HOSTING
TAKING YOU FURTHER

SCALING DATABASE APPLICATIONS IN THE CLOUD FOR PEAK PRICE AND PERFORMANCE

Software developers and Software-as-a-Service (SaaS) providers, who have long benefited from the substantial economies of scale available in the cloud, now have a wealth of options for configuring public, private and hybrid cloud infrastructures. Indeed, the permutations and combinations of the many options available can make it difficult to configure a cloud for peak price and performance.

Achieving satisfactory performance at scale while preserving the many cost-saving advantages of the cloud can be even more difficult in a multi-tenant environment. This is the case at the application level for a SaaS provider, as well as at the infrastructure level for a cloud service provider (CSP).

Perhaps the most difficult of all applications to optimize for peak price and performance are those that utilize a database. The reason these applications present such a challenge is not technical; it is instead the way database licensing arrangements often make high-performance configurations very expensive—sometimes prohibitively so.

This white paper can help software developers and SaaS providers maximize price and performance even in this most challenging of use cases: a database application for a multi-tenant SaaS offering that is running in a hosted cloud environment.

SCALE OUT VS. SCALE UP

At the risk of over-simplification, there are two fundamental options for scaling performance in a cloud infrastructure: out and up. Scaling out performance involves adding more processing power by adding more virtual machines (VMs), more processor sockets or cores, and/or more servers.

A major problem with scaling out performance for almost all applications is the need for additional software licenses. This is obviously not the case for those applications that utilize only open source software, which is why hyper data center operators normally use Linux and Xen, and often write all or most of their own application software. Scaling out does become inevitable at some point, of course, but should be used only after exhausting all of the more cost-effective scale up options.

Scaling up might or might not involve additional licensing fees (although virtualization has made it increasingly common for commercial software). One notorious exception to the "free" scale up option is databases, which normally have licensing based on the number of VMs, sockets and/or cores.

There are, however, ways to scale up application performance without incurring an additional database license fee. Increasing processor speed, for example, can have a beneficial effect on most applications. But this popular technique offers virtually no improvement in transaction-intensive database applications where the usual bottleneck is storage I/O.

Adding more memory for caching can improve performance for most applications, but this also often delivers only a marginal improvement with the "I/O blender" created on virtualized servers owing to the orders of magnitude difference in scale (i.e., Gigabytes of RAM caching Terabytes of storage). With its Terabytes of capacity, however, flash cache can address the I/O blender problem, including for transaction-intensive database applications, making this a viable option.

The use of direct-attached storage is often considered as yet another way to improve performance for most applications. But because such configurations undermine scalability and the cost-saving advantages that derive from using shared storage, direct-attached storage might not offer a satisfactory improvement in overall price and performance.

There is one other "license-free" way to boost performance, including for databases, and that involves addressing the fundamental cause of the I/O bottleneck directly.

SCALING UP STORAGE I/O

The hard disk drive (HDD) has served the information technology industry well for over 50 years, but its I/O throughput is limited by the very nature of the mechanical design required for magnetic media. The primary cause is the rotational latency of a spinning disk platter, and while this is not a problem for sequential reads and writes, it has a significant adverse impact on performance for the random access needed in database applications. So even the fastest-spinning HDDs offer only a modest improvement in database I/O operations per second (IOPS).

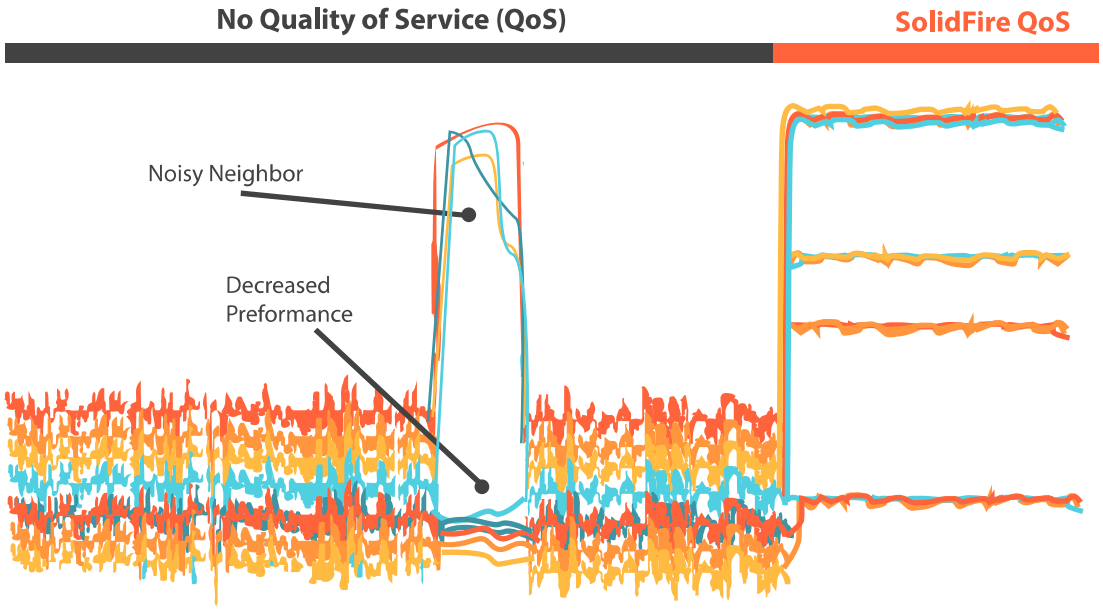
IOPS can be improved dramatically, however, by using a different storage medium: the flash memory used in solid state drives (SSDs). The difference between HDD and SSD performance is quite profound. For HDDs spinning at 7200 or 10,000 RPM, IOPS is typically in the range of 100-200. For an SSD, IOPS can increase by up to four orders of magnitude to over 1,000,000. This "license-free" option becomes an obvious choice for improving price and performance with I/O-intensive database applications.

The ability to deliver dramatic improvements in performance has led to a growing number of SSD and all-flash storage array solutions becoming available. While a comparison of all of these different solutions is beyond the scope of this discussion, there is one capability that does warrant consideration here: support for Quality of Service (QoS) to assure performance in a multi-tenant environment where "noisy neighbors" can undermine the potential for improvement.

According to Henry Baltazar, Senior Analyst of Infrastructure & Operations at Forrester Research, "'Noisy neighbors disrupt cloud and enterprise storage delivery… This limitation is a major reason why it is so difficult for cloud storage providers to create consistent multi-tenant environments using traditional storage systems which lack Storage QoS." This is why Baltazar calls storage QoS a "must-have" capability in a multi-tenant cloud.

Storage QoS works in ways that are similar to the ways network QoS works, but some work better than others. The best solutions go beyond simple rate-limiting and "best effort" prioritization mechanisms because these are unable to assure predictable and consistent levels of performance under all circumstances. The only way to ensure that business-critical operations always get the performance they require is to take a Min/Max/Burst approach to managing IOPS:

- Min IOPS is needed to ensure every volume gets a guaranteed minimum level of performance regardless of system conditions or application activity.
- Max IOPS is needed to set a maximum level of performance that each volume receives over time as a means to enforce fairness among all volumes.
- Burst IOPS is needed to accommodate the occasional spikes in demand that occur in virtually all database applications.



Storage QoS enables applications to be prioritized, as shown here in a SolidFire all-flash storage array with a 4-level hierarchy, to ensure that each gets the performance it needs at all times under all circumstances.

The ability to effectively manage IOPS with a Min/Max/Burst approach to QoS requires some underlying capabilities in the storage array, including:

- the use of only SSDs (to ensure consistently low latency)
- a true scale-out architecture (to assure predictable performance gains as capacity grows)
- RAID-less data protection (to ensure predictable performance under any failure condition)
- a balanced load distribution (to eliminate "hot spots" that can create unpredictable I/O latency)

"Storage QoS is a must-have feature for the cloud."

— Henry Baltazar, Senior Analyst of Infrastructure & Operations at Forrester Research

SCALING UP PERFORMANCE WITH A CLOUD SERVICE PROVIDER

In their public, private and hybrid configurations, cloud service providers normally offer the same scale out and scale up options that are available in an on-premises private cloud. But only a CSP can achieve the economies of scale needed to be able to produce the same or greater level of high-availability performance at the lowest possible cost.

Not all CSPs can accommodate the demands of a SaaS provider with a multi-tenant database application, however. For example, some CSPs offer great software development tools but lack the architectural strength to handle a complex, large-scale SaaS application. Others might lack the expertise needed to ensure SaaS applications remain up and running despite the inevitable system failures, ongoing security threats and potential for disasters. And others simply lack the ability to support the latest Dev/Ops best practices.

So in addition to usual criteria used to evaluate CSPs (e.g., high availability and disaster recovery backed by service level agreements, robust security, scalable capacity, responsive support, etc.), SaaS providers should use a CSP that also offers:

- Storage QoS to ensure that the IOPS performance gains will be predictable and consistent over time under changing configurations and conditions
- A development/test environment to characterize SSD vs. HDD performance under varying loads and circumstances
- Storage encryption that does not compromise performance to help ensure appropriate controls over security or privacy
- Support for hybrid architectures that combine the best of public clouds for cost-effective scalability, private clouds for security, and dedicated servers and/or colocation for special needs (including database clusters)

CONCLUSION

Achieving peak price and performance for database applications requires SaaS providers and software Dev/Ops teams pay equal attention to both price and performance. The price portion of the equation can depend just as much on licensing fees as it does on the choice of technology used to improve performance.

Of the two fundamental options used to scale performance, only scaling up can be done in a way that avoids additional licensing fees, and for I/O-intensive database applications, the best improvements in price and performance normally come from the use of solid state storage. But even with all-flash storage arrays, the performance gains in a multi-tenant SaaS environment are often not predictable or consistent without some means of managing the IOPS effectively.

To learn more about how you can get peak price and performance for your database applications by using all-flash storage arrays with Min/Max/Burst QoS management, please visit HOSTING at www.**HOSTING**.com or call (888-894-4678).

SOLIDFIRE ALL-FLASH STORAGE AT HOSTING

The HOSTING SolidFire storage solution with QoS gives customers the ability to guarantee IOPS for their applications on a 100% solid state drive platform. Performance and capacity can be scaled dynamically, as needed and without disruption to the application. Traditional RAID technologies are replaced with a robust hashing algorithm that does not sacrifice performance under both normal and rebuild conditions.

Performance with the SolidFire all-flash storage array increases four-fold when compared to traditional mechanical drive based arrays, and at a cost that is comparable to the HOSTING Tier 0 HDD storage offering. This dramatic increase in performance is what enables HOSTING to offer an always-on, QoS-backed service level agreement for business-critical applications.