Sun microsystems

TAKE YOUR BUSINESS TO A

# HIGHER LEVEL

Sun cloud-computing technology scales your infrastructure to take advantage of new business opportunities.

# TABLE OF CONTENTS

# CLOUD COMPUTING AT A HIGHER LEVEL

**In many ways, cloud computing is simply a metaphor for the Internet, the increasing movement of compute and data resources onto the Web. But there's a difference: cloud computing represents a new tipping point for the value of network computing. It delivers higher efficiency, massive scalability, and faster, easier software development. It's about new programming models, new IT infrastructure, and the enabling of new business models.**

For those developers and enterprises who want to embrace cloud computing, Sun is developing critical technologies to deliver enterprise scale and systemic qualities to this new paradigm:

**Interoperability** — While most current clouds offer closed platforms and vendor lock-in, developers clamor for interoperability. Sun's open-source product strategy and Java™ principles are focused on providing interoperability for large-scale computing resources. Think of the existing cloud "islands" merging into a new, interoperable "Intercloud" where applications can be moved to and operate across multiple platforms.

**High-density horizontal computing** — Sun is pioneering high-power-density compute-node architectures and extreme-scale Infiniband fabrics as part of our top-tier HPC deployments. This high-density technology is being incorporated into our large-scale cloud designs.

**Data in the cloud** — More than just compute utilities, cloud computing is increasingly about petascale data. Sun's Open Storage products offer hybrid data servers with unprecedented efficiency and performance for the emerging data-intensive computing applications that will become a key part of the cloud.

These technology bets are focused on driving more efficient large-scale cloud deployments that can provide the infrastructure for next-generation business opportunities: social networks, algorithmic trading, continuous risk analysis, and so on.

# WHY CLOUD COMPUTING?

> *"The rise of the cloud is more than just another platform shift that gets geeks excited. It will undoubtedly transform the IT industry, but it will also profoundly change the way people work and companies operate."*
>
> —The Economist, "Let it Rise," 10/23/08

## Clouds: Much More Than Cheap Computing

Cloud computing brings a new level of efficiency and economy to delivering IT resources on demand — and in the process it opens up new business models and market opportunities.

While many people think of current cloud computing offerings as purely "pay by the drink" compute platforms, they're really a convergence of two major interdependent IT trends:

**IT Efficiency — Minimize costs** where companies are converting their IT costs from capital expenses to operating expenses through technologies such as virtualization. Cloud computing begins as a way to improve infrastructure resource deployment and utilization, but fully exploiting this infrastructure eventually leads to a new application development model.

**Business Agility — Maximize return** using IT as a competitive weapon through rapid time to market, integrated application stacks, instant machine image deployment, and petascale parallel programming. Cloud computing is embraced as a critical way to revolutionize time to service. But inevitably these services must be built on equally innovative rapid-deployment-infrastructure models.

To be sure, these trends have existed in the IT industry for years. However, the recent emergence of massive network bandwidth and virtualization technologies has enabled this transformation to a new services-oriented infrastructure.

Cloud computing enables IT organizations to increase hardware utilization rates dramatically, and to scale up to massive capacities in an instant — without constantly having to invest in new infrastructure, train new personnel, or license new software. It also creates new opportunities to build a better breed of network services, in less time, for less money.

> *"By 2011, early technology adopters will forgo capital expenditures and instead purchase 40% of their IT infrastructure as a service . . .. 'Cloud computing' will take off, thus untying applications from specific infrastructure."*
>
> —Gartner Press Release, "Gartner Highlights Key Predictions for IT Organisations and Users in 2008 and Beyond," 1/31/08

### IT Efficiency on a Whole New Scale

Cloud computing is all about efficiency. It provides a way to deploy and access everything from single systems to huge amounts of IT resources — on demand, in real time, at an affordable cost. It makes high-performance compute and high-capacity storage available to anyone with a credit card. And since the best cloud strategies build on concepts and tools that developers already know, clouds also have the potential to redefine the relationship between information technology and the developers and business units that depend on it.

**Reduce capital expenditures** — Cloud computing makes it possible for companies to convert IT costs from capital expense to operating expense through technologies such as virtualization.

**Cut the cost of running a datacenter** — Cloud computing improves infrastructure utilization rates and streamlines resource management. For example, clouds allow for self-service provisioning through APIs, bringing a higher level of automation to the datacenter and reducing management costs.

**Eliminate overprovisioning** — Cloud computing provides scaling on demand, which, when combined with utility pricing, removes the need to overprovision to meet demand. With cloud computing, companies can scale up to massive capacities in an instant.

For those who think cloud computing is just fluff, take a closer look at the cloud offerings that are already available. Major Internet providers Amazon.com, Google, and others are leveraging their infrastructure investments and "sharing" their large-scale economics. Already the bandwidth used by Amazon Web Services (AWS) exceeds that associated with their core e-tailing services. Forward-looking enterprises of all types — from Web 2.0 startups to global enterprises — are embracing cloud computing to reduce infrastructure costs.

*The New York Times* needed to convert 11 million articles and images in its archive (from 1851 to 1980) to PDF. Their Internal IT said it would take them seven weeks. In the meantime, one developer using 100 Amazon EC2 simple Web service interface instances running Hadoop (an open-source implementation similar to MapReduce) completed the job in 24 hours for less than $300.

—open.blogs.nytimes.com, "Self-service, Prorated Super Computing Fun!" 11/1/07, open.blogs.nytimes.com/2007/11/01/self-service-prorated-super-computing-fun/

### Faster, More Flexible Programming

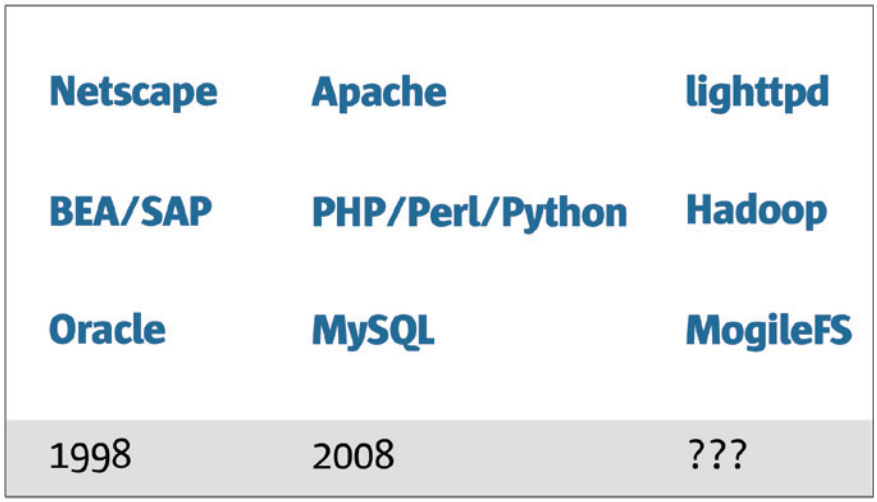
Cloud computing isn't only about hardware — it's also a programming revolution. Agile, easy-to-access, lightweight Web protocols — coupled with pervasive horizontally scaled architecture — can accelerate development cycles and time to market with new applications and services. New business functions are now just a script away.

**Accelerated cycles** — The cloud computing model provides a faster, more efficient way to develop the new generation of applications and services. Faster development and testing cycles means businesses can accomplish in hours what used to take days, weeks, or months.

**Increase agility** — Cloud computing accommodates change like no other model. For example, Animoto Productions, makers of a mashup tool that creates video from images and music, used cloud computing to scale up from 50 servers to 3,500 in just three days. Cloud computing can also provide a wider selection of more lightweight and agile development tools, simplifying and speeding up the development process.

The immediate impact will be unprecedented flexibility in service creation and accelerated development cycles. But at the same time, development flexibility could become constrained by APIs if they're not truly open. Cloud computing can usher in a new era of productivity for developers if they build on platforms that are designed to be federated rather than centralized. But there's a major shift underway in programming culture and the languages that will be used in clouds.

*What's the Next Web Stack?*

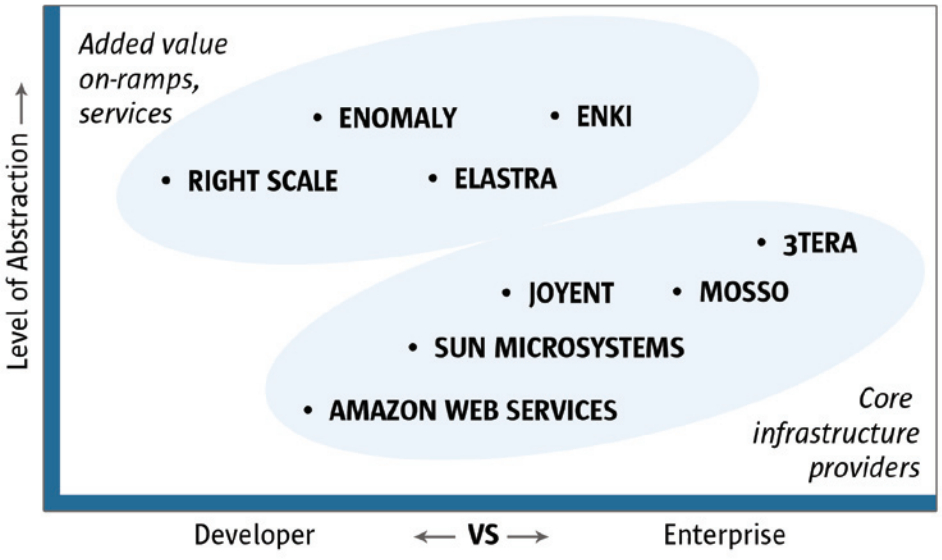


Today, the integrated, optimized, open-source Apache, MySQL, PHP/Perl/Python (AMP) stack is the preferred platform for building and deploying new Web applications and services. Cloud computing will be the catalyst for the adoption of an even newer stack of more lightweight, agile tools such as lighttpd, an open-source Web server; Hadoop, the free Java software framework that supports data-intensive distributed applications; and MogileFS, a file system that enables horizontal scaling of storage across any number of machines.

## Compelling New Opportunities: The Cloud Ecosystem

But cloud computing isn't just about a proliferation of Xen image stacks on a restricted handful of infrastructure providers. It's also about an emerging ecosystem of complementary services that provide computing resources such as on-ramps for cloud abstraction, professional services to help in deployment, specialized application components such as distributed databases, and virtual private datacenters for the entire range of IT providers and consumers.

These services span the range of customer requirements, from individual developers and small startups to large enterprises. And they continue to expand the levels of virtualization, a key architectural component of the cloud that offers ever-higher abstractions of underlying services.



# How Did Cloud Computing Start?

At a basic level, cloud computing is simply a means of delivering IT resources as services. Almost all IT resources can be delivered as a cloud service: applications, compute power, storage capacity, networking, programming tools, even communications services and collaboration tools.

Cloud computing began as large-scale Internet service providers such as Google, Amazon, and others built out their infrastructure. An architecture emerged: massively scaled, horizontally distributed system resources, abstracted as virtual IT services and managed as continuously configured, pooled resources. This architectural model was immortalized by George Gilder in his October 2006 *Wired* magazine article titled "The Information Factories." The server farms Gilder wrote about were architecturally similar to grid computing, but where grids are used for loosely coupled, technical computing applications, this new cloud model was being applied to Internet services.

> *"In this architecture, the data is mostly resident on servers 'somewhere on the Internet' and the application runs on both the 'cloud servers' and the user's browser."*
>
> —Eric Schmidt in 'Information Factories' by G. Gilder

Both clouds and grids are built to scale horizontally very efficiently. Both are built to withstand failures of individual elements or nodes. Both are charged on a per-use basis. But while grids typically process batch jobs, with a defined start and end point, cloud services can be continuous. What's more, clouds expand the types of resources available — file storage, databases, and Web services — and extend the applicability to Web and enterprise applications.

At the same time, the concept of utility computing became a focus of IT design and operations. As Nick Carr observed in his book *The Big Switch*, computing services infrastructure was beginning to parallel the development of electricity as a utility. Wouldn't it be great if you could purchase compute resources, on demand, only paying for what you need, when you need it?

For end users, cloud computing means there are no hardware acquisition costs, no software licenses or upgrades to manage, no new employees or consultants to hire, no facilities to lease, no capital costs of any kind — and no hidden costs. Just a metered, per-use rate or a fixed subscription fee. Use only what you want, pay only for what you use.

Cloud computing actually takes the utility model to the next level. It's a new and evolved form of utility computing in which many different types of resources (hardware, software, storage, communications, and so on) can be combined and recombined on the fly into

the specific capabilities or services customers require. From CPU cycles for HPC projects to storage capacity for enterprise-grade backups to complete IDEs for software development, cloud computing can deliver virtually any IT capability, in real time.

Under the circumstances it is easy to see that a broad range of organizations and individuals would like to purchase "computing" as a service, and those firms already building hyperscale distributed data centers would inevitably choose to begin offering this infrastructure as a service.

## Harnessing Cloud Computing

So how does an individual or a business take advantage of the cloud computing trend? It's not just about loading machine images consisting of your entire software stack onto a public cloud like AWS — there are several different ways to exploit this infrastructure and explore the ecosystem of new business models.

### Use the Cloud

The number and quality of public, commercially available cloud-based service offerings is growing fast. Using the cloud is often the best option for startups, research projects, Web 2.0 developers, or niche players who want a simple, low-cost way to "load and go." If you're an Internet startup today, you will be mandated by your investors to keep you IT spend to a minimum. This is certainly what the cloud is for.

### Leverage the Cloud

Typically, enterprises are using public clouds for specific functions or workloads. The cloud is an attractive alternative for:

**Development and testing** — This is perhaps the easiest cloud use case for enterprises (not just startup developers). Why wait to order servers when you don't even know if the project will pass the proof of concept?

**Functional offloading** — You can use the cloud for specific workloads. For example, SmugMug does its image thumbnailing as a batch job in the cloud.

> *"We really don't want to operate datacenters anymore. We'd rather spend our time giving our customers great service and writing great software than managing physical hardware."*
>
> —Don MacAskill, CEO, SmugMug

**Augmentation** — Clouds give you a new option for handling peak load or anticipated spikes in demand for services. This is a very attractive option for enterprises, but also potentially one of the most difficult use cases. Success is dependent on the statefulness of the application and the interdependence with other datasets that may need to be replicated and load-balanced across the two sites.

**Experimenting** — Why download demos of new software, and then install, license, and test it? In the future, software evaluation can be performed in the cloud, before licenses or support need to be purchased.

### Build the Cloud

Many large enterprises understand the economic benefits of cloud computing but want to ensure strict enforcement of security policies. So they're experimenting first with "private" clouds (see section 1.4), with a longer-term option of migrating mature enterprise applications to a cloud that's able to deliver the right service levels.

Other companies may simply want to build private clouds to take advantage of the economics of resource pools and standardize their development and deployment processes.

### Be the Cloud

This category includes both cloud computing service providers and cloud aggregators — companies that offer multiple types of cloud services.

As enterprises and service providers gain experience with the cloud architecture model and confidence in the security and access-control technologies that are available, many will decide to deploy externally facing cloud services. The phenomenal growth rates of some of the public cloud offerings available today will no doubt accelerate the momentum. Amazon's EC2 was introduced only two years ago and officially graduated from beta to general availability in October 2008.

Cloud service providers can:

- Provide new routes to market for startups and Web 2.0 application developers
- Offer new value-added capabilities such as analytics
- Derive a competitive edge through enterprise-level SLAs
- Help enterprise customers develop their own clouds

If you're building large datacenters today, you should probably be thinking about whether you're going to offer cloud services.

## Public, Private, and Hybrid Clouds

A company may choose to use a service provider's cloud or build its own — but is it always all or nothing? Sun sees an opportunity to blend the advantages of the two primary options:

**Public clouds** are run by third parties, and jobs from many different customers may be mixed together on the servers, storage systems, and other infrastructure within the cloud. End users don't know who else's job may be me running on the same server, network, or disk as their own jobs.

**Private clouds** are a good option for companies dealing with data protection and service-level issues. Private clouds are on-demand infrastructure owned by a single customer who controls which applications run, and where. They own the server, network, and disk and can decide which users are allowed to use the infrastructure.

But even those who feel compelled in the short term to build a private cloud will likely want to run applications both in privately owned infrastructure and in the public cloud space. This gives rise to the concept of a hybrid cloud.

**Hybrid clouds** combine the public and private cloud models. You own parts and share other parts, though in a controlled way. Hybrid clouds offer the promise of on-demand, externally provisioned scale, but add the complexity of determining how to distribute applications across these different environments. While enterprises may be attracted to the promise of a hybrid cloud, this option, at least initially, will likely be reserved for simple stateless applications that require no complex databases or synchronization.

# > CLOUD COMPUTING DEFINED

*"It's one of the foundations of the next generation of computing. . .. It's a world where the network is the platform for all computing, where everything we think of as a computer today is just a device that connects to the big computer we're building. Cloud computing is a great way to think about how we'll deliver computing services in the future."*

—Tim O'Reilly, CEO, O'Reilly Media

## Cornerstone Technology

While the basic technologies of cloud computing such as horizontally scaled, distributed compute nodes have been available for some time, virtualization — the abstraction of computer resources — is the cornerstone technology for all cloud architectures. With the ability to virtualize servers (behind a hypervisor-abstracted operating system), storage devices, desktops, and applications, a wide array of IT resources can now be allocated on demand.

The dramatic growth in the ubiquitous availability of affordable high-bandwidth networking over the past several years is equally critical. What was available to only a small percentage of Internet users a decade ago is now offered to the majority of Internet users in North America, Europe, and Asia: high bandwidth, which allows massive compute and data resources to be accessed from the browser. Virtualized resources can truly be anywhere in the cloud — not just across gigabit datacenter LANs and WANs but also via broadband to remote programmers and end users.

Additional enabling technologies for cloud computing can deliver IT capabilities on an absolutely unprecedented scale. Just a few examples:

**Sophisticated file systems** such as ZFS can support virtually unlimited storage capacities, integration of the file system and volume management, snapshots and copy-on-write clones, on-line integrity checking, and repair.

**Patterns in architecture** allow for accelerated development of superscale cloud architectures by providing repeatable solutions to common problems.

**New techniques** for managing structured, unstructured, and semistructured data can provide radical improvements in data-intensive computing.

**Machine images** can be instantly deployed, dramatically simplifying and accelerating resource allocation while increasing IT agility and responsiveness.

## The Architectural Services Layers of Cloud Computing

While the first revolution of the Internet saw the three-tier (or n-tier) model emerge as a general architecture, the use of virtualization in clouds has created a new set of layers: applications, services, and infrastructure. These layers don't just encapsulate on-demand resources, they also define a new application development model. And within each layer of abstraction there are myriad business opportunities for defining services that can be offered on a pay-per-use basis.

### Software as a Service (SaaS)

SaaS is at the highest layer and features a complete application offered as a service, on-demand, via multitenancy — meaning a single instance of the software runs on the

provider's infrastructure and serves multiple client organizations. The most widely known example of SaaS is Salesforce.com, but there are now many others, including the Google Apps offering of basic business services such as e-mail. Of course, Salesforce.com's multi-tenant application has preceded the definition of cloud computing by a few years. On the other hand, like many other players in cloud computing, Salesforce.com now operates at more than one cloud layer with its release of Force.com, a companion application development environment, or platform as a service.

**Platform as a Service (PaaS)**

The middle layer, or PaaS, is the encapsulation of a development environment abstraction and the packaging of a payload of services. The archetypal payload is a Xen image (part of Amazon Web Services) containing a basic Web stack (for example, a Linux distro, a Web server, and a programming environment such as Pearl or Ruby).

PaaS offerings can provide for every phase of software development and testing, or they can be specialized around a particular area, such as content management.

Commercial examples include Google App Engine, which serves applications on Google's infrastructure. PaaS services such as these can provide a great deal of flexibility but may be constrained by the capabilities that are available through the provider.

**Infrastructure as a Service (IaaS)**

IaaS is at the lowest layer and is a means of delivering basic storage and compute capabilities as standardized services over the network. Servers, storage systems, switches, routers, and other systems are pooled (through virtualization technology, for example) to handle specific types of workloads — from batch processing to server/storage augmentation during peak loads.

The best-known commercial example is Amazon Web Services, whose EC2 and S3 services offer bare-bones compute and storage services (respectively). Another example is Joyent whose main product is a line of virtualized servers which provide a highly scalable on-demand infrastructure for running Web sites, including rich Web applications written in Ruby on Rails, PHP, Python, and Java.

## > INSIDE THE CLOUD

A key attraction of cloud computing is that it conceals the complexity of the infrastructure from developers and end users. They don't know or need to know what's in the cloud — they only care that it delivers the services they need. But those who choose to build clouds for private use or as a business in itself have critical technology decisions to make in abstracting and managing underlying resources. This section takes a closer look at the key architectural attributes and underlying technologies of virtualization.

# Virtualization

Virtualization is a cornerstone design technique for all cloud architectures. In cloud computing it refers primarily to platform virtualization, or the abstraction of physical IT resources from the people and applications using them. Virtualization allows servers, storage devices, and other hardware to be treated as a pool of resources rather than discrete systems, so that these resources can be allocated on demand. In cloud computing, we're interested in techniques such as *paravirtualization*, which allows a single server to be treated as multiple virtual servers, and *clustering*, which allows multiple servers to be treated as a single server.

As a means of encapsulation of physical resources, virtualization solves several core challenges of datacenter managers and delivers specific advantages, including:

**Higher utilization rates** — Prior to virtualization, server and storage utilization rates in enterprise datacenters typically averaged less than 50% (in fact, 10% to 15% utilization rates were common). Through virtualization, workloads can be encapsulated and transferred to idle or underused systems — which means existing systems can be consolidated, so purchases of additional server capacity can be delayed or avoided.

**Resource consolidation** — Virtualization allows for consolidation of multiple IT resources. Beyond server and storage consolidation, virtualization provides an opportunity to consolidate the systems architecture, application infrastructure, data and databases, interfaces, networks, desktops, and even business processes, resulting in cost savings and greater efficiency.

**Lower power usage/costs** — The electricity required to run enterprise-class datacenters is no longer available in unlimited supplies, and the cost is on an upward spiral. For every dollar spent on server hardware, an addition dollar is spent on power (including the cost of running and cooling servers). Using virtualization to consolidate makes it possible to cut total power consumption and save significant money.

**Space savings** — Server sprawl remains a serious problem in most enterprise datacenters, but datacenter expansion is not always an option, with building costs averaging several thousand dollars per square foot. Virtualization can alleviate the strain by consolidating many virtual systems onto fewer physical systems.

**Disaster recovery/business continuity** — Virtualization can increase overall service-level availability rates and provide new options for disaster recovery solutions.

**Reduced operations costs** — The average enterprise spends $8 in maintenance for every $1 spent on new infrastructure. Virtualization can change the server-to-admin ratio, reduce the total administrative workload, and cut total operations costs.

### Operating System Virtualization

The use of OS-level virtualization or partitioning (such as LPARs, VPARs, NPARs, Dynamic System Domains, and so on) in cloud architectures can help solve some of the core security, privacy, and regulatory issues that could otherwise hinder the adoption of cloud computing.

For example, OS virtualization such as that provided by Solaris™ Containers makes it possible to maintain a one-application-per-server deployment model while simultaneously sharing hardware resources. Solaris Containers isolate software applications and services using software-defined boundaries and allow many private execution environments to be created within a single instance of the Solaris OS. Each environment has its own identity, separate from the underlying hardware, so it behaves as if it's running on its own system, making consolidation simple, safe, and secure. This makes it possible to reduce the administrative overhead and complexity of managing multiple operating systems and improve utilization at the same time.

### Platform Virtualization

Platform virtualization allows arbitrary operating systems and resulting application environments to run on a given system. There are two basic models for this system virtualization: full virtualization, or a complete simulation of underlying hardware, and paravirtualization, which offers a "mostly similar" model of the underlying hardware. These are implemented as Type 1 hypervisors, which run directly on hardware, and Type 2 hypervisors, which run on top of a traditional operating system.

Each of the top virtualization vendors offers variations of both models. It's important to realize that there are design and performance trade-offs for any model of system virtualization. Generally, the more abstract the OS is made from the underlying hardware, the less hardware-specific features can be accessed. Increased OS abstraction can also increase the potential for performance reduction and limitations.

### Network Virtualization

Load-balancing techniques have been a hot topic in cloud computing because, as the physical and virtual systems within the cloud scale up, so does the complexity of managing the workload that's performed to deliver the service.

Load balancers group multiple servers and services behind virtual IP addresses. They provide resource-based scheduling of service requests and automatic failover when a node fails. While hardware balancers may outperform software-based balancers, their flexibility is always limited. Engineers end up either writing software that interacts with hardware via a suboptimal user interface or using a large number of computers to solve the problem.

A significant challenge in cloud computing networking is not just the provisioning of individual virtual network interfaces to a given virtual environment, but also the increasing need of cloud infrastructures to offer a more complicated virtual private datacenter, which provisions a set of different system roles and the logical interconnections between those roles.

### Application Virtualization

There is also a software angle to "containers" within the cloud. The Web container technology implemented in the cloud greatly impacts developer productivity and flexibility.

The Web container is the part of the application server that manages servlets, JavaServer™ Page (JSP) files, and other Web-tier components. But not all Web container technologies are created equal. Apache Tomcat, for example, is a popular open-source Web container technology, but it has limitations for developers who want to go beyond Web-tier applications. If an application needs to use persistence, clustering, failover, messaging, or Enterprise Java Beans (EJB™), these capabilities have to be added to Tomcat one by one, whereas the GlassFish™ Project provides an integrated collection of Java EE containers that delivers all of these capabilities.

Today, most cloud computing offerings concentrate on platform virtualization, and the developer chooses the OS and development platform. But increasingly public clouds and certainly private clouds will offer higher-level development environment programming abstractions. Over time, we might expect the level of abstraction that the developer interfaces with to move gradually upward as more functionality percolates down into the platform.

## Software Deployment

With cloud computing offering increasing abstraction of the underlying hardware, a related, but separate, set of decisions must be made concerning how the software and applications are deployed on cloud infrastructure. The cloud computing model is flexible enough to accommodate applications of all types and sizes, at all phases of development and deployment. Cloud architectures can be the delivery platform for monolithic, proprietary applications such as ERP and CRM; the development and deployment platform for a new breed of lightweight, dynamically typed applications built on open source software; or a source of IDEs and testing resources.

### Software Packaging

The software-based packaging of software components, data, server and storage pools, and other cloud resources makes efficient resource allocation, re-use, and management possible.

The packaging system is essentially a software delivery mechanism that simplifies and accelerates the installation of everything from operating systems to applications to end-user data. The image packaging system (IPS) for the OpenSolaris™ OS, for example, makes it possible to create images and install, search, update, and manage packages in the image. The IPS can also be used to create custom packages and repositories and to publish and manage packages to the repositories. Increasingly, cloud operators and datacenters are moving away from installing systems software on each server, choosing to deploy golden images on farms of servers. In any case, basic software configurations must be provisioned on the system resource pools.

### Machine Images

Increasingly, a similar image-based deployment model is becoming the primary mechanism for deploying application development payloads on virtual resource pools. Machine images contain user-specific applications, libraries, data, and associated configuration settings and are hosted within the cloud. Perhaps the best-known examples are Xen images. This model of deployment is the basis of Amazon Machine Images (AMIs), which are built around a variety of kernels. You can select among a range of public AMIs (preconfigured, templated images) or build your own custom/private AMI.

Most AMIs are built on some form of Linux, such as Fedora or Ubuntu. They're easy to modify and share, and tools are provided by Amazon. Paid AMIs can be created by ISVs and stored on Amazon Simple Storage Service (S3). Amazon Machine Images are available for OpenSolaris (32-bit) and Solaris Express (32-bit and 64-bit) operating systems.

## > SUN'S CLOUD PHILOSOPHIES

It's Sun's goal to combine the systems and software to build a cloud, the architectural expertise to maximize cloud capabilities, and the technologies to take cloud computing to a higher level. Our approach is to deliver all the components that enterprises, developers, and end users need to build cloud environments, through our own or partners' offerings.

## Open Source and Interoperability

While some clouds are closed platforms with vendor lock-in, Sun's open-source philosophy and Java principles are the basis of our strategy: providing interoperability for large-scale computing resources and distributing applications across multiple cloud infrastructure components.

Ideally, users of cloud computing would be able to move their applications among a variety of standardized providers who offer open-source interfaces to common services. Today, most clouds are proprietary, and even where the components offered are open source, cloud operators cultivate significant lock-in through their underlying services, such as storage and databases.

Private clouds created by individual enterprises certainly have the advantage of offering (and requiring) adherence to corporate standards, but even here the desire for enterprises to be able to "flex" their private clouds with public-cloud capacity on demand calls for increasing levels of open standards to emerge in the cloud computing milieu.
Think of the existing cloud islands merging into a new, interoperable "Intercloud." The Intercloud would take the basic concept of the Internet up another level, essentially a global cloud of clouds, united by a set of protocols and software, yet segmented (for security and predictability) into clusters and "intraclouds."

Sun is working toward the vision of the Intercloud by expanding research and development efforts in four key open-source areas:

**Software** — Providing the open-standards-based tools that developers and architects need to build agile services that can be deployed in the cloud — from Sun's Web stack to software elements from other vendors

**Systems** — Delivering compute, storage, and networking systems that interoperate with each other and integrate with systems from other vendors, whether they're based on AMD™, Intel®, or SPARC® architectures

**Microelectronics** — Pushing the envelope for chip multithreading (CMT) and multicore computing; moving to ever-higher compute densities within the cloud

**Services** — Supporting development efforts through a broad range of professional services, network services, and value-added service offerings from partners (ISVs, OEMs, channel partners, and systems integrators)

## Comprehensive Product Portfolio

Sun is specially positioned to bring cloud computing to fruition because we have a top-to-bottom solution to support the entire stack — from microprocessors (and servers) offering unique multithreaded power/performance capability to innovative Open Storage solutions to a full complement of application development software technologies, including virtualization, identity management, and Web 2.0 programming platform tools.

Sun products are integrated across all the layers of technology involved and can be integrated with standards-based technologies from other vendors. And many of Sun's products and technologies are on-ramps to cloud computing, including virtually all of Sun's server and storage systems, the Solaris OS, the ZFS file system, the Sun xVM portfolio, and Sun Ray™ desktops.

## Enterprise-grade Systemic Qualities

The unpredictable nature of cloud computing workloads requires that clouds be architected for extremely high levels of efficiency, service-level availability, scalability, manageability, security, and other systemic qualities.

Initially, cloud computing platforms are attractive for their low-cost development and deployment capabilities. But as firms increasingly use cloud platforms for actual production environments, they will require enterprise-level SLAs.
Maximizing systemic qualities requires integrating the development of these qualities into the design process of large-scale architectures. For cloud computing, the focus of systemic qualities is different from the host-based, client-server models and Web-based models of the past. In some ways the challenge to achieve systemic qualities is more complex. On the other hand, if these architectures are properly designed from the beginning, this can contribute to, and not challenge, the achievement of systemic qualities.

Sun has introduced a number of innovations that deliver enterprise-grade systemic qualities in cloud computing architectures. These innovations are primarily in the areas of efficiency and economy, reliability and availability, density and scalability, agility, and security.

### Efficiency/Economy

- Pioneered the "green"-computing movement with energy-efficient CoolThreads™ technology and the use of printed circuit boards using far less hazardous materials — which has saved companies hundreds of millions of dollars in energy costs alone
- Low-cost innovator with offerings that span datacenter design, hardware, OS, and software components; leading advocate of open-source software; using virtualization technologies in all aspects of product design and development in order to achieve greater power efficiencies
- Enables large numbers of servers to function more efficiently and saves costs on energy, cabling, HVAC, and so on; minimizes capital expenditures (infrastructure owned by the provider)

### Reliability/Availability

- Service-level availability through the built-in RAS features of the Solaris OS and OpenSolaris OS and sophisticated hardware-level availability features from failover to clustering to dynamic reconfiguration
- Reliability by way of multiple redundant sites, which makes it suitable for business continuity and disaster recovery

### Density/Scalability

- Extremely high density; large number of cores per rack and transactions per rack unit
- Cloud nodes in the form of Sun™ Modular Datacenter systems and the Sun Constellation System cloud computing environment; virtualization and dynamic reconfiguration for efficient scaling on demand without having to engineer for peak loads

### Agility

- Multiple hardware architectures to customize systems to workloads
- Multitenancy, enabling sharing of resources (and costs) among a large pool of users, allowing for:
    - Centralization of infrastructure in areas with lower costs such as real estate and electricity
    - Peak-load capacity increases without engineering for highest possible load levels
- Sun Grid Engine software to request and reserve resources for specific amounts of time (see sun.com/software/gridware.)

### Security

Typically, security improves with the centralization of data and increased security-focused resources, so cloud computing raises concerns about loss of control over certain sensitive data. Accesses are typically logged, but accessing the audit logs themselves can be difficult or impossible. Sun addresses the challenges with a range of innovations. For example:

- The Solaris 10 OS includes Process and User Rights Management, Trusted Extensions for Mandatory Access Control (MAC), and the Cryptographic Framework and Secure By Default Networking that allows developers to safely deliver new solutions, consolidate with security, and protect mission-critical data.
- Sun Identity Manager software is the market leader, delivering the only complete user-provisioning and metadirectory solution that enhances enterprise security.
- The Java Composite Application Platform Suite (Java CAPS) contains everything an enterprise needs to develop and deploy an SOA platform for the reuse of existing applications, the delivery of new services, and enabling legacy and packaged applications to rapidly integrate within an existing infrastructure. The suite is SOA-based, is fully integrated, and delivers a rich set of integration and composite application capabilities, including business process management (BPM), industry-leading messaging, rich transformation, and a broad array of connectors.
- Sun is positioned in the Leaders Quadrant of Gartner's Magic Quadrant for Web Access Management for our governance, reporting, and compliance software, providing controlled and role-based access management to back-line resources or federated partner services, based on unique-ID, role, IP-address, group, or per-asset entitlements.

# NEW SUN TECHNOLOGIES RELEVANT TO THE CLOUD

## Virtualization

Sun is one of the few companies with the ability to address all of the different kinds of cloud virtualization: hypervisor (Sun xVM Server), OS (Solaris Containers), network (Crossbow), storage (COMSTAR, ZFS), and applications (GlassFish and Java CAPS technologies).

As a vertically integrated system company with two decades of experience in virtualization technologies — from the Network File System (NFS) that Sun introduced in 1985 to Dynamic System Domains, chip multithreading (CMT), and Solaris Containers — Sun has the experience and expertise to take virtualization a new level.

Our virtualization platform is the Sun xVM portfolio, which provides comprehensive virtualization capabilities, interoperability across heterogeneous environments, and integrated management of both virtual and physical resources.

Sun xVM Server is Sun's datacenter-grade, Xen-based, bare-metal Type-1 hypervisor that uses the enterprise-scale Solaris Operating System as the OS core (as opposed to a restricted Linux core), providing access to OS-level network virtualization/optimization. xVM Server includes both the hypervisor and the relevant management infrastructure to monitor and manage the running of multiple different OS guests, including Windows, Linux, and Solaris guest operating systems, on a single physical server at the same time. It also provides live migration and works well with VMware and Microsoft virtual machines. That makes xVM Server an excellent foundation for larger virtualization solutions, which can then be managed and orchestrated by xVM Ops Center software, Sun's virtualization management product. But unlike other Type-1 hypervisors using a bare Linux core, xVM Server is built within a Solaris OS container providing unique hardware capabilities: multithreaded CPUs, 10GbE links, and quality-of-service control to improve I/O performance. xVM Server is also able to extend the advanced technologies in the Solaris OS, such as ZFS, Predictive Self-Healing, DTrace, advanced networking, and security to Windows and Linux guests (in addition to any Solaris guest instances). Additionally, unlike other virtualization platforms, Sun xVM Server draws on open-source and community involvement through the OpenSolaris and OpenxVM communities to provide an open and interoperable offering.

Sun xVM Server, coupled with Sun's OpenSolaris project, provides the most innovative and advanced building blocks for cloud infrastructure:

- Network virtualization with Crossbow
- Storage virtualization based on COMSTAR and ZFS
- OS virtualization based on Solaris Containers
- Virtualization based on OpenxVM
- Device and location independence, enabling users to access systems regardless of their physical location or type of access device (PC, PDA, mobile phone, and more)
- Desktop virtualization via Sun xVM Virtual Desktop Infrastructure (VDI)

## Modular Systems

Large-scale datacenters are using increasingly modular approaches to provision and manage pools of standard servers, storage systems, and network resources.

Points of delivery (PODs), for example, provide environments that are optimized for specific workloads, such as HTTP or HPC, or specific capacities, such as numbers of users or transactions. They encapsulate storage, networking, management, and servers.

The POD hardware platform layer consists of compute hardware, networking, and storage. The availability and scalability requirements and the service tier that the hardware is intended to support often drive the specifications of the servers. Applications can scale independently. As applications need more resources than are available in a POD, additional PODs can be added, providing more capacity. Both horizontal and vertical scaling can be used as appropriate for each application.

One example of a POD is the Sun Customer Ready HPC Cluster, a platform that allows IT organizations to deploy a standard set of preintegrated servers, switches, and storage devices with rack-at-a-time granularity. These HPC clusters can be built out of standard rackmount servers, such as the Sun Fire™ X4150 server, or Sun Constellation Systems built with Sun Blade™ X6000 series blade modules. The Sun Constellation C48 racks (four Sun Blade X6000 systems) offer 7 TFLOPS from 768 cores, but with 17% improved power efficiency. Systems such as this, while providing unprecedented power efficiency, are typical of extreme power density associated with cloud computing datacenters. Thus, most cloud datacenters reject traditional underfloor cooling, opting for more efficient overhead services and hot isle/cold isle layouts.

Another well-known example of POD design is the Sun Modular Datacenter S20, a comprehensive datacenter delivered in a shipping container. The 20-foot enhanced container can be loaded into almost any transportation system and delivered to a customer's site, ready to be installed by Sun or a Sun partner. Inside is an integrated power, cooling, and rack system that can be populated with any 19-inch rackable, front-to-back-cooled equipment that fits the customer's specific computing needs.

The Sun Modular Datacenter has proven to be ten times faster to deploy than a conventional datacenter. Plus, it reduces capital expenses with incremental expansion capabilities, and it provides four times higher density per rack compared with a typical datacenter — with 40% lower cooling costs in one-eighth the space.

## Open Storage

Open Storage enables cloud computing to be done at a lower cost and at larger scale than traditional, proprietary storage. Open Storage is about using industry-standard components, including x64/x86 servers as storage controllers and flash memory, to accelerate low-cost, high-capacity disk drives, with enterprise-class open-source software to deliver inexpensive, high-capacity, highly scalable architectures.

Open Storage also enables new architectural models for data management. With the open-source storage stack running on industry-standard hardware, including x64 and SPARC systems, we're able to get the data closer to the processors. This simplifies data-intensive computing by eliminating the need to move data around a network. Devices or servers that fit this model include Sun Fire X4540 servers, which are available in 12 to 48-TB configurations, all in a 4RU platform with four-core x64 processors. Shared JBODs benefit from the Open Storage stack and enable highly available storage architectures. Open Storage enables enterprise-class architectures to be built out of industry-standard components.

Customers, developers, and consumers can download this stack and build their own storage appliances. But those customers who would rather buy a fully integrated storage appliance can choose the Sun Storage 7000 family of unified storage systems. Following the Open Storage model, these systems are built out of industry-standard servers, leverage the capacity advantages of industry-standard SATA II drives, and integrate flash-based SSDs in a hybrid storage model, all with a simple-to-use and elegant user interface. By leveraging general-purpose hardware and software, a new breed of system becomes possible. For example, the Sun Storage 7000 line has the ability to observe what's going on inside storage devices at a level that has previously not been possible.

Open Storage provides a new and unique business model as well. Capabilities such as snapshot, replication, and compression are all included and there are no additional costs for the data services. This open source stack also includes protocols such as NFS, CIFS, iSCSI, and FC.

Open Storage architectures benefit from innovation in the information technology industry. Being built on industry standard components enables quicker adaptation of new processors and new interconnects (such as 1GigE and 10GigE), as well as incorporating new technologies like flash-based SSDs.

Sun's breakthrough Sun Fire X4540 series data servers are redefining storage density. By integrating state-of-the-art server and storage technologies, the Sun Fire X4500 server delivers the performance of a four-way x64 server and up to 48 TB in 4U of rack space. This system also delivers incredibly high data throughput (about three times that of competitive systems) for about half the cost of traditional solutions.

> *"[The Sun Fire X4500 server ] is the Web 2.0 server. . . . I really think it is the category of the future. Now companies can get hardware like this and build next-generation applications."*
>
> —Tim O'Reilly, CEO, O'Reilly Media

Moreover, Sun's storage servers are the vanguard of refactoring storage into general-purpose server appliances. They combine a server with disk, networking capabilities, and native metadata and query capabilities. Specialized software enables these general-purpose systems to provide high-performance data services, making possible compute-on-storage strategies for avoiding the high-latency movement of extreme-scale data for data-intensive clouds.

## WHAT YOU CAN DO

As you can see, cloud computing changes everything. It abstracts the software application platform from the underlying hardware infrastructure, freeing developers and users from becoming locked in to specific hardware. In cloud computing, the user's data and software execution are in the cloud (a.k.a. the Internet).

With a singular vision—The Network Is The Computer™—and the research, product portfolio, and communities that this vision has created, Sun is uniquely positioned to help enterprises build and use cloud computing deployments.

This is a vision that everyone can participate in. So here's how you can help continue the development of this architecture and take advantage of cloud computing:

- Evaluate your business and technology requirements – Sun can help, by conducting a Cloud Computing Workshop or Datacenter Assessment sun.com/service/assess. A few key questions to help you get started:
    - What are the different layers at which you might leverage cloud services — Infrastructure as a Service (IaaS), Platform as a Service (PaaS), Software as a Service (SaaS)?
    - What are the business models under which you would operate and use the cloud — public, private, hybrid?
    - What are the different types of applications you want to put into the cloud — Web, HPC and analytics, regulated applications?

- Join the Sun Cloud API community at http://kenai.com/projects/suncloudapis to join the discussion and influence the direction of Sun's open APIs for the cloud.

- Sign up for the Sun Cloud public beta sun.com/cloud and start building and testing your applications and services in the cloud.

Lit. #GNHT14877-0 03/09