructure and operations [I&O] organizations in large enterprises will have achieved Level 4 service alignment or above, based on our IT I&O maturity model (see "Introducing the Gartner IT Infrastructure and Operations Maturity Model"). For small or midsize businesses, the numbers are significantly lower at just a fraction of large enterprises.

From a technology perspective, the primary longer-term inhibitors to RTI include root cause analysis tools (to know the underlying cause of impending outages or performance slow-downs so that they can be corrected dynamically) and the service governor (to understand IT services, SLAs and resources and to optimize them dynamically to achieve the SLAs). Provisioning and configuration management technologies have advanced for servers; however, the lack of standardized software configurations and release management processes have prevented widespread, mainstream deployment of these technologies. Despite all the inhibitors, RTI (and several key technologies in the architecture) is considered transformational, with significant

benefits, including faster time-to-market for new functionality/releases, greater agility in making changes more frequently and faster, higher quality of services and lower costs.

Figure 2. Priority Matrix for Real-Time Infrastructure, 2008



Source: Gartner (July 2008)

On the Rise

Model/Policy-Based Software Architecture Design Tools

Analysis By: Donna Scott; Milind Govekar; Will Cappelli

Definition: Based on concepts derived from model-driven software architecture disciplines, these tools enable application developers and architects to represent the logical architecture of IT services, that is, the dependencies among the various software components that support IT services. In addition, policies are defined to enforce the quality of service delivery, scalability, access constraints and hardware/hosting requirements. The IT service logical architecture becomes a "living" and accurate view of IT service dependencies and policies, as opposed to the one-time, nonmaintained view typically performed in Microsoft Visio.

Position and Adoption Speed Justification: The success of the IT operations quality mission (near-100% availability and consistently high performance) is only partly under the control of the IT department. The application development community also plays a critical role early in the life cycle of a new project, ensuring that the application and infrastructure components introduced during planning, building and testing fit smoothly into the environment to further incorporate

quality and security goals. Unfortunately, operations and application development communities typically work largely in isolation from each other. This is due to two factors: first, there is no broadly accepted meta model with the sophistication capable of supporting a development view and an operational view of IT components and services. Second, a profound cultural gap separates the two communities, with developers thinking in terms of projects, with specified lifetimes and abstract formal structures and operations professionals thinking in terms of perpetual support and meeting required service-level agreements.

The ongoing maturity of IT organizations aligned toward Level 4 service requires greater integration between the application development and IT operational roles; this will be a major driver for these tools. Moreover, the desire to implement real-time infrastructure (RTI) architectures, enabling more-streamlined, policy-based automation (with lower labor costs) is also a key driver. However, the lack of maturity and technology are barriers, thus, the tools will take five to 10 years to mature to mainstream. So, although the emergence of standard proposals, such as service modeling language and the operational focus of configuration management database as a consolidation point for configuration data to support IT services, some progress is being made toward the articulation of a cross-community meta model, it is slow going. In the end, the ability to meet increasingly stringent quality and security objectives will require negotiating the cultural divide. Thus, developers must design, plan and build with operational requirements in mind, and operations professionals must think increasingly of the systems they manage in terms of shared environments, with policies and defined dependencies driving automated provisioning and management of IT services.

Some limited progress is being made in modeling and design tools with server virtualization as the deployment platform. However, it will be five to 10 years at least before general heterogeneous application design tools supporting physical and virtual environments become mainstream.

User Advice: Get involved in application standards and design, ensuring that manageability is considered early in the phases. As modeling tools become available, assess them from an IT operations perspective, and ensure the interoperability of the models with other modeling tools.

Business Impact: Modeling represents the future foundation of RTI and its ability to offer significant benefits in cost savings, quality of service and agility.

Benefit Rating: Transformational

Market Penetration: Less than 1% of target audience

Maturity: Embryonic

Sample Vendors: 3Tera

Data/Information Services

Analysis By: Mark Beyer

Definition: Data/information services consist of processing routines that provide data manipulation pertaining to data storage, access, delivery, semantic interpretation, stewardship and governance. Unlike point-to-point data integration solutions, data services decouple data storage, security and mode of delivery from each other, as well as from individual applications, to deliver them as independently designed and deployed functionality that can be connected via a registry or composite processing framework. Specifically, the metadata interpretation between business process models, semantic usage models and logical/physical data models will enhance the overall adaptiveness of IT solutions. Data services can be used in a networked fashion that is

orchestrated through a composite processing model or designed separately, then reused in various, larger-grained processes.

Position and Adoption Speed Justification: Data/information services are, by their nature, a new style of data access strategy that replaces the data management, access and storage duties currently deployed in an application-specific manner with an abstracted and invoked style that includes local interpretation based on the use case or context. Data services architecture is a subclass or category of service-oriented architecture (SOA) that does not form a new architecture, but brings emphasis to the varying services that exist within SOA. Most of the large vendors have announced road maps and plans to pursue some variant of the data service approach. This is an evolutionary architectural style that does not warrant "rip and replace" at this time, and will coexist with current application design techniques. Disillusionment will occur as organizations realize the wide variation in services grain, ranging from possibly thousands of very fine-grained services performing rudimentary tasks through coarse-grained and composite services.

User Advice: Data services are a granular service category under a wider SOA. Organizations should focus on delivering a semantic layer to support the use of granular data services. Data services can begin with data quality tools used to deploy reusable, loosely coupled data stewardship and governance services that run as process calls from within operational application interfaces and data integration efforts. In 2007, this technology class continued its focus on information in traditional data only (formerly "structured"). Similarly, initial advances in using model-to-model (M2M) language communication via metadata operators were blended into this technology. In addition, data quality tools, profiling tools and metadata tools started to contribute to this approach. The M2M introduction caused a temporary retrograde in the technology position in 2008, but will eventually accelerate the movement along the cycle. Integrated and accessible metadata is mandatory for this technology and the key to interoperability. Shared and common metadata is one approach, but as this technology progresses, services within these tools could pass data between them with accompanying metadata as well. Organizations should avoid vendor solutions and development platforms that do not support interoperability with third-party tools.

Business Impact: Data services are not an excuse for each organization to write its own, unique database management system, as most database management systems store data and provide ready access. Data services can sever the tight links between application interface development and the more infrastructure-style decisions of database platforms, operating systems and hardware. This will create a portability of applications to lower-cost repository environments when appropriate, and create a direct corollary between the cost of information management and the value of the information delivered, by delivering semantically consistent data and information to any available presentation format. This is opposed to the current scenario in which monolithic application design can drive infrastructure costs up because of their dependence on specific platform or database management system capabilities.

Benefit Rating: Transformational

Market Penetration: 1% to 5% of target audience

Maturity: Emerging

Sample Vendors: Ab Initio; Business Objects; IBM; Informatica; Oracle

Recommended Reading: "The Emerging Vision for Data Services: Becoming Information-Centric in an SOA World"

"Data Integration Is Key to Successful Service-Oriented Architecture Implementations"

"Service-Oriented Business Applications Require EIM Strategy"

Service Billing

Analysis By: Milind Govekar

Definition: Service-billing tools differ from IT chargeback tools in that they use resource usage data to calculate the costs for chargeback and aggregate it for a service. Alternatively, they may offer service pricing options (such as per employee, per transaction) independent of resource usage. When pricing is based on usage, these tools are able to gather resource-based data across various infrastructure components, such as servers, network, storage, databases and applications. Service-billing tools perform proportional allocation — based on the amount of resources (including virtualized resources) allocated and used by the service — for accounting and chargeback purposes. Service-billing costs include infrastructure and other resource use (such as people) costs, based on service definitions. These tools will also evolve to work with service governor(s) to set a billing policy that takes cost as a parameter, and ensure that the resource allocation is managed based on cost and service levels.

These tools will first emerge and be deployed in a service provider environment due to their business imperative, as opposed to in internal IT environments.

Position and Adoption Speed Justification: Shared sets of resources require the ability to track usage for billing purposes. As the IT infrastructure moves to a shared-services model, resource-oriented chargeback models will evolve into end-to-end collection and reporting approaches. Furthermore, these tools will work with service governor(s) to proactively manage cost of services and their associated service levels.

User Advice: This is a new and emerging area. Currently, there are no tools available. Service-billing tools will take a life cycle approach to services, will perform service cost optimization based on underlying technology resource usage optimization over the entire life cycle and will enable granular cost allocation.

The initial emergence of these tools, to accommodate the vision of dynamic automation of real-time infrastructure (RTI), will start to integrate and work with virtualization automation tools that dynamically move virtual environments based on resource or performance criteria to assess the cost effect of such movement. Enterprises that want to incorporate, for example, virtual server movement automation in their environments should assess these tools, as they emerge, to assist with controlling costs in their data center.

Service chargeback tools that are available today do the aggregation of costs mainly for a static environment where there is no automation or dynamic control of resources. These tools will not only emerge from new startups, but also from traditional chargeback vendors, asset management vendors, virtualization management vendors and software stack vendors.

Business Impact: These tools are critical to running IT as a business, by determining the financial effect of sharing IT and other resources in the context of services.

Benefit Rating: Moderate

Market Penetration: Less than 1% of target audience

Maturity: Embryonic

Sample Vendors: CA; Digital Fuel; Lontra

Release Management Tools

Analysis By: Ronni Colville; Donna Scott; Kris Brittain

Definition: Today, most IT organizations' release management processes and automation vary in terms of formalization and documentation from development to deployment. Some organizations have leveraged application developers' release management tools, others have leveraged IT service desk suites, and yet others have adapted e-mail or workflow tools to automate aspects of release management. Whatever methods are used, release management, as defined by IT operations, addresses the coordinated effort that manages the introduction of significant technology changes into the production environment, whether the technology is a newly developed application, a newly introduced piece of infrastructure, or a modification to an existing technology that could consist of multiple bundles or release activities.

To govern the release process, newly emerging tools will provide workflow functionality to support the release process activities of documenting, planning, building, testing and scheduling. Embedding policy rules within the tools provide for standardization of procedures, as well as the ability to report on performance and compliance (such as release success/failure and releases compliant with release windows). To support release analysis, release management tools must support grouping of changes into release deployment activities; thus, these new tools will demand integration with change and configuration management tools. Postimplementation review of releases deployed also will require integration with incident, problem and change management tools.

Position and Adoption Speed Justification: There has been a renewed and broadened appreciation for production-focused release management rigor due to two main factors: Information Technology Infrastructure Library (ITIL) process improvement initiatives and the increase in deployment of service-oriented architecture (SOA)-based applications. Many IT organizations have ITIL initiatives that are focused on several processes (such as problem, incident, change and configuration), where release is often one of the processes being addressed — but often later in the timeline of the initiative.

Additionally, with the increased number of composite applications being deployed (such as Java EE and .NET) into production, release management becomes a critical complement to configuration management as a means of ensuring the "piecemeal" updates do not have a negative effect on existing application components. This is in contrast to more-traditional applications where developers manage the development and release to production process, and then "hand over" the update to production control to roll out. Whereas traditionally major updates are scheduled a few times a year, in contrast, these updates are made more often.

User Advice: IT organizations need to ensure that release management for release planning and release distribution activities are defined and mapped to the specifics of the particular stakeholder. Because releases occur throughout and across all the IT application and infrastructure organization, release management will require the participation of many IT groups and may be considered parts of development, operational change management, production acceptance/control, and the tail-end of the IT delivery life cycle. With the addition of SOA-based applications, the granularity and frequency will add to the fluency with which releases will need to occur. Successful release management will likely require a competency center approach to enable cross-departmental skills for release activities.

Release management takes on a greater significance, much as production control departments did in the mainframe era. Here, planned changes to applications, infrastructure and operations (such as system and application management, documentation and job scheduling) processes are integration-tested with the same rigor that occurs today on the development side. In addition, working with architecture groups, more-rigorous policies are put in place for maintenance

planning (such as patches and upgrades) and retirement of software releases and hardware platforms — in adherence to standard policies. The release management group will also be responsible for preproduction testing and postproduction validation and adjustments to any policies regarding dynamic capacity management and priorities for real-time infrastructure (RTI).

Business Impact: Since new releases are the first opportunity IT customers have to experience IT's capabilities, success or failure of the release management process will greatly affect the future business-IT relationship. It is important, therefore, that releases are managed effectively from inception to closure, and that all IT groups work in concert to deliver quality releases as consistently as possible.

Benefit Rating: Moderate

Market Penetration: 1% to 5% of target audience

Maturity: Emerging

Sample Vendors: BMC Software; CA; HP; IBM Tivoli

Service Governor

Analysis By: Donna Scott

Definition: A service governor is a runtime execution engine that has several inputs: business priorities, IT service descriptions (and dependency model), service quality and cost policies. In addition, it takes real-time data feeds that assess the performance of user transactions and the end-to-end infrastructure, and uses them to dynamically optimize the consumption of real and virtual IT infrastructure resources to meet the business requirements and service-level agreements (SLAs). It performs optimization through dynamic capacity management (that is, scaling resources up and down) and dynamically tuning the environment for optimum throughput given the demand. The service governor is the culmination of all technologies required to build the real-time infrastructure (RTI), and it's the runtime execution management tool that pulls everything together.

Position and Adoption Speed Justification: The service governor is the "brain" behind the RTI, with knowledge of the policies, resources, services and SLAs, and which executes the services to achieve the SLAs. Service governors have emerged for homogeneous siloed infrastructure stack environments — for example, for virtual server environments and Java EE environments. However, it will take five to 10 years for rich, hierarchically organized service governors supporting cross-silo infrastructure stacks to emerge. Ultimately, service governors will be decentralized and hierarchical in nature, with intelligence emanating from the bottom through physical and virtualized resources, followed by applications, IT services and business processes.

User Advice: It's too early to plan for the service governors that will manage across infrastructure stacks. However, IT organizations (including IT operations and application development) can lay the groundwork by ensuring that manageability is a fundamental design goal for all application and infrastructure components, that it's planned early in their respective life cycles and includes automated responses (that is, designing "sense, reason and respond" functionality, where possible, using available technology). Organizations with a large homogeneous infrastructure stack environment (such as x86-based virtual servers and service-oriented-architecture-based applications) can benefit from sharing the resources and implementing early service governors to increase use, reduce cost and increase agility.

Business Impact: This "future" technology has the potential to choreograph and direct the realization of optimal relationships among cost, risk and service quality in real time. It can also be viewed as the key element in the massive automation of IT operations, which will enable IT

services to be delivered similarly to how highly automated factories deliver products. As such, service governor technology will be a key element of a modern, heterogeneous "lights out" data center environment of the future.

Benefit Rating: Transformational

Market Penetration: Less than 1% of target audience

Maturity: Emerging

Sample Vendors: DataSynapse; Novell; VMware

IT Workload Automation Broker Tools

Analysis By: Milind Govekar

Definition: IT workload automation broker (ITWAB) technology is designed to overcome the static and manual nature of scheduling jobs. It's capable of managing mixed workloads based on business policies, where resources are assigned and unassigned in an automated fashion to meet service-level objectives. ITWAB tools automate processing requirements based on events, workloads, resources and schedules. They manage dependencies across applications and infrastructure platforms, within and among companies. ITWAB technology uses service-oriented architecture (SOA) because it enables standards-based integration — for example, using Web services to integrate with various platforms and services. Interoperability and integration standards between ITWAB tools and job scheduling tools will be primarily based on proprietary standards or Web services, thereby introducing some form of vendor lock-in.

Position and Adoption Speed Justification: Not all ITWAB characteristics are known at this time. For example, a policy engine that takes business policies and converts them into a technology service-level agreement (SLA) is three to five years away from being deployed in production environments. It will emerge in vertical industry segments — for example, insurance, where a given set of standardized processes is driven by common taxonomy business policies. An intermediate stage will be the manual conversion of business policies into technology SLAs. Visibility, discovery and optimization of resource pools across the entire computing environment isn't possible today. Intermediate solutions based on targeted environments, such as server resource pools through the use of virtualization management tools, will emerge first. These tools will work with run book automation (RBA) tools to automate IT operations processes, particularly for provisioning, change and configuration management.

User Advice: Although the visibility of ITWAB tools is high, enterprises should be careful when choosing them because commercial off-the-shelf tools in this category aren't yet fully mature. Enterprises that need these tools today will have to manually integrate various commercial job scheduling tools with other tools.

Business Impact: ITWAB tools will have a big impact on end-to-end automation, and will be involved in the initial stages of implementing the service governor concept of real-time infrastructure.

Benefit Rating: High

Market Penetration: Less than 1% of target audience

Maturity: Emerging

Sample Vendors: BMC Software; CA; IBM Tivoli; Tidal Software; UC4 Software

Recommended Reading: "IT Workload Automation Broker: Job Scheduler 2.0"

Physical Resource and Infrastructure Management Tools

Analysis By: Rakesh Kumar

Definition: Physical resource and infrastructure management tools consist of software products that collate, manage and model data center hardware components such as IT hardware, cables, power distribution units and generators, and space. They are used to design new data centers as well as to optimize existing data centers, such as for thermal, electrical and spatial characteristics. Additionally, they are designed to be used in the day-to-day operations of data centers by providing change, configuration and asset management capabilities.

Position and Adoption Speed Justification: These tools are maturing steadily to capitalize on the changing design points of computer equipment and facilities components. For example, modern intelligent power strips and servers gather real-time energy consumption information that can be used by these tools. The tools then give a reasonably accurate picture of how and where energy is being used in different parts of the data center. Moreover, they can give a good prediction of future needs and enable "what if" analysis. Additionally, they integrate with event monitoring tools to assist with the management of service delivery.

As data center designs lead to the requirements for a more-integrated software management environment, linking traditional system management tools to building management systems, these tools will become more essential and loosely integrated into automated IT processes (such as in automated provisioning of IT services based on facilities monitoring and availability information). Real-time infrastructure (RTI) architectures will use information from these tools to make key decisions regarding IT service provisioning, scaling and cost optimization.

User Advice: Users looking to make significant changes to their data center environment (such as refurbishment or rebuilding) should evaluate these tools before they start making changes. In addition, users should examine if there is any overlap with existing configuration management database (CMDB) change management tools, and adjust their operational processes to take full advantage of the new tools.

Business Impact: Use of these tools will optimize energy consumption and change management, and should lead to lower data center operational costs. Other effects will be better use of data center space and more-efficient asset tracking.

Benefit Rating: Moderate

Market Penetration: 1% to 5% of target audience

Maturity: Adolescent

Sample Vendors: American Power Conversion; Aperture; Global Data Center Management; Rackwise

At the Peak

Automated Storage Provisioning

Analysis By: David Russell

Definition: The automated creation, addition and modification of storage capacity, from the array, through the network, to the host and application in a structured and repeatable workflow that can be triggered by external events and completed in a fast and reliable manner. Automated storage provisioning can integrate storage management into IT change orchestration to make IT infrastructure management more responsive to changing business requirements. The goal is to

more efficiently deliver storage resources to improve business agility and customer service quality.

Position and Adoption Speed Justification: Robust and complete device application programming interfaces (APIs) are becoming available to software developers. However, customers must organize for effective storage management to take advantage of these tools. Storage provisioning requires a well-tested, process-based approach, which will take clients time to develop and refine before they can successfully implement it in their solutions. Although APIs and scripting via a command-line interface help with adoption, impediments to adoption and Time to Plateau include: distributed buying centers cross systems and storage management, lack of a budget line item for these tools and storage domain concerns over automation in general.

User Advice: IT organizations looking to implement a holistic real-time infrastructure (RTI) capability should investigate products that provide automated storage-provisioning capabilities and a robust solution. Storage provisioning capabilities are often integrated into or delivered in connection with storage resource management (SRM) products. This capability can also be scripted into some enterprise system management and workflow products; however, SRM offerings have the advantage of understanding the storage domain and being better-positioned to respond to storage capacity requests. Major system management vendors have improved the linkage between their systems and storage toolsets, offering better integration; requiring less scripting; offering a more-unified management console for drawing an end-to-end topology of a line-of-business view of the IT assets, including storage; and have made strides toward offering combined policy, alerting and reporting. Storage virtualization in the form of block-level, file-level and storage area network (SAN) virtualization are often building blocks used to implement automated storage provisioning.

Business Impact: Automated storage provisioning, as part of an overall RTI, is important to ensure that the appropriate amounts and quality of storage capacity are made available in a timely manner and enable organizations to meet the agreed-on service levels. Without automated storage provisioning, organizations must respond to storage capacity requests by manually provisioning storage, which can be error-prone and ad hoc, thus prolonging the time to delivery and, therefore, not truly supporting an RTI environment, or overallocating capacity by making standard quantities and qualities of storage preallocated in pools of reserve capacity that can be assigned, which can be inefficient because it is static.

Benefit Rating: High

Market Penetration: 1% to 5% of target audience

Maturity: Emerging

Sample Vendors: EMC; HP; IBM; Symantec

Recommended Reading: "Magic Quadrant for Storage Resource Management and SAN Management Software, 2007"

"Storage Resource Management and SAN Management in 2012"

IT Service Provisioning

Analysis By: Donna Scott

Definition: IT service provisioning is the automation technology underpinning business or IT service modeling, and enabling automated provisioning, configuration, activation and deployment of all resources associated with the IT service in production.

Position and Adoption Speed Justification: Early technology is beginning to emerge for IT service modeling and provisioning for production environments; however, mature technology exists for test lab automation (virtual server provisioning only). Service provisioning technology must enable deployment and configuration across all platforms underlying an IT service, and across physical and virtual environments.

User Advice: IT organizations should focus on improving their project and release management processes, including the use of automation to drive changes in configurations and new releases. Moving to IT service provisioning will require mature IT management processes — or Level 4 Service Aligned in our IT infrastructure and operations maturity model — as well as orchestration of provisioning and configuration management tasks across technology domains. Although provisioning and configuration management technologies exist in various stages in each technology domain (for example, server, storage and network), tools that cross the heterogeneous domains are just emerging, often with x86-based virtualization as their underpinning (thus, they can only provision resources inside a virtual machine and not both virtual and physical resources).

Business Impact: Future technology has the potential to automate IT management processes more fully and completely across the application development and service delivery life cycles. As such, release provisioning, configuration and deployment across physical and virtual environments will be completely automated, increasing service quality (that is, reducing unplanned downtime through greater accuracy in automation) and reducing labor costs.

Benefit Rating: High

Market Penetration: Less than 1% of target audience

Maturity: Embryonic

Sample Vendors: 3Tera; VMware

Run Book Automation

Analysis By: David Williams

Definition: IT operations run book automation (RBA) or IT operations process automation products automate IT operations processes across different IT management disciplines. RBA products have an orchestration interface to design, administer and monitor processes; a workflow to support the processes; and integration to many IT operations tools needed to support the processes (such as fault and problem management, configuration and change management, and disaster recovery). Unlike IT operations management products, which have internal processes to deliver their value (such as change management tools, which define a process that makes device configuration changes), RBA products have the capability to integrate and orchestrate multiple IT operations management tools to support a process that may need multiple tools, and spans many IT operations management domains. Today, two types of RBA vendors have emerged: specific and generic. Specific RBA vendors focus on specific processes (for example, server provisioning); generic RBA vendors are designed to meet the needs of a wide range of IT operations processes.

Position and Adoption Speed Justification: IT operations process automation has become a key focus for IT organizations looking to improve and measure IT operations efficiencies, and to find the mechanisms that can help them integrate their disparate IT operations tool portfolios in support of best practices. The RBA market continues to attract new players from a wide range of technology focus areas, including event and fault management, change management, configuration management, job scheduling, and business process management. The combination of enterprise need and a continual flow of new RBA products continues to fuel market awareness

and adoption. New RBA products will continue coming to market from established software vendors as well as startups. The visibility of these tools has increased significantly during the past year, and this will continue to grow, although process improvement is always slow to progress.

User Advice: Specific RBA tools provide a defined process framework that can aid in achieving rapid value; however, gaining sustainable, measurable value from a general RBA tool requires mature, documented processes. Clients should first develop and document their IT operations management processes before implementing general RBA tools. For specific RBA tools, clients should first understand the embedded process to be sure it fits their process structures. IT operations managers who understand the challenges and potential benefits of automating key IT management processes should consider RBA tools as a way to achieve this objective — only after removing the issues of complexity through standardizing processes and technologies. In addition, IT operations processes that cross different IT management domain areas will require organizational cooperation and support, and the establishment of process owners.

Business Impact: RBA tools will have a significant effect on end-to-end IT operations process automation and running IT operations as a business by providing consistent, measurable, better-quality services at optimal cost, reducing the human factor and associated risks by automating safe, repeatable processes, and increasing IT operations efficiencies by integrating and leveraging the IT management tools needed to support IT operations processes across IT domains.

Benefit Rating: High

Market Penetration: Less than 1% of target audience

Maturity: Adolescent

Sample Vendors: BMC Software; CA; Enigmatec; HP (Opsware); IBM; LANDesk; NetIQ; Opalis Software; Optinuity; Stratavia; UC4 Software; Unisys

Recommended Reading: "Polling Results Show IT Operations Process Automation Needs and Expectations"

"BMC Buys RealOps to Strengthen Process Automation Support"

"IT Operations Management Run Book Automation Market Consolidation Will Accelerate"

"IT Operations Run Book Automation: The Evolving Vendor Landscape"

"Opsware to Broaden Process Automation With iConclude Buy"

"Cool Vendors in IT Operations Process Automation, 2007"

"IT Operations Run Book Automation Experiences Continued Market Growth"

"Run Book Automation: The Drivers and Inhibitors"

"IT Operations Run Book Automation: Automated Operations Revisited"

Root Cause Analysis Tools

Analysis By: Donna Scott; Debra Curtis

Definition: Through correlation and analytics, root cause analysis tools identify performance and availability problems in the infrastructure and have the ability to identify the root cause of the underlying problem so that an automated corrective action can be taken. These tools may take

advantage of many IT operations management technologies, such as event correlation and analysis, business service management, application transaction profiling and diagnostic tools.

Position and Adoption Speed Justification: Root cause analysis is enabled through instrumentation, which is being embedded in component hardware and software. The instrumentation, along with availability and performance management tools, will enable an end-to-end view of user activity and a dynamic understanding of bottlenecks, permitting automatic correction of them. Increased complexity and lack of instrumentation standards challenge the ability to achieve root cause analysis in a heterogeneous environment. Evolving technologies in the area of application transaction profiling and Java EE management, which provide dynamic application instrumentation and use Web services standards to collect statistics and dynamically effect optimization, have shown promise for service-oriented architecture (SOA) and Web-based application architectures.

User Advice: Technology that enables automatic detection and diagnosis of projected failures and bottlenecks is not available today. Today's incident and problem resolution process, as well as the technology, requires a great deal of manual effort. As technology matures in this area, it will take time to implement because administrators must trust the new technology before they are willing to take automated actions based on the root cause advice. This technology will first emerge for homogeneous or standardized application and infrastructure stacks prior to being available for heterogeneous environments.

Business Impact: This technology has the potential to massively improve business service performance and availability by dynamically determining root cause of an underlying incident or problem, versus today's requirement of manual collaboration across multiple application and infrastructure engineering groups. If root cause can be dynamically determined, corrective action can also be automated, thus improving service quality through shorter resolution time. Dynamic root cause analysis is a prerequisite to achieving a real-time infrastructure.

Benefit Rating: High

Market Penetration: 1% to 5% of target audience

Maturity: Adolescent

Sample Vendors: CA-Wily; EMC Smarts; HP-Bristol Technology; OpTier; Splunk

Standardized Infrastructure Stacks

Analysis By: Donna Scott

Definition: Standardized infrastructure stacks reduce the complexity of hardware and software architectures by systematically reducing the diversity of the environment. Tactics include limiting the components or vendors in the architecture, as well as standardizing and testing the configurations of the components so they can be changed and assembled quickly in response to new business and IT requirements.

Position and Adoption Speed Justification: Just as a car company drives efficiency through parts standardization, reusability and repeatable processes, so too must the IT organization. IT services are defined in conjunction with business requirements, but are predicated on architecture and standard configurations and practices. IT organizations that lack architectural standards and configurations (in hardware, software/applications and operations) will have higher costs and less agility than those with standards that reduce the diversity of the environment and are enforced. Standards enable rapid deployment, because services are based on repeatable processes that can be more-easily automated. Architectural standards help IT operations sustain

high rates of change with low error rates, because the behavior of the underlying systems is well-understood, tested and predictable.

Despite the benefits of standardization, few companies have been able to standardize and reduce the complexity of their data center infrastructures, because of:

- The cultural bias toward enabling business customers to have choice and flexibility
- The cost and time required to modernize legacy systems, which make standardization difficult to justify (vs. other projects)
- The introduction of innovative technologies that drive businesses to adopt nonstandard technologies

Nonetheless, increasing pressures in the areas of compliance, cost reduction, faster delivery of IT services and changes in IT services are causing many IT organizations to reassess their commitment to architecture and standards.

Standardized infrastructure stacks are necessary for a real-time infrastructure (RTI), because they enable common, shared pools of resources. These resources can then be dynamically optimized by service governors based on IT service-level agreements (SLAs) and business priorities.

User Advice: Enterprises should move to standardized infrastructure stacks as a prerequisite to sharing resources across IT services and implementing RTI architectures.

To move toward standardization:

- Define standard infrastructure configurations and architecture building blocks, publish the standard configurations and start requiring compliance with new projects.
- Migrate established applications to new standard infrastructures using an opportunistic approach when updates are planned or funding is available.
- Set software release support standards and policies, and enforce them through audits and the active management of software assets.
- Create standard builds by layers in the software stack, so that automation can be leveraged and costs can be reduced, while consistency and reliability rise.
- Implement an ongoing set of processes that evaluates new technology benefits and risks, as well as develops standard configurations, as appropriate.
- Determine the risks and trade-offs, as well as the cost impacts for deviation from standards, and minimize the implementation of nonstandard configurations.

Business Impact: Infrastructure stack standardization reduces complexity, and is the foundation on which RTI architectures become viable. Moreover, standardization lowers current and future costs, reduces risks, and improves service levels and speed of service delivery. Standardization has a significant business impact, but requires long-term commitment from the organization to maintain.

Benefit Rating: High

Market Penetration: 5% to 20% of target audience

Maturity: Adolescent

Sliding Into the Trough

Business Service Management Tools

Analysis By: Debra Curtis; Will Cappelli

Definition: Business service management (BSM) is a category of IT operations management software products that dynamically links the availability and performance events from underlying IT infrastructure and application components to the business-oriented IT services that enable business processes. To qualify for the BSM category, a product must support the definition, storage and visualization of IT service topology via an object model that documents and maintains parent-child relationships and other associations among the supporting IT infrastructure components.

The BSM product must gather real-time operational status data from the underlying applications and IT infrastructure components via its services or through established monitoring tools, such as distributed system- and mainframe-based event correlation and analysis (ECA), job scheduling and, in some cases, application performance monitoring. BSM products then process the status data against the object model, using potentially complex service health calculations and weightings, rather than straightforward inheritance, to communicate real-time IT service status. Results are displayed in graphical business service views, sometimes referred to as "dashboards."

Position and Adoption Speed Justification: Every company wants to assess the impact of IT infrastructure and applications on its business processes to match IT to business needs. However, only 10% of large companies have developed their IT operational processes to the point where they're ready to successfully deploy a BSM tool to achieve this. Adoption speed will continue to be slow, but steady, as IT organizations improve their IT management process maturity. BSM is starting to sink toward the Trough of Disillusionment. IT organizations are discovering that BSM tools aren't easy to deploy, because a manual effort is required to identify the IT service relationships and dependencies, or implementation requires that a configuration management database (CMDB) already be in place, which is not the case in most companies.

User Advice: Clients should choose BSM tools when they need to present a real-time, business-oriented dashboard display of service status, but only if they already have a mature, service-oriented IT organization. BSM requires a good understanding of the logical linkages between IT components and the IT services they enable, as well as good instrumentation and monitoring for these components.

Clients should not implement BSM to monitor individual IT infrastructure components or technology domains. At its core, BSM provides the capability to manage technology as a business service, rather than as individual IT silos. Thus, BSM should be used as IT organizations attempt to become more business-aligned in their IT service quality monitoring and reporting.

Business Impact: BSM tools help the IT organization present its business unit customers with a business-oriented display of how well IT services are performing in support of critical processes. BSM tools identify the IT services affected by IT component problems, helping to prioritize operational tasks and support efforts relative to business impact. By following the visual representation of the dependencies from IT services to business applications and IT infrastructure components (including servers, storage, networks, middleware and databases), BSM tools can help the IT department determine the root causes of IT service problems, thus shortening mean time to repair, especially for critical business processes.

Benefit Rating: High

Market Penetration: 1% to 5% of target audience

Maturity: Adolescent

Sample Vendors: ASG; BMC Software; CA; Compuware; HP; IBM Tivoli; Interlink Software; Managed Objects; Nimsoft; Quest Software

Recommended Reading: "Toolkit: How to Begin Business Service Management Implementation"

Network Configuration and Change Management Tools

Analysis By: Debra Curtis; Will Cappelli; David Williams

Definition: Network configuration and change management (NCCM) tools focus on discovering and documenting network device configuration files, auditing changes, comparing configurations to the policy or "gold standard" for that device, and deploying configuration updates to multivendor network devices.

Position and Adoption Speed Justification: NCCM has primarily been a manual process involving typing commands into vendor-specific command-line interfaces, or creating homegrown scripts to ease retyping requirements. Little consideration was given to rigorous configuration and change management, compliance audit or disaster recovery rollback processes when executing network configuration alterations. A new generation of NCCM vendors created tools that operate in multivendor environments and bring a more-rigorous configuration and change management process, as well as compliance audit capability. The market has progressed to the point that many of these startups have been acquired and new vendors, including open-source alternatives, have entered the market.

User Advice: Replace manual processes with automated NCCM tools to monitor and control network device configurations, thus improving staff efficiency, reducing risk and enabling enforcement of compliance policies. Prior to investing in tools, establish standard network device configuration policies to reduce complexity and enable more-effective automated change. NCCM tends to be a discipline unto itself, but in the future, it must increasingly be considered part of the configuration process for an end-to-end IT service. This will require participation in the change management system and integration with configuration management tools for other technologies such as servers and storage.

Business Impact: These tools provide an automated way of maintaining network device configurations, offering an opportunity to lower cost, reduce human error and improve compliance with configuration policies.

Benefit Rating: Moderate

Market Penetration: 1% to 5% of target audience

Maturity: Adolescent

Sample Vendors: AdventNet; AlterPoint; BMC Software (Emprisa Networks); EMC (Voyence); HP (Opsware); Intelliden; SolarWinds

Recommended Reading: "Network Managers Face Pressures and Conflicting Investment Priorities"

Server Virtualization (Hypervisors)

Analysis By: Philip Dawson; Thomas Bittman

Definition: Server virtualization technologies support the operation of multiple operating-system instances concurrently on a single physical server, without using a general-purpose host operating system for primary access to the hardware. It enables hardware resources to be allocated on a fractionalized basis. Mainframe virtualization is not included because it is already well-established and mature. However, other non-x86-based hypervisors are included for technologies such as Integrity, Sun T1, T2 and Power.

Position and Adoption Speed Justification: These technologies are maturing, and competition is expanding. Embedded hypervisors were introduced in 2008 for x86 architecture servers. Input/output (I/O) virtualization improvements in hardware continue to be important in 2008 and beyond. VMware has been the dominant vendor in the x86 market for the past seven years, and Microsoft is delivering Hyper-V in the third quarter of 2008 (packaged with Windows Server 2008). In 2007, Citrix acquired XenSource, Oracle delivered a Xen-based hypervisor, and Sun announced a Xen-based hypervisor to be delivered in the third quarter of 2008. Competition and the embedded offerings have driven the average price of hypervisor technology to near zero. To be noted as well are the Red Hat and kernel-based virtual machine (KVM) efforts as an open-source software (OSS) alternative to Xen that may undermine the value proposition of Xen-based virtualization products from Citrix and Novell. Additionally, non-x86-based hypervisors are improving with better I/O virtualization and heterogeneous workload management, but have less market impact due to a smaller market base (and a larger x86 server sprawl issue).

User Advice: Hypervisor-based virtualization is a strategic technology, and will become the default for most servers within four years. While hypervisors are becoming free (or near free), there are still functionality and maturity differences between hypervisors, and vendors today require their specific layered virtualization management tools (which are usually not free) to manage them. Choosing the right management toolset is critical. Special focus should be put on the ability to manage virtual machine (VM) life cycles (to avoid or monitor VM sprawl, as well as understand and control offline creation), the ability to possibly manage large numbers of VMs, and tools to enable management flexibility, deployment and live migration of VMs.

Business Impact: Server virtualization technology reduces hardware costs through server consolidation and by increasing hardware use. It lowers the barrier to entry for disaster-recovery solutions and modestly reduces server administration costs (perhaps by 5% for x86 servers), especially for server deployment. Most importantly, server virtualization technologies enable rapid deployment, portability and resource flexibility. In the long term, this technology will become the basis for processes such as completely changing server chargeback, planned downtime management, capacity planning and disaster recovery.

Benefit Rating: Transformational

Market Penetration: 5% to 20% of target audience

Maturity: Early mainstream

Sample Vendors: Citrix; HP; IBM; Novell; Oracle; Parallels; Sun Microsystems; Virtual Iron; VMware

IT Management Process Maturity

Analysis By: Debra Curtis

Definition: IT management process maturity is one dimension of the IT Infrastructure and Operations (I&O) Maturity Model. The process dimension of the IT I&O Maturity Model shows how IT organizations evolve from having defined processes for IT service support (such as incident and problem management) at the "committed" level, to the "proactive" level, where processes (such as configuration and change management) are repeatable and individually

automated, to the service-aligned level at which multiple IT management processes are integrated and automated (such as change management feeding configuration management, and performance management linking to problem management). Attaining the service-aligned level of the model, defining end-to-end IT services and service-level agreements (SLAs), and automating integrated IT management processes are prerequisites to achieving a real-time infrastructure (RTI), where IT services and underlying IT resources are allocated dynamically based on business priorities and service-level policies. IT organizations that have implemented end-to-end IT service management and automated IT management processes will find it easier to deploy new RTI technologies as they become available.

Position and Adoption Speed Justification: As part of the IT organization's efforts to align itself more effectively with business needs, IT operations departments are being pressured to move from a component orientation (such as managing networks, servers, storage, databases and applications) to managing business-oriented, end-to-end IT services. Because IT operations departments have been so segregated in IT component-oriented organizational structures and metrics, this transition will be challenging for most of them. We estimate that fewer than 9% of large enterprise IT organizations have made the transition to achieve the service-aligned or business partnership levels of the IT I&O Maturity Model, including setting end-to-end IT SLAs and managing IT service delivery to those SLAs.

Gartner's Strategic Planning Assumption states that by year-end 2012, fewer than 14% of IT I&O organizations in large enterprises will have reached the service-aligned level or above. Thus, we place IT management process maturity as having market penetration of 5% to 20% of the target audience. Based on the challenges involved, we estimate that the transformation from the committed level of IT management process maturity through the proactive level, to achieve the service-aligned level, will take at least two to three years (and often longer), and will require significant cultural change, as well as strong leadership and vision. This realization has pushed the concept of IT management process maturity toward the Trough of Disillusionment as the profound change it requires becomes abundantly clear to many enterprises. However, the benefits of improved IT service quality, greater levels of agility, reduced risk and lower costs, as well as the desire to achieve RTI, will require enterprises to mature and continually optimize their management processes. Although organizations can be at varying levels of IT management process maturity, we are assessing the relative aggregate measure of process maturity, as needed, to implement RTI architectures for the position on this Hype Cycle.

User Advice: Understand your position in the process discipline of the IT I&O Maturity Model, and set a goal as to the level you must reach to best support the business. You'll need to reach the proactive level, at a minimum; ideally, you'll need to reach the service-aligned level of the model before you can be successful with RTI. Define and document IT management processes so they can be instrumented and automated to improve process performance. Leverage industry-standard best practice guidance, such as the Information Technology Infrastructure Library (ITIL).

Realize that improving IT management process maturity will affect your organizational design, because it generally requires a matrix management approach due to process managers whose responsibility crosses multiple IT component-oriented technology domains. Maturity models provide an effective means of systematically and continually improving IT infrastructure and operational processes. Establishing the right personnel performance metrics and rewards will encourage IT operations staff to look beyond their reactive "firefighting" modes and devote the time and training necessary to document repeatable processes, become proficient in their execution and achieve predictable service quality.

Business Impact: Moving to higher levels of IT management process maturity not only improves service quality, but also lowers labor costs, which affect the efficiency and effectiveness of IT operations, as well as the value derived from IT.

Benefit Rating: Transformational

Market Penetration: 5% to 20% of target audience

Maturity: Adolescent

Recommended Reading: "Introducing the Gartner IT Infrastructure and Operations Maturity Model"

"Survey Results: IT Infrastructure and Operations Management Maturity"

Server Provisioning and Configuration Management

Analysis By: Ronni Colville; Donna Scott

Definition: Server provisioning and configuration management tools provide one or more of the following functions:

- **Server provisioning:** Multiplatform bare-metal discovery and unattended (that is, scripted) and/or image-based operating system (OS) installation.
- **Application provisioning and configuration management:** A broad suite of multiplatform functionality to discover and provision (that is, package, deploy and install) OSs and application software; these tools also can make ongoing updates to OSs or applications (for example, patches, new versions and new functionality), or update configuration settings.
- **Inventory/discovery, configuration modeling, audit and compliance** enable the discovery of software, hardware and virtual systems; some can discover dependency relationships across servers and applications. Using modeling of application and OS configuration settings (that is, the desired state or "gold" standard), these tools can report or remediate variations by modifying the actual state back to whatever the model requires, or the desired state for applications, as well as security configuration settings, such as the U.S. National Institute of Standards and Technology (NIST), the Center for Internet Security (CIS) and the U.S. National Security Agency (NSA).
- **Repurposing and orchestrating servers and other devices (switches, storage), based on events or policies,** dynamically increases or decreases the capacity to enable more-efficient resource use, and to meet service-level agreements. This is the foundation of a real-time infrastructure strategy.

Position and Adoption Speed Justification: Server provisioning and configuration management tools continue to mature in depth of function for server management as well as integration with adjacent technologies (such as change management, dependency mapping and run book automation). Products are being deployed more widely for automated server provisioning and configuration management, with a focus on compliance, improved efficiencies and repeatability.

User Advice: With an increase in the frequency and number of changes to servers and applications, IT organizations should emphasize the standardization of technology and processes to improve and increase availability, as well as to succeed in using server provisioning and configuration management tools. Besides providing increased quality, these tools can reduce the overall cost to manage and support patching and rapid deployments, as well as provide a mechanism to monitor compliance. Evaluation criteria should include technologies that provide active capability (installation and deployment) as well as auditing and reporting. They should also assess functionality to manage the life cycle of software contained in virtual machines. Only

clients that have standardized on server infrastructures, mature imaging, change and configuration processes should purchase repurposing and orchestration tools.

Business Impact: Server provisioning and configuration management tools help IT operations automate many server provisioning tasks, thereby lowering the cost of IT operations, increasing application availability and increasing the speed of modifications to software and servers.

Benefit Rating: High

Market Penetration: 5% to 20% of target audience

Maturity: Early mainstream

Sample Vendors: BMC Software (BladeLogic); CA; HP (Opsware); IBM

Recommended Reading: "Vendor Landscape: Server Provisioning and Configuration Management"

"Market Dynamics: Server Provisioning and Configuration Management"

Grid Computing

Analysis By: Carl Claunch

Definition: Grid computing involves using computers managed by more than one organization to collectively accomplish large tasks, such as derivative risk analysis, candidate drug screening and complex simulations. The management domains can be separate companies, separate divisions of one company, or different data centers or operating organizations inside one company.

Position and Adoption Speed Justification: Grid computing is an extension of cluster computing, and its use is well under way in financial services and pharmaceutical organizations that have made substantial progress in applications, algorithms and new research processes. Grid application candidates are compute-intensive, can be parallelized so that parts of the processing can be done across a distributed set of systems and can combine results in a central location. Because of the shared resource pools and massively scalable architectures with dynamic capacity adjustment, grid computing is an example of a real-time infrastructure architecture.

User Advice: Conceptually, grid computing could be used in two ways. It can help lower costs or increase the efficiency of a fixed amount of work. But, more importantly, it can offer business advantage by accomplishing what was infeasible when using more-traditional approaches. Often, this means increasing the accuracy of a model, producing results in an unprecedented short time, looking for interactions earlier, reducing the time it takes to search libraries of compounds as drug candidates or enabling a new business model. When business advantage can be gained from scaling up compute-intensive or data-intensive processing, add grid computing to the list of potential implementation approaches. When the objectives are mainly to reduce costs or improve efficiency, consider alternatives that are more mature and have fewer issues to overcome.

Business Impact: Investment analysis, drug discovery, design simulation and verification, actuarial modeling, crash simulation and extreme business intelligence tasks are all areas where grid computing may enable business advantage.

Benefit Rating: Transformational

Market Penetration: 1% to 5% of target audience

Maturity: Early mainstream

Sample Vendors: Appistry; DataSynapse; Digipede; Platform Computing; Univa UD

IT Chargeback Tools

Analysis By: Milind Govekar

Definition: IT chargeback tools help IT operations groups recover their costs from the business units or align costs with the business units. These tools help allocate costs incurred based on the IT services provided and a cost allocation model.

Position and Adoption Speed Justification: Most enterprises use spreadsheets and other homegrown tools, and their expectations from commercial chargeback tools have been enormous. However, most of the commercial tools' capabilities have been restricted largely to IT component resources use. The demand for these tools is growing, particularly because of increasing deployment of virtualization tools. These tools have become more visible and will continue to do so, due to virtualization technology that requires enterprises to share infrastructure.

Most commercial chargeback tools focus on the infrastructure components (for example, server, storage and network). However, this is changing because these tools aggregate chargeback data from various infrastructure components for IT service-based chargeback. The expectations from these tools were higher than what they delivered, and the tools are beginning to match user expectations.

User Advice: Users should increasingly look at evaluating commercial, off-the-shelf tools for chargeback, especially in a virtualized environment or where different business units are demanding support of various chargeback models and want to adopt a shared-service delivery model where the complexity of the computing environment is high. Most organizations where the business users are not discerning or demanding enough can continue to use spreadsheets or other tools. However, users must have an agreed chargeback or cost allocation process in place before implementing these tools.

Business Impact: IT chargeback tools mainly help run the IT organization as a business and have an impact on the IT organization's ability to perform accurate cost allocation and on the value of the services it provides. A major benefit is that chargeback tools enable enterprises to fairly apportion IT service costs based on differentiated levels of business unit service consumption and show how the IT organization contributes to business value.

Benefit Rating: High

Market Penetration: 5% to 20% of target audience

Maturity: Early mainstream

Sample Vendors: Evident Software; IBM Tivoli; Satoritech (Provment); SAS; VKernel

Recommended Reading: "Server Virtualization Forces Rise in Chargeback Interest"

Entering the Plateau

High-Performance Computing Clusters

Analysis By: Carl Claunch

Definition: High-performance computing (HPC) clusters involves using multiple computers owned by one organization to accomplish very large computer tasks.

Position and Adoption Speed Justification: Cluster computing has become the dominant mechanism for HPC workloads, displacing vector machines, massively parallel processors and large symmetric multiprocessors for intensive workloads. Cluster candidates are compute-intensive and can be parallelized such that parts of the processing can be done across multiple nodes in the cluster. Because of the shared resource pools and massively scalable architectures with dynamic capacity adjustment, cluster computing is an example of a real-time infrastructure architecture.

User Advice: Cluster computing is the mainstream and most cost-effective approach for requirements where software exists to scale in parallel processing across multiple machines, or where the code can be written in-house, based on a suitable algorithm. Exceptions will exist in cases with high intercommunication requirements, for problems that are inherently single-threaded and where ease of implementation favors a shared memory system. However, the majority of potential applications are suitable for deployment on clusters.

Business Impact: Cluster computing shifts HPC work to cost-effective, high-volume, industry-standard technologies. This permits organizations to tackle computing challenges that were otherwise economically infeasible. In addition, clusters are more fluidly scalable than fixed supercomputers, improving business ability to react to changing needs. Those with a fixed and bounded need for computing resources can leverage clusters to drive down expenditure for HPC hardware, while those that can connect business advantage to the use of larger systems will use clusters to increase capacity to fairly high levels. Those that need more extreme capacity might move to grid computing to harness more computing power.

Benefit Rating: Transformational

Market Penetration: 20% to 50% of target audience

Maturity: Mature mainstream

Sample Vendors: Cray; Data Synapse; Dell; HP; IBM; Platform Computing; SGI; Sun Microsystems; Univa UD

Mainframe Service Governor: z/OS Workload Manager

Analysis By: John Phelps

Definition: A service governor is a runtime execution engine that has several inputs: business priorities, IT service descriptions (and dependency models), service quality and cost policies. A service governor also takes real-time data feeds that assess the performance of user transactions and the end-to-end infrastructure, and uses them to dynamically optimize the consumption of real and virtual IT infrastructure resources to meet business requirements and service-level agreements (SLAs). The service governor performs optimization through dynamic capacity management (that is, scaling resources up and down) and dynamically tunes the environment for optimum throughput based on demand. For z/OS, the service governor is the z/OS Workload Manager running in goal mode.

Position and Adoption Speed Justification: z/OS workload management has achieved the Plateau of Productivity based on years of evolution stemming from prioritizing workloads in a consolidated environment. Online transactions will meet response time goals, and batch jobs will meet throughput goals within the bounds of available resources, because capacity can be allocated dynamically based on SLAs. To do this, the workload manager uses a large number of algorithms, such as dispatching priority, input/output (I/O) dispatching priority, capping, donor and

receiver determinations, central storage occupancy, swap-protect time, multiprogramming level targets and performance index. z/OS workload management provides real-time feedback for these algorithms, with extensive built-in instrumentation throughout the system and major subsystems, such as Customer Information Control System, Information Management System, DB2 and WebSphere. With these algorithms, z/OS workload management manages not only CPU use but also dispatch priority, storage used and I/O priority queuing. IBM is looking to move this capability into a cross-system and cross-operating-system environment in plans.

User Advice: All z/OS users employ the workload manager in goal mode to ensure that service levels are met by the largest number of tasks in times of high resource use (goal mode use has been required since z/OS v.1.3 and later). Users should examine all the functionality that is provided to enhance management of newer workloads, such as WebSphere and Linux. Also look to increase capabilities outside a single z/OS environment as IBM continues to add functionality to the product.

Business Impact: For a consolidated environment like the IBM System z, it is critical that the large number of tasks executing in the system be controlled easily and optimized based on business goals and that the system be able to handle large variations in load. The large number of tasks requires that the enterprise's business goals be applied automatically to all the z/OS workload to dynamically scale and optimize the shared resources to meet the business goals. This ensures that important workloads achieve their performance goals without degrading the performance of lower-priority workloads by preventing higher-priority tasks from using more resources than are needed to achieve their service levels.

Benefit Rating: High

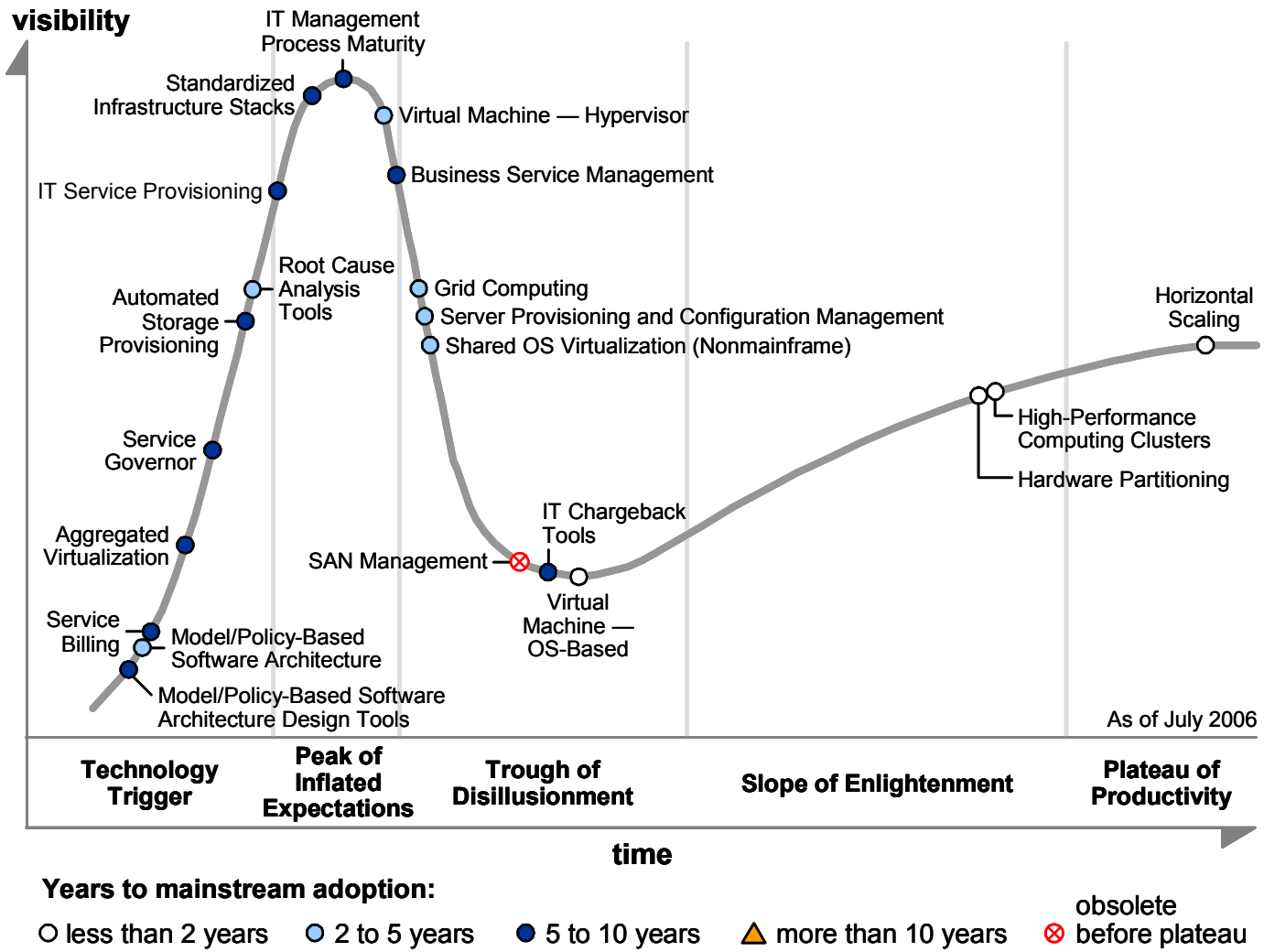
Market Penetration: More than 50% of target audience

Maturity: Mature mainstream

Sample Vendors: IBM System z

Appendixes

Figure 3. Hype Cycle for Real-Time Infrastructure, 2006



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 2008)

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 2008)

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 2008)

RECOMMENDED READING

"Understanding Gartner's Hype Cycles, 2008"

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