CLOUD COMPUTING AND OPEN DATA CENTERS

Contributors

Kai X. Miao
Cloud Infrastructure Group,
Intel Corporation

Jackson He
Software and Services Group,
Intel Corporation

Cloud computing, as an emerging paradigm in the IT industry, promises significant gains in efficiency and flexibility at a time when demands on data centers are growing exponentially. As the tools, building blocks, solutions, and best practices for cloud computing are evolving, challenges exist in the industry to develop sound cloud solutions to meet the requirements of different end users. The technology and industry leadership that Intel brings to this environment is broader and deeper than most realize. Intel's Open Cloud Vision is that cloud computing is federated, automated, and client-aware. Moving the industry toward that promise will require a focus on three industry-wide pillars of cloud computing—efficiency, simplification, and security—and on developing solutions that are open, multi-vendor, client-aware, and interoperable. The nature of diverse applications and services requires diverse cloud architecture and instances. This article also addresses the challenges of building towards a distributed cloud-based environment on Open Cloud Vision.

Introduction

Cloud computing is an evolution in which IT consumption and delivery are made available in a self–service fashion via the Internet or internal network, with a flexible pay-as-you-go business model, which requires a highly efficient and scalable architecture. In a cloud computing architecture, services and data reside in shared, dynamically scalable resource pools, often virtualized. Those services and data are accessible by any authenticated device over the Internet or an internal network. The key attributes that distinguish cloud computing from conventional computing are:

- Compute and storage functions are abstracted and offered as services
- Services are built on a massively scalable infrastructure
- Services are delivered on demand through dynamic, flexibly configurable resources
- Services are easily purchased and billed by consumption
- Resources are shared among multiple users (multi-tenancy)
- Services are accessible over the Internet or internal network by any device

Rather than a revolution, cloud computing is an important transition, a paradigm shift in IT delivery—one that has broad impact and significant challenges to consider. Cloud computing offers the potential for a transformation in the design, development, and deployment of next-generation technologies—technologies that enable flexible, pay-as-you-go business models that will alter the future of computing from mobile platforms and devices to the data center.

The impetus behind cloud computing is the ever-increasing demands placed on data centers that are near capacity and resource-constrained. These demands include growing needs to manage business growth and increase IT flexibility. In response to these challenges, cloud computing is evolving in the forms of both public clouds (deployed by Internet companies, telecommunications companies, hosting service providers, and others) and private or enterprise clouds (deployed by enterprises behind a firewall for an organization's internal use).

Public clouds are being driven by explosive growth of Internet data and traffic as the Internet matures and Internet-based services proliferate. By 2015, over 3 billion people with more than 15 billion devices will access the Internet[1]—over twice today's demand. The monumental requirements associated with the data center build-outs needed to satisfy this growing demand can only be met with the increased efficiency, performance, and flexibility of cloud architectures.

Private clouds are being driven by the expanding business demands on enterprise IT. More and more data centers find themselves facing real limits, whether based on lack of power, lack of room, lack of server capacity, or lack of network bandwidth. Expanding traditional infrastructures to meet these challenges quickly uncovers multiple inherent inflexibilities.

Cloud computing is far beyond data center virtualization. Initially, virtualization technologies allowed data centers to consolidate server infrastructure to save cost. Next, flexible resource management technologies added the ability to more dynamically allocate data center resources. This further reduced costs and also increased data center flexibility and performance, ushering in a new era of technology development and deployment. Software vendors have begun to design robust management features and technology optimizations for enterprise and public clouds based upon virtualization. Hardware vendors have extended their management tools and reliability features to include increased flexibility. The era of cloud computing can be seen as the next natural step, where significant automation and scalability become possible. Cloud computing offers a path to optimized use and rapid deployment of resources, improved operational efficiency, and potential for significant cost savings. When fully realized, cloud computing infrastructures can provide competitively significant IT agility, flexibility, and adaptability through systems that are efficient, simplified, and secure.

Today's workers are increasingly likely to utilize multiple devices, including smart phones, tablets, and PCs to access information. They embrace new applications and devices in their personal life, and expect those same capabilities to be available at work. Yet today, most of these devices function on a standalone basis, requiring users to juggle multiple independent devices. In addition, when it comes to the ability to access, display, manipulate, or secure data, clearly some devices are more capable than others. Yet today, most Internet services are "dumbed down," perhaps being able to recognize screen size or display, but with limited ability to take advantage of enhanced security or performance on more capable devices.

Though cloud computing can be viewed as an evolutionary step, it is a fundamental shift and there are challenges to consider:

- Maintaining the stability of mission-critical applications as you transition into cloud environments is paramount.
- Intellectual property protection, data security, and privacy all require additional attention and new tools if shared resources in a public cloud are to be used.
- The automation and flexibility of resource pools will be imperfect while cloud computing tools evolve.
- Solutions must be selected that provide for flexibility and interoperability.
- Cloud-based applications must enable (rather than negatively impact) user productivity, regardless of the device used or connectivity.

Intel's Open Cloud Vision

Cloud computing technology is evolving at a fast pace and many cloud services and vendors are entering the market to enable the development of private clouds for enterprise IT. Several public cloud providers are expanding their services to support enterprises and small and midsize businesses. In Intel's numerous conversations with vendors, analysts, and customers, we've identified key themes that emerge as critical to what customers want from cloud computing infrastructures and solutions. Intel's vision for cloud computing over the next five years centers on three themes that are essential to help overcome key challenges and realize the full potential and value of cloud computing, *Federated*, *Automated*, and *Client-aware*, as shown in Figure 1.[2]

Figure 1: Intel's Open Cloud Vision
(Source: Intel Corporation, 2011[2])

Federated means communications, data, and services can move easily within and across cloud computing infrastructures. To accomplish truly federated systems, smooth interoperability across many platforms and solutions must be a reality. Today, the industry is just reaching the point that enterprises can move or migrate workloads within and between their own data centers. Data center operators are far from being able to have data and services seamlessly and securely scale beyond their borders to span public and private clouds when desired. Intel's cloud vision calls for a level of federation that enables the movement of workloads and data from one service provider to another; burst implementations between internal private cloud and public cloud providers if additional capacity is needed; and secure and reliable data flow across vendors, partners, and clients. The first element mentioned above, the one related to the "federated" clouds, will probably be the most difficult to develop, but this task will represent the primary area of interest of the Open Data Center Alliance[3], who'll ultimately draw up the necessary standards for creating public and/or private clouds that can be easily interconnected without affecting their level of security and/or functionality. Plus, getting the various entities out there to share precious computing resources will also be a difficult task, especially if we're talking about direct competitors.

Automated means that cloud computing services and resources can be specified, located, and securely provisioned with very little or zero human interaction. Today, the industry faces many gaps in automation. According to IDC's Data Center Survey in 2009, virtualization thus far has failed to reduce complexity. The number of server instances that can be managed by the average systems administrator has increased from 37 to only 41 comparing nonvirtualized servers to virtualized servers. Moreover, virtual machines are generally statically provisioned rather than automatically responding to user needs. Data center management remains very manual today—patching of servers doesn't scale reliably. Intel's cloud computing vision calls for automation that dynamically allocates resources to agreed-upon service levels and optimizes the data center for maximum resource utilization and power efficiency. This includes automation of provisioning, resource monitoring, reporting of consumption for bill-back, and workload balancing.

Client-aware means that cloud-based applications are able to dynamically sense and take advantage of the capabilities of the end point device to optimize application delivery in a secure fashion while enhancing the experience of the end user. Today, there are certain frameworks that allow for some level of data center intelligence and scaling to support the client being served, but they are neither consistently applied nor ubiquitous. Many of today's Internet services default to the lowest common denominator device even if the user is accessing the service with a more capable device such as a PC. Conversely, other services are difficult to use on a handheld device because they were written for a PC. Intel's cloud computing vision calls for the data center and service provider to enable secure access and optimal experience across a range of devices, by enabling the cloud to sense and dynamically adjust to take advantage of the attributes and capabilities of the client device. These attributes include items

such as a device's remaining battery life, policies, and connectivity. At the same time, client device capabilities can affect the overall performance of cloud solutions: taking advantage of local performance on the client device can enable a better end-user experience, and security capabilities on the client device can ensure security policies are applied at the device.

Pillars to Enable Intel's Open Cloud Vision

Building open, interoperable solutions that embrace standards evolving the infrastructure to realize the full potential of cloud computing will not be easy. It will require cooperative development and specific focus by many providers and customers across the IT landscape. We believe that to move towards this vision of cloud computing, individual organizations and the IT industry as a whole need to focus on three key areas:

- *Efficiency:* While the need for computing across different industries increases exponentially, resources are limited. These resources include space, power, cooling capacity, qualified IT professionals, money for infrastructure, and money for operations. To meet the increasing demand of computing with existing or available resources will require better efficiency both compute infrastructure and processes.
- *Simplification:* Generally, the growth of a system inherently increases complexity, and this is certainly true of IT infrastructures. Multiple architectures complicate management. Increased server utilization raises network bandwidth requirements. And systems from different vendors typically present integration complications. For cloud computing environments to deliver on their promise, simplification must underlie cloud architectures and practices.
- *Security:* Both business risk and compliance requirements make data security paramount. In an environment with abundant traditional security issues, cloud computing creates new challenges because it moves data in new ways, often outside of traditional physical boundaries. The successful implementation of cloud computing requires new security models to meet new challenges.

In addition to the three pillars described, to achieve the vision, delivery of open, interoperable solutions that embrace industry standards is essential. When multiple providers (of solutions, hardware, software, integration, or processes) act independently, poor interoperability and lack of flexibility are the natural results, which are in direct contradiction to the main promises of cloud computing.The evolution of cloud computing requires open, interoperable solutions that embrace standards.

Open Data Center Alliance

Intel is driving an Open Data Center initiative. The Intel® Open Data Center Initiative is Intel's comprehensive engagement with ecosystem partners and end customers to help speed the delivery of technology that enables more

secure, efficient, and simplified cloud data centers that preserve IT flexibility and choice. Intel is working directly with leaders in global IT for enterprise and service providers and is an advisor to the Open Data Center Alliance. The Alliance will define a roadmap of the highest priority usage models for cloud and next-generation data centers and lay out the requirements to address with multivendor, interoperable solutions that embrace standards. Intel responds to these usage models and others that we identify through our end-user engagements to deliver products and technologies that meet the requirements of the usage models. We then engage and rally leading systems and solution providers to deliver products and solutions and enable deployment via reference architectures and best practices through Intel® Cloud Builders[4]. Intel Cloud Builders brings together leading systems and software solutions vendors to provide best practices and practical guidance on how to deploy, maintain, and optimize a cloud infrastructure based on Intel architecture. In short, Intel Cloud Builders provides the industry a central point for cloud innovation based on the IT requirements defined by the Open Data Center Alliance and other IT end users. Intel Cloud Builders publishes detailed reference architectures, success stories, and best practices that you can use right now to deploy and enhance your cloud. Using this guidance and interaction with cloud leaders, IT managers can begin utilizing proven solutions to improve cloud security and efficiency while simplifying data center management and operations.

The Open Data Center Alliance is an independent organization of leading global IT managers who have come together to amplify their collective voice to set data center requirements for today and the future that are best-of-breed and enable flexibility and choice. Their mission focuses on delivering next-generation data center and cloud requirements to meet the challenges facing IT today and tomorrow and delivering them in an open, industry-standard, and multivendor fashion. Intel serves as the technical advisor to this organization. Open Data Center Alliance is a coalition of more than 300 leading businesses that together represent more than 100 billion US dollars (USD) in annual IT investment and that have cloud research or projects underway. Alliance Steering Committee members include BMW, China Life, Deutsche Bank, J.P. Morgan Chase, Lockheed Martin, Marriott International, Inc., National Australia Bank, Shell, Terremark, and UBS. The alliance will lay out future hardware and software requirements that lead to more open and interoperable cloud and data center solutions. Intel plays a unique advisory role within the alliance, whose initial membership was purposely focused on end user companies rather than technology providers.

Intel Cloud Builders

The goal of the Intel Cloud Builders program is to provide a path to the Cloud 2015 vision. Intel announced a significant expansion of this program that brings together leading system and software solution partners to provide proven cloud building recipes and practical guidance on how to deploy, maintain and optimize a cloud infrastructure.

While the alliance will determine future requirements for cloud infrastructure, the Intel Cloud Builders program will help bring these requirements to life with full solutions. The program now has a total of 90 Cloud reference architectures and solutions with more on the horizon. It represents a community of the most critical providers of technology in the cloud, including Canonical, Cisco, Citrix, Dell, EMC, Huawei, Enomaly, Eucalyptus Systems, Neusoft, Gproxy, HP, IBM, Joyent, Microsoft, NetApp, NetSuite, Novell, Parallels, Red Hat, and VMware.

Cloud 2.0: Balanced Cloud Architecture for Scalable Cloud Services

It is a myth that many people believe that cloud computing means all the complicated computing tasks will be centralized on the server side in the cloud, while on the device side, all it takes is simple terminals with a browser—this is a typical thin client story that has been around for the last 15 years. It also reminds people the mainframe days that all the computing was done by the "big iron"; users were connected via IBM T3270 terminals over a dedicated line. As cloud data centers become bigger and bigger, it further enhanced the concept that cloud computing equals big data center build-out and concentrate compute tasks to data centers and thin client will be the future of client compute model of choice in the era of cloud computing. Is that the true meaning of cloud computing? Will cloud computing lead us to thin client model? The answer is clearly "No". Cloud services are not going back to the mainframe days. Users prefer diverse and compelling services that require different types of service content, delivery methods, and customized service experience.

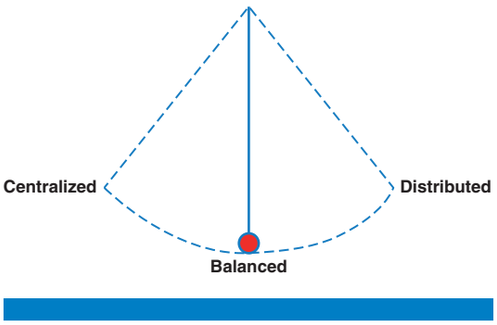

Figure 2: Computing models: centralized versus distributed

To further understand this debate, let's go back to history of the philosophical discussions of thin versus thick client and centralize versus distributed compute models, shown in Figure 2. The debate around different compute models could go back as early as the beginning of computer science. In the early days, mainframes were very expensive and few. There was not intelligence on the terminal side and all the terminal could do was simple text string display and transmission. That was represented by such terminals as the IBM T3270. This was the extreme case of the centralized compute model where all the compute tasks happened on the mainframe side, while the terminals were really dumb. With the emergence of the PC in the 1980s, more and more compute power was built up on the PC side, but in the early days of PC, they were not connected. It was the extreme case of distributed computing, where all applications were distributed on PC and they were running standalone. With the advance of network technology, especially Ethernet and TCP/IP-based Internet, PCs were connected with each other and connected to servers and mainframes. This was the start of client-server model, where data was stored at server side for data integrity and management. In addition, with the minicomputer and X-86 server development, they replaced the mainframe. Servers are no longer mainframes at a few central locations,

but rather distributed at all levels—at central offices, branch offices, even at small businesses. At the height of client/server computing, most servers were distributed. PCs were the primary forms of client, and most of the personal data processing was happening on the PC side, while the servers were working in coordination with clients to provide services to the many PCs connected to the servers.

The Internet became more popular, especially when services from the server side became available beyond the PC, such as cell phones, tablets, and other connected devices. These new devices provide broader services and further enhanced user experience. As the Apple iPhone* and iPad* become more popular, the call for user-experience–centric computing has become the key. Therefore, we should put centralized versus distributed and thin client versus thick client aside, but rather focus on user experience when we consider the right implementation architecture for cloud—we should have the right balance needed for best user experience.

So what is the user experience we should support in the age of cloud computing? The promise of cloud computing is to provide seamless computing experience, anytime, anywhere, on any devices. By definition, users' needs are different with personal preferences, different context, and different locations. It will not be possible to have a one-size-fits-all device to meet all the needs. It will have to be multiple devices, whether it is a PC, TV, smart phone, tablet, or even embedded devices inside cars, boats, and smart buildings. In addition, there are great varieties of network connectivity; some devices have broadband connections and some have 3G wireless connections, while others may not have consistent connections at all. Cloud computing has to provide services to all types of devices, at different connectivity context, and to different people. There have been over 3 billion connected PCs already. With more and more smart phones, tablets, and embedded devices connected to the Internet, there will be over 15 billion connected devices using cloud services. Cloud implementation architecture has to keep this in mind to provide balanced compute experiences to best serve users in the most cost effective way—this is the essence of what we called Cloud 2.0: the next generation of intelligent cloud services infrastructure built on top of distributed and balanced compute systems architecture for scalable and customized service delivery to the masses (billions of users) with best user experience for given devices and network environment.

Figure 3 shows a high level architecture illustration of Cloud 2.0. The large centralized data centers that drive the growth of Cloud 1.0 will still play a major role. That is where most cloud services are hosted and managed. However, Cloud 1.0 focuses only on compute, storage, and network resources at the data center. It assumes that the wide area network does not have much intelligence or service capabilities. Cloud 1.0 assumes the wide area network's bandwidth is abundant and can scale to customer needs. This is not true, especially with a large number of users using different devices with diverse network conditions. In fact, a number of myths exist in the industry today as

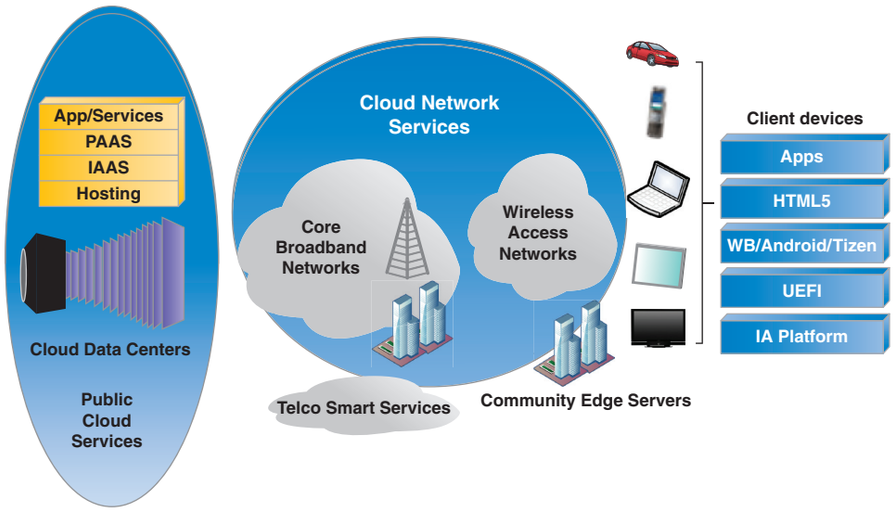

Figure 2: Cloud 2.0 architecture components

shown in Table 1 on the next page. The grand vision of cloud computing will not be realized if network bandwidth is not stable or scalable to millions of users. The additional components of Cloud 2.0 are to enhance the network pipelines with distributed intelligent services that are an extension to the cloud services from the data center. This could be traditional data cache services, it could also be new services routing data intelligently to make the best use of given bandwidth and running applications services in the network or at the edge of the network. In this way, user experience could be greatly improved by having intelligence and cloud services distributed from data center throughout the network.

So, user experience is it! We should design a cloud architecture that will support multiple types of devices. We should also design cloud services that will be user-aware, device-aware, and context-aware.

- User-aware: a cloud service needs to know who accesses the service and whether the person has the right authority. The cloud service needs to verify the user identify before the service is provided.
- Device-aware: a cloud service needs to know what device is consuming the service: what capability the device has, whether it is safe to execute the service on the device, should the service be customized for best experience, and so on.
- Context-aware: a cloud service needs to know what the connectivity condition at the point of service, whether the location is allowed for the service delivery, and whether there is another service provider (cache server) nearby that can provide better services.

Myth	Fact
Cloud computing is a technology.	Cloud computing is a new service model based on several technologies evolved over time, such as virtualization, service orientation, and dynamic resource management.
Cloud computing equals thin client.	Service orientation can be implemented using Web services, but using Web services will not necessarily result in a service orientation solutions.
Cloud computing means all services are from the cloud data center.	No quite true. There will still be services that are distributed along the network infrastructure or even at local storage devices when the network bandwidth is limited or not available.
Cloud computing services are homogenous, one-size-fits-all.	Cloud services have to be customized and diverse to meet the needs of different people, different devices, and different network connectivity conditions.
Cloud computing is one-way service flow from the data center to devices.	Services could be bidirectional and peer-to-peer in a cloud environment.
Endpoint devices in a cloud environment are all interactive with user interfaces to access services.	Endpoint devices do not have to interactive. They could be one-way data gathering nodes without any user interface, such as surveillance cameras.
The world could be covered by a few big clouds from a few big cloud service providers.	The needs for cloud services are diverse, sensitive to culture and usage patterns, some private, some public. There is no way a few big companies can cover these diverse needs. There will be many clouds and they need to work together to provide the best user experience.

Table 1: Cloud Computing Myth

In addition, cloud services have to have the following basic architecture considerations:

- Designed for manageability: cloud services should be built to be managed with sufficient interfaces to expose information for an independent management system to ensure that services are managed in an orderly fashion, despite that the services could be developed by different developers from different places at different times. There should be clear version control and change management for service and application publishing. Service usages should be tracked and logged.

- Designed for scalability: cloud services are meant to scale. They should facilitate any target number of different service consumers defined as required by service consumers. Architecturally, services should take advantage of the latest programming technologies for multithreading and have clear definition of service boundaries (functionalities and interfaces). In addition, management facilities have to be in place to measure the quality and health of services. Automatically extend service capacity to ensure quality of services.
- Designed for federated solutions: cloud services may span typical boundaries, such as network administrative boundaries, organizational and operational boundaries, and the boundaries of time and space. Services will have the ability to cross corporate or transnational boundaries, requiring a high degree of built-in security, trust, and internal identity, so that they can negotiate and establish federated service relationships with other services following given policies administrated by the management system.
- Obviously, cohesiveness across services based on standards in a network of service is essential for service federation.
- In a federated environment, it is possible for technology to run ahead of legal and regulatory boundaries. For instance, although it might be possible to encrypt and package a data set for remote execution in another country to any desired level of tamperproof strength, laws in the originating country may make it illegal to move the data. In this case any legal statutes trump technological options. Flexibility from service orientation still allows outsourcing to countries where these operations are allowed.

Key architecture tenants of balanced cloud architecture:

- Service environment discovery
- Data synchronization and federation within cloud
- End-to-end security
- Energy efficiency
- Manageability from data center to devices
- Service distribution close to the user

Conclusion

All in all, we'd dare to say that Intel has some pretty big plans for the future of cloud computing, but the journey ahead will not an easy one. Intel is fully committed to making Intel's Open Cloud Vision a reality.

References

[1] Cisco: Internet traffic to quadruple by 2015. ***http://news.cnet.com/8301-1023_3-20067979-93.html***

[2] Intel Cloud 2012 Vision: ***http://www.intel.com/content/www/us/en/cloud-computing/cloud-computing-intel-cloud-2015-vision.html***

[3] Open Data Center Alliance: ***www.opendatacenteralliance.org***

[4] Intel Cloud Builders: ***www.intel.com/cloudbuilder***

Author Biographies

Kai X. Miao, PhD, principal engineer, Intel China Ltd., has held several technical and management positions at Intel in the past 14 years. Kai Miao received his MS and PhD in Electrical Engineering from University of Cincinnati and BS degree in Communications from Northern (Beijing) Jiaotong University.

Jackson He, PhD is General Manager of Intel APAC R&D Ltd., He has been with Intel for 17 years and has led many technical evangelizations. In the past a few years, Dr. He has been driving cloud technology innovations in both the US and China.