Cloud Computing: Virtualization and Resiliency for Data Center Computing

Valentina Salapura
IBM T. J. Watson Research Center
Yorktown Heights, NY, USA
salapura@us.ibm.com

***Index Terms* — Cloud computing, data center management, data center optimization, virtualization, Infrastructure as a service (IaaS), Platform as a service (PaaS), Software as a service (SaaS), high availability, disaster recovery, virtual appliance.**

INTRODUCTION

Cloud computing is being rapidly adopted across the IT industry, driven by the need to reduce the total cost of ownership of increasingly more demanding workloads. Within companies, private clouds are offering a more efficient way to manage and use private data centers. In the broader marketplace, public clouds offer the promise of buying computing capabilities based on a utility model. This utility model enables IT consumers to purchase compute resources on demand to fit current business needs and scale expenses associated with computing resources. Thus, cloud computing offers IT to be treated as an ongoing variable operating expense billed by usage rather than requiring capital expenditures that must be planned years in advance. Advantageously, operating expenses can be charged against the revenue generated by these expenses directly. In contrast, capital expenses incurred by the purchase of a system need to be paid at the time of purchase, but can only be depreciated to reduce the taxable income over the lifetime of the system.

THE MAIN ATTRIBUTES OF CLOUD COMPUTING

The main attributes of cloud computing are scalable, shared, on-demand computing resources delivered over the network, and pay-per-use pricing. This offers flexibility in using as few or as many IT resources as needed at any point in time. Thus, users do not need to predict future resources they might need, and to commit to capital investment in hardware. This is especially advantageous for start-ups, and small and medium businesses which might otherwise not be able to afford the IT infrastructure they need to support their growing business. At the same time, redirecting capital investment from IT infrastructure to the core business is attractive even for large and financially strong businesses.

From a technical perspective, cloud computing brings the benefits of virtualization and multi-tenancy to scale-out systems. Virtualization techniques allow multiple system images to share the same hardware resources: CPU virtualization techniques create multiple virtual hardware systems, while network virtualization gives each of these systems its independent network presence. Multi-tenancy allows multiple users to share the same hardware, while providing strong isolation between workloads to ensure data security and privacy.

Cloud computing is usually associated with central management and a highly standardized environment. This allows for simplified provisioning and administration, and can reduce the burden of software installation and upgrades, and maintaining the environment.

For system design, virtualization efficiency and modularity in provisioning are becoming of paramount importance. Metrics such as virtual machines per core, per socket and per Watt are the new metrics that must be optimized for. Where dividing scale-up systems into multiple virtual machines was previously the sweet spot for optimizing virtual machine management, cloud computing must address the creation, operation and migration of virtual machines across system boundaries.

DELIVERY MODELS OF CLOUD COMPUTING

Cloud computing can be offered in three delivery models today, depending on the amount of integration. Infrastructure as a service (IaaS) offers virtual machines and other virtualized resources, such as storage and network access. Platform as a service (PaaS) offers a computing platform which typically includes the operating system, a programming language execution environment, a database, and web server. Software as a service (SaaS) offers application software in the cloud, and cloud users access to the software from cloud clients.

In the future, the cloud computing delivery model for business applications will move into a more integrated virtual appliance delivery model to simplify delivery and deployment of complex software systems. A Virtual Appliance is a self-contained package with pre-installed software and described by metadata which specify the required computing resources as an ensemble of several virtual machines, storage and network configuration. The delivered appliance is pre-configured and contains a customized and tested software stack, from operating systems and middleware to application software. Metadata describe goals and constraints for the virtual appliance, such as performance and availability goals, placement constraints (e.g., for security isolation and for legal

978-1-4673-3052-7/12/$31.00 ©2012 IEEE

data compliance), and configuration variables. The entire software solution is managed (deployed, updated, etc.) as a unit.

As consumer computing is moving into the post-PC era, the importance of personal computers is diminishing, and most computing interaction is moving towards mobile computing. In this space, we increasingly see a separation of user interaction and data storage and processing which occurs transparently to the user. While user interaction and limited processing is performed on mobile devices, persistent storage and processing intensive applications are moving into the cloud.

A BRIEF HISTORY OF VIRTUALIZATION

The first computer system with system virtualization was released exactly 40 years ago in 1972, when IBM released the Virtual Machine Facility/370 (VM/370) for the S/370 mainframes.

For the first time, VM/370 provided multiple users with seemingly separate and independent computing systems, partitioning a single mainframe into several virtual machines. This offered the same advantages cloud computing is offering today: multi-tenancy, centralized management, over the network delivery, and better hardware utilization. To this date, mainframes offer the highest level of centralized, virtual system integration in the enterprise based on System z virtualization.

In the 1990s, POWER systems servers started to offer virtualization with PowerVM hypervisors implementing live partition mobility and active memory sharing. Also in the 1990s, the DAISY and BOA research projects at the IBM T.J. Watson Research Center demonstrated how to fully virtualize instruction sets. Virtualization techniques became widely available for x86 systems in late 1990s based on software solutions to overcome the intrinsic limitations in Intel x86 architecture that had previously prevented virtualization of the Intel x86 architecture by using instruction set virtualization techniques.

HIGH AVAILABILITY AND DISASTER RECOVERY

The standardization, virtualization, modularity and cross-system management capabilities of cloud computing offer a unique opportunity to provide highly resilient and highly available systems. As virtual machines are becoming first class citizens, resilience techniques can build on a well-defined framework for providing recovery measures for replicating unresponsive services, and recovering failed services to respond to disaster scenarios.

Since virtualization allows packaging of workloads — operating system, applications, and data — in a portable virtual machine image container, it enables easy transfer of workloads from one server to another. High availability features can migrate a VM image seamlessly from one physical server to another within the same data center if the original server suffers any failure, performance loss, or to perform scheduled maintenance. This portability and migration of workloads allows building of highly available systems – systems where a customer expects that the most important workloads are continuously available.

Cloud computing is radically changing the way how a disaster recovery (DR) of a data center can be implemented. In traditional disaster recovery solutions, a data center has to have a disaster recovery site. Traditional disaster recovery typically uses data backup and stores data onto storage tapes which are then transported to a recovery data centers, or stored in a vault until they are needed for recovery. For recovery, the failover servers need to be loaded with the operating system images and application software and patched to the last configuration used in production before the data can be restored. This process is lengthy and can take hours and days until the workloads are available again.

Cloud computing and virtualization make recovery of VM images – and their restarting on another server in a different data center for disaster recovery – more cost effective, and enable significantly faster recovery times. The ability to restart VMs from one data center in another data center in a matter of minutes is redefining how disaster recovery can be implemented.

Cloud computing is also changing the economics of disaster resilient computing. Traditional disaster recovery solutions of a data center are costly, and rely on recovery data centers, which can be either dedicated as a recovery data center of a single data center, or can be shared between multiple data centers to lower cost. With cloud computing, virtual workloads from a data center affected by a disaster scenario can be restarted and distributed across multiple unaffected data centers by using their available capacity. This significantly lowers the cost and affordability of a disaster recovery solution.

In addition to high availability techniques at the hypervisor and control system level of cloud computing, a new class of distributed, stateless applications is emerging. These applications are designed from ground up to run in a distributed architecture across multiple sites in a cloud environment, being stateless and resilient to failure of any individual server or site.

These changes at the system, software and applications levels have the potential to move computing from disaster resilient and highly available systems to disaster avoiding, continuously available systems.

CONCLUSION

Virtualization technology is the key enabling technology for cloud computing, which transcends static view of a system as a piece of hardware, but views systems as dynamic entities.

Cloud computing increases flexibility, and opens new ways to think of computing systems by separating computing capabilities and services from the hardware used to provide these functions. Thus, servers are no longer bound to specific racks – instead, systems can migrate from one rack to another, or one data center to another. This flexibility allows systems to be more dynamic, and be optimized to be more efficient for any number of metrics, e.g., energy use, capacity, reliability, availability, serviceability in ways not possible before, by focusing on how systems are used and thereby making IT more consumable.