Enterprise Cloud Computing: Transforming IT

A Platform Computing White Paper
July 2009

{ Platform
Computing

Contents

Enterprise cloud computing: Transforming IT

Cloud computing is a new, important IT delivery model that provides infrastructure and computer resources as a service. This new paradigm, the next phase in the evolution of distributed computing, enables enterprises to transform IT by implementing private clouds. It leverages the power of sharing and enables increased utilization rates of IT resources, more rapid and efficient delivery of IT services, faster time to market, and reduced IT capital and operating expenditures.

This white paper examines the evolution of private clouds and discusses how many organizations that have implemented grid or virtual machines are moving towards private clouds. It also describes how organizations can successfully implement this new IT delivery model using a phased approach that starts with cloud management software. In addition, the white paper highlights common use cases featuring Platform Computing customers that are on the forefront of cloud computing. It underlines the fact that, "Cloud is built, not bought."

1. Introduction to cloud computing

In the sixty-plus years since business computing began, there have been two major trends: first there was mainframe computing, followed by client-server. Now we have a new paradigm. Cloud computing is the third generation IT model and the next phase in the evolution of distributed computing and the data center.

Whatever the model, IT departments have always been challenged to quickly and cost-effectively deliver IT resources to support user applications. Client-server architecture greatly increased IT's ability to implement a dependable computing infrastructure that supported diverse lines of business and their applications.

However, client-server came with a price. It has contributed to server sprawl, skyrocketing capital and operating expenditures, and complex data centers that are difficult to manage and reconfigure quickly to address to changing demands. Servers dedicated to individual applications are usually underutilized due to fluctuating application workload demands. The IT technology stacks on these servers are often customized, requiring individual attention by IT administrators. To provision a business request for a new application, one or more new servers complete with management software have to be procured and deployed, often resulting in weeks or months of delay.

For over 10 years, distributed computing architectures such as clusters and grids have provided business users with access to shared and scalable IT infrastructure with high resource utilization. While primarily deployed to handle technical and scientific applications – often called high performance computing (HPC) applications – the lessons learned from implementing clusters and grids are relevant to general business applications. Specifically, the knowledge gained over the last decade can contribute directly to the new cloud computing model – a new paradigm in which sharing benefits both end users and IT.

Like software that has evolved from a proprietary, centralized model to an open, distributed structure, cloud computing is also a rapidly evolving trend that will have a profound impact on the way IT is delivered to business users. Cloud computing is the logical next step in the evolution of the data center from mainframe to client-server, and now to clouds.

> **So just what is cloud computing, particularly in the context of the enterprise?**
>
> Cloud computing is an IT delivery model that offers large-scale, shared infrastructure and computing resources as a service through self-service pay-per-use access. Although it leverages recently developed technology, cloud computing is a business, not a technical trend.

2. Features of cloud computing

Cloud computing has several key features including: infrastructure sharing; scalability; self-service; and pay-per-use:

- **Infrastructure sharing** –Today's enterprise data centers are characterized by fluctuating resource demands from a variety of users. Cloud computing enables dynamic sharing of these resources so that demands can be met cost effectively.

- **Scalability** – To handle ever increasing workload demands and support the entire enterprise, cloud computing must have the flexibility to significantly scale IT resources. Scalability and flexibility allow the cloud provider to fulfill, or at least come close, to the promise of unlimited IT services on demand.

- **Self service** – Cloud computing provides customers with access to IT resources through service-based offerings. The details of IT resources and their setup are transparent to the users.

- **Pay-per-use** – Because cloud resources can be adding and removed according to workload demand, users pay for only what they use and are not charged when their service demands decrease.
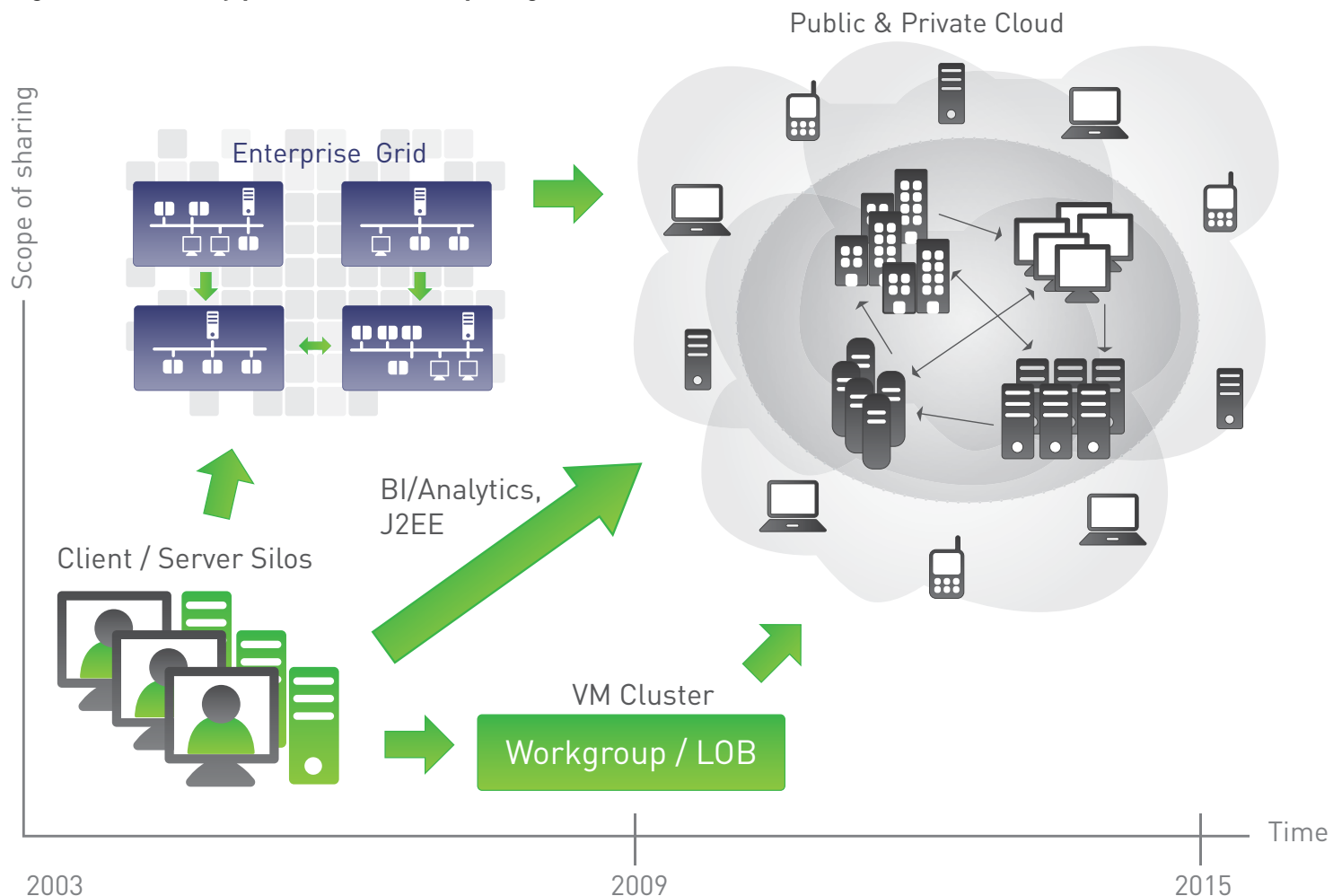
3. Paths to cloud computing

While there are many roads to cloud computing from existing client-server infrastructures, there are at least three major paths open to business users. These include:

• **VM to cloud** – For users already running applications hosted on VMs (virtual machines), their virtual servers can be brought together to form a VM cluster. As VM clusters proliferate, automatic resource allocation is required to manage the virtual machines within a VM cluster in order to handle load balancing across the clusters, and for self-service access to resources. This approach leads to implementing a private cloud run by the organization's internal IT department.

• **Grid to cloud** – Some organizations are already running grids. These are distributed systems managed by IT staff and shared by technical applications that are typically compute or data intensive. VM technologies are not used in grids because each application can easily consume all of the resources on a server. Also, many servers are often harnessed together to run a parallel application. By deploying cloud management software, grids can be generalized to support more types of applications. Incorporating VM technologies and provisioning tools from data center automation products allows IT to transform an enterprise grid into a private cloud.

Figure 1. The many paths to cloud computing

- **Desktop to cloud** – Public clouds are another option. This approach allows users to access applications hosted in cloud computing centers run by external service providers directly from their desktops or other client devices.

An organization may choose any of these paths, or even several at the same time for different applications or parts of its business. To ensure success, the adoption of cloud computing should follow a sequence of evolutionary steps rather than an overnight revolution. Progress is driven by the requirements of specific applications for scalability and cost-effectiveness, or by the needs of capabilities such as business process automation.

4. Cloud management software

An often-quoted definition of cloud computing comes from Charles Brett, a principal analyst at Forrester Research. Says Brett, cloud computing is "a pool of abstracted, highly scalable, and managed infrastructure, capable of hosting end-customer applications and billed by consumption."[1]

While the definition of what a cloud delivers is quite clear, what is unclear is what IT needs to do to deploy a cloud. Just what products are required in order

Figure 2. Cloud management software – the engine that makes the cloud function
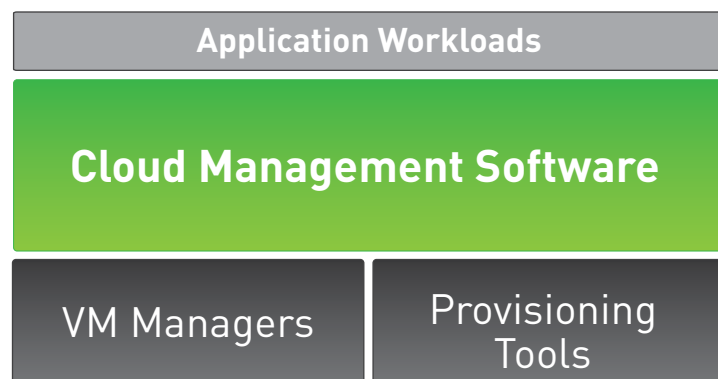
to build a "pool of abstracted, highly scalable and managed infrastructure?"

The hardware components are apparent: servers, storage and networking – all of which exist in today's data centers. What does not exist – and this is the key to a successful cloud deployment – is the software layer that creates a shared computing infrastructure from physical and virtual resources in order to deliver cloud computing services upon request. This new layer is called cloud management software.

Server virtualization and provisioning tools are key enabling technologies for clouds but do not provide cloud management. Server virtualization, with its ability to distribute workloads across servers, has been a major development in the path to cloud computing. Hypervisors abstract hardware dependencies and enable application portability. But hypervisors and VM management tools – typically used manually by system administrators in response to user requests and application performance problems – do not make for cloud computing on their own. In addition, not all applications can be virtualized, so some applications supported in the cloud may need to run on physical servers.

Similarly, provisioning tools for servers, storage and networking is another key enabling technology used for setting up application environments within the cloud's resources. Provisioning tools should be invoked automatically without involvement by system administrators whenever a user makes a cloud service request.

Cloud management software initiates automated resource provisioning and creates VMs as required. In addition, this software layer:

- Integrates cloud resources
- Incorporates management tools for cloud operators
- Provides self-service access for end-users to directly request compute resources

1 Charles Brett, "Grids, Clouds or Ultra Modular Computing for the Enterprise," Forrester Research, presented at BeinGrid, Barcelona, June 20, 2008.

- Keeps track of the usage of resources for reporting, billing and auditing purposes

While the functions and capabilities of cloud management software vary depending on the nature and requirements of the specific type of cloud, in all cases the cloud management software functions at the core of the infrastructure – it provides the control layer needed to pull resources into a unified cloud environment and deliver them as services.

5. Public cloud vs. private cloud

Cloud computing can take various forms, leading to the current confusion about how to define the term. The two major types of clouds are public and private clouds.

Public cloud

With public clouds, a service provider makes computational, storage, and other IT resources available over the Internet to customers that range from individuals to organizations as diverse as enterprises, government, and academia. Public clouds provide capital and operating expense reductions by allowing customers to pay for use instead of owning assets. In addition, public clouds can provide a potentially wider range of hardware, scalability, and expertise than may be available in-house. Some of the providers of public clouds are Amazon Elastic Compute Cloud (EC2), IBM's Blue Cloud, Sun Cloud, and Google App Engine.

Public clouds are located outside the corporate firewall and are managed by a 3rd party hosting provider. These clouds have their own set of services that may not meet the internal needs of an organization. Public clouds are generalized compute platforms that are designed to handle the generic types of user demand, often in the form of a limited selection of virtual machines. While this is sufficient for some individuals and applications, enterprises with established IT operations may find such public clouds unusable for many applications. Service level agreements (SLAs)

are harder to manage and enforce when working with external providers. Because the public cloud is outside the firewall, there are concerns over security, data access, and privacy for enterprise customers. Public clouds also find it difficult to meet auditing, regulatory, and compliance requirements.

Because of these concerns, public clouds are most frequently used for web applications and disaster recovery. Public clouds can be ideally suited for small and medium businesses (SMBs) that may not have capital resources to purchase and configure servers, or for enterprises that need short term bursting of compute resources for specific tasks due to spikes in demand.

Private cloud

Private clouds are used by organizations to pool internal IT resources for shared use. Private clouds not only have all the features described above, but also have the cloud management software capabilities required by enterprise data center environments – software that is missing in public clouds.

Because private clouds are owned and managed by an enterprise's IT team, control and security can be handled internally. Departments can negotiate pricing contracts with internal IT based on their requirements, or skip billing altogether when the cloud is treated as a shared corporate resource. Private clouds also allow organizations to have greater visibility and control over mission critical SLAs. In addition, they support IT's obligation to oversee fundamental corporate requirements including, governance, compliance, business continuity, cost reduction, and risk management.

The value of a private cloud management platform is independent of location and ownership of the resources. In addition to using existing IT data center resources, the private cloud may also incorporate resources from external service providers. In this

case, the private cloud will use its cloud management software capabilities to ensure consistent service delivery and integrity of the external resources. Such cloud bursting can help an enterprise avoid the necessity of provisioning for peak aggregate resource demands. It is one of the economic drivers behind the evolution of the enterprise data center from a client-server model to the private cloud.

6. Economic drivers for private clouds

Cloud computing can take various forms, leading to the current confusion about how to define the term. The two major types of clouds are public and private clouds.

Faster delivery of IT services

A cloud infrastructure allows IT to respond to user demands in minutes rather than weeks. Because infrastructure is shared, the required computing resources needed to support an application are often already present in the cloud pool. This means the resources can be provisioned on the fly, rather than being requisitioned, purchased, installed and configured – a time consuming and expensive process. By removing or streamlining the organization's entrenched silos, IT moves from an inflexible legacy infrastructure to one that responds to new business needs in real time.

Automation, self-service, and the delivery of application services put control in the hands of the business owners. Based on a utility model, users can request the precise resources they need for a specific application, and IT is able to more efficiently balance all its resources to meet the constantly changing demands of corporate customers. The bottom line for the enterprise is faster delivery of essential IT services, which speeds up time to market and allows the company to be more competitive.

Lower IT Costs

Both capital and operating expenditures (CapEx and OpEx) can be reduced by the adoption of a cloud infrastructure. Increased utilization of IT resources through sharing means fewer servers are required, and the data center's footprint and power and cooling requirements are reduced. This, in turn, reduces OpEx. Through configurable automation of repetitive IT administration tasks, cloud management software enables each IT administrator to manage hundreds, or even thousands of servers, which also substantially reduces OpEx.

CapEx is reduced because IT no longer needs to provide over-provisioned infrastructure silos to meet peak demand for each application. Applications can be added without adding servers. In fact, cloud adoption can lead to the overall reduction in the number of an enterprise's data centers.

7. Requirements of private cloud management software

For IT, the successful deployment of a private cloud depends on the use of cloud management software specifically designed for this type of cloud. The cloud management software must be integrated with other enterprise data center hardware and software components.

There are six key requirements of private clouds and their cloud management software. They include:

• **Heterogeneous systems support** – The private cloud needs to support the organization's existing heterogeneous infrastructure as well as resources from external providers. This includes server, storage and networking hardware, operating systems, hypervisors, storage systems, and file systems. These resources will continue to evolve over time because of advances in technology and the changing needs of the business and its applications.

The provider of private cloud management software should have a track record of supporting both current and new types of resources. This ensures that all resources will continue to be incorporated in the shared cloud pool as it grows and changes.

- **Integration with management tools** – Enterprises use a variety of IT management tools for security, provisioning, directory, reporting, billing, data management, internal regulation, compliance and management consoles. Cloud computing does not replace these tools. Instead, properly designed private cloud management software easily integrates with existing tools and invokes them as needed during cloud operations. Most 3rd party management tools should be supported out-of-the-box.

- **Configurable resource allocation policies** – The cloud must be workload-aware as well as resource-aware. This means that the cloud management software can determine the most efficient placement of user requested workloads. The cloud management software guarantees resource reservations to its customers based on well-defined policies. And, when demands peak, the software is able to arbitrate resources based on business priorities of various parts of the cloud workload to cost-effectively meet SLAs.

- **Integration with workload managers, middleware and applications** – Clouds exist to run applications. In addition to a self-service portal for users to request one or more virtual or physical machines, private cloud management software provides flexible programmable interfaces to enable easy integration with the enterprise's essential workload managers, middleware, and applications.

- **Support IT and business processes** – Clouds provide support for various IT and business processes and allow IT to automate many of its operations. In fact, cloud management enables the definition and on-going modifications of many IT management processes that had been performed manually.

- **Enterprise, not workgroup, solution** – An organization usually consists of multiple departments and locations, often distributed internationally. A flexible cloud scales to meet their diverse needs in real time. While cloud computing may be adopted initially within an individual line of business (LOB) or location, it enables the integration of IT across the enterprise by reconfiguring rather than replacing the private cloud management software. Therefore, a private cloud can be an enterprise-wide IT services delivery system that provides transparent and consistent access to global resources.

As the saying goes, "Cloud is built, not bought." Just like other mission-critical enterprise business systems and services, a private cloud is built by the IT organization, not bought from a vendor. Private cloud management software is the key to enabling IT to configure its data center resources, integrate its management tools, and support its applications and business processes. It is the glue that binds together enterprise data center operations as we move into the era of cloud computing.

8. Platform Computing and Platform ISF

Founded in 1992, Platform Computing has been a leader in management software for distributed clusters, grids, and HPC workloads deployed at enterprises. Over the years Platform has developed the skills and technologies required to efficiently provision and share heterogeneous hardware resources as well as orchestrate application processes.

As the global leader in management software for grid computing, Platform has accumulated extensive technology and expertise in enterprise IT infrastructure sharing and workload scheduling. Both are key capabilities for the adoption of private clouds. Platform is also experienced in deploying multi-thousand node compute environments that leverage both physical

Figure 3. Platform ISF for private cloud management and it's three layers
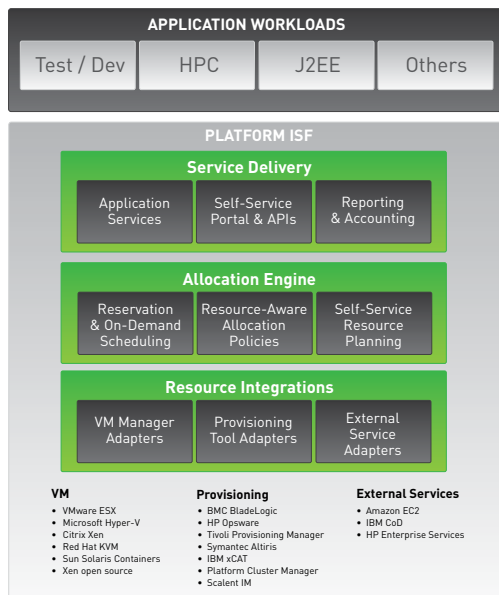
and virtual resources, clusters, and grids. For over a decade, these deployments have been in production with increasing scale in leading enterprises across numerous industries such as financial services, manufacturing, high tech, government, education, energy, telecommunications, and healthcare.

Platform ISF is the leading end-to-end private cloud management software. Platform ISF creates a shared computing infrastructure from physical and virtual resources to deliver application environments according to workload-aware and resource-aware policies.

Platform ISF meets all six requirements of private cloud management discussed above.

- **Resource integration** – This foundation layer of Platform ISF integrates distributed and heterogeneous IT resources to form a shared system. Resource integration is the opposite of server virtualization – instead of creating multiple VMs on one physical server, this capability creates

one shared computer out of many heterogeneous servers, storage devices, and interconnects. All major industry standard hardware, operating systems (including Linux and Windows) and VM hypervisors (including VMware ESX, Xen, Citrix XenServer, Microsoft Hyper-V and Red Hat KVM) are supported. The resource integration layer also uses provisioning tools to set up application environments on demand. It integrates with many 3rd party tools for various systems management tasks out-of-the-box, including directory services, security, and monitoring and alert. Its extensible framework of resource and management adaptors enables IT to fit Platform ISF into their existing data center systems environment. This layer can transparently integrate resources from external providers while maintaining its private cloud management environment.

- **Allocation engine** – Once a pool of shared resources is formed, a set of site-specific sharing policies is configured in the allocation engine layer to ensure that applications receive the required resources. The policies also make certain that the organization's resource sharing priorities are applied, and that the quota constraints applicable to business groups sharing the cloud are reinforced. The allocation engine matches IT resource supplies to their demands based on resource-aware and application-aware policies. This private cloud "brain" is critical for IT agility.

- **Service delivery** – This top layer of Platform ISF provides interfaces to users and applications as well as supporting the lifecycle of cloud service management. A self-service portal enables users to request and obtain physical servers and VMs in minutes instead of days or weeks. Platform ISF has a set of APIs that can be called by applications, middleware and workload managers to request and return resources without human intervention. Templates can be configured for simple and complex N-tier business applications to automate their lifecycle management. Platform ISF allows for the

starting of all the components of an N-tier application, the adding or removal of a resource, and monitoring and failure recovery. Platform ISF supports middleware such as J2EE, SOA, CEP and BPM, and workload schedulers such as AutoSys and Platform LSF. No change to the application supported by this software is needed – cloud computing should be transparent to users and applications. The service offerings can be structured as: complete application environments (e.g., application packages, CPU, memory, storage and networking); as bare metal servers with an operating system installed; or as virtual machines. SLAs can be associated with each service offering.

Platform ISF collects all resource usage data and provides reports and billing information. Alternatively, the cloud administrator may choose to feed the usage data into site-specific reporting and charge-back tools.

9. Use cases

As a general-purpose, end-to-end private cloud management platform, Platform ISF is highly configurable and can run in diverse operating and application environments. Platform ISF aggregates heterogeneous resources across data centers and dynamically provisions them to a diverse collection of applications with fluctuating resource demands.

Some representative use cases include: test/dev, test/dev/production, engineering cloud and a hosted services public cloud.

Test/Dev
The software test/dev environment supports the entire application lifecycle task flow including: software design, coding, unit testing, functional testing, system scalability testing, and benchmarking.

Figure 4. Platform ISF allows heterogeneous applications to share private cloud resources

DEMAND

HPC Clusters

LSF / Symphony

Test & Dev

J2EE / Analytics

WLM / Middleware

Platform ISF

Private cloud management software

SUPPLY
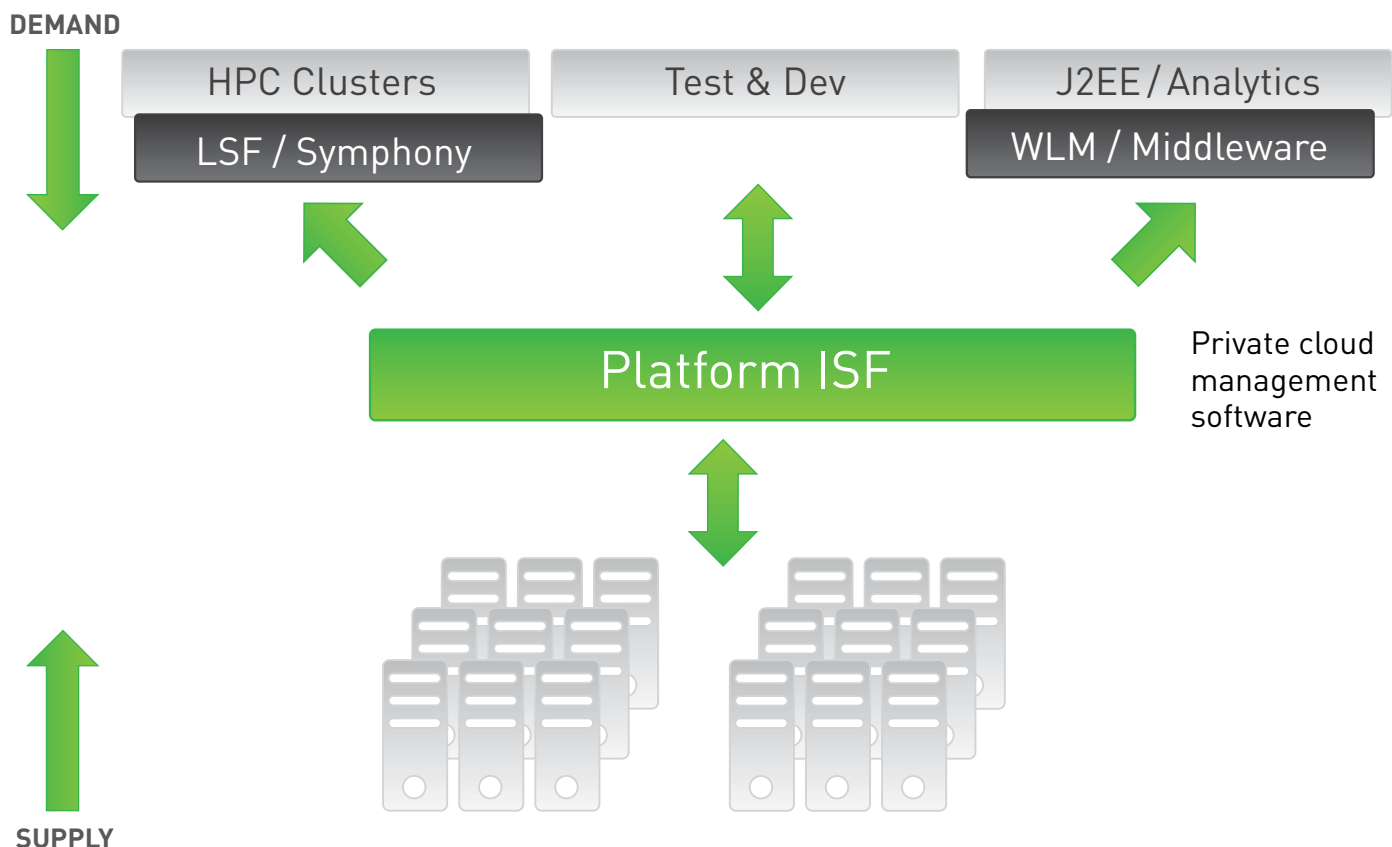
Typically, multiple development teams have specific infrastructure requirements that compete for available resources. Development and testing teams have limited visibility into resource usage and their costs. Also, new applications must be developed and tested while, simultaneously, patches for existing

applications must be tested for scalability, compliance and compatibility. Workflow in a test/dev environment is almost impossible to predict, and the result is an infrastructure that is severely over-provisioned or dedicated to specific applications, making resources unavailable to other developers even when those resources are not being used. Over-provisioning and lack of visibility leads to VM sprawl and complex, laborious multi-machine management.

With Platform ISF, development teams are able to share virtual and physical servers from a pool of available resources. A self-service portal enables developers and testers to make requests for resources on demand. Platform ISF also offers advance reservation of resources. Its automated policy-driven placement of VMs and physical servers ensures that project priorities and demands are met while enabling optimum utilization of resources. Capacity waste resulting from siloed environments is significantly reduced.

Platform ISF test/dev teams can obtain application environments in minutes instead of weeks. The specific set of resources for software testing and benchmarking can be configured in Platform ISF as application templates and initiated automatically to ensure proper environment set up and repeatability. Development managers are able to control quotas of these resources among their users, and can dynamically read–just as needed to ensure correct prioritization of resource usage. Developers can obtain error-free, fully configured resources that adhere to standards and compliance policies. IT maintains control over the infrastructure and benefits from the elimination of manual setup and repurposing tasks.

Test/Dev/Production

Using Platform ISF, application teams can obtain resources faster, eliminating delays due to server setup and reconfiguration, and reduce the time it takes to bring an application into production. Developers are

able to maintain consistency of standard builds through production. Because more application environments can be delivered with fewer resources, application service levels are improved. Resources can be dynamically added as needed.

The infrastructure management team starts with a test/dev cloud deployment enabled by Platform ISF before running production applications on cloud resources. Using the Platform ISF's self-service portal, the application manager requests capacity for his test/dev environment. When resource quotas are allocated to the application manager, he simply partitions his capacity among his test/dev/production environments using the self-service resource plan. When an application has been proven in the test/dev environment, it can then move to production. Application managers can make on-demand adjustments to redistribute compute resources to production environments to address higher priority requests.

Engineering cloud

Private clouds allow multiple engineering and HPC production environments to share IT resources. Even heterogeneous clusters running different types of workload managers – such as multiple LSF or Symphony clusters – can dynamically share resources managed by Platform ISF. Lower costs can be achieved by leveraging complementary application usage patterns to increase and optimize overall resource utilization. Service levels are also improved because resources can be dynamically added to higher priority applications when needed.

Hosting services public cloud

Although primarily targeting private clouds for enterprises, Platform ISF also allows hosting providers to offer IaaS clouds to external customers. By providing hosted infrastructure services and automating resource allocation for their customers, these providers are able to focus on core competencies to sharpen their business competitiveness.

CUSTOMER STORY
Alatum: Singapore Telecom public cloud

Alatum is the cloud service provider arm of Singapore Telecom for multi-national enterprises, SMBs and public sector agencies. Founded on a partnership between SingTel and a strong field of technology providers, Alatum allows enterprises and consumers to benefit from affordable, secure, on-demand, pay-per-use access to high performance computing, software, and data storage services.

Platform Computing is the cloud management software vendor powering Alatum. Platform Computing's technology enables access to significant IT resources and applications through the cloud, greatly reducing the costs of ownership and operation.

Platform also opened a Cloud Innovation Centre (CIC) in Singapore, a facility that provides Singapore-based companies with access to free cloud software, technical consulting, and training programs. Located at Platform Computing Singapore, the CIC's goal is to train enterprises, government organizations, software vendors and start-ups on cloud technology, and encourage the creation of new products and services based on this rapidly emerging computing paradigm.

Hosting providers can create service offerings for virtual and physical server resources on multiple OS platforms on a monthly recurring or pay-per-use, on-demand basis. Using Platform ISF as the underlying hosting platform for shared resources, more application environments can be hosted with fewer servers. Platform ISF's guaranteed reservation feature enables customers to reserve resources for anticipated peak demand periods. With its workload and resource-aware allocation policies, ISF optimally balances the supply and demand of hosted resources, resulting in maximized utilization. Lower capital and operational costs translate into the ability to offer competitive pricing for external customers.