Business & Financial Aspects of Server Virtualization

ABSTRACT

Server virtualization falls within the domain of IT operations but it can very much be a business matter with a measurable impact on the bottom line. Esoteric as it may seem from a business perspective, the server platform, powered by the processor, can make a significant difference in the financial outcome of a virtualization effort. Technological factors such as processing power, memory management and power consumption can translate into tangible business results. This paper showcases the business advantages of the AMD approach to virtualization at the processor level, focusing on the use cases of consolidation, agility and business continuity.

INTRODUCTION

Virtualization and cloud computing have arrived. The global cloud computing market is expected to grow to $270 billion by 2020[1]—cloud computing is a basic mandate for most IT organizations. Indeed, a recent AMD survey of over 1,000 IT professionals revealed that 37% of organizations worldwide are currently operating in the cloud and 63% of global cloud customers estimate they store more than $250,000 worth of data in the cloud.[2] No doubt, the cloud is happening.

However, as with any other serious information technology trend, there should be a well understood business reason for undertaking a cloud migration. The virtualization/cloud mandate can affect IT investment decisions your organization is going to make in the next few years. If you are responsible for evaluating virtualization/ cloud technology investments, you will likely want to justify them in economic terms. his paper discusses several approaches to thinking about the financial and business impact of virtualization and the cloud that go beyond basic estimates of return on investment (ROI).

THE BUSINESS PERSPECTIVE ON VIRTUALIZATION AND CLOUD COMPUTING

Getting the economics of virtualization right means going through a comprehensive decision making process that leads to an optimal virtualized environment. There is no such thing as a turn-key virtualization/cloud solution that will work for everyone. For each cloud workload, there will be the right mix of hardware, software, infrastructure, network, server platform, and processor. All of these factors count when you want the best business outcomes from virtualization. Yet, more than one correct choice often presents itself for each category.

The Virtualization-Cloud Connection

It's worth taking a moment to review the connections between virtualization and the cloud. Imprecise use of these terms among business managers can detract from a productive discussion. It's well understood that virtualization is the ability to run multiple sets of server software on a single physical machine. When we talk about virtualization here we mean Advanced, Enterprise-grade Virtualization that demands much more high-performance capability. Virtualization in this paper refers to a technology that allows you to provision numerous, heterogeneous virtual machines, or "VMs", in a variety of datacenters. It's about being able to monitor and maintain those VMs, by changing configurations, patching systems, managing load, and more.

While "cloud computing" generically connotes the placement of IT assets in remote, abstract locations that are accessed online and on-demand, an IT manager needs a more accurate definition. A more refined definition for our purposes contains several facets: Regardless of the IT resource you need—a server, database, storage—you don't have to own it (though you might). Hardware and software become abstract services that run "in the cloud." You connect to them when you need them, for as much or little of the service as you need. Implicit in this construct is the utility computing model. Cloud computing resources are available to meet whatever load you might require. They should scale automatically to meet demand. Additionally, when we talk about "the cloud," we are actually referring to a number of different but related modes of computing. These include:

- **Public cloud –** A commercial service, such as Windows Azure™, which enables you to set up and run VMs in a cloud environment that you can configure but which you do not own or control.
- **Private cloud –** Cloud computing infrastructure that you own and control. A private cloud may run in your own datacenter, but system managers can access its resources on demand as they would with a public cloud provider.
- **Private/public hybrid –**A combination of private and public cloud instances that are used to run integrated systems. For example, a customer relationship management (CRM) application could run in a public cloud but be integrated with an enterprise resource planning (ERP) system that runs on-premise in a private cloud architecture.

Virtualization and the Cloud in the Context of Computing Costs

Getting your head around the economics of virtualization and the cloud involves grasping a set of concepts that are at once quite simple but also detailed and potentially complex. Sounds tricky, and it is. For one thing, while the "how" of virtualization economics is fairly easy to comprehend, the "why" usually involves digging through many layers of seemingly arcane computing technology. How does virtualization save money? Simply put, it combines previously underutilized servers onto fewer physical machines, thus saving money on equipment, facilities and energy. Figure 1 provides a simple example. Astonishingly, more than 80% of servers are running at less than 50% of utilization[3], meaning that half of the money spent on running and cooling them is wasted. Consolidating servers saves money on energy, cooling and facilities expense. As well, fewer person-hours are required to maintain a smaller physical base.

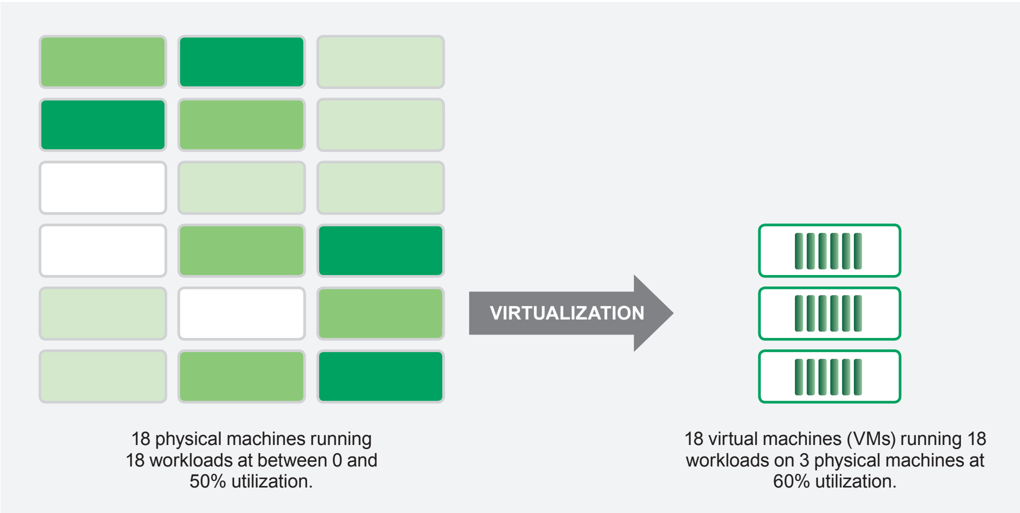

Figure 1 – An example of the power of virtualization to consolidate underutilized physical servers onto highly utilized virtual machines that consume less power and require less cooling capacity, equipment and facility expense to run. (Note, this example is for illustrative purposes only and is not based on actual server performance data.)

If you ask how it actually works, or why do virtual machines achieve such economies, that's a bit more involved, and it's worth getting into depth on this topic. If you think of virtualization simply as a "black box" process that helps you shift computing capacity from physical to virtual machines, you will miss the opportunity to understand how it can really save you money. It might be helpful to use the following analogy when thinking about the costs of computing and benefits of virtualization: The compute capacity of your IT department is like a factory. Servers are like manufacturing plants that process materials, taking in data on one end and discharging modified data at the other. Regardless of your IT department's mission and the software it uses, the microprocessor-level bit crunching that goes on will be the same. The faster and cheaper the server platform can run bits of data through the processor cores, the more economic benefit you will derive. Conversely, underused assets are seen as a waste of money, both in terms of capital expenditure and overhead.

Understanding how virtualization can successfully make the compute "factory" run better requires getting acquainted with the ways that computing actually costs money. A computer, such as a server in your data center, is designed to process "threads" of computing instructions and data as directed by the operating system (OS). Think of the threads as production runs of a product. The processor core inside the server is like the assembly line that completes the production run. Memory on the processor chip is like a buffer of manufacturing inventory – holding data/thread "product" until it's ready to be processed. The network is like the conveyer belt moving production materials all around the factory.

The server itself is like a physical manufacturing asset. It costs money to buy and assemble. It costs money for the electricity that runs it and cool it down because processors generate heat. A computer's CPU is like an electric heater with millions of tiny pulses of electricity heating up as each thread is pushed through the transistors. The resistor in the microchip's transistor is not that different from the heating coil in your toaster. It's just a lot smaller. You have to buy or rent the space for your servers just as you would with a factory building. And, there are people required to maintain the factory.

The Virtual Machine Power Equation, in Dollars and Cents

Energy consumption by data centers is typically a big line item in corporate IT budgets. In 2007, U.S. data centers consumed 61 billion kilowatt hours at a cost of $4.5 billion, or 7 cents per kilowatt hour. The 2011 figures are now closer to 100 billion KwH and $7 billion.[4] The following simplified model puts the server consolidation example shown in Figure 2 into hard dollar terms:

	PRE-VIRTUALIZATION	POST VIRTUALIZATION
Energy Costs		
Number of servers	18	3
Server Power (In KW)	500	500
Cost per KW/Hour*	0.07	0.07
3 Year Energy Use Cost per Server**	$ 1,839.60	$ 1,839.60
Total 3 Year Energy Use Cost**	$ 33,112.80	$ 5,518.80
PUE***	2.00	2.00

* National average cost per KwH for data centers (US Government/EnergyStar)

** Source: 3 year energy use calculation cited in Electronics Cooling, 2006.

*** PUE = Power Usage Effectiveness, a measure that compares "IT Electrical Load" i.e. server use vs. "Data Center Electrical Load", which is power use for air conditioners and facilities. In many cases in North America, that is a 1:1 ration, so the PUE = 2. 1 kW of electricity for the server requires 1 kW to power the rest of the data center. Source: White Paper # 6 from Green Grid Data Center Power Efficency Metrics: PUE and DCIE – www.greengrid.org.

Table 1 illustrates how virtualization can result in substantial savings as the physical servers required to process a workload are reduced from 18 to 3. Energy costs drop as the number of machines decreases, as does the cost of cooling. Overall, the $33,112 needed to pay for power and cooling over a three year period drops to $5,518, an 83% savings. Of course, this example is only meant for informational purposes. In reality, numerous factors, such as the type and age of hardware, the kind of workloads being processed, and others, will influence the potential savings from virtualization. In addition, this model does not include the costs of running a network or staffing the datacenter. However, the basic savings concept should be clear and hold up in most scenarios.

Envisioning an Economically Optimal Virtualized/Cloud Environment

The economically optimal virtual environment means more than having enough VM computing power at the lowest possible cost. It's not a pure muscle competition, though. Attaining ideal economics in a virtual environment involves balancing the four interdependent factors shown in Figure 2. You likely will want virtual and cloud-based computing resources that are as fast and responsive as anything you have on premises. You will likely 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. At the same time, your virtual environment needs to have the capacity to be cost-effectively managed. Cost effective management of virtual environments is both a human and technology issue. People manage the systems, of course, but they do so using specialized software. The people and system management software need to be in alignment with the server platform. Finally, the energy and cooling costs need to be as low as possible.

Understanding how to arrive at the right mix of performance, flexibility, manageability and cost means diving deeply into the server platform technology. With the factory analogy in mind, the key to understanding both the "how" and "why" of virtualization economics is to be familiar with how the costs of computing are affected by the elaborate handshake between hardware, software, silicon and system management. Virtualization is a software phenomenon, but the hardware, including the silicon processor, plays a critical role in making virtual environments function as envisioned.

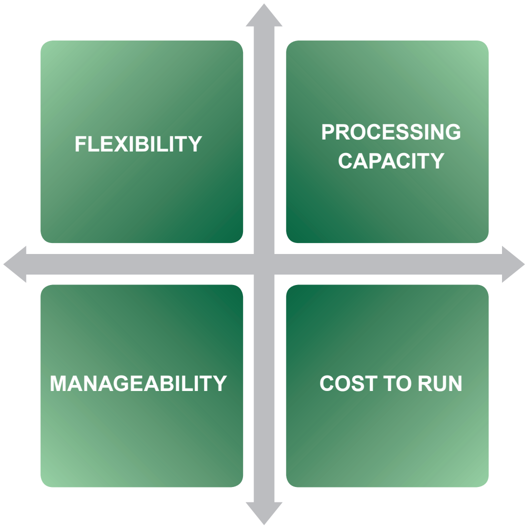

Figure 2 – The four interdependent factors that affect the economics of virtualization.

- **Hardware –** Servers can often be configured for best financial performance for specific workloads. For example, data intensive applications such financial transaction processing may be best handled by one type of server while less intense activity, such as Website page views, are processed by another. The amount of memory on a server, the processor type, and hard-drive all have an impact on how much processing power the machine can deliver for a given level of investment, energy cost and cooling requirement.
- **Software –** The way an operating system, database and software application (such as Enterprise Resource Planning or ERP) function on a server depends on the hardware configuration and choice of processor. Given that efficient processing of threads and workloads is a driver of economic success, software is a critical factor in the server platform's financial performance.
- **Silicon –** It can be said that Virtualization originates in the silicon. As a result, so does the financial impact of virtualization on a given server platform. The processor influences the economic effect of each of the following aspects of virtualization:
    - **Data processing efficiency –** The more threads your processor can handle quickly and efficiently, the more processing capacity you will have on the machine.
    - **Virtual Machine Density –** The more VMs you can run on a physical machine, the more cost effective the machine will be.
    - **Core density –** Having more cores means more flexibility in processing capability and dynamic power consumption.
    - **Memory management –** The way that the processor manages memory affects the amount of computing that the VM can perform in a given time period. The better the memory management, the better the processing outcome.
    - **Network/processor connections –** The processor must juggle requests for processor capacity coming over the network. The better the processor is at managing connections between the more efficient the processor will be.
- **System Management and Overhead –** Server platform affects data center operating costs:
    - **Cooling and energy –** By enabling multiple VMs running on fewer machines, the processor is a major factor in keeping data center cooling energy costs down. The more VMs that can be packed onto a server, the more you can consolidate machines and cut on power and cooling expense. This may seem like a trivial matter, but studies have shown that a single server can cost as much $1800 to power and cool over a three year period.[5]
    - **Facilities utilization –** The facility itself costs money to own (or rent) and maintain. Floor space can be expensive and the reality is you may simply be running out of it. Virtualization can save money by compressing the compute capacity into a smaller, high server density environment.
    - **Headcount –** People run data centers. The more machines that have to be maintained, the more people you will require. If VMs can be efficiently managed, virtualization will help you cut head count or deploy it more productively. System management software, such as Microsoft System Center, needs to be able to handle large numbers of VMs.

Going deeper, the ideal server platform for virtualization economics needs to provide the best outcomes in the following areas:

- **Predictable performance –** Cloud computing workloads tend to be different from those handled by traditional data centers. Cloud work is "spiky" in nature, as Figure 3 shows. You must be able to account for both the peaks – typically with more cores - and the valleys, with more power efficiency. Your cloud-based CPUs should offer predictable performance to handle peaks in traffic.
- **Scalability –** On-demand scaling may be the essence of cloud computing. Your VMs need to be able to scale easily and rapidly, without excessive management overhead required to oversee the process. This is 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 requires highly efficient management of processor workflows. As multiple applications and operating systems run threads through multiple processing cores, your virtualization solution should provide dedicated compute resources to each VM, enabling optimal assignment threads to processor resources.
- **Security –** Risk management intensifies as IT assets are moved off premise. Your cloud solutions must have robust security to protect data and processing resources. Hardware assisted security helps keep intruders out and safeguards sensitive data. Security is both a direct and indirect cost factor for virtualization. The best server platforms should enable cost effective security monitoring and policy enforcement. However, if security is deficient, there can be high costs if there is an incident.

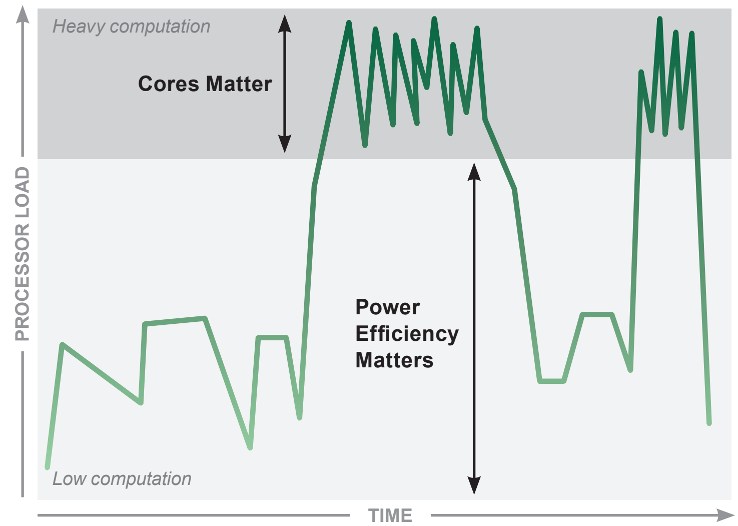

Figure 3 – Cloud server workloads, which tend to "spike" at busy times, need processors that can handle the load predictably.[6]

THE AMD APPROACH TO ECONOMICALLY OPTIMAL VIRTUALIZATION

AMD offers the industry's most financially advantageous solution for virtualization and cloud computing.[7] AMD Opteron processor-based server platforms enable you to migrate to the cloud with the mix of performance, flexibility, maintainability and cost that suit your business requirements. AMD's key to economic success with virtualization is the incredible density of VMs that you can run on a single physical machine running an AMD Opteron processor. Greater core density in the new AMD Opteron processor chip architecture translates into more VMs with dedicated resources per server. You get the most out of your investment in compute resources. AMD Opteron processor-based server platforms provide superior scalability, memory management, energy efficiency cooling, economical operation, and better price/performance, at a price that can mean great things for your ROI. There's also great flexibility to balance workload requirements and respond to changes in a business.

Collaboration with ISVs and OEMs

AMD processors alone cannot affect breakthrough financial outcomes in virtualization. AMD's achievements 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 Opteron 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:

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

From the business perspective, AMD gives you the ability to deploy a solution - hardware/software/processor combination - that will suit specific business needs. AMD partners with ISVs across the entire software ecosystem, from compilers through middleware to applications, with vendors such as Microsoft, Red Hat, VMWare, Citrix, Apache Foundation and MySQL. OEM hardware manufacturers, such as HP and Dell, complete AMD's virtualization and cloud solution. If your workload is best handled by, for example, Microsoft® SQL Server® running on a VM powered by Microsoft Windows Server® Hyper-V®, there will be an ideal server to make that run as cost effectively as possible. If you require MySQL running on a Java/Linux machine, AMD and its OEM partners also have a category-leading solution for that. Specific examples include:

- The HP ProLiant server series is designed for scale-out and power efficiency. 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.[8]
- Dell's PowerEdge C6105, intended 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.[9]
- The Dell PowerEdge C6145, built 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.[10]

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. Nearly 2 million AMD processors are engaged in cloud computing clusters worldwide[11]. Several high-profile cloud initiatives are built off of the AMD stack. These include:

- **Microsoft Windows Azure™ –** Microsoft's cloud services platform uses servers with AMD Opteron processor technology.
- **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

AMD's hardware-assisted virtualization helps you get the economics of virtualization right. With maximum core density and the compute power per processor, AMD virtualization delivers high levels of throughput, performance and scalability.
AMD Opteron processors are designed with features to provide consistent performance across all types of demanding workloads even when system utilization is at its peak.
With the dynamic constraints of performance, flexibility, maintainability and cost in mind, consider the following AMD Opteron processor characteristics:

Performance/Watt Improvement per Processor Generation

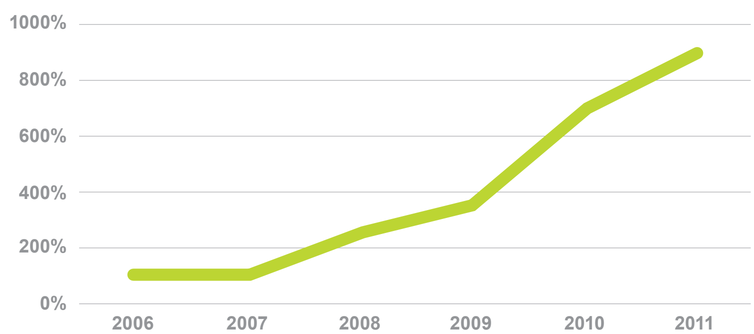

Figure 4 – AMD Opteron™ processors have demonstrated gains in performance/watt steadily, with the latest generation reaching a 900% improvement over the last 5 years.

- **Transaction processing performance –** With AMD Opteron 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 technology automatically increases clock speed by up to 1 GHz+, which is useful for such virtualization/cloud workloads as financial, database, and business intelligence applications.[12,13]

- **Predictable performance –** AMD helps you achieve consistent, predictable performance and improved throughput with dedicated cores for each virtual machine. The high core count makes possible a super high-density cloud environment with better overall application responsiveness. AMD's Direct Connect Architecture with HyperTransport™ 3.0 Technology provides enhanced interconnect speeds to handle fluctuating workload demands. Application performance tends toto be predictable and consistent during normal and peak processing times.
- **Scalability –** Compared to earlier generations of the processor, the AMD Opteron™ 6200 Series processor-based server platform has 33% more cores and 50% 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.[14,15]
- **Workload management –** AMD-V™ and AMD-Vi technologies reduce the overhead of virtualization software and help improve performance for I/O intensive workloads for near-native application performance.
- **Memory management –** AMD Opteron processors support up to 12 DIMM memory chips and 384 GB memory per CPU. This can give 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 generations of AMD processors.
- **Network/processor connections –** HyperTransport Technology Assist increases 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.
- **Server density –** By offering the world's first 16-core x86 processor, AMD Opteron processors provide up to 100% more cores than competing products.[16] This means you get more compute capacity in the same physical footprint while minimizing contention for compute resources. The higher core density gives you more flexibility and lets you run more VMs, allowing your systems to achieve more efficient server utilization that should result in a 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 remote server management even in "lights out" conditions when a machine has failed and will not reboot.
- **Security –** AMD's AES–NI instructions 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 Opteron processors from exceeding their thermal design guidance. With AMD Opteron processors, you can have the capacity you need along with a major improvement in performance per watt and overall power usage. AMD chips 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 reduce processor energy consumption, operating with separate power planes for cores and memory controller.
  - **C1E –** Reduces 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.

USE CASE: BUSINESS AGILITY

Agility is a frequently-used term that can mean different things to different people. Business agility refers to a company’s ability to respond to changes in business conditions quickly enough to benefit from them. In practical terms, agility is a matter of changing strategy and introducing new products on a cycle that keeps you ahead of the market. Sometimes, agility means doing mergers and acquisitions as companies restructure to compete better in a changing market.

The execution of an agile business move usually flows along a predictable path, though each company’s version of it will be slightly different. As depicted in Figure 5, three parallel work tracks typically coalesce as a company changes direction or operations in response to a perceived shift in business conditions. There is the change initiative itself, which might be the introduction of a new product line or service offering, for example. Then, in almost every case, IT has a key role in making the change a reality. A website might have to be created and launched. You can expect the business process changes will have to be implemented across multiple systems. Consequently, new IT infrastructure has to be deployed, as you might have with a regional logistics hub being established, and so forth. Finally, there is the operational launch of the change initiative, which also involves IT.

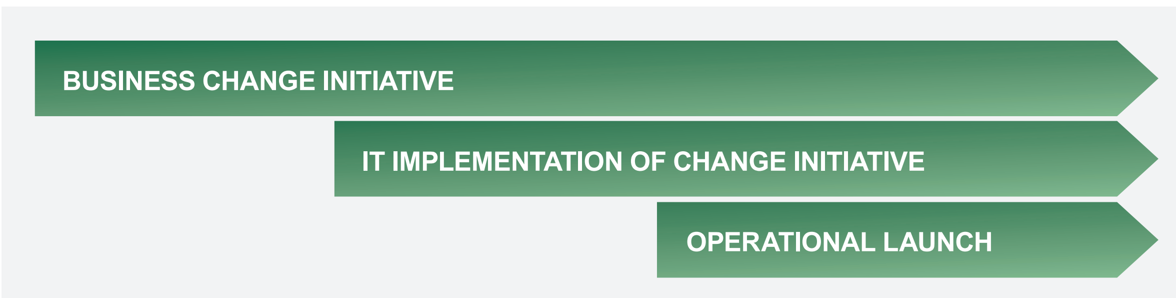

Figure 5 – Execution of an agile move as three parallel tracks involving the change, IT and launch.

Many factors can slow down the execution of an agile business move, and in many cases IT can bear the fault of such delays. To be fair, it’s not always IT’s fault, but there can be a financial penalty for the business when an agile move gets slowed down. The financial repercussions of sluggish execution arise on both the cost and revenue sides of the profit and loss (P&L) statement. In terms of cost, there may be an actual discretionary expenditure involved, meaning that the longer the execution takes, the more the move will cost. Imagine, for example, if you were paying an IT consultant $85 an hour to stand up new equipment and deploy a new solution. If the project took six weeks, you’d be out about $18,000. If you could do it in two weeks, you’d save $12,000. That’s a clear savings. However, when the IT department handles an agile move, there does not have to be an actual increase in out-of-pocket spend, so the savings from moving faster might not be apparent. But, there is a big opportunity cost factor to consider. What else could the IT department be doing for the business if it were not bogged down in a slow execution cycle for a business change initiative?

You can also lose revenue when agile moves are sluggish to execute. In a simple sense, if you could earn $10,000 a month on a new product, if you miss two months, your revenue is off by $20,000. A more realistic scenario, however, is one where an agile move is intended to take advantage of a new situation, one where prices might be high initially but then fall as competition heats up. Table 2 shows what this might look like. If the new product can be sold in the first month, it will fetch $10. By month 4, it’s down to $7, and so forth. The faster the new product can come online, the more high priced business you can capture. If you wait too long, you’re probably getting the worst pricing.

	MONTH											
	1	2	3	4	5	6	7	8	9	10	11	12
Volume	0	0	1000	1500	2500	5000	10000	15000	15000	15000	20000	20000
Unit Price	$10.00	$9.00	$8.00	$7.00	$6.50	$6.00	$5.50	$5.00	$5.00	$5.00	$5.00	$5.00
Revenue	$ –	$ –	$8,000	$10,500	$16,250	$30,000	$55,000	$75,000	$75,000	$75,000	$100,000	$100,000
Total	**$544,750**											

	MONTH											
	1	2	3	4	5	6	7	8	9	10	11	12
Volume	0	1000	1500	2500	5000	10000	15000	15000	15000	20000	20000	20000
Unit Price	$10.00	$9.00	$8.00	$7.00	$6.50	$6.00	$5.50	$5.00	$5.00	$5.00	$5.00	$5.00
Revenue	$ –	$9,000	$12,000	$17,500	$32,500	$60,000	$82,500	$75,000	$75,000	$100,000	$100,000	$100,000
Total	**$663,500**											
Delta	**$118,750**											

Table 2 – Example: The revenue effect of a one month gain in agile the execution process. Example for illustrative purposes only.

Figures 6 and 7 illustrate the financial difference a single month could make in a product launch. The model assumes, as is usually the case that volume starts small and then ramps. Getting to market a month earlier means that this business can sell more product at a higher price. For instance, in month 4, the more rapidly introduce product sells 2,500 units at $7 each, compared to 1,500 in the slower case. By capturing the market at the higher point on the price curve, the business can earn $118,000 in incremental revenue. The cumulative effect of this increase in business is shown in figure 7.

Cumulative Revenue (month by month)

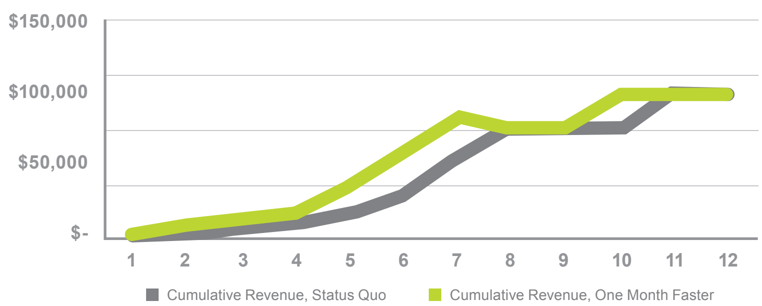

Figure 6 – Comparative revenue for a one month difference in product launch, given price declines featured in Table 2. Example for illustrative purposes only.

Cumulative Revenue Comparison

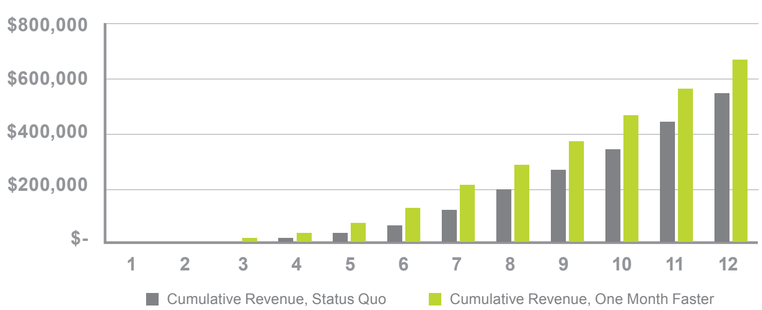

Figure 7 – Cumulative revenue difference between a product launched one month earlier, with figures from Table 2. Example for illustrative purposes only.

IT managers generally want to partner for agility and execute projects 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

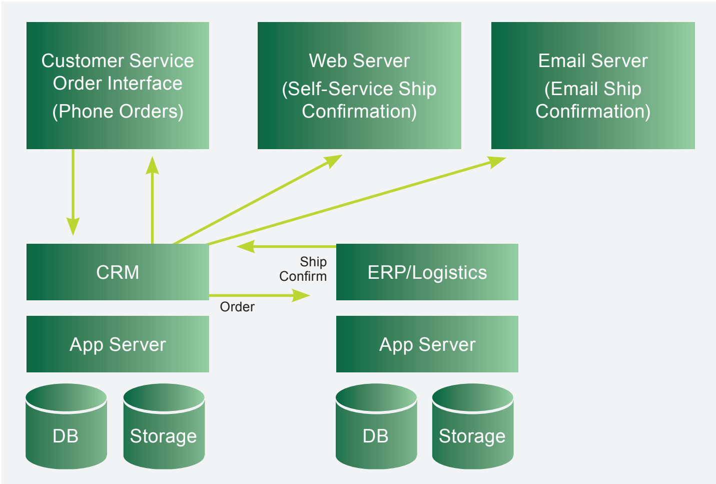

Figure 8 – Reference architecture for retailer that combines CRM and ERP for order processing.

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 8 shows an example of how the two systems might 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 the scenario depicted in Figure 8 takes two months, of which five weeks are needed to acquire and test new hardware and then install and configure the software. Figure 9 depicts the IT project Gantt chart view of the same sample 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" traditionally inherent in server deployment.

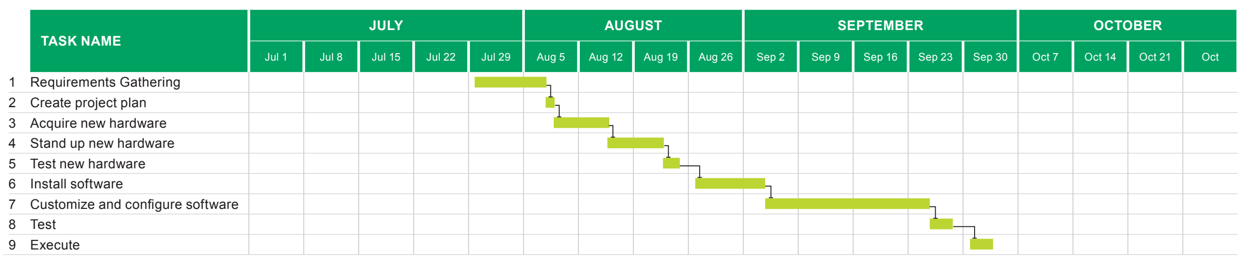

Figure 9 – Gantt chart showing the calendar impact of acquiring and standing up hardware, deploying software, and conducting a migration of software into a new, expanded data center environment. Example for illustration only.

Migrating the CRM-ERP solution to a virtualized or cloud-based environment enables greater agility by speeding up the process of getting server capacity online. As shown in the sample scenario of Figure 10, 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. The AMD-powered server platform enables this capability on several levels. On the front end Web tier, the server platform combines 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 AMD-powered server platform leverages 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 is efficient with memory caching and serving data to the app layer.

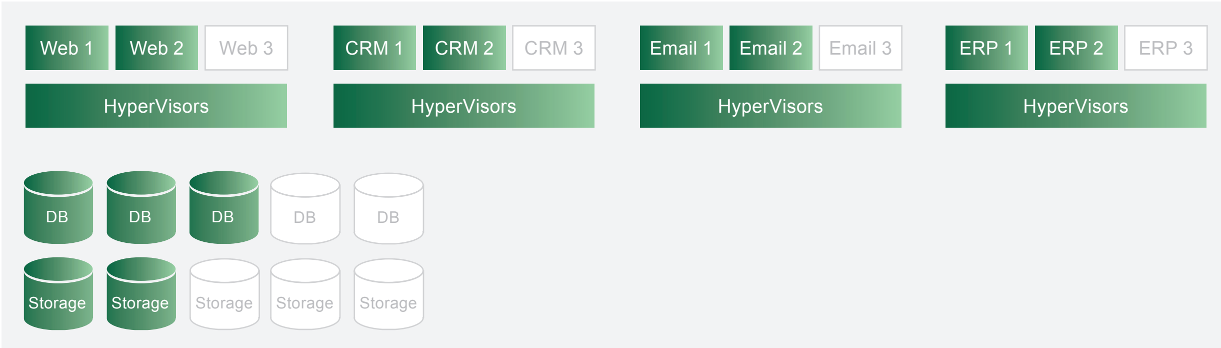

Figure 10 – The retailer can be more agile with a virtualized environment that enables rapid spinning up and spinning down of servers and other IT assets on demand. Example for illustration only.

The whole process moves faster with virtualization. In our revised Gantt chart (Figure 11) reveals, the overall capacity expansion cycle now takes a little over 4 weeks, instead of 8, and the time elapsed between creating the project plan and customizing the software is cut from 3 weeks to 4 days. The software customization and configuration time is 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 productive and less reactive. With less dead time, the department becomes a better partner for the business in achieving agility.

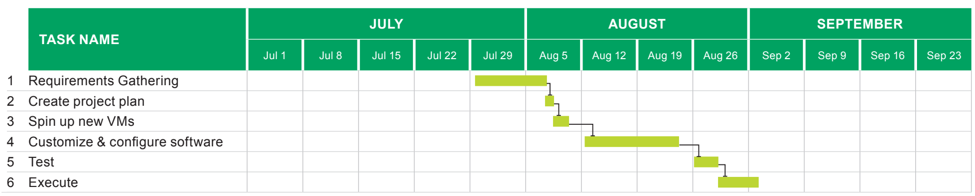

Figure 11 – The project timeframe for adding server capacity is now reduced from 8 weeks to 5. Example for illustration only.

USE CASE: CONSOLIDATION

Virtualization enables the consolidation of underused physical servers by turning them into compact, highly utilized VMs on a reduced number of physical machines. In the example shown in Figure 12, a company wants to consolidate three applications that are running on 100 physical machines due to load levels and their need to have separate, dedicated databases and storage. With the AMD Opteron processor-based server platform virtualization, you can move those 100 machines onto a single 42U rack.[17] The architecture includes a private cloud-based database-as-a-service (DBaaS) and storage area network (SAN). In addition, two system management consoles, which limited IT staff productivity, can be reduced to one. This example is based on the approach holding that an AMD Opteron-based server can easily run 32 VMs per physical machine when one VM is assigned to a single processor core in a standard 2 socket commercially available server[15]. In this case, each physical (1U) machine would be running between 2 and 3 VMs.

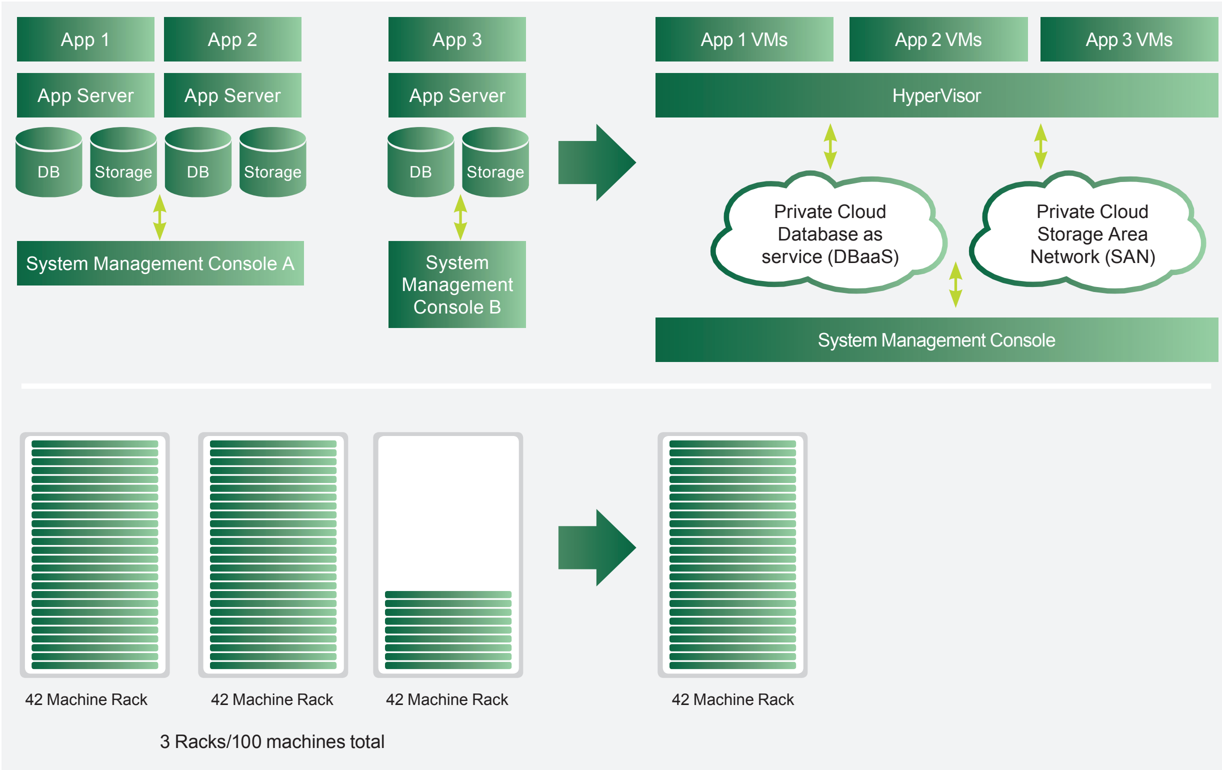

Figure 12 – Example of server consolidation through virtualization. Separate database and storage systems for each application are merged into cloud-based database-as-a-service (DBaaS) and storage area network (SAN). One system management console can monitor the new virtualized environment. Chart for illustration only.

Consolidation sounds great, but, like many revolutionary IT concepts, it's easier to envision than to execute. It is not a push-button process. Virtualizing like the example described in Figure 12 requires using the right tools. When it's done right, consolidation should result in savings on numerous levels of IT operations without compromising on system availability and performance. Consider what happens 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. Alternatively, maintenance is costly if you have to use multiple management consoles to oversee the sprawling IT operation. Consolidation through virtualization should be coupled with comprehensive management tools or it may actually create new problems as it solves others. The AMD Opteron processor-powered server platforms, including software and hardware from AMD partners, makes optimal virtualization and cloud computing possible.

ASSUMPTIONS		
Cost per KwH of electricity*	$0.07	
Data Center facilities cost (Cost per Square Foot/year)	$50.00	
Number of Servers per Rack	42	
Square feet of floor space required per rack	30	
Full time employee (FTE) cost per year (fully loaded)	$100,000	
New physical server cost	$2,000	
Migration project cost, i.e. the cost to perform consolidation	$25,000	
CONSOLIDATION PARAMETERS	**AS IS**	**AFTER CONSOLIDATION**
Energy use – Individual server power use (Watts)	500	500
Power cost per server/year (including cooling)**	$613.20	$613.20
PUE***	2.00	2.00
CONSOLIDATION RETURN ON INVESTMENT (ROI)	**AS IS**	**AFTER CONSOLIDATION**
Number of physical servers (Assume 1U server)	100	42
Power cost	$61,320	$25,754
Racks required to house servers	3	1
Floor space (Square feet)	90	30
Facilities Cost	$4,500	$1,500
FTEs to maintain system	1	0.5
FTE cost	$100,000	$50,000
Total Cost of Ownership (TCO)	$165,820	$77,254
Delta	$88,566	
Equipment cost	$84,000	
Project Cost	$25,000	
Total investment	$109,000	
Project payback time (Years)	1.23	
ROI	81%	

* National average cost per KwH for data centers (US Government/EnergyStar)
** Source: Energy use calculation cited in Electronics Cooling, 2006.
*** PUE = Power Usage Effectiveness, a measure that compares "IT Electrical Load" i.e. server use vs. "Data Center Electrical Load", which is power use for air conditioners and facilities. In many cases in North America, that is a 1:1 ration, so the PUE = 2. 1 kW of electricity for the server requires 1 kW to power the rest of the data center. Source: Electronics Cooling, 2006.

Table 3 – Consolidation return on investment (ROI) estimate.

It is possible to estimate the return on investment (ROI) for consolidation though your numbers will inevitably be unique given your specific circumstances. Table 3 provides a quantitative view of the consolidation depicted in Figure 12. 100 physical machines on 3 racks migrate to 42 machines on a single 42U rack. Cooling and facilities costs are expected to decrease in turn. Reducing the management consoles from two to one frees half of a full time employee (FTE). This example of consolidation saves over $88,000 a year. With an up-front investment of $109,000, this model project yields an ROI of 81% and about a year and a quarter payback time.

AMD has seen a number of clients benefit financially from consolidation. For example, Union Pacific Railroad used 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. 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.[18]

AMD itself also faced a similar challenge. Its compute power was in isolated pockets and software was not standardized. To make its compute resources more flexible and easily accessible, they transformed their existing infrastructure into a private cloud and updated to use newer, more power-efficient AMD Opteron processor technology. The 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 AMD CPU cores. 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.[19]

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 potentially losing customers, revenue and reputation. If your instant messenger system goes down at two o'clock on a Saturday morning, you'll live, even if it takes a whole day to get it back online.
- **The most critical systems need the fastest recovery time objective (RTO) –** It makes sense that a good business continuity plan will feature RTOs for systems deemed critical to the business. RTOs vary with the type of business. If you are a manufacturer, a 30-minute RTO for an ERP system might be adequate. If you're a stock trader, a 10-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 practically impossible to ensure the same level of continuity for all systems. Instead, you would probably choose to focus your resources on the most critical systems.

These fundamental rules are based on economics. If you wanted a system to failover 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 often popular with banks and other 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 usually 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 runs 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 contains the equipment you need, switched on and running. A cold site has the equipment on the shelf, available for setup. Recover Time Objectives got longer as the "temperature" went down. Table 4 shows how an information security manager might assess risk and assign financially realistic recovery times.

SYSTEM	BUSINESS IMPACT OF FAILURE	LIKELIHOOD OF FAILURE	RISK RATING	COUNTERMEASURE	RTO
ERP	High	Medium		Mirror	1 Minute
CRM	Medium	Low		Failover site	10 Minutes
Email	Low	Low		Warm site	4 Hours

Table 4 – sample business impact analysis with RTOs.

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

The potential business and financial gains from virtualization are easy to see but not always simple to realize. 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 economic value for server consolidation, IT agility and business continuity. However, getting the right results from virtualization – cost savings, flexibility, performance and manageability - involves bringing together hardware, software, networks and infrastructure in a complete server platform solution.

The economic value of virtualization flows from deep within the technology, including from the silicon itself. Getting there is a total process, a total investment decision that will yield benefits in proportion to the extent its well thought out. The AMD Opteron processor-powered server platform, with its blending of "virtualization in the silicon" and unique software and hardware partnerships, delivers a true breakthrough for virtualization and the cloud. High core density translates into a greater number of VMs per physical machine. Innovative connections with system management tools make it possible to manage and scale your AMD Opteron processor-based virtualization solution cost effectively, no matter how complex your business requirements might be.

FOOTNOTES

1 http://www.marketresearchmedia.com/2012/01/08/global-cloud-computing-market

2 AMD online survey, March 2011—1513 respondents from orgs with 100+ staff:1000 USA, 259 Asia (China, India, Singapore) 254 Europe (UK, France, Germany)

3 IDC WW Server Workloads 2009—Final Market Model (06.30.09)

4 US Government Energystar data at http://www.energystar.gov/index.cfm?c=prod_development.server_efficiency

5 Calculation based on 7 cents per KwH and formula described in Electronics Cooling, 2006 based on metrics developed by Green Grid. (500 Watt server running for three years, or 26,280 hours X a PUE of 2 = 26,280 KwH. @ 7 cents/KwH = $1839.)

6 Based on internal AMD engineering projections of AMD Opteron 6200 Series processors with up to 500 MHz in P1 boost state and up to 1.4 GHz in P0 boost state over base P2 clock frequency.

7 It is assumed that the HP DL385 G8 will be similarly priced to the HP DL385 G7. HP ProLiant DL385 G7 with 2 x AMD Opteron™ processor Model 6282 SE (1ku price $1019) with 32GB RAM, 146 GB 15K hdd, DVD, and 3yr base warranty is $5,143 as of 4/2/12 at www.hp.com. The 1kU list price per AMD Opteron processor Model 6284 SE is $1265, which is $246 more than the 1kU list price for the AMD Opteron processor Model 6282 SE ($1,019). Therefore, $492 has been added to give an estimated server price of $5635. HP ProLiant DL380 G7 with 2 x Intel Xeon processor Model E5-2690 with 32GB RAM, 146 GB 15K hdd, DVD, and 3yr base warranty is $9,127 as of 4/2/12 at www.hp.com. SVR-129

8 http://h18004.www1.hp.com/products/servers/platforms/

9 http://www.dell.com/us/enterprise/p/poweredge-c6105/pd

10 http://www.dell.com/us/enterprise/p/poweredge-c6145/pd

11 Source: AMD internal estimates as of Q4 2011

12 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). SVR-27

13 AMD Opteron™ 4200 Series processors experience all core boost of up to 300 MHz (P2 base to P1 boost state) and up to 1.2 GHz max turbo boost (half or fewer cores boost from P2 to P0 boost state). SVR-63

14 Comparison of 12-core AMD Opteron™ 6100 Series processors and 16-core AMD Opteron™ 6200 Series processors.SVR-5

15 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. SVR-64

16 Comparison of 16-core AMD Opteron™ 6200 Series processor with 8-core Intel Xeon E5-2600 Series processor See http://www.intc.com/pricelist.cfm as of 3/14/12. SVR-133

17 AMD Opteron™ 6200 Series-based 2P servers can support up to 32VMs each, assuming 1 VM per core. 21 servers fit in a rack x 32 VMs = 672 VMs per rack. Intel Xeon 5600 Series-based 2P servers can support up to 12VMs each, based on core counts listed at www.intc.com/pricelist.cfm as of 10/24/11, which equates to 21 servers x 12 VMs = 252 VMs. SVR-90

18 http://www.amd.com/us/Documents/47561A_unionpacific_casestudy_WEB.pdf

19 http://www.em.avnet.com/en-us/design/featuredpromotions/Documents/Featured%20Products/AMD/Develop%20a%20private%20cloud%20for%20AMD_Case%20Study.pdf

Sponsored by AMD