IT@Intel White Paper
Intel Information Technology
Business Solutions
June 2010

An Enterprise Private Cloud Architecture and Implementation Roadmap

The private cloud is a shared multi-tenant environment built on a highly efficient, automated, and virtualized infrastructure.

Sudip Chahal
Principal Engineer, Intel IT

Jay Hahn-Steichen
Systems Engineer, Intel IT

Das Kamhout
Engineer, Intel IT

Rick Kraemer
Enterprise Architect, Intel IT

Hong Li
Principal Engineer, Intel IT

Chris Peters
Manager, Intel IT

Executive Overview

Intel IT has defined an architecture and implementation roadmap for a private enterprise cloud designed to increase agility and IT efficiency.

The private cloud is a shared multi-tenant environment built on a highly efficient, automated, and virtualized infrastructure. Other key elements of the cloud include standardized application platforms provided as a service and a self-service portal that enables business groups to request and manage capacity for their applications.

During the design of our private cloud architecture, we were able to take advantage of the extensive experience gained while building the Intel IT design grid, a large multi-tenant environment used for silicon design.

The expected benefits of our office and enterprise private cloud include:

- Increased agility, including significantly reduced provisioning times.
- Greater efficiency, including energy savings, due to better resource utilization.
- High availability with minimal incremental cost, by taking advantage of enhancements to industry-standard hardware and software.
- Improved capacity management, taking advantage of new business intelligence tools.

Because of the extensive scope of this initiative, we plan to deliver private cloud capabilities in phases over the next three or more years. As we add these capabilities, we expect that the cloud will become capable of hosting highly demanding, mission-critical business applications.

Contents

IT@INTEL

IT@Intel is a resource that enables IT professionals, managers, and executives to engage with peers in the Intel IT organization—and with thousands of other industry IT leaders—so you can gain insights into the tools, methods, strategies, and best practices that are proving most successful in addressing today's tough IT challenges. Visit us today at www.intel.com/IT or contact your local Intel representative if you'd like to learn more.

Intel IT Cloud Computing Terminology

We see cloud computing as a highly available computing environment where secure services and data are delivered on-demand to authenticated devices and users utilizing a shared, elastic infrastructure that concurrently supports multiple tenants.

We identify several attributes that distinguish cloud computing from conventional computing. These attributes are:

- On-demand self-service
- Broad network access
- Resource pooling
- Rapid elasticity
- Measured service
- Sharing by multiple tenants

Currently there are three primary categories of cloud computing service:

Infrastructure as a service (IaaS). Computing infrastructure, such as servers, storage, and network, delivered as a cloud service, typically through virtualization.

Platform as a service (PaaS). Platforms that can be used to develop and deploy applications.

Software as a service (SaaS). Software deployed as a hosted service and accessed over the Internet.

BACKGROUND

Today, Intel IT operates a massive, worldwide computing environment that supports about 78,000 Intel employees and includes approximately 100,000 servers.

About 20 percent of our servers are used to provide a broad range of office and enterprise computing services to Intel's employees, customers, and partners. This office and enterprise environment includes applications for online collaboration, e-mail, and calendaring, as well as large business applications, such as enterprise resource planning software.

Though this environment has met Intel's needs to date, the accelerating pace of business is driving a need to respond more quickly to changing business demands. At the same time, Intel IT is continually challenged to reduce cost.

Conventional approaches to computing have constrained our ability to meet these needs. For example, in traditional enterprise computing, servers are dedicated to specific applications, and each server is sized to support application growth and spikes in demand. This results in low physical server utilization and limits the ability to quickly provision new server capacity. In addition, capacity planning to support new IT initiatives has been complicated by the need to manually gather configuration, historical purchasing, and other information.

To meet business requirements mandating increased agility and efficiency, Intel has moved to a new enterprise architecture based on a cloud computing approach.

INTEL IT CLOUD COMPUTING STRATEGY

The Intel IT cloud computing strategy, shown in Figure 1, is designed to deliver benefits including increased agility and more efficient resource utilization.

Our primary strategy is to grow the cloud from the inside out. We are building a private cloud for office and enterprise computing, based on a highly virtualized, energy-efficient, and flexible environment. This approach offers many of the benefits of public clouds, such as increased agility and efficiency, without the risks associated with hosting Intel's sensitive applications and data outside the firewall. It also positions us to take advantage of public clouds over time, as standards emerge, the technology matures, costs are lowered, and security concerns are overcome.

At the same time, we are opportunistically taking advantage of specific public cloud services offered by external providers when they provide benefit to Intel. For example, we have already deployed several software as a service (SaaS) applications, including expense and time card tools, health benefit applications, and social media applications.

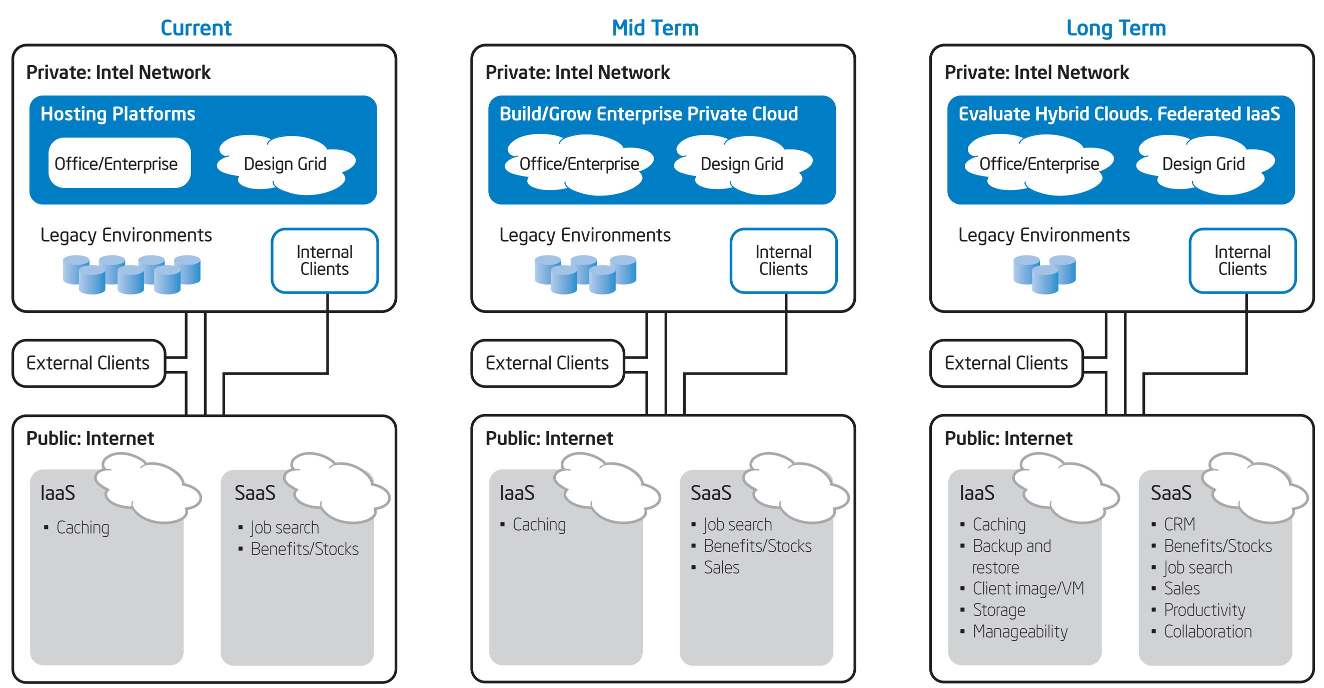
Because of the significant scope of our private cloud project, we are planning to implement the private cloud in phases over several years.

INTEL IT PRIVATE CLOUD EXPECTED BENEFITS

We anticipate that our enterprise private cloud will deliver key benefits both to Intel IT and to Intel's employees. These benefits include increased agility, improved infrastructure efficiency, and high availability.

Increased Agility

Our goal is to meet business needs more quickly. The private cloud will enable business groups and developers to rapidly acquire and manage their own cloud capacity, and—within pre-defined limits—dynamically scale resources to meet their application needs.



CRM - customer relationship management; IaaS - Infrastructure as a service; SaaS - Software as a service; VM - virtual machine

Figure 1. The Intel IT cloud computing strategy grows the cloud from the inside out.

The private cloud is evolving from our current virtualization initiatives, which have already demonstrated that we can reduce provisioning time to weeks, compared with months in the conventional computing environment.

We expect to further reduce provisioning time using a self-service portal and automated workflows. Ultimately, we expect that this will enable users to obtain infrastructure capacity within minutes.

To help ensure that we can quickly adjust overall cloud capacity to match business requirements, we are developing business intelligence (BI) capabilities that facilitate strategic capacity planning and better monitoring of near-term demand signals and long-term trends.

Improved Infrastructure Efficiency

The foundation of the private cloud is a shared virtualized infrastructure: Computing resources are virtualized and pooled to serve all business groups using a multi-tenant model. We anticipate that this will result in increased efficiency by driving higher levels of resource utilization within each pool. This enables us to reduce power consumption overall by consolidating the workloads from older, less-efficient servers onto a smaller number of more power-efficient, new servers. This can reduce costs by lessening the need to add data center capacity.

High Availability and Security

By building a private cloud, we can deliver the benefits of public clouds without incurring the risks associated with hosting Intel's sensitive applications and data outside the firewall. We expect that our private cloud will enable us to extend even higher levels of availability to all applications without the need for costly specialized hardware and software. This is due to new high-availability capabilities that virtualization software will support over time, as well as the availability of mission-critical features, such as Machine Check Architecture Recovery, in higher-end industry standard servers.

INTEL IT PRIVATE CLOUD ARCHITECTURE

We defined a private cloud architecture to establish the overall direction of our private cloud and to provide a foundation for further development and innovation.

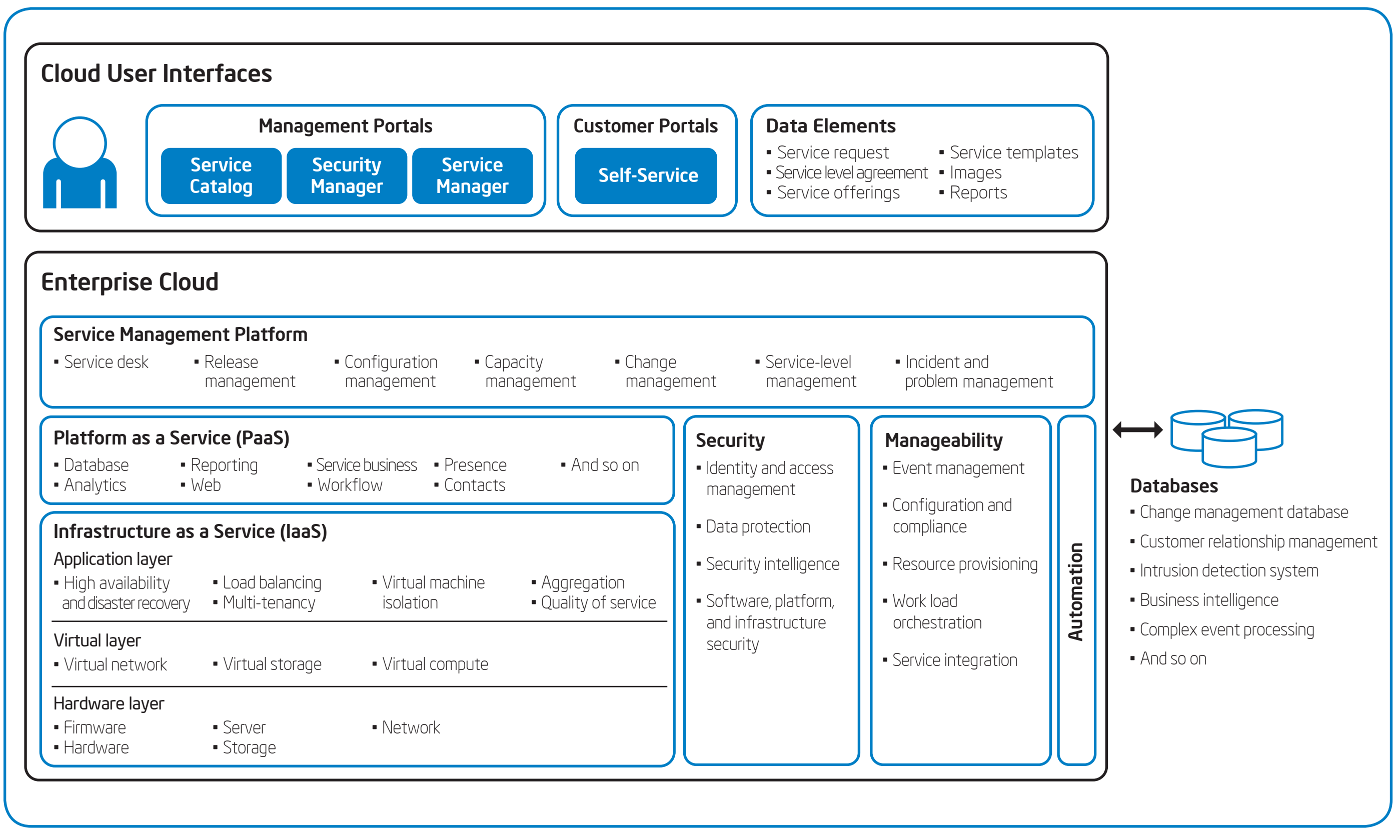


Figure 2. The primary elements of Intel IT's private cloud architecture.

The primary elements of our architecture include infrastructure as a service (IaaS), which provides dynamically scalable resources based on virtualized infrastructure; platform as a service (PaaS), which builds on IaaS and simplifies application development by adding standardized stacks of services used by a wide range of applications; and a self-service portal that enables business groups to request, manage, and track cloud resources. The cloud will also include some internally hosted SaaS multi-tenant applications. This SaaS element could grow in the future to support increased collaboration between Intel and other companies, and hybrid internal-external cloud hosting services.

Our architecture is shown in Figure 2. The primary focus areas are described below.

Infrastructure as a Service

IaaS is the virtualized multi-tenant infrastructure underpinning our private cloud. IaaS infrastructure delivers compute services, typically in the form of a virtual machine (VM) with associated storage and network connectivity. This enables multiple applications, owned by different business groups, to transparently share common underlying physical resources, such as servers and storage. Rather than purchasing physical servers, software, data center space, and network equipment, our internal customers will receive these resources as VMs.

To build this infrastructure, we are accelerating our adoption of virtualization. Today, we have virtualized 15 percent of the servers used for office and enterprise computing. Our goal is to increase this rate to 70 to 80 percent of our office and enterprise servers within about two years.

We plan to achieve this through server refresh in combination with consolidating the workloads of multiple older servers into VMs onto new, more powerful servers based on Intel® Xeon® processors. At the same time, we are aggregating and consolidating physical resources, such as servers, storage frames, and network bandwidth, into large pools, as shown in Figure 3.

This enables us to achieve increased efficiency by driving higher levels of resource utilization within each pool. It also enables implementation of more advanced services, such as balancing workloads across physical servers and storage frames. Workload balancing is achieved with VM live migration, which migrates virtualized applications between physical resources within a resource pool in a way that is transparent to users and does not interrupt the service provided by the application.

We also anticipate that we will be able to extend high availability to most applications by using techniques such as automated VM restart in conjunction with networked storage design.

We plan to implement a common disaster recovery architecture, independent of the OS and applications. In the event of a disaster, this architecture is designed to enable rollover of all designated applications within a resource pool to another site.

Over time, we expect to take advantage of continuing technology improvements to further increase security and quality of service for the applications running in this shared multi-tenant environment.

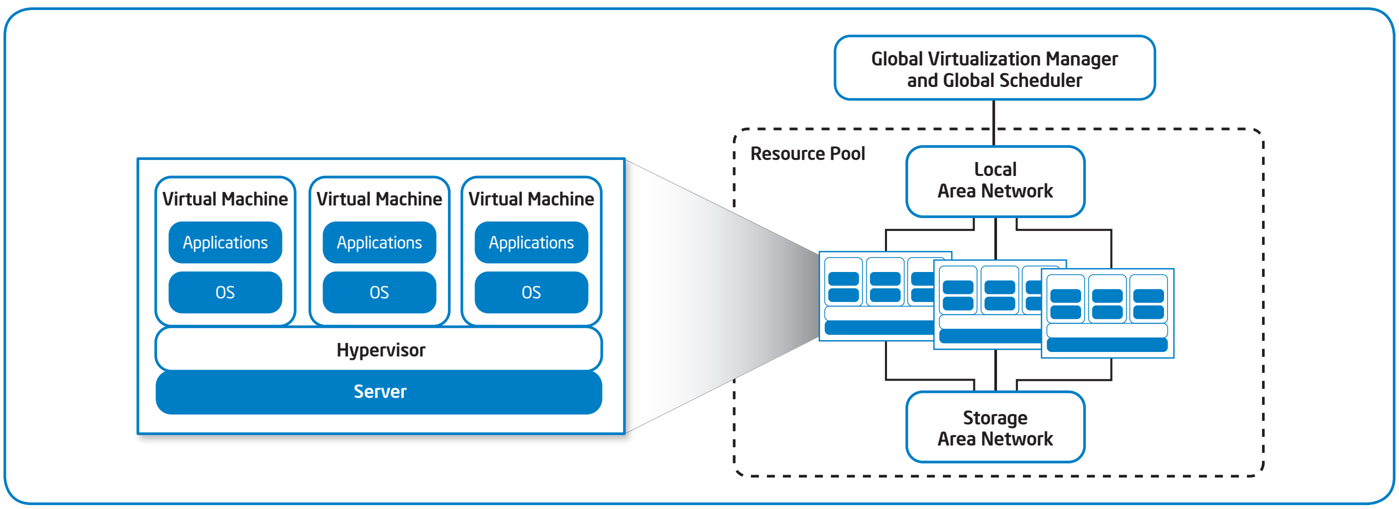


Figure 3. Private cloud infrastructure as a service: Large resource pools based on virtualized infrastructure provide greater flexibility and efficiency. Resources of each physical host are virtualized and presented as multiple virtual machines to run multiple OS and application instances. Private Cloud IaaS consists of pools of virtualized resources (compute, memory, storage, bandwidth) spanning multiple hosts and storage frames. Multi-tenancy (different resource pools for different customers) are on shared physical infrastructure.

Learning from Our Design Grid Environment

During the development of Intel IT's office and enterprise private cloud, we were able to take advantage of the extensive experience we gained between 2006 and 2008 building a large cloud-like grid computing environment to run silicon design jobs. This environment already has delivered benefits, such as increased efficiency and agility, similar to those we are aiming to achieve with our office and enterprise cloud.

The design grid creates a pool of compute resources, including tens of thousands of servers, located across multiple sites. By sharing these resources across sites, we can use our capacity more efficiently and apply Intel's global computing resources to individual silicon design projects.

Today, Intel's designers submit jobs to the grid; these run on servers wherever there is available capacity across the grid. Both the location and the type of hardware used are transparent to the user.

By using the grid, we have increased design compute server utilization to an average 80 percent worldwide. This, together with proactive server refresh, has reduced the need to add data center capacity, delivering about USD 200 million in business value to Intel over several years by enabling us to avoid expensive data center construction.

In addition, the approaches and techniques pioneered during the development of the design grid have provided key learnings that we are applying to our office and enterprise cloud. These include comprehensive business intelligence capabilities used to analyze utilization and plan capacity, a self-service approach that enables users to provision capacity and manage resource utilization, and extensive automation to ensure standardization and data quality across the environment. We are also using key IT personnel that worked on the design grid to help drive our internal cloud project.

Platform as a Service

One of our goals is to enable our developers to spend more time on creating applications and less on systems engineering tasks. To achieve this, our PaaS solution provides developers with standard platforms for application development while relieving them of most of the tasks traditionally associated with maintaining a server OS, such as patching, configuration, and monitoring.

We have built our PaaS solution on top of the IaaS base environment. This allows the platform to respond dynamically to demand by taking advantage of IaaS capabilities. The platform is a standard service that applications can utilize; each platform component can grow and shrink as necessary to meet the needs of the application at each phase of its life cycle.

Our PaaS architecture is intended to promote and facilitate standardization of our most important application environments. We will provide at least two primary PaaS platforms: one based on our industry-standard enterprise computing stack and one based on open-source technologies. Each stack provides developers with a standard set of capabilities, initially including a database, Web server, authentication and authorization services, and an application server role. We plan to over time add other services, such as analytics and reporting.

Application developers request and manage their PaaS application platforms through the self-service portal described in the following section. They select server platforms or complete software stacks based on the requirements of each application.

To fully benefit from all PaaS capabilities, developers need to adapt their applications to be able to react to changing conditions, which may require them to invoke the appropriate IaaS and PaaS services via API calls from within their applications. For example, if the application receives a greater number of requests for Web data than expected, it can request a rapid increase in the number of Web front-end servers.

Self-Service Portal

The self-service portal provides business groups with a graphical interface that they can use to directly request, manage, track, and retire private cloud services and capacity to meet their business needs and demands.

We first implemented a basic self-service portal for our office and enterprise application developers in 2008. This custom portal, developed using off-the-shelf tools, has enabled developers to rapidly create VMs for short-term use during application development. The portal uses automated workflows to accelerate and streamline the provisioning process.

To meet the broader requirements of our enterprise private cloud, in late 2009 we began enhancing the portal to support the entire application life cycle. This year, we completed the first iteration of this new environment, in which developers can acquire, with a few clicks, our standardized platforms. The portal also reports consumption of each cloud resource relative to the amount allocated.

The portal enables users to acquire capacity much more quickly than with previous methods. During the first year of use, our conservative goal is to enable developers to obtain fully configured set of VMs within three hours of submitting a request, compared to a time frame of weeks in our current virtualized environment. Over time, we plan to progressively reduce this to minutes rather than hours. Other key benefits include the ability to shorten development cycles by creating multiple test environments and conducting testing activities in parallel.

Later this year, we plan to extend use of the portal to all private cloud development and test environments, and then to establish the portal as the standard interface for requesting IaaS capacity for production applications. Application owners will be able to move an application through the life cycle from development to production, eliminating laborious manual processes; the underlying environment will optimize the platform based on the requirements of the life cycle stage, and provide information about how capacity is being used. We also are planning other new features, such as the ability to create VM templates that can be reused and shared among users.

The portal is also enabling other important business process changes. Initially, the most important of these is the ability to directly monitor incoming demand rather than relying on manually produced forecasts. As we extend the use of the portal to the full cycle of development, test, and production, we will begin to collect data showing the average throughput of specific applications and business groups. We will be able to use the portal as our primary method for understanding incoming demand signals and establishing a more complete supply chain for IaaS capacity.

IT Service Management and Manageability

The Intel IT private cloud is a highly dynamic, virtualized, and automated environment. Managing this environment will require significant changes in the area of IT service management, which we define as the IT business processes, policies, and roles we use to operate IT services. We also anticipate significant changes in the manageability area, including the tools that we use to monitor and manage infrastructure and applications.

IT SERVICE MANAGEMENT

Intel IT uses an IT service management framework, based on the Information Technology Information Library approach, to create business policies and processes for customer services, guide manageability design for those services, and receive information about the health of each service.

We are continuing to use this framework as we implement our private cloud. For example, creation and deletion of virtual environments are handled as standard changes and documented in our existing change management database.

However, we anticipate significant changes to business processes in several areas.

Capacity Management

Capacity management in our private cloud differs radically from capacity management in a conventional computing environment.

Since early 2010, a centralized cloud infrastructure team has been responsible for managing the overall capacity of the entire cloud, maintaining a buffer of unused capacity that can be assigned to individual applications as needed. Via the self-service portal, business groups request and manage capacity for individual applications, up to pre-defined limits based on business demand. These limits will be specified using policies established in service level agreements.

The infrastructure team will add private cloud infrastructure as necessary to ensure that capacity is not a constraint to the business and to ensure optimum utilization. Over time, we plan to implement manageability automation to enable the infrastructure team to further increase responsiveness and efficiency. A key goal is to minimize infrastructure costs by maintaining a very thin overhead of unused capacity across the private cloud and adding infrastructure capacity on a just-in-time basis.

Business Intelligence

The private cloud will include comprehensive BI capabilities, enabling us to quickly and automatically gather data that previously may have required extensive manual effort. These capabilities will enable us to deliver some of the more advanced capabilities. For example, we will need reliable BI tools to ensure, based on utilization data and demand signals, that we maintain the optimum cloud capacity buffer.

BI tools that enable us to quickly analyze historical consumption and purchasing information, performance and utilization trends, and summaries of alerts and security-related events, will provide us with actionable information that we can apply to a broad range of scenarios.

Costing Information

When users request capacity through the self-service portal, we will provide them with information about the cost of that

Mobile Business PCs and Cloud Computing

For many enterprises, cloud computing poses new questions about the optimal business client computing strategy. With more and more services being delivered from the cloud, which combination of client platforms and service delivery models best meets the needs of users and IT organizations?

Today, the Intel IT environment contains a mixture of conventional and cloud computing services. These are delivered primarily to mobile business PCs, as well as some handheld devices. We have found that whether services are kept in-house or outsourced to the cloud, mobile business PCs offer the best user experience and the flexibility to run different types of applications across our diverse employee population. Only mobile business PCs support the full range of service delivery methods in our environment. In addition, they deliver full mobile computing capabilities for our users, including the ability to work offline.

In addition, due to the continuing consumerization of IT and to employee requests, we are evaluating the use of a growing number of mobile companion devices as core business tools. These will complement our mobile business PC standard, allowing employees to securely access cloud-based information and services from a range of devices.

We are currently evaluating client-side virtualization technologies that will enable us to combine the performance advantages of local execution on mobile business PCs with the benefits of centralized management, providing users with access to both locally installed and cloud-based applications. These technologies take advantage of hardware capabilities in PCs with Intel® Core™ processors, such as Intel® Virtualization Technology, to provide increased security and virtual machine isolation.

capacity to Intel. This will not generate a chargeback to the user; instead it will be a reporting detail that lets Intel employees and their managers understand how well or poorly they are utilizing shared Intel assets. We have found that reporting the cost data is enough to help people make decisions based on the information. Business groups use cost data during project planning when having to choose between options and during operations when seeking ways to reduce expense.

Manageability

A dynamic virtualized, multi-tenant environment results in many new manageability requirements and possibilities. Key requirements include optimal runtime placement of virtualized workloads and comprehensive VM performance monitoring and diagnostics. We will need tools that extend today's capabilities to allow us to compare different execution alternatives across the entire IT environment, based on priority and efficiency policies defined in workload metadata, and if necessary move the workload to another location with minimal human intervention and without interrupting service.

Identifying the causes of application performance problems is particularly important during initial adoption of the private cloud, to help ensure that users have confidence in the new environment and do not request excess capacity in order to ensure adequate performance. However, because the private cloud will be built on a virtualized infrastructure with built-in high availability, detection of faults below the OS level won't have the same urgency that it has had historically. In the past, if part of the infrastructure failed, applications tightly linked to that infrastructure would fail with it.

Security

The security of Intel's data and applications remains a critical focus as we develop and implement our cloud strategy. We are acutely aware of our responsibility as an IT organization to maintain the security and integrity of both corporate intellectual property and personal information, regardless of where it resides or is being used.

Private and public clouds create new security challenges in areas such as resource isolation, security event management, and data protection. In a non-virtualized environment, the separation provided by physical infrastructure is assumed to provide a level of protection for applications and data. As we increase the use of a shared multi-tenant environment based on virtualization, business groups will require differentiated security policies based on data classification and mission criticality, and more visibility into secure data flow in the cloud, and how business-specific security policies are enforced.

Key security focus areas include data encryption and segregation, VM isolation, secure VM migration, virtual network isolation, and security event and access monitoring. Externally facing applications, accessible by business partners or consumers, are an area of particular concern; we anticipate providing further detail about our security approach for externally facing applications in the future.

PRIVATE CLOUD IMPLEMENTATION PLAN

We plan to phase in private cloud capabilities over the next three-plus years; as the private cloud matures, we plan to progressively migrate applications from our conventional environment to the cloud.

Infrastructure Capability Phasing

We are taking a pragmatic approach to implementing the enterprise private cloud. Like other IT organizations, we have limited resources and must prioritize and phase in the desired capabilities over time, while continuing to support our legacy environment.

Our phasing takes advantage of key new enabling technologies in industry-standard hardware and software. These technologies, which make it possible to build a large private enterprise cloud, include major improvements in the scalability of virtualization software and servers based on Intel Xeon processors. The addition of features to Intel® Virtualization Technology and off-the-shelf virtualization software will make it feasible to provide cost-effective high availability, increase VM isolation, and increase application quality of service.

Selected near-term, mid- and long-term technical capabilities are described below and shown in Table 1.

Table 1. Private Cloud Infrastructure Capability Phasing

	Near Term	Mid Term	Long Term
Business Transformation	▪ On-demand self-service ▪ Measured services	▪ Automated workflows ▪ Capacity planning transformation	▪ IT business intelligence solutions for enabling business decisions
Compute, Resiliency	▪ Default to virtualized ▪ Automated virtual machine restart	▪ Cross-site disaster recovery ▪ Machine Check Architecture Recovery	▪ Lockstep virtual machines ▪ Near-native virtualization performance
Storage	▪ Thin-provisioning ▪ Data deduplication ▪ Consolidated backup and restore	▪ Storage resource pools and quality of service ▪ Incremental forever backups and recovery	▪ Solid-state data center ▪ Continuous data protection
Network	▪ 10 GbE ▪ Distributed virtual switch	▪ Unified fabric – compute – storage	▪ 40 GbE
Security	▪ Non-production virtual machines in demilitarized zone ▪ Event and access monitoring	▪ Secure live virtual machine migration ▪ Virtual machine isolation	▪ Public cloud federation ▪ Pervasive encryption
Management	▪ Infrastructure inventory and health ▪ Basic business intelligence - capitalization, performance, and health ▪ Automated patch/provision	▪ Auto end-to-end life cycle management	▪ Cloud brokerage and federation ▪ Private-public cloud live migration
Data Center	▪ Energy savings via virtualization	▪ Cross-platform power and data center management	▪ Near-linear power scaling ▪ Power usage effectiveness improvements
Clients	▪ Client virtualization ▪ Mobile business PCs plus handhelds	▪ Expanded small form-factor support	▪ Client-aware services optimized across a range of clients

Table 2. Intel IT Enterprise Private Cloud Application Phasing

	Internal or External Use	Near Term	Mid Term	Long Term
Infrastructure services, including network, hosting, security, and manageability	Internal and external use	Pre-production and selected production	All	All
Line of business and departmental applications	Internal use only	All	All	All
	Externally facing applications	All	All	All
Mission-critical	Internal use only	Pre-production and selected production	All	All
	Externally facing applications	NA	Pre-production	All

All = Production and pre-production

NEAR TERM (CURRENT TO 18 MONTHS)

We will continue to broaden the capabilities of the self-service portal, as we expand its use to the production environment. We anticipate increased energy and cost savings through server refresh as we accelerate implementation of virtualization across our environment. We anticipate being able to provide high availability as a default capability across the environment.

MID TERM (18 MONTHS TO THREE YEARS)

The private cloud will become capable of running our most critical applications. We expect to implement just-in-time centralized capacity planning, using new BI capabilities. To safeguard the private cloud, we plan to implement cross-site, application-independent disaster recovery. New security capabilities will include secure live VM migration and increased VM isolation. Other planned capabilities include a unified compute and storage fabric designed to reduce complexity and cost.

LONG TERM (MORE THAN THREE YEARS)

As private and public clouds mature, we expect to be able to take advantage of services providing increased efficiency and flexibility, such as cloud brokerage and federated identity management with public clouds. Other anticipated capabilities include continuous data protection, solid-state storage, near-native virtualization performance, and increased security using pervasive encryption. We plan to continue to automate areas that require manual effort.

Application Phasing

In general, applications that are mission-critical, have demanding requirements, provide competitive advantage or include sensitive data will continue to be hosted within Intel rather than outsourced to public clouds.

As the technology capabilities of our private cloud increase, we anticipate moving an increasing number of these applications from our conventional computing environment to the private cloud. Over time, the private cloud will be able to host production instances of some of our most demanding internal applications, including externally facing applications, as shown in Table 2.

At the same time, we will continue to outsource some applications to external clouds. In general, we expect that candidates for outsourcing will be applications that do not provide competitive advantage, can be run at a lower cost outside Intel, are not mission-critical, and do not contain sensitive information.

CONCLUSION

Our enterprise private cloud is designed to deliver critical business benefits, including reduced provisioning times, higher resource utilization, high availability, and improved capacity management.

Because of the extensive scope of this initiative, we plan to deliver private cloud capabilities in phases over the next three or more years. As we add these capabilities, we expect that the cloud will become suitable for hosting highly demanding, mission-critical business applications.

FOR MORE INFORMATION

Find additional IT@Intel white papers at www.intel.com/IT.

- "Developing an Enterprise Cloud Computing Strategy"
- "Architecting Software as a Service for the Enterprise"
- "Better Together: Rich Client PCs and Cloud Computing"
- "Intel Cloud Computing Taxonomy and Ecosystem Analysis"

CONTRIBUTORS

Uttam Shetty

Ravi Subramaniam

ACRONYMS

BI	business intelligence
IaaS	infrastructure as a service
PaaS	platform as a service
SaaS	software as a service
VM	virtual machine

For more straight talk on current topics from Intel's IT leaders, visit www.intel.com/it.

Printed in USA
0610/KC/KC/PDF

♻ Please Recycle
323007-001US