GIGAOM RESEARCH

The IT benefits of bare metal clouds

Paul Burns

December 15, 2014

GIGAOM RESEARCH

TABLE OF CONTENTS

Executive summary

Although public infrastructure-as-a-service (IaaS) clouds have been widely adopted to drive IT agility and deliver new business value, confusion still exists about their defining characteristics. Many in the IT industry believe that virtualized infrastructure is a requirement for cloud computing. In truth, however, both virtualized and bare metal infrastructures are well suited for building clouds. And both types of infrastructure enable the defining characteristics of cloud computing, including resource pooling, metered consumption, self-service, and elasticity. The real question is how virtualized and bare metal clouds are different, and in which cases the bare metal option provides an advantage.

CIOs and CTOs want to know which type of cloud will help them meet their strategic IT objectives. Infrastructure architects want to understand where applications will run best. DevOps teams, meanwhile, want to learn which environment will help them deliver new applications quickly, and also manage them efficiently after deployment.

Virtualized clouds offer a convenient, ubiquitous source of cloud infrastructure but they don't always deliver the performance, consistency, or control that IT organizations need in order to support their businesses' mission-critical requirements. IT organizations frequently need bare metal clouds to run large relational databases, meet single-tenant data governance requirements, and build customized environments that match their on-premises configuration.

This report shows IT leaders how bare metal clouds help deliver greater value and efficiency than virtualized clouds alone through highly consistent performance, complete hardware customization, security and risk management, and hybrid workloads.

Key highlights in this report include:

- Virtualized clouds are one option to consider, but IT leaders must not overlook the significant advantages of bare metal.

- Prospective customers for bare metal clouds must be selective, particularly when determining whether they will receive complete management, automation, and control through an API.

- Adopting bare metal clouds can resolve many of today's challenges: poor performance, lack of control and customization, and complex risk management.

The value of public clouds

Public IaaS clouds are vastly popular for a reason: They deliver tremendous value. IT organizations from all industry verticals are adopting them to improve agility, economics, and focus.

Agility. Typical hardware procurement can take up to 12 weeks, while provisioning public cloud services take only a matter of minutes. With on-demand access to infrastructure, initiating new IT projects is dramatically expedited, which improves staff productivity and leads to faster project completion. Cloud computing's low cost also enables experimentation and testing, so staff can spend more time on experimentation and innovation.

Public clouds give IT the agility to keep pace with the business and its customers. Since resources are elastic, consumption levels can be adjusted quickly and to meet demand, ensuring that service-level objectives are met, end users have a better quality experience, and IT spending is aligned with demand.

Economics. Infrastructure ownership is expensive, and many IT leaders have adopted public cloud services to reduce total cost of ownership (TCO). On-premises servers often utilize just 8 to 15 percent of their total capacity, indicating a high rate of overprovisioning and overspending. Virtualization can help increase utilization from 25 to 40 percent, but this means the majority of capacity still goes unused. With public cloud services, customers pay for only the resources they need, leading to greater efficiency, reduced consumption, and lower capital expenditures (CapEx).

Public cloud services can also lower operational expenditures (OpEx), which is a critical cost factor since yearly OpEx associated with in-house hardware can be as much as 70 percent of the initial hardware purchase cost. With hardware lifecycles typically lasting between three and five years, TCO can be three to four times the initial hardware acquisition costs. While reduced OpEx is not guaranteed for every application, savings can be quite large for applications that only run periodically or that scale up and down according to demand.

Focus. IT organizations have historically spent a large percentage of their time and resources designing, operating, and maintaining the infrastructure needed to deliver applications and services. However, most of the business value IT delivers is derived from the applications and services. Public-cloud services allow IT organizations to offload infrastructure responsibilities to a service provider — including upgrades, maintenance and repairs — so they can leverage the providers' skills and expertise, save time, and increase their focus on more valuable tasks.

Developing new applications is one of the most important ways for IT to add additional business value. At the same time, a significant amount of overhead, such as setting up development and test environments, distracts developers and slows the development process. Public clouds let developers focus on what they do best by automating resource provisioning to better support DevOps and continuous application-lifecycle management.

The importance of cloud architecture

Different clouds deliver different types and levels of value. Much of what drives these differences comes down to the underlying architecture. Some providers emphasize convenience and consistency while others enable granular control, customization, and higher levels of service.

Architectural building blocks

Public clouds utilize several core architectural building blocks. At the center is the cloud-management platform, which orchestrates the pooling and provisioning of resources in multitenant environments. Cloud-management platforms provide the essential capabilities providers need to design and sell cloud services. They also offer APIs and automation that enable users to control and manage resources.

Network, storage, and compute infrastructure choices have a direct impact on attributes such as performance and reliability. For example, the specifications and design of the local-area network (LAN) determine the latency and throughput of data transmitted within the cloud. Similarly, wide-area network (WAN) connectivity options, from private lines to peering, determine how well cloud-hosted systems can interact with on-premises systems.

Service providers must carefully assess the pros and cons of individual building blocks and consider how each component will affect the resulting cloud services. For example, hardware specifications can greatly impact performance. Slow LANs inevitably lead to service issues, such as inconsistent performance, unexpected downtime, and limited scale. Software-related decisions are also critical. In fact, the most far-reaching consideration in public cloud design may be whether or not to use a hypervisor.

Multitenancy and virtualization

The vast majority of public clouds available today use server virtualization — in the form of a hypervisor — as a foundation to deliver multitenancy and resource sharing. Hypervisors allow multiple tenants to share compute resources on a single server by partitioning the server into virtual machines and providing secure isolation between them. Each virtual machine (VM) then operates as an independent server. The hypervisor also enables secure, partitioned access to block storage, whether the storage is directly attached to the server or is located within a storage-area network (SAN). Finally, the hypervisor plays an important role in network multitenancy by providing virtual-switching that delivers network traffic to and from the right virtual machine.

Beyond the hypervisor

While hypervisor-based virtualization is commonly used in public clouds, there is another valuable option: bare metal servers. With this approach, individual tenants have exclusive use of dedicated physical servers. When tenants provision bare metal servers, they can install an operating system of their choice or, if they want to run multiple VMs of their own, they can still choose to install a hypervisor. The important distinction is that customers have complete control over whether or not they use a hypervisor.

The lack of a service-provider-imposed hypervisor has a significant impact on architecture. Unlike virtualized clouds, a hypervisor is not required since there is no need to share access to CPU, RAM, or NIC cards, or directly attached storage (DAS). Similarly, there is no need for virtual switches since physical switches managed by the service provider forward only the packets intended for that server. Additionally, since the authorized tenant is the only one with access to the server, only that tenant can install applications.

The limitations of virtualized clouds

Despite their pervasiveness, virtualized clouds are not the perfect choice for every application or workload. Virtualized clouds can and often do provide all the benefits of public clouds, including agility, economic efficiency, and focus. However, they also have inherent limitations that can prevent IT organizations from achieving their mission-critical service objectives.

Inconsistent performance

Hypervisor-based virtualization allows service providers to allocate more than the total capacity of each physical server to tenants. This is called "oversubscription" and is common in public cloud environments. For example, five tenants could each purchase a virtual instance with 2 cores and 2 gigabytes (GB) of RAM, but the total capacity on the host server might only be 8 cores and 8GB of RAM. Oversubscription becomes a problem when tenants run heavy workloads simultaneously. When this happens, the VMs and associated applications attempt to use more resources than are physically available. As VMs compete for resources, the result is slow and inconsistent performance, or the "noisy neighbor problem" where activities of neighboring tenants impact the performance of others.

Performance penalty

Hypervisors have a measurable impact on performance even when resources are not oversubscribed. While the code in today's top hypervisors is highly tuned, even the best hypervisors create processing overhead. For example, every time a hypervisor shifts control of physical-server resources from one VM to another, it must execute instructions to clear CPU registers and caches. Otherwise, the next VM to run could access information left behind by the previous VM. Executing these instructions wastes CPU cycles that could otherwise be utilized by a VM. Since control is transferred from one VM to another many times per second, this overhead adds up quickly. Hypervisors enable VMs to share other system resources beyond the CPU, including the system bus, memory bus, and input/output (I/O) controller. Since these resources are shared, individual VMs must take turns using them. If an application on a VM is able to fully consume one of these resources —an I/O intensive application that may constantly access the I/O controller by performing reads and writes, for example — it will be slowed by other VMs sharing these resources. Applications can also become temporarily blocked while waiting for actions on other VMs to complete.

Lack of control and customization

Virtualized clouds abstract away underlying hardware details and present users with a simplified catalog of compute, network, and storage services. Most providers offer a relatively small number of pre-

configured virtual-machine images. Some images are better for compute-intensive applications while others are better for I/O-intensive applications. With so many possible combinations of image parameters, users often end up with under and/or over-provisioned CPU, RAM, and I/O resources, so they must either pay for more than they need or risk performance and capacity issues. Virtualized clouds with pre-configured images may simplify the configuration process, but they also remove the ability to customize the underlying hardware to meet specific application requirements. For example, some workloads may require a GPU-powered server, direct-attach SSDs instead of SAN storage, and multiple network cards for higher I/O rates. Unfortunately, virtualized clouds don't support hardware customization.

Rip and replace

Most applications are designed to run with a target infrastructure in mind, and many have not been designed to run in virtualized, multitenant environments. When customers simply rip and replace on-premises infrastructure with a virtualized public cloud, they may have performance, reliability, and security issues. In order to achieve the same level of performance, reliability, scale, and security as on-premises, these applications typically must undergo some amount of redesign or reimplementation. Since this can be both expensive and time consuming, many applications that could benefit from cloud attributes such as pay-per-use are left behind in on-premises data centers.

Governance and risk management

Cloud providers implement and enforce security in a variety of ways, and it's important to be aware of the distinction between perimeter security and secure multitenancy. Providers establish hardened perimeter security in the interest of protecting customers from external threats. In some cases, perimeter security can be comparable or even better than what IT organizations have onsite.

Secure multitenancy differs in that it involves isolating and securing tenants against threats from within the environment. In a virtualized cloud, hypervisors enable most of the secure multitenancy on the host server. While hypervisors are generally considered secure, any vulnerability to a hypervisor represents a vulnerability to all VMs on the host. Bare-metal clouds avoid this risk by not allowing multiple tenants on the same server. Regardless of the potential for a hypervisor vulnerability, businesses in industries with data isolation requirements can risk breaking regulatory compliance if they host certain applications or data on virtualized, multitenant servers.

Bare metal overcomes public cloud limitations

Bare-metal clouds deliver the full benefits of cloud computing with the additional advantages of dedicated servers. They also overcome the challenges found in virtualized clouds.

A rare breed of cloud

Authentic bare metal clouds are hard to find. Most public clouds are VM-based environments where physical resources are divided and shared among multiple tenants. Some virtualized clouds claim the benefits of bare metal, but most of them simply keep tenants from sharing the same physical server. These "dedicated virtual servers" are not bare metal servers. The hypervisor still exists, along with the performance penalty, and hardware customization is not an option. Some public cloud providers do offer bare metal servers, but most fail to provide the same level of API support as their virtualized cloud services.

The benefits of bare metal

Bare-metal clouds can deliver significant improvements over virtualized clouds. For example, bare metal servers achieve much greater performance consistency than shared servers since all resources are dedicated to a single tenant. They can also be configured specifically for applications that have unique hardware requirements, such as an extremely large amount of RAM. By right sizing bare metal instances, users can achieve higher rates of utilization that lead to greater cost efficiency.

Uncompromised performance

Bare-metal servers do not require a hypervisor layer since resources are not being partitioned among multiple tenants. Therefore, each server's entire processing power is available to applications, resulting in better performance than a comparable virtual server.

Bare-metal clouds can deliver roughly four times the performance levels of traditional VM-based clouds. Without multiple tenants competing for shared resources, high-I/O workloads, such as large database applications, can run at peak performance. Bare metal is also highly consistent since resources are not oversubscribed among tenants.

Peace of mind

Dedicated physical servers provide complete isolation and protection of an organization's applications and data from both outside intrusions and other tenants using the same provider. Users get assurance that workloads will run consistently and that provisioned resources will always be available. They can also

implement custom security configurations, such as additional hardware-based firewalls, to give applications the specific protection they require while still benefiting from the provider's hardened security perimeter. Bare-metal clouds also help IT organizations meet essential compliance checks in industries that require data isolation without sacrificing the benefits of the public cloud.

Complete customization

Customization is essential if customers are to deliver high rates of utilization while ensuring the ongoing performance and availability of their workloads. Bare-metal clouds allow users to custom-design the underlying hardware to meet the specific needs of their applications. This helps maximize utilization since no resources are wasted.

Bare metal can be the optimal solution for IT organizations that want to move on-premises systems to the cloud. Unlike VM-based clouds, users can take advantage of bare metal to mirror their on-premises environment. This lets them effectively shift applications to the cloud while avoiding the risks of ripping and replacing.

No cloud is perfect

There are no perfect technologies in IT or in cloud computing. Different applications will have different requirements. Some applications are simply best suited to an on-premises environment, and others do not require the additional benefits of bare metal.

IT organizations should only consider deploying applications that gain tangible benefits from running on bare metal servers. For example, applications and workloads that require direct access to physical hardware (e.g., databases and calculation-intensive applications) gain additional performance benefits from running on bare metal. Workloads that should not be virtualized are also good candidates since bare metal servers can be configured to match an application's existing environment.

Bare metal can provide additional benefits for application tiers that rely on physical hardware for consistent performance, security, or isolation. For example, a user might deploy a database tier on bare metal to receive performance and isolation benefits while still running the full application in a public cloud environment. Some of these workloads and application tiers need to scale up and down quickly with fluctuations in demand. Bare-metal cloud allows users to provision additional resource capacity for their dedicated servers in a matter of minutes, delivering the agility and flexibility of cloud with the performance and reliability of dedicated hardware.

Bare-metal cloud use cases

There are many real-world use cases in which bare metal clouds provide significant advantages over virtualized clouds.

Migrate existing applications to the cloud

Many IT organizations are still trying to determine how best to move their existing applications to the cloud. Migration teams can use bare metal services to mirror their on-premises environment and migrate applications with ease. Granular customization helps them meet application dependencies and gain benefits such as scalability, flexibility, and agility. Migration teams can avoid the costs, complexity, and risks associated with ripping and replacing application infrastructure, and can eliminate the time-consuming process of re-architecting applications for a public cloud environment.

Multitier applications with virtual and bare metal components

Multitier applications can have diverse requirements that necessitate the use of both bare metal and virtual infrastructure. For example, database tiers typically require the high performance and complete isolation of bare metal. Other components, such as web tiers, will instead benefit from the agility and elasticity of a virtualized environment. Users can achieve an ideal blend of benefits by deploying the right application tiers to the right form of infrastructure.

Performance-intensive applications that require consistency

Noisy neighbors can significantly impact performance in virtualized environments, particularly if server hardware has been oversubscribed among tenants. Applications that require a high degree of performance consistency, such as applications that have not been designed to scale in and out according to fluctuations in demand, can also benefit from bare metal clouds.

Key takeaways

In today's competitive market, IT budgets are constantly under the microscope. Businesses are eager to graduate to more attractive products and services that give them fundamental advantages in IT. They also expect the IT organization to assume a more significant role in driving cost efficiency and delivering value to the business. Public clouds offer a number of benefits that are hard to ignore, including agility, flexibility, and cost savings.

Virtualized clouds tend to be emphasized for their convenience, widespread popularity, and ease-of-use. However, virtualized clouds are just one option to consider. It is important for IT leaders not to overlook the significant advantages of bare metal. Bare metal clouds provide all the expected benefits of cloud computing while going beyond the inherent limitations of virtualized clouds, such as resource contention and hypervisor performance penalties.

IT leaders must be selective when choosing a bare metal cloud. Like virtualized clouds, there are differences among bare metal clouds. First and foremost, prospective customers should look for full automation and control through an extensive API, not just the existence of bare metal. They should also determine response times for setting up new custom bare metal configurations, and know how long changes and additions will take. IT organizations gain greater value when they combine both bare metal and virtualized servers in the same environment, so management and connectivity should be seamless across both types of services.

Bare-metal clouds create important opportunities for IT. Many of the challenges cloud users face today, such as poor and inconsistent performance, lack of control and customization, and complex risk management, can be resolved with the adoption of bare metal clouds. Bare metal provides the granular customization and consistent performance users need to maximize resource utilization and deliver exceptional quality of service to end users.

With the combination of bare metal and virtualized public cloud services, IT organizations can achieve unprecedented efficiency, performance, and flexibility, all while delivering the competitive value that businesses demand.

About Paul Burns

Paul Burns is a Gigaom Research Analyst and the president and founder of Neovise, an IT industry analyst firm launched in 2009 to focus on cloud computing. Burns also writes articles for industry publications, speaks at industry events, and is quoted by a variety of media organizations.

Burns has over 25 years of experience in the IT industry, driving strategy for enterprise software solutions through product management, competitive analysis, and business planning. He has held a series of leadership positions in marketing and R&D, and he was research director of the IT service management practice at another industry analyst firm. Burns earned both a B.S. in Computer Science and an M.B.A. from Colorado State University.

About Gigaom Research

Gigaom Research gives you insider access to expert industry insights on emerging markets. Focused on delivering highly relevant and timely research to the people who need it most, our analysis, reports, and original research come from the most respected voices in the industry. Whether you're beginning to learn about a new market or are an industry insider, Gigaom Research addresses the need for relevant, illuminating insights into the industry's most dynamic markets.

Visit us at: research.gigaom.com.