Getting Virtualization Right

Understanding the Totality of the Move to Virtual Machines and Infrastructure

ABSTRACT

Virtualization has gone way past the point of pilot projects and limited scope implementations. Virtualization is ***driving cloud migration***. The moment has arrived to focus on a thorough, optimal approach to virtualization. Getting virtualization right, however, involves overcoming a number of interlocking challenges related to hardware and software. The choice of server platform is critical in ensuring a successful virtualization outcome on several levels, including memory, processing performance, network/processor connections, server density, energy consumption and cooling factors. This paper will discuss the totality of the move to virtualization across three use cases: business agility, consolidation, and business continuity.

OVERVIEW: THE VIRTUALIZATION AND CLOUD MANDATES

Virtualization and cloud computing have matured well beyond the "emerging technology" designation and the "thinking about it" approach to their implementation. The global cloud computing market is expected to grow to $270 billion by 2020. In government computing, a driver of industry trends, Market Media Research forecasts U.S. government spending on cloud computing will reach $10 billion by 2018[1].

Virtualization and the cloud have effectively become mandates for IT. Though we are still somewhat early in the virtualization and cloud lifecycles, it's quite clear that these two technologies are going to become major, if not dominant models of IT asset deployment. If you manage IT, you are going to be implementing virtualization and cloud computing, if you are not already. Consider the following findings from a 2011 AMD survey of 1000 IT professionals:

- 37% of organizations worldwide are currently operating in the cloud.
- 63% of global cloud customers estimate they store more than $250,000 worth of data in the cloud.
- 92% of cloud customers stated that the infrastructure was an important part of the decision to implement cloud computing.[2]

The virtualization/cloud mandate affects IT managers on multiple levels. The maturing of the technology means you can have confidence that you will find right solution for your particular IT requirements. The momentum of the industry's move to virtualization and the cloud, however, creates both challenges and opportunities for infrastructure planning. The choices you make today about virtualization and cloud computing will affect your IT operations for the foreseeable future. The good news is that although the stakes are high, you have the chance to get it right.

Getting virtualization right means understanding the comprehensive decision making process that leads to an optimal virtualized environment. You may wonder what matters most in the transition to virtualization and the cloud: hardware, software, infrastructure, network, server platform, or processor? The answer is "yes". All of these factors count when you want to get virtualization right, though there is often more than one correct choice for each category. What you will need in your organization will depend on many things, including your IT department's core mission, business strategy, and so forth. What's imperative to understand is that there will be an ideal alignment of hardware, software, server and processor choices that will have an impact on how well virtualization and the cloud work for you.

OBJECTIVE: OPTIMAL VIRTUAL AND CLOUD ENVIRONMENTS

Given your mandate to undertake virtualization and some degree of cloud computing, what does it mean to "get virtualization right"? One answer is that "getting it right" involves arranging to have more than enough VM computing power at the lowest possible cost but also with the highest degree of control and flexibility. So, without breaking the bank, you will want to have enough depth of processing capacity that you won't worry about getting overstretched. You will want virtual and cloud-based computing resources that are at least equivalent of an on-premises set-up. You will want enough flexibility in your virtual/cloud resources to be able to modify your deployment scenarios in the future without causing undue stress or new hardware acquisition. Sound simple, but it isn't.

Much of your success will depend on your server platform. Having the right kind and amount of hardware, network bandwidth and storage capacity should be assumed. Those are "table stakes" for virtualization. If you don't have those, you aren't going to be getting close to an optimal virtualization or cloud environment. If you break it down further, you will see the importance of having deep, connected technological capabilities. Many of these capabilities exist at the processor level or in the choreography between processor and server platform. Virtualization is a software phenomenon, but the hardware plays a critical role in making virtual environments function as envisioned. In that context, getting virtualization right means providing the right mix of the following:

- **Hardware assisted virtualization** – For best results, you should run your VMs with reduced processing overhead and in increased isolation.

- **Transaction processing performance** – Your VMs should provide mind-bending compute power if you need to handle large numbers of users and vast amounts of data. Processor cores, memory, and I/O capabilities should be able to execute huge volumes of instructions at high speed, particularly if your workloads are data- or compute-intensive.

- **Predictable performance** – Cloud computing workloads tend to be different from those handled by traditional data centers. Cloud work is "spiky" in nature, as Figure 1 shows. You should be able to account for both the peaks—typically with more cores—and the valleys, and do so in an efficient manner. Your cloud-based CPUs should offer predictable performance to handle peaks in traffic.
- **Scalability** – On-demand scaling is the essence of cloud computing. Your VMs should be able to scale easily and rapidly, without excessive management overhead required to oversee the process. This can be both a hardware and software issue, but ideally your virtualization hardware and software will be designed to work collaboratively for efficient scaling.
- **Workload management** – Running multiple VMs on a single machine may require highly efficient management of processor workflows. As multiple applications and operating systems run threads through multiple processing cores, you may find that the best results can be attained when your virtualization solution provides dedicated compute resources to each VM, enabling optimal assignment threads to processor resources.

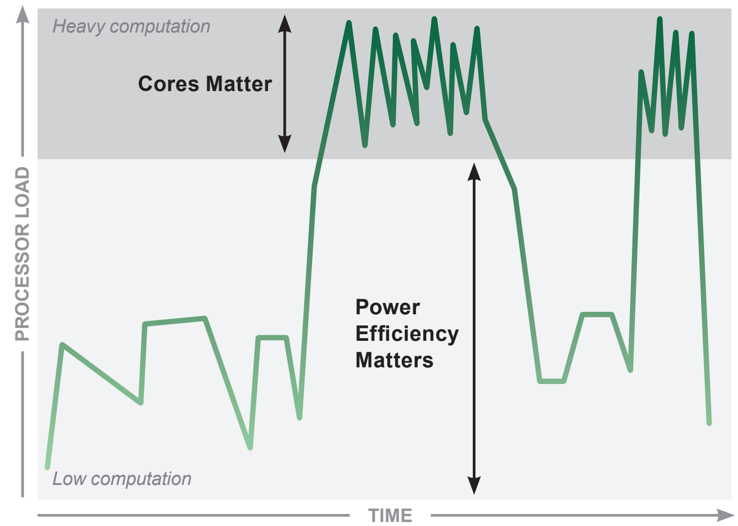

Figure 1 – Cloud server workloads, which tend to "spike" at busy times, thus requiring processors that can handle the load predictably.

- **Memory management** – The way your processor manages memory will have an effect on the VM's ability to process workloads. You will want sufficient high-capacity memory management at the processor level to power high-performance cloud computing.
- **Network/processor connections** – The queuing and assignment of network packets to processor resources can have a big impact on workload throughput and overall computing efficiency. You should have a processor that excels at balancing network traffic and core utilization.
- **Server density** – The more VMs you can put onto a single physical machine, the more attractive you will find the economics of virtualization and the cloud. A high level of server density often translates into reduced floor space and energy consumption, both of which can lead to big savings for the IT facilities and operational budgets. Improving density means having more VMs running per physical machine. However, you have to be able to balance density with performance. Without the right virtualization technology on the server platform, you run the risk of degrading performance by overloading a machine with VMs.
- **Maintainability** – Virtualization and the cloud lose some of their appeal if they are not economical to maintain. Virtual environment maintenance should ideally be simple, efficient and low-cost. The server platform, along with related management systems, should minimize the need for personnel allocated to maintenance.
- **Security** – Risk management often intensifies as IT assets are moved off premise or hosted on virtualized, multi-tenant systems. Your virtual/cloud solutions should have robust security to protect data and processing resources. Hardware assisted security can help keep intruders out, and safeguard sensitive data from unauthorized access.

- **Cooling and energy** – The green movement has created greater emphasis of servers' energy use and cooling needs. As a result, your optimal virtualization and cloud solution may be the one that operates with energy efficiency and attractive cooling characteristics. As Figure 2 shows, 82% of servers operate at less than 50% utilization. Underutilized servers waste energy as they sit idly while consuming power and generating heat. These may seem like subtle issues but they can in fact be drivers of substantial financial savings. And, in some cases, the physical data center facility simply cannot add more electric and cooling capacity. Even if IT assets are cloud-based, good energy and cooling qualities in the server platform can translate into reduced prices for cloud services.

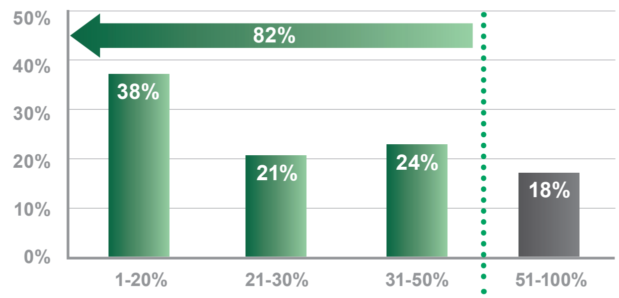

Figure 2 – Average server utilization. (DATA SOURCE– IDC WW Server Workloads 2009—Final Market Model (06.30.09)

OPTIMIZING VIRTUALIZATION AND THE CLOUD WITH AMD

AMD offers industry leading overall solutions for virtualization and cloud computing. AMD's key to success with virtualization is the incredible density of VMs that you can run on a single physical server powered by an AMD Opteron™ 6200 Series processor. Virtualization technologies are built "into the silicon" with the AMD Opteron 6200 Series processors. The CPU architecture allows for dedicated compute resources (cores) per virtual machine. Also, the AMD Opteron 6200 Series processors have unique core density, power management and memory handling functionality. These capabilities can translate into multiple VMs with dedicated resources per server. There's also great flexibility to balance workload requirements and respond to changes in a business.

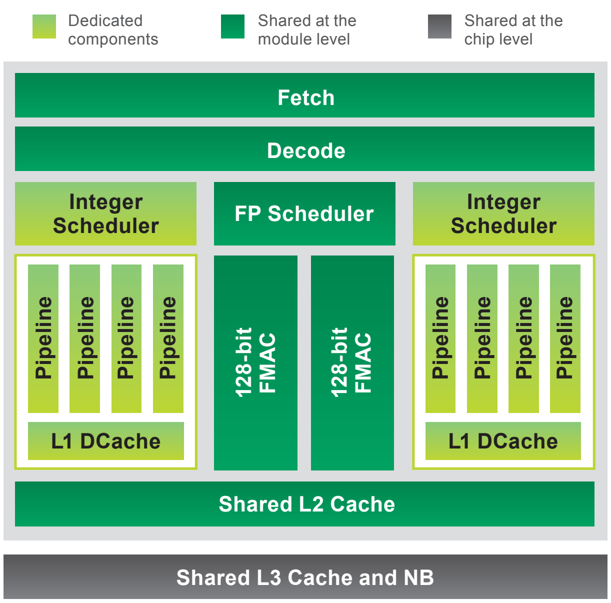

Figure 3 – AMD Opteron 6200 Series processor design

The first 16 core x86 CPU processor chip ever created, the AMD Opteron 6200 Series processor has a revolutionary power design. As depicted in Figure 3, shared resources in the core help enable maximum power efficiency and full power gating and power management at the system level. Shared resources also help increase performance and scalability. In terms of processing, the AMD Opteron 6200 Series processor excels in workload handling with straight-through computing and individual integer cores. Flex FP, a dynamic, flexible floating point complex drives exceptionally high processing throughput in combination with AMD Turbo CORE technology, for up to 1GHz of additional performance, depending on the processor and type of boost[3].

Collaboration with ISVs and OEMs

The AMD chipset cannot effectively break through virtualization outcomes on its own. AMD's breakthroughs in the cloud are based on numerous successful partnerships with other technology leaders, including software vendors, hardware manufacturers and industry organizations. The success of the server platform is based on AMD's processors working in concert with independent software vendor (ISV) and original equipment manufacturer (OEM) partner technologies. The following table summarizes AMD's key ISV relationships.

OEM hardware manufacturers, such as HP and Dell, supplement AMD's virtualization and cloud solution. In most cases, these manufacturers offer server products that are configured for specific virtualization and cloud use

ISV PARTNERS	COMPILERS/PROCESSORS	OS/HYPERVISORS	MIDDLEWARE/APPS
▪ Microsoft ▪ Red Hat ▪ VMWare ▪ Citrix ▪ Apache Foundation ▪ MySQL	▪ Microsoft for Visual Studio and .NET ▪ The Portland Group (PGI) ▪ GCC ▪ Java	▪ Windows Server ▪ Windows Server Hyper V ▪ RedHat Linux ▪ Novell SuSE Linux ▪ VMWare	▪ Microsoft SQL Server ▪ Microsoft IIS ▪ Hadoop ▪ Memcached ▪ MySQL ▪ Apache Web Server

cases. For example, the HP ProLiant SL165s G7 server is designed for scale-out and power efficiency. It features maximum expansion with AMD Opteron™ processors, 24 DIMM slots, and 6 hard disk drives (HDDs). The HP ProLiant SL335s G7, in contrast, has maximum density and efficiency with AMD Opteron™ processors and single node serviceability. It's suitable for virtual and cloud scenarios where performance and cost per watt need to be at the lowest possible levels, such as in a Web front-end.

Dell's PowerEdge C6105 is designed for cloud front-end use, where again best performance per watt and performance per dollar is a necessity. In this case, Dell has designed a machine that can be deployed with 4 2P servers in only 2Us of rack space, but with 24 total hard drives. The net effect of the PowerEdge C6105 is shared infrastructure that uses less floor space, power and cooling. In addition, the server is designed to enable servicing of individual nodes to increase uptime.

The Dell PowerEdge C6145 is intended for heavy virtualized hosting workloads and "big data" in the cloud. The PowerEdge C6145, with AMD Opteron processors, can host up to 2016 virtual servers per 42U rack. Each virtual server can contain up to 10.6GB of memory and up to 48TB storage. When combined with an Intelligent Platform Management Interface (IPMI), version 2.0 and shared power and cooling infrastructure, the machine can deliver massive transaction processing at low cost.

The nuanced choreography between AMD and its ISV and OEM partners results in virtualized environments that optimize performance, power and cost in a variety of task-specific scenarios. Industry adoption has been strong, with nearly 2 million AMD processors engaged in cloud computing clusters worldwide.[4] Several high-profile cloud initiatives are built off of the AMD stack. These include:

- **Microsoft Windows Azur**e – Microsoft's cloud services platform uses servers with AMD Opteron processor technology.
- **Facebook Open Compute** – Custom-engineered technology in Facebook's data center is focused on delivering increased energy efficiency at a lower cost.
- **Rackspace OpenStack Project** – Open source technologies deliver a massively scalable cloud operating environment.
- **Xen Open Cloud Project** – Open source enterprise-ready server virtualization and cloud computing platform, delivering the Xen Hypervisor with support for a range of guest operating systems.

Performance in the Silicon

Virtual and cloud environments running on AMD-powered server platforms can deliver the performance characteristics you need to achieve an optimal virtual and cloud environment, to "get virtualization right". With maximum core density and industry-leading compute power per processor, AMD virtualization delivers high levels of throughput, performance and scalability. The AMD-powered server platform can fulfill the established objectives for optimal virtualization in the following ways:

- **Hardware assisted virtualization** – AMD processors are designed with features to provide consistent performance across all types of demanding workloads even when system utilization is at its peak.

- **Transaction processing performance** – With AMD processors, you can tackle highly concurrent single and multi-threaded cloud workloads and deliver rapid response times with robust compute capacity. You can boost performance and speed up response times and throughput for applications that need it most. AMD Turbo CORE helps improve performance by turning unused power headroom into compute power. Turbo Core can automatically increase clock speed by up to 500MHz[5] which can be useful for such virtualization/cloud workloads as financial, database, and business intelligence applications. If your work involves highly technical and specialized cloud applications, such as those for scientific, financial, and engineering, AMD gives you significant performance improvements. The Flex-FP FMAC units can improve throughput and increase computational power, bringing 256-bit floating point processing to mainstream computing. FMA4 instructions execute in half cycles of competing processors.[6]

Performance/Watt Improvement per Processor Generation

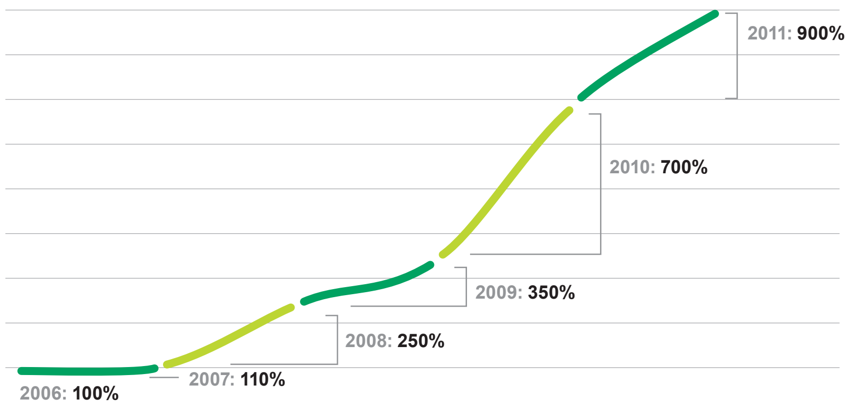

Figure 4 – AMD processors have demonstrated gains in performance/watt steadily, with the latest generation reaching a 900% improvement over the last 5 years.

- **Predictable performance** – AMD helps you achieve consistent, predictable performance and improved throughput with dedicated cores for each virtual machine. Depending on your chosen processor, you may be able to run more than 16 VMs per CPU, which can be the backbone of a super high-density cloud environment with superior application responsiveness. AMD's Direct Connect Architecture with HyperTransport™ 3.0 Technology helps increase interconnect speeds to handle fluctuating workload demands.
- **Scalability** – Compared to earlier generations of Opteron processor, the 6200 Series based server platform has 33%[7] more cores and 50%[8] more memory capacity for handling increasing user and data loads. This enables you to scale on demand to handle shifting, expanding workloads, as found in high volume cloud datacenters and complex, unstructured data types.
- **Workload management** – AMD-V™ and AMD-Vi technologies reduce the overhead of virtualization software and improve performance for I/O intensive workloads for near-native application performance.
- **Memory management** – AMD Opteron™ processors support up to 12 DIMMs and 384 GB memory per CPU. This gives you more resources for memory-intensive workloads and opens up more throughput and scaling possibilities. Quad-channel memory doubles the memory channels that were available with earlier six-core AMD Opteron™ processors. Support for LR-DIMM provides more capacity per DIMM and increased bandwidth. AMD Opteron processors also can support DDR3-1600, DDR3-1866 for improved overall systems performance.
- **Network/processor connections** – HyperTransport Technology Assist helps increase coherent memory bandwidth by reducing cache probe traffic between cores. Less probe traffic can translate into better performance for cache-sensitive applications such as database, virtualization, and compute-intensive applications. AMD-based cloud server platforms can provide tremendous I/O connectivity to the latest 10 gigabit Ethernet (10 GbE), Infiband, FCoE, and SAS interfaces, resulting in improved I/O performance and datacenter simplification.

- **Server density** – The latest AMD Opteron™ Series processors have the world's first 16-core x86 processor, up to 160% more cores than competing products.[9] This means you could get more compute capacity in the same physical footprint while minimizing contention for compute resources. Having a higher core density often means you will have more flexibility and the ability to run more VMs, thereby making your systems more efficient and achieving better server utilization and higher ROI per rack.
- **Maintainability** – AMD Opteron™ processors are designed to work with system management solutions, such as Microsoft System Center, for economical maintenance and "lights out" remote management.
- **Security** – AMD's AES–NI instructions can provide a performance boost and improved security and data protection for high volume encryption and decryption activities.
- **Cooling and energy** – Traditionally, you have faced a tradeoff between having more cores to handle spikes in load and power efficiency. That is no longer the case. Under normal operating conditions many of the AMD-P technologies help keep processor power consumption down when not all of the processor logic is in use by a given workload. Real world workloads do not exercise all of the processor logic at once, which typically keeps AMD processors from exceeding their thermal design guidance. With AMD Opteron™ processors, however, you can have the capacity you need along with a major improvement in performance per Watt and overall power usage. Many AMD processors have cooling and energy efficiency built in through multiple unique capabilities:
  - **AMD PowerNow!™ Technology with Independent Dynamic Core Technology** – Allows processors and cores to operate dynamically at lower power and frequencies, depending on usage and workload.
  - **Low Power U/RDDR3 memory** – Supports DDR3 1.5v and low power DDR3L 1.35v memory technologies.
  - **Dual Dynamic Power Management** – Enables more granular power management capabilities to help reduce processor energy consumption, operating with separate power planes for cores and memory controller.
  - **C1E** – Can reduce memory controller and Hypertransport technology links' power.
  - **Advanced Processor Management Link** – Allows advanced power control and thermal policies.
  - **AMD CoolSpeed Technology** – Highly accurate thermal information and thermal protection.
  - **AMD CoolCore™ Technology** – Can reduce processor power consumption by dynamically turning off sections of the processor when inactive.
  - **AMD Smart Fetch Technology** – Can reduce power consumption by allowing idle cores to enter a "halt" state.
  - **TDP PowerCap Manager** – Set a fixed limit on a server's processor power consumption.

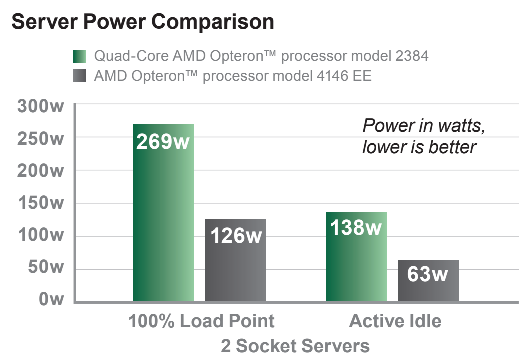

Figure 5 – AMD-based server power consumption continues to drop.

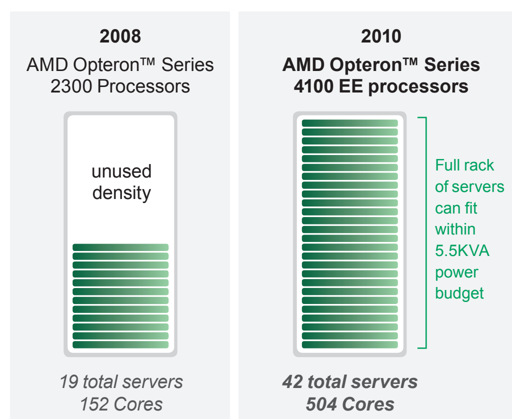

Figure 6 – AMD-based servers enable tremendous compute density with low power needs.

USE CASE: BUSINESS AGILITY

Agility is a frequently-used term that can mean different things to different people. Business agility, however, can be thought of as a company's ability to respond to changes in business conditions quickly enough to benefit from them. In practical terms, business agility is often referring to the ability to change strategy and introduce new products on a cycle that keeps you ahead of the market.

From the IT perspective, business agility is mostly about being able to make things happen rapidly—rapidly enough to be a partner for strategic strength rather than an obstacle to change. IT managers generally want to be able to execute projects for business clients quickly and efficiently, but circumstances often conspire against them. Many factors inhibit agility, including slow requirements gathering, prolonged and political project review cycles, and IT staff allocation. These issues are largely outside of the IT department's control. However, one of the biggest culprits, which IT can potentially manage, is the time required to acquire and "stand up" new hardware for an IT project.

Consider what happens at a business that experiences seasonal growth. Some retailers, for example, find that they have to add computing capacity every year at the holiday time. In this example, assume that the retailer has an integration between its customer relationship management (CRM) system, which is used to process incoming orders and its enterprise resource planning (ERP) and logistics solution. The two systems are connected so that new orders can quickly be matched with supply chain management to ensure that merchandise is available and shipped on time to customers.

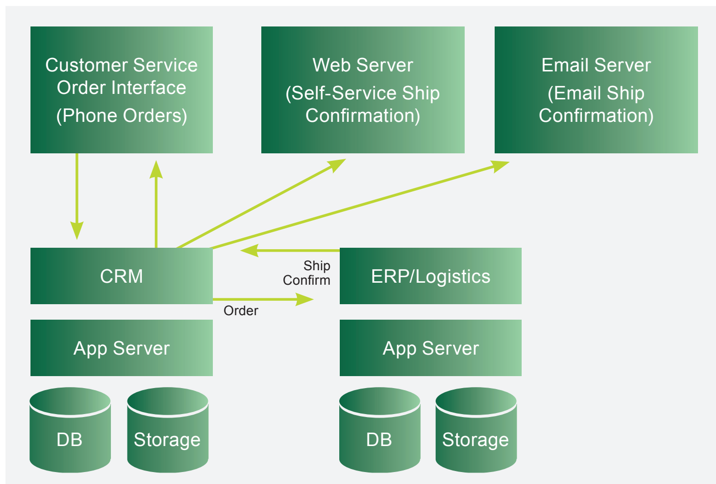

Figure 7 – Reference architecture for retailer that combines CRM and ERP for order processing.

Figure 7 shows how the two systems connect with one another, as well as with a customer service order interface, a Web self-service interface and an email server. The CRM and ERP systems have their respective application servers, databases and storage. It's a common and quite adequate way to manage the retail operation. Yet, scaling it is troublesome.

Adding capacity to the combined CRM-ERP solution shown in Figure 7 could take months, of which a large chunk of that time would be needed to acquire and test new hardware and then install and configure the software. Figure 8 depicts the IT project Gantt chart view of this process. Each phase of the project is dependent on the previous one. For instance, you can't stand up the hardware until you've bought it. You can't install the software until you've tested the hardware, and so forth. The drawn out capacity adding process hinders agility. The IT department loses productivity as it waits out the "dead time" inherent in server deployment.

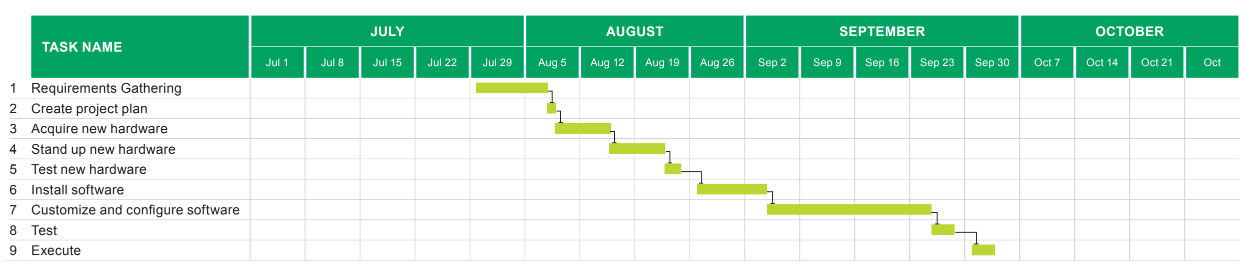

Figure 8 – Gantt chart showing the calendar impact of acquiring and standing up hardware, deploying software, and conducting a migration of software into a new, expanded data center environment.

Migrating the CRM-ERP solution to a virtualized or cloud-based environment can enable greater agility by speeding up the process of getting server capacity online. As shown in Figure 9, with VMs running the CRM, ERP, Web and email servers, as well as the databases and storage, the retailer can spin up machines as they are needed without having to buy and stand up new equipment. An AMD-powered server platform enables this capability on several levels. On the front end Web tier, the server platform can combine low power and VM density to provide servers that cost effectively handle requests and content caching. At the application tier, where CRM and ERP are functioning, the server platform can leverage low power and compute strength to deliver scalable systems with enough core density to handle high volume computational actions. At the backend, with the databases, the server platform can be structured to be efficient with memory caching and serving data to the app layer.

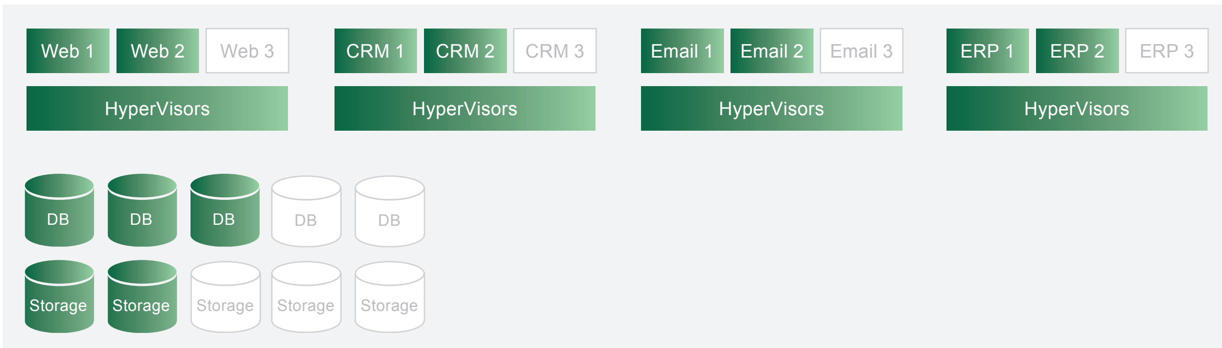

Figure 9 – The retailer can be more agile with a virtualized environment that enables rapid spinning up and spinning down of servers and other IT assets on demand.

Thus, the upgrading process can move faster with virtualization. As the revised Gantt chart in Figure 10 reveals, the overall capacity expansion cycle could take a little over 4 weeks, instead of several months, and the time elapsed between creating the project plan and customizing the software could be cut from weeks to days. The software customization and configuration time could also be reduced because the application server VMs can be preconfigured to spin up as clones of existing servers. Some configuration time is always needed, of course, but the speeding up of this capacity addition process lets IT people spend more time on more strategic activities. The department can be more proactive instead of reactive. With less dead time, the department can become a better partner for the business in achieving agility.

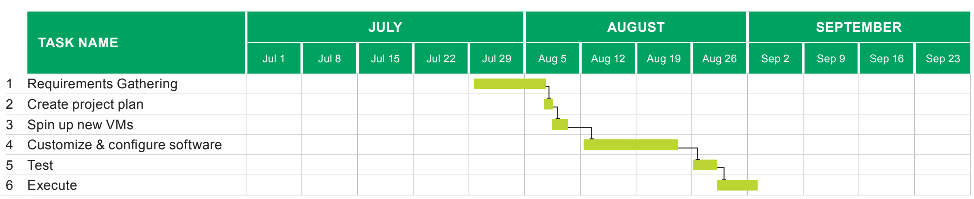

Figure 10 – The project timeframe for adding server capacity is now reduced from 8 weeks to 5.

USE CASE: SERVER CONSOLIDATION

Consolidation of servers is one of the great uses of virtualization and increasingly, cloud computing. Almost every IT manager can relate to the problem of server sprawl, where facilities, rack space and overhead is allocated to underutilized physical machines. With up to 82% of servers running at less than 50% utilization, you can imagine how much money is wasted on idle machines.[10] Virtualization gives you the ability to transform underused physical servers into VMs and host them on compact, highly utilized physical machines. The result of this consolidation is often more cost effective use of IT resources. To take a simple example, if you had three servers running at 30% capacity, you could consolidate them into three VMs on a single physical machine running at 60% capacity. Your physical space needs could but cut by two thirds and your energy consumption and cooling needs could plummet as well.

Maintenance can also be costly if your department is suffering from server sprawl. If you have to use multiple management consoles to oversee the sprawling IT operation, that will make things even more inefficient. Routine activities such as patching and scheduled maintenance exact an excessive toll on IT operations and budgets.

The Union Pacific railroad presents a successful case of using virtualization to consolidate and reduce IT overhead. Originally, the company had approximately 2,500 servers, nearly 1,000 of which were running the Windows operating system. The company's "one application per server" deployment model had resulted in a proliferation of servers, with many of the servers' resources barely utilized. To solve this problem, Union Pacific created 380 virtual machines on 40 Hyper-V host servers[11]. These were Dell PowerEdge R905 and 6950 servers powered by Quad-Core and Six-Core AMD Opteron processors. In addition to migrating to VMs, the company replaced older, large servers that take up half a rack with more powerful blade servers that take up a third of the space. They are packing 16 servers in the space previously consumed by one. The company's Hyper-V virtual machine density averages 12:1 (12 virtual machines on one physical host server), although some development and test hosts contain up to 50 virtual machines per host.[12]

AMD itself also faced a similar challenge. Not too long ago much of our compute power was in isolated pockets and software was not standardized. To make our compute resources more flexible and easily accessible, we transformed our existing infrastructure was transformed into a private cloud and updated to use newer, more power-efficient AMD Opteron™ processor technology. Our IT organization worked to standardize software throughout the emerging cloud, and selected The Red Hat Enterprise Linux operating system as the foundation. AMD's new cloud infrastructure runs more than 115,000 AMD CPU cores and over 4 Petabytes of storage. Our base compute resource capacity has increased by 20% while we have simultaneously achieved a sustained utilization rate greater than 90%. Overall, the transition to VMs has saved over $6 million through in-place upgrades.[13]

Consolidation sounds great, but, like many revolutionary IT concepts, it's easier to envision than to execute. To us, getting virtualization right means being able to consolidate servers without compromising on system availability and performance. For instance, if you consolidate physical servers but you lack an effective way to balance load, you may be creating processing bottlenecks as you reduce your physical footprint. Consolidation through virtualization has to be coupled with sophisticated management tools or it may actually create new problems as it solves others.

It can also be hard to combine machines for architectural reasons. For example, if you want to consolidate servers running two applications that each require dedicated databases and storage, you may be able to make the application servers into VMs and run those on a single machine but you may still be forced to run the databases and storage inefficiently. A solution to this might be to create a private cloud-based database-as-a-service (DBaaS), along with a storage area network (SAN) that also runs in the cloud on VMs. With these two additional steps, you could virtualize your entire application setup and reap significant efficiency gains.

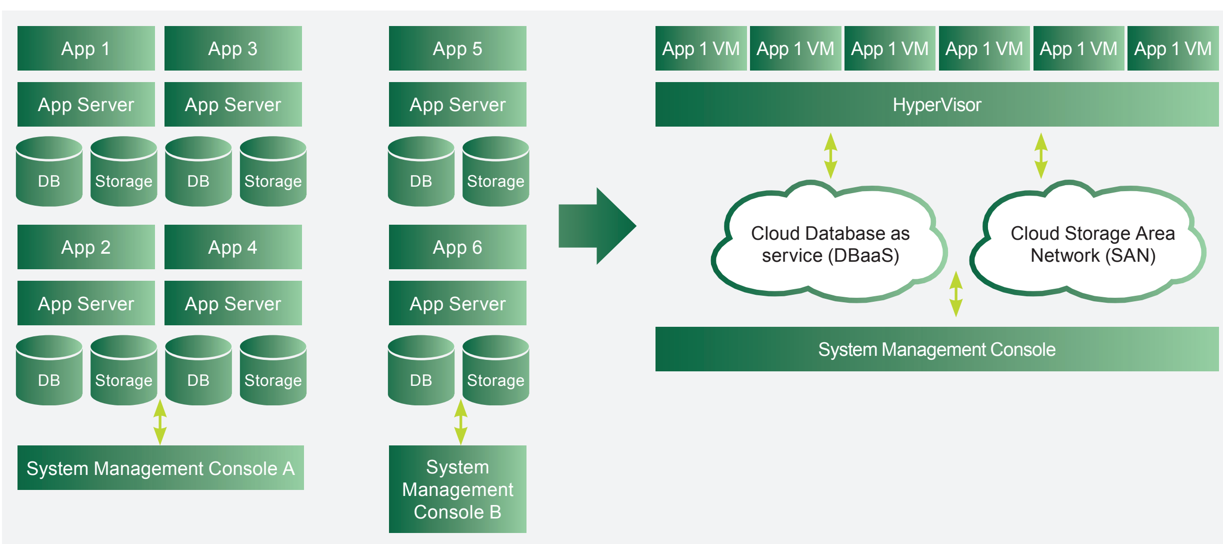

Figure 11 – Example of server consolidation through virtualization. Separate database and storage systems for each application are merged into cloud-based database-as-a-service (DBaaS) and storage area network (SAN). Six applications are transformed into VMS on a single physical machine. One system management console can monitor the new virtualized environment.

Figure 11 show how this transformation might look. It is not a push-button process, however. Having the ability to make these kinds of virtualization moves requires using the right tools, the right server platform—getting virtualization right. The AMD-powered server platform, including software and hardware from AMD partners, enables this level of virtualization and cloud computing. In this example, the AMD-powered server platform enables you to manage diverse VMs, some of which are based in a private cloud environment, together with on-premise VMs.

USE CASE: BUSINESS CONTINUITY

There's an old joke that goes like this. What's the only thing worse than a sub-standard system? Answer: A sub-standard system that's not available… If you manage IT, you are, at some basic level, in the system availability business. When computing capacity that is your responsibility fails to do it job, you're in trouble. As a result, most IT managers spend some of their time focused on business continuity, a collection of policies and practices aimed at providing the highest possible level of system availability for business critical IT assets.

Business continuity tends to be a fairly broad and blurry domain, but it does operate on some fundamental principles:

- **The risks associated with high business impact need the greatest attention** – If your e-commerce website goes down the Friday after Thanksgiving, you're in hot water, no matter how quickly you can get it back. Every minute that site is down, you're losing customers, revenue and reputation.
- **The most critical systems need the fastest recovery time objective (RTO)** – A good business continuity plan will feature RTOs for systems deemed critical to the business. RTOs vary with your business. If you are a manufacturer, a thirty minute RTO for an ERP system might be adequate. If you're a stock trader, a ten second trading system RTO might be too long.
- **Resource limitations preclude providing the same level of continuity for all systems** – Business continuity tends to be expensive. Historically, as a result, it was impossible to ensure the same level of continuity for all systems. You would focus your resources on the most critical systems.

These fundamental rules are based on economics. If you wanted a system to fail over to a second instance, you had to build that failover machine and run it in parallel. This was a costly proposition, so the IT profession settled on different levels of failover. You could have a failover instance running next to your original machine. Or, if you were concerned about a disaster, you could create a duplicate of the original system in a separate location. Your disaster recovery site could be a "mirror," meaning that it was running perfectly in unison with the original. This formula, which is popular with banks and other super critical customers, enables virtually instantaneous shifting of computing from one data center to another that might be hundreds of miles away. Of course, mirrors are the most expensive way to ensure business continuity.

Alternatively, you traditionally could have disaster sites that were either "hot," "warm" or "cold." A hot site would be running an identical twin of your original system, though usually not completely in parallel. Switching capacity to a hot site was relatively quick. A warm site contained the equipment you needed, switched on and running. A cold site had the equipment on the shelf, available for setup. RTOs got longer as the "temperature" went down.

Virtualization and cloud computing have not altered the fundamentals, but they do have the potential to change the economics and practical aspects of the whole situation. By making it comparatively simple and economical to spin up identical machines in multiple datacenters—without having to acquire and stand up equipment—you can accelerate RTOs at every level. What might have been an expensive mirror site can now be a straightforward second virtual instance of a system (including all the complicated database connections, and so forth.) The system that previously only merited a warm site backup, or nothing at all, can have a failover VM instance and a lightning fast RTO.

Scheduled maintenance is a related concept. Before virtualization and the cloud, a server invariably had to be taken offline for maintenance. This downtime was usually scheduled over a weekend, but it was still a hassle for the IT department and system users. Now, you can spin up a virtual clone of the system and divert traffic to it while you maintain the original instance. When you're finished with maintenance, you spin down the clone.

CONCLUSION

There are no magic bullets when it comes to virtualization and the cloud. The benefits of the technology are well understood at this point. Being able to spin up VMs when needed and spin them down when finished has a clearly demonstrated value for server consolidation, business agility and business continuity. However, getting from this basic kind of virtual solution to one that delivers optimal results involves bringing together hardware, software, networks and infrastructure in a complete server platform solution.

It's a total process, a total investment decision that will yield benefits in proportion to the extent it's well thought out. The AMD-powered server platform, with its blending of a design intended to enhance virtualization efforts and unique software and hardware partnerships, delivers a true breakthrough for virtualization and the cloud. High core density can translate into a greater number of VMs per physical machine. Coupled with unique power management capabilities, the result can be a more cost effective and high performing virtualization and cloud solution. Innovative connections with system management tools make it possible to manage and scale your AMD-based virtualization solution cost effectively no matter how complex your business requirements might be.

FOOTNOTES

1 http://www.marketresearchmedia.com/2009/05/20/us-federal-cloud-computing-market-forecast-2010-2015/

2 https://www.eiseverywhere.com/file_uploads/d4d4c4c48ef3003ce8bc5c444442a6ea_ITR_DC_AMD_Opteron_Processors_in_the_Cloud_Pat_Patla_AMD.pdf

3 Certain AMD Opteron™ processors experience all core boost of up to 500 MHz (P2 base to P1 boost state) and up to 1.3 GHz max turbo boost (half or fewer cores boost from P2 to P0 boost state).

4 AMD internal estimates as of Q4 2011

5 AMD Opteron™ 6200 Series processors experience all core boost of up to 500 MHz (P2 base to P1 boost state) and up to 1.3 GHz max turbo boost (half or fewer cores boost from P2 to P0 boost state).

6 FMAC can execute an FMA4 execution (a=b+c*d) in one cycle vs. 2 cycles that would be required for FMA3 or standard SSE floating point calculation

7 Comparison of 12-core AMD Opteron™ 6100 Series processors and 16-core AMD Opteron 6200 Series processors.

8 AMD Opteron™ 6200 Series supports up to 1.5 TB memory capacity in a four processor configuration using LR DIMMs.
AMD Opteron 6100 Series supports up to 1TB memory capacity in a four processor configuration using RDIMMs

9 Comparison of 16-core AMD Opteron™ 6200 Series processor with 8-core Intel Xeon E5-2600 Series processor and 10-core Intel Xeon E7 Series processors according to www.intc.com/pricelist.cfm as of 4/12/12.

10 See Figure 2, above.

11 Case study taken from "AMD Virtualization Plan" Presentation (Slide #22)

12 www.microsoft.com/casestudies/ServeFileResource.aspx?4000018221

13 https://www.eiseverywhere.com/file_uploads/23b618b0e87f5db758c823a15207802c_Cloud_Computing_case_study_Rd5.pdf