WHITEPAPER

Flip the Switch: A Storage Architect's Guide to Accelerating Database Migration to the Public Cloud

The Register®

EXECUTIVE SUMMARY

SQL and NoSQL databases such as those offered by Oracle, MongoDB, PostgreSQL and others – and the high-performance storage infrastructure they require – sit at the heart of the world's biggest enterprises and organizations.

Storage architects and database administrators (DBAs) have amassed years of expertise building and tuning high performance, resilient on-premises systems to keep them working at optimal levels to support storage-intensive database workloads.

At the same time, a challenging economy and a punishing pace of technical evolution mean that virtually every organization on the planet is considering whether to move some – or all – of their operations to the cloud, in the hope of driving down costs and accelerating innovation.

Migrating applications or development pipelines to the cloud can be a relatively straightforward process. But replicating in the cloud the performance and density of the underlying storage infrastructure that SQL databases require is a much more complicated challenge.

In this white paper, we will detail the context and rationale for those high performance SQL database cloud migrations, and explain the pain points and challenges around database storage infrastructure that can frustrate or delay them – or prevent them from ever getting off the drawing board in the first place.

The paper will also illustrate how the unique architecture of Lightbits disaggregated cloud storage can deliver similar performance to local flash on the public cloud — enabling a high-performance, low-latency storage target for Amazon Web Services Cloud (AWS Cloud) and Microsoft Azure instances that allows DBAs to meet or exceed their SLAs.

Ultimately, Lightbits is able to remove the storage bottlenecks that can prevent enterprises from taking their existing storage intensive applications to the cloud, and leave them better positioned to innovate going forward.

MARKET CONTEXT

The early to mid-2020s could go down as one of the most challenging eras for organizations yet. The Covid-19 pandemic forced the pace of digitization and digital transformation. This was followed by a booming economy combined with supply chain shocks. A deteriorating geopolitical situation created economic instability and fueled the use of cyberattacks to sow further chaos.

The upshot has been a global tightening of economic policy, which has created general volatility and a direct effect on organizations' IT infrastructure investment plans.

Cloud Pressure

All of these factors have contributed to further complicating what has always been a rapidly changing technology landscape. In addition to maintaining existing workloads, organizations are under pressure to deliver innovation and gain a competitive edge, particularly by leveraging the data they collect, store, and process to feed cutting-edge analytics systems. This is coupled together with the desire to move IT infrastructure expenses from the CapEx to the OpEx budget column.

This budgetary pressure can push organizations towards the purported benefits of the cloud, with its promises of configurability, scalability, and pay-per-use economics.

This is why, according to Gartner, "Cloud computing will continue to be a bastion of safety and innovation, supporting growth during uncertain times due to its agile, elastic and scalable nature." Specifically, cloud application infrastructure services will show the biggest rate of growth in 2023.

Amidst all this flux, there's one thing that doesn't change. At the center of most enterprises sits a database - often SQL - that supports both general enterprise workloads and the cutting-edge applications that are associated with full-scale digital transformation.

The software and hardware infrastructure, and the expertise, supporting an organization's SQL databases have been built up over many years. Running an Oracle database for example requires extremely high performance and resiliency - in terms of bandwidth, latency, and raw IOPS. The hardware that enterprises rely on to run SQL databases on-premises will be precisely tuned towards specific workloads.

Given seasonality, some capacity might be underutilized for at least part of the year if the organization is to ensure that it is prepared for the surges it knows it will face in peak periods. And data, being data, will continue to grow, meaning that over time the organization will need more storage, more compute, and more bandwidth if it is to keep delivering their required Service Level Agreements (SLAs).

This can be necessarily expensive and hard to reconcile with flat or shrinking IT budgets.

But the fact is that rehosting SQL workloads in the cloud is not as easy as it might seem. Applications might need to be refactored if they are to run on native cloud storage platforms. Even if it is possible technically, it might not be feasible financially, at least not with native cloud storage solutions.

In the following sections, we'll examine six specific pain points that SQL DBAs and their storage architect counterparts must confront if they want to successfully migrate high-speed database workloads to public clouds - whether Amazon AWS or Microsoft Azure.

PAIN POINTS ON THE WAY TO THE CLOUD

Organizations have been migrating workloads into the cloud, whether on a SaaS, IaaS, or PaaS basis, for well over a decade. Many younger organizations have even lived their entire existence in the cloud.

But there are a number of technical issues, particularly around storage provisioning, that can cause enterprise-scale SQL migrations to fail. Even if storage architects find ways to overcome these challenges, there are forbidding cost implications.

Cloud Performance Limits

As we've discussed, performance is almost everything when it comes to running SQL databases like Oracle or mySQL and Postgres. Providing the sort of capacity needed to support critical workloads requires a delicate orchestration of compute power, bandwidth, and - crucially - blisteringly fast and resilient block-based storage.

High-performance databases can demand IOPS measured in the hundreds of thousands per volume and latencies in microseconds. It can be difficult if not impossible to achieve such performance requirements using native cloud storage solutions without breaking your budget. Such limitations around latency and IOPS have been the main barrier to migrating storage intensive workloads like Oracle to the cloud for quite some time now.

Performance Comes at a Cost

Low-tier, cloud-native storage offerings may be cost-efficient, but they might not provide the high IOPS and low consistent latency required to power SQL databases at scale in the cloud. This forces DBAs or Data Engineers to architect solutions that could end up having high and unpredictable costs on the public cloud.

In the battle to reach sufficient performance, Storage Architects and DBAs have to examine Amazon AWS' and Microsoft Azure's top tier, and costly, block storage offerings. This may end up being more expensive than on-premise environments.

On top of this, customers are not only charged for the capacity. Beyond a basic allowance, they are typically charged for snapshots, provisioned IOPS, and in some cases provisioned throughput. They may also need to increase the number of compute instances running Oracle, requiring more cloud infrastructure and more Oracle licenses.

The Cloud's Limited Flexibility

Theoretically, Storage Architects may be able to craft a cloud infrastructure capable of supporting a Tier One, critical database workload, for example. But they will need a high degree of foresight, as they will have to specify the number of IOPS per volume that they need when provisioning storage. If they go above this threshold, they may face the dilemma of having to overprovision IOPS, even if they don't need the additional capacity.

Accurately predicting business demand at the time of provisioning compute and storage services is difficult. But one misformed, badly-timed, or simply long-running query can suck up resources and throw all calculations out the window.

Last but not least, storage requirements could also change over time, invalidating all of the estimations previously done during the provisioning phase.

Essential Features Become Paid Extras

We've seen how running a database in the cloud can be much more expensive than Storage Architects might predict. But there's another price issue as well.

When deploying storage systems on premises, DBAs typically rely on data services such as data reduction, thin provisioning, and snapshots/clones. With hyperscalers these are typically delivered as separate billable add-ons or are not supported, further driving up potential costs and management overhead.

Be Persistent

Local storage is typically recommended for its high performance and low latency characteristics. However, choosing instance types that leverage directly-attached NVMe comes with some risk of sacrificing persistence.

Local NVMe is short-lived. It lacks data protection, data services, and doesn't scale. If the service is deactivated or if the node fails, all data is gone. One remedy, of course, is to copy or back up the data (preferably across multiple availability zones), but this strategy could be a drain on the IT budget.

A Question of Provisioning

Amazon AWS or Microsoft Azure defaults can be a problem for DBAs in other ways. Three copies of data spread across AZs makes sense for a Tier One application. But what if an organization is also running a service like Oracle Data Guard as well? Or Oracle RAC, which allows a single database to run across multiple servers?

Alternatively, with a Tier Two or Three database, two copies might be sufficient. This can enable savings on capacity overall. Moreover, cloud comparative inflexibility means that customers are under pressure to over-provision capacity,

even if they are not actually going to be using it continuously. This lack of thin provisioning options can present another pain - and cost - point for customers.

It's quite clear then that a number of obstacles often stands in the way of a smooth migration of an SQL database to the cloud, with storage being a particularly tricky element. So, is there a way to address this particular obstacle, freeing up organizations to enjoy the undoubted benefits of the cloud in terms of scalability and flexibility?

LIGHTBITS ARCHITECTURE AND BENEFITS

Lightbits is a data storage platform designed to power high-I/O workloads on the public cloud with the same capabilities typically found in on-premises SANs, such as unified storage provisioning and bulk volume management. It's designed to deliver speed, flexibility, and efficiency on both Amazon AWS and Microsoft Azure.

Lightbits helps IT organizations with a cloud-first strategy to move critical workloads like SQL and mySQL databases to the cloud, with no compromises and while still meeting or exceeding their SLAs. As such, it closes a gap that native cloud solutions cannot - delivering better performance with consistent low latency, enterprise-grade features, high availability with multi-zone capabilities, and hybrid and multi-cloud models in one scheme. And all of this at a lower cost than cloud native storage.

As a software-defined, disaggregated block storage system that works with common orchestration environments such as Kubernetes and VMware, Lightbits plugs directly into Amazon EC2 instances (Figure 1) and/or Azure Virtual Machines (VMs) (Figure 2).

The minimum cluster size is three nodes, but a three-node cluster can deliver 1M IOPs/volume and low <500 µsec avg latency, and it's six times faster than native cloud storage solutions.

The cloud data platform also supports hybrid deployments by offering the flexibility to port the software license between on-premises storage servers and the public cloud.

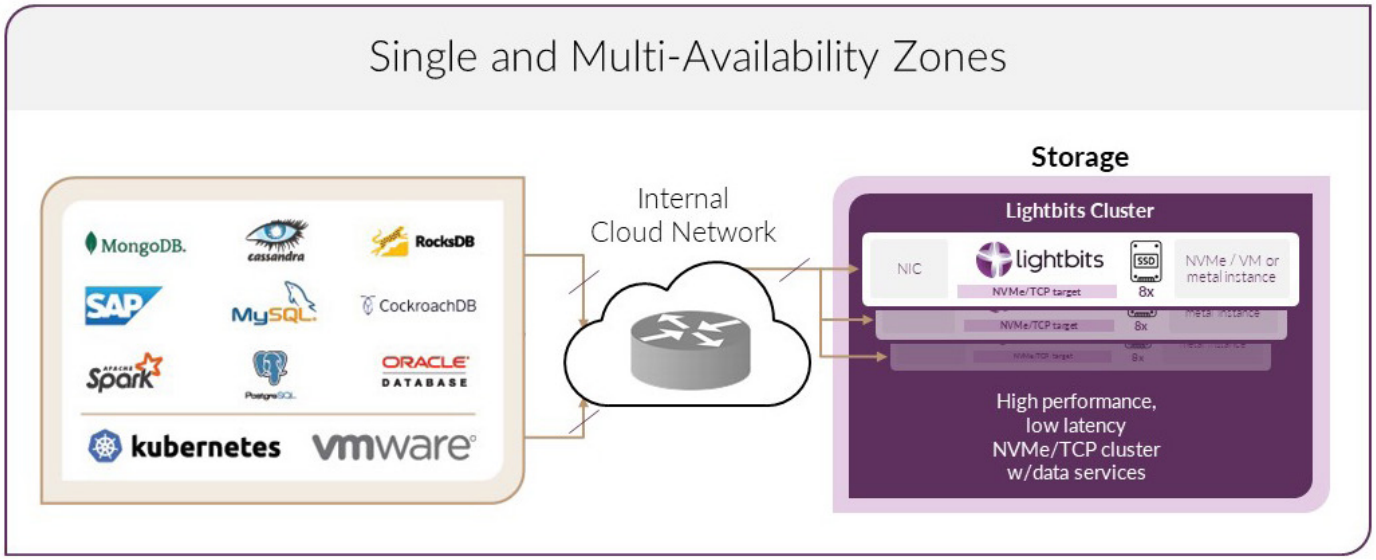

Figure 1: *Lightbits' Elastic, High-Performance Block Storage on Amazon AWS.*

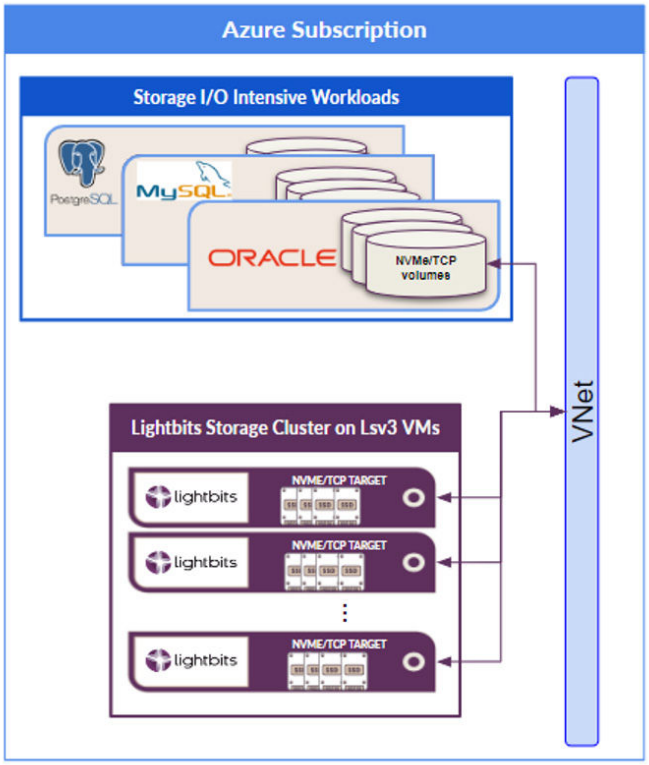

Figure 2: *Lightbits SAN on Microsoft Azure.*

The benefits of Lightbits include:

- **High performance/IOPS with consistent low latency:** Lightbits performance is comparable to the instances with local NVMe devices, and there is no limitation with regard to the IOPS per gigabyte; IOPS can be in a single volume or split across hundreds of volumes.
- **Predictable and lower cost:** Lightbits cost on the public cloud is fixed/predictable.
- **Data services:** With Lightbits, enterprise data services and self-managed capabilities are included in the license: thin provisioning, compression, QoS, snapshots, clones, self-healing, auto-scaling, incremental backups, rolling upgrades, events monitoring, and more.
- **Simple:** DBAs and storage architects only need to know the required capacity to provision the right number of Lightbits storage instances. After that, they just provision the block volumes. It's that simple, and it's future-proofed for any changing business demands.
- **Scalable:** Including disaggregated and linearly-scalable architecture (capacity and performance).
- **Multi-Availability Zone (AZ)**: Cross AZ synchronous replication significantly reduces the risk of unplanned downtime.

PERFORMANCE FIGURES

SLOB (Silly Little Oracle Benchmark) is the de facto method of testing storage with Oracle databases. The benchmark is not targeted at the CPU and memory elements of the infrastructure. Instead, it provides a way of stressing the storage part of the infrastructure, by forcing SQL functions – queries, selects, updates – into using the storage. As such, it can be used to provide a best-case scenario of what can be achieved with various configurations.

Figure 3 on the next page shows the performance benefits of running Oracle databases on Amazon AWS with Lightbits storage compared to io2.bx, the high-end AWS EBS storage option. This test illustrates the best performance results of io2.bx versus Lightbits using Oracle and SLOB (16 schemas) running on R6in.24x instances. It shows the total number of IOPs per database with read-only (0% updates) through to write-only (100% updates). In these tests, the R6in.24x instance using io2.bx maxed out at a consumption rate of an average of 256k IOPS, while the Lightbits numbers were significantly higher.

Performance can be increased by adding clusters or instances – either automatically or through a number of clicks – and varying the number of replicas being used.

The results provide a starting point for seeing what can be achieved in terms of storage efficiency and resulting cost advantages.

We all know that cloud pricing can change. However, previous benchmarks running Oracle on two of AWS' r5n16xlarge instances - with a Lightbits cluster running on three i4i.metal instances - show that the cost per hour to deliver 1.6 million physical reads was $46.10, compared to an equivalent EBS IO2-based configuration, which was $293.59 per hour. These figures did not take into account the potential for reduced Oracle licenses or compression.

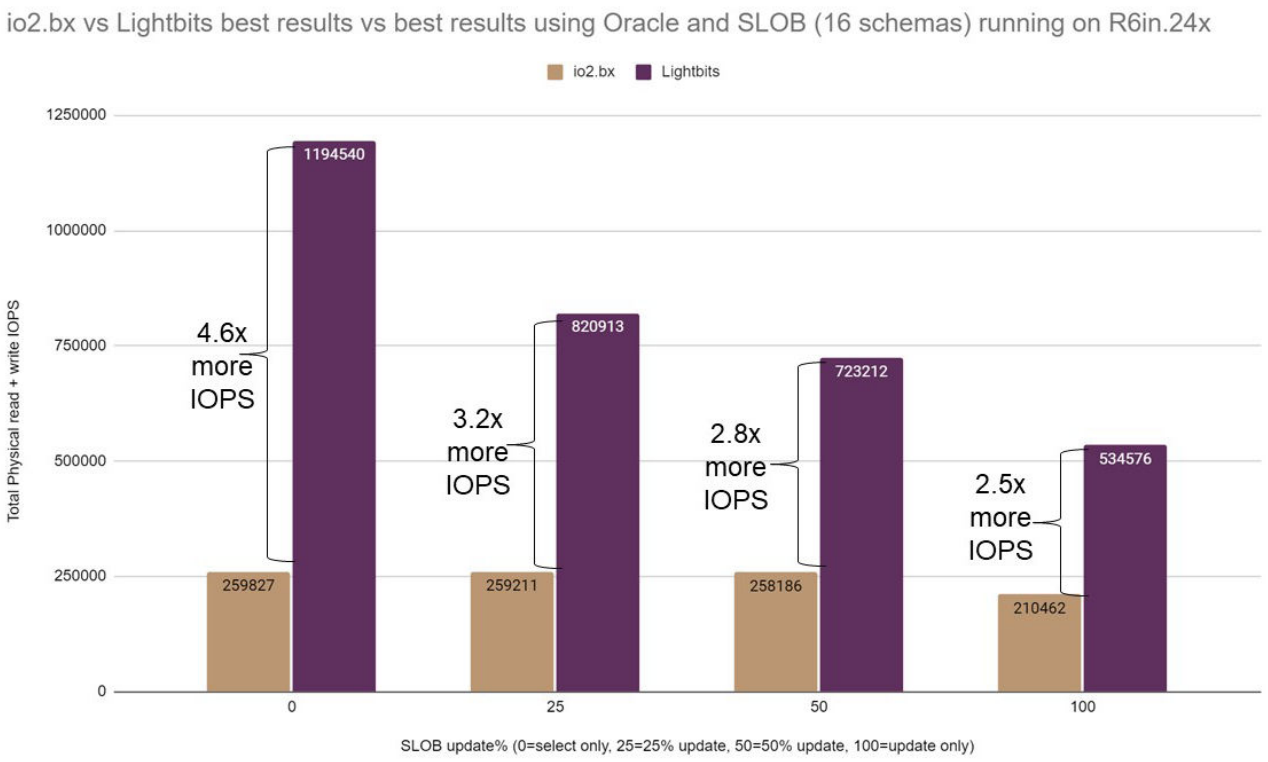

Figure 3: *SLOB Benchmark Performance Comparison of Oracle Using Lightbits on AWS Versus AWS io2.bx*

Figure 4 below shows the performance benefits of Lightbits storage for an Oracle database on Microsoft Azure - compared to Ultra Disk, the fastest cloud native block storage option available today in Azure. The bars correspond to the total number of IOPS measured by SLOB for different update ratios (0%, 25%, 50%, and 100%).

As illustrated below, Lightbits on Microsoft Azure can deliver almost 1M sustained IOPS of 8KB, while Ultra Disk is limited to only 80K IOPS of 8KB.

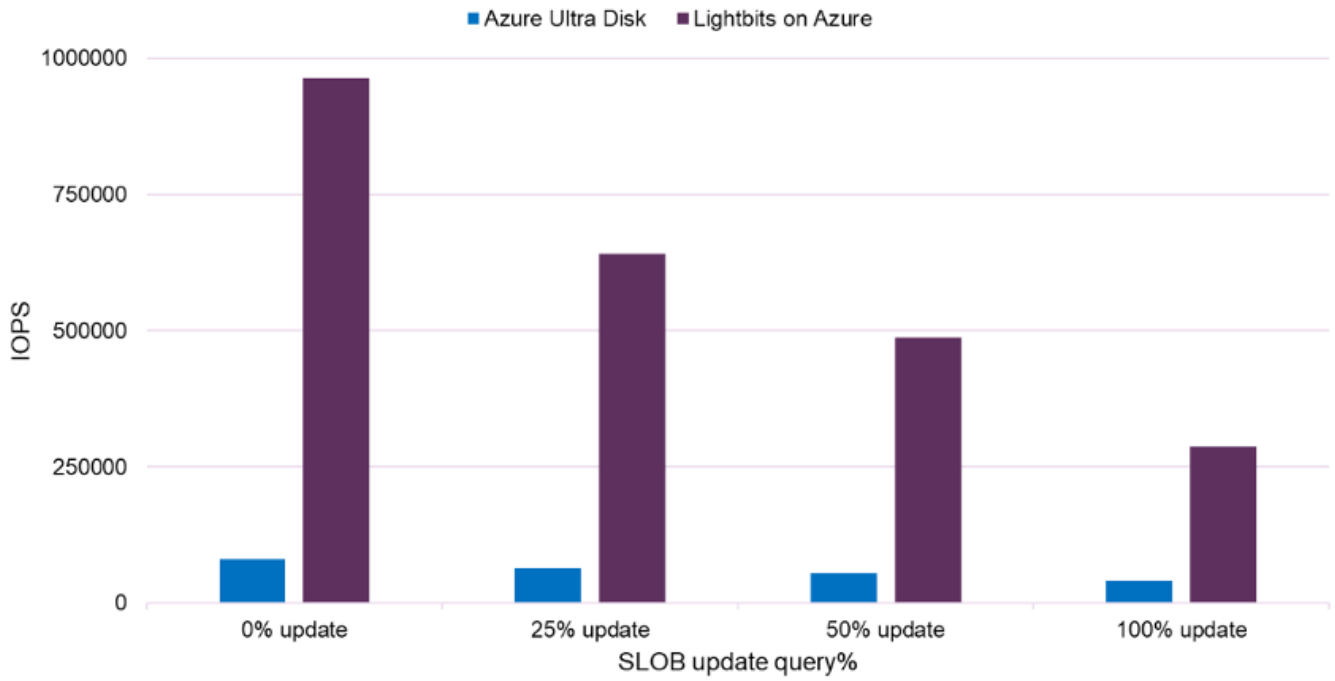

Figure 4: *SLOB Benchmark Performance Comparison of Oracle Using Lightbits on Azure Versus Azure Ultra Disk*

CONCLUSION

As more and more workloads move off-premises and into the cloud, it will be harder to justify why SQL databases such as Oracle should be any different. But cloud providers' managed services, and even traditional bare metal approaches, are not going to be able to offer the performance or flexibility that Tier One SQL database installations need.

Storage architects and DBAs have to find a way to work together to deliver the high performance required by these demanding databases. Implementing disaggregated, software-defined storage that delivers the necessary capacity - with the required flexibility and at a competitive price point - will be one of the most challenging elements of this effort going forward.

Lightbits' solution offers the uncompromising performance, flexibility, and industry standard features that DBAs and storage architects supporting critical workloads expect.

And as the figures above suggest, it can do so in a highly cost-effective way.

Once their database installation has been migrated, DBAs can take full advantage of all of the flexibility that the cloud can offer. This means that the database installation doesn't just remain at the heart of the enterprise, but can become a platform for future innovation and transformation.

To learn more about the Lightbits solution, visit our website at www.lightbitslabs.com.

Additional resources:

- Lightbits on AWS Marketplace
- Lightbits on AWS Product Overview
- Lightbits on AWS Administrator's Guide
- Lightbits on Azure
- Running Oracle with Lightbits

Sponsored by Lightbits.

About Lightbits Labs™

Lightbits is a robust data storage platform for any cloud with unmatched speed, flexibility, and efficiency. It facilitates organizations that are migrating to the cloud to shift IO-intensive workloads without any compromises while still meeting or exceeding their SLAs. It fills a gap that native cloud solutions cannot deliver for these workloads delivering the best and highest performance with consistently low latency, enterprise-grade features, high availability with multi-zone capability, hybrid and multi-cloud models in one scheme, and predictable and lower costs. Built from the ground up to deliver SAN capabilities for the public cloud era.