Lightbits Cloud Data Platform

Migrating Storage-Intensive Workloads to the Cloud

Table of Contents

Abstract

Enterprises have migrated some – but not all – of their workloads to public cloud platforms. In most cases, the workloads that have migrated to the cloud are non-critical workloads with predictable input/output (I/O) requirements. Factors holding them back from migrating their critical, highly variable workloads with spiking I/O density to the cloud include:

- Unpredictable costs
- Inadequate performance
- Lack of flexibility

The true promises of the cloud in terms of lower costs, simplicity, efficiency, and agility have not been achievable with cloud-native storage alone. The Lightbits Cloud Data Platform fills a void that native cloud solutions cannot deliver for storage-intensive workloads — by delivering the 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. The Lightbits solution is built from the ground up, to deliver SAN capabilities for the evolving public cloud era.

The scope of this paper will not include performance comparisons. For more detailed information on how Lightbits performance compares to native cloud storage, see the following papers:

- Running Oracle on AWS with Lightbits
- A Storage Architect's Guide to Accelerating Database Migration to the Public Cloud

Introduction

By now, most enterprises have successfully migrated their non-critical workloads to the public cloud, but have held back on making the switch for their critical workloads. Typically it's because these workloads have high and variable I/O requirements. While native cloud storage solutions, Amazon Web Services Elastic Block Storage (AWS EBS), and Microsoft Azure managed disks are excellent storage solutions for less storage-intensive workloads, they have limitations for supporting high-I/O workloads.

Inadequate performance, complexity in provisioning services, and unpredictable costs all contribute to the obstacles impeding migration. In most cases, enterprises will cite efficiency and cost as the biggest challenges to making the switch to the cloud. It's too expensive to run I/O-intensive workloads at scale, and they don't want to give up everything they love about their Storage Attached Network (SAN) on the cloud.

The Lightbits cloud data platform overcomes the limitations of native cloud storage by leveraging local NVMe drives provided by AWS EC2 instances and Azure VMs. It delivers high performance, predictable and lower costs, and enterprise-grade features, enabling seamless migration of high-I/O workloads to the cloud. The solution delivers the features you would expect from an on-premises Storage Attached Network (SAN) while leveraging the auto-scaling functionality of the cloud to minimize excess compute usage and costs. It also offers the ultimate flexibility in deployment methods and delivers on the true promises of the cloud: simplicity, lower costs, lower risk, better efficiency, and more flexibility.

Today's Storage Challenges

Lightbits conducted a survey of storage professionals in 2023 to gauge their experiences and expectations for cloud storage. The individuals surveyed run diverse storage workloads (Figure 1), with 23% supporting Apache Spark databases, 28% supporting MongoDB, and 55% supporting MySQL. An additional 38% run storage for Oracle databases. 60% store their data on the public cloud, while 40% rely on SANs, and 39% use Network Attached Storage (NAS), as shown in Figure 2. 52% use both AWS and Azure (Figure 3).

Figure 1: Cloud Survey - Supported Workloads

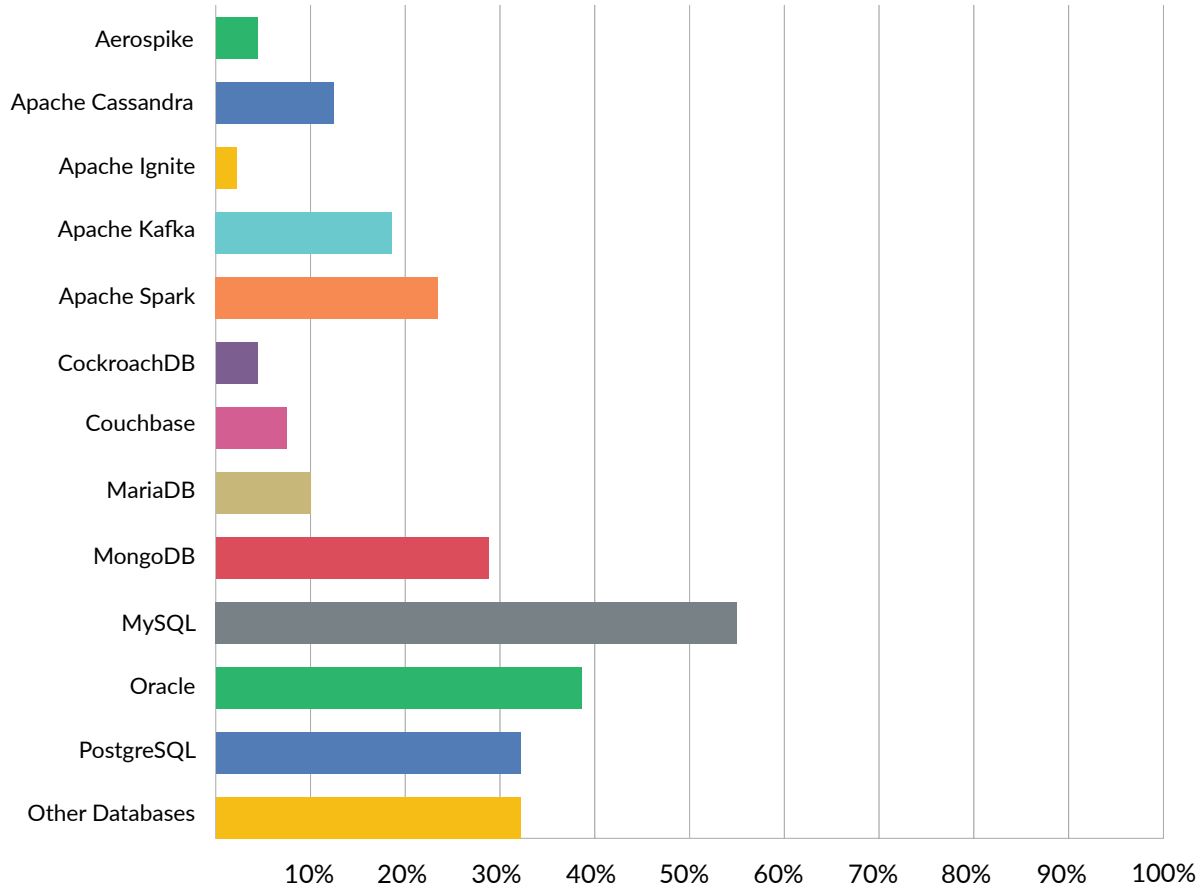


Figure 2: Cloud Survey - Data Storage

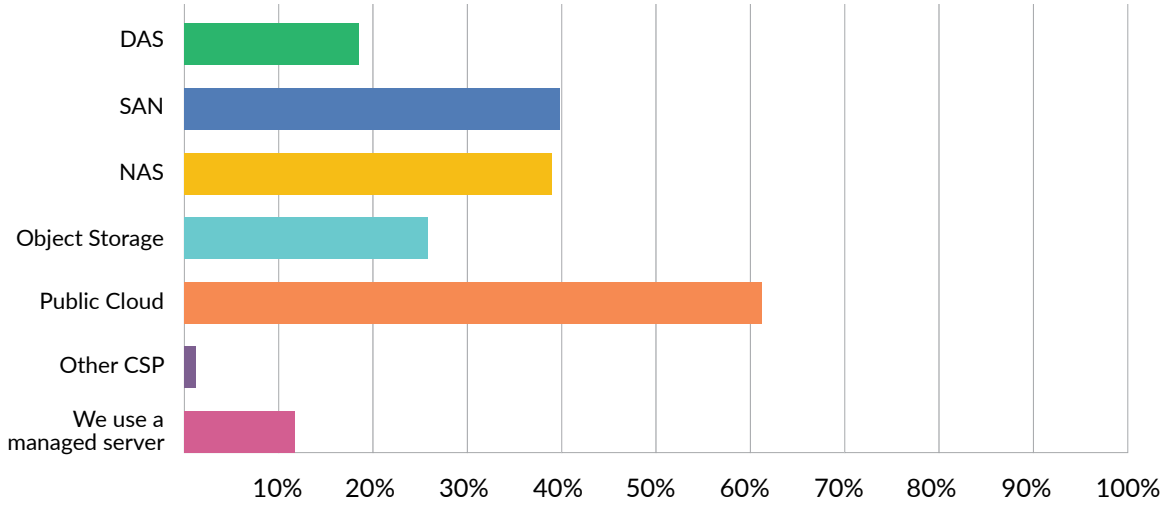
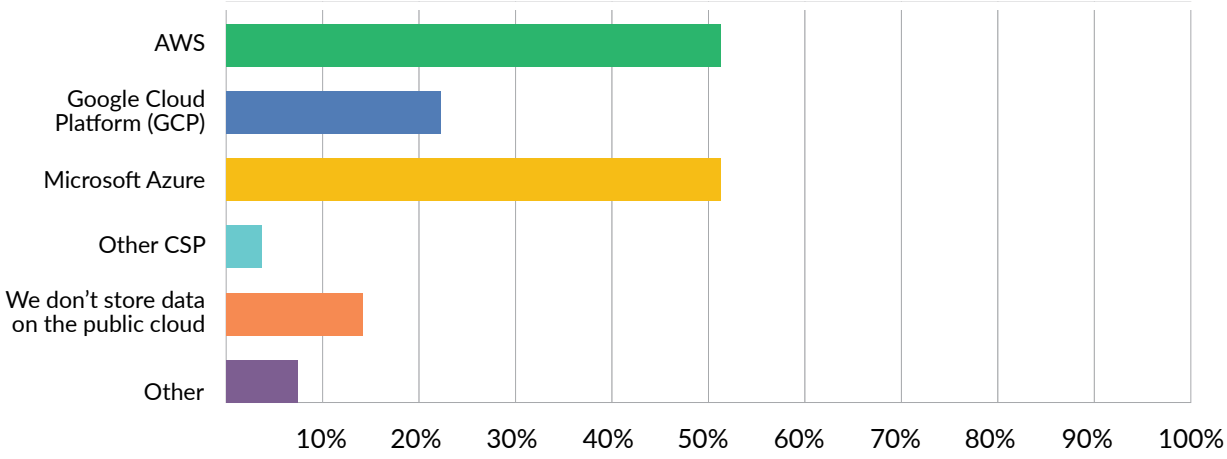


Figure 3: Cloud Survey - Public Cloud Data Storage

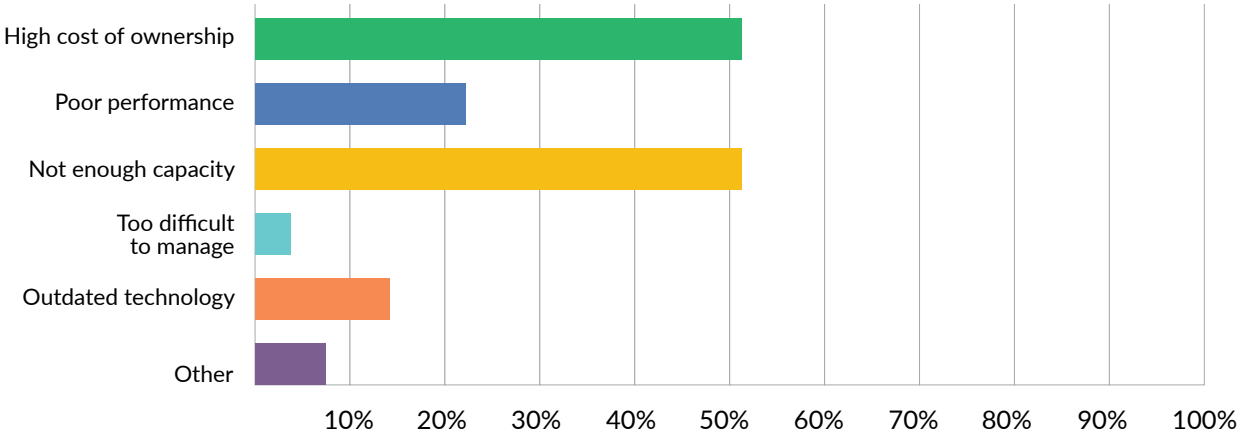


When asked, "What do you consider to be your biggest storage challenge today?," 49% cited "High cost of ownership." A further 28% said "Poor performance," and 20% replied "Not enough capacity." This latter finding is quite significant, as it likely relates to a common problem of on-premises storage procurement and provisioning.

Storage managers typically overprovision storage resources to accommodate unpredictability in business demand. If they estimate or calculate incorrectly too far in either direction, it can cost the business too much money or introduce risk. If the storage is overprovisioned, the company has spent too much money and resources sit idle. Conversely, if storage is underprovisioned, compute jobs and workloads may fail to run, causing a poor digital experience for customers.

In Figure 4 below, we see that 20% of those surveyed claimed that their storage was "Too difficult to manage/too complex," and 23% shared that they suffered from "Outdated technology."

Figure 4: Cloud Storage Challenges



Benefits and Challenges for Storage in the Public Cloud

The public cloud, of course, does have its appeal. The cloud is infinitely scalable, organizations can move IT expenses from the CapEx column to the OpEx, and minimize data center management overhead. The need for specialized storage managers is also lower than what is required to run complex storage technologies on-premises.

However, the public cloud's storage advantages are not endless. 26% of survey respondents answered, "Too expensive," when asked, "If you don't store data on the public cloud, why?" 32% indicated, "We won't get the same performance as we do with our on-premises systems."

There are many challenges for storage in the public cloud, mostly arising from several foundational factors in the way public cloud platforms design and offer storage solutions. Public cloud storage originated as a service to complement infrastructure-as-a-service (IaaS) and platform-as-a-service (PaaS) offerings. It tended to be architected for many general purpose use cases so it can serve the masses, not for critical workloads that have highly variable I/O requirements.

Unpredictable costs comprise another issue affecting the adoption of public cloud storage as well. Customers often find themselves dealing with unwanted extra charges for data services such as snapshots and clones. As a result, they may have to provision and pay for I/O operations per second (IOPS) – even if they don't fully utilize it. Or, they may have to provision storage on a per-compute instance basis, which leads to overpaying for cloud storage that they don't use.

On top of that, they have to calculate their I/O requirements when the services are provisioned and for each individual workload. This, therefore, makes the price/performance ratio more complex with cloud-native storage. The highest-performing storage on AWS, io2 Block Express (io2.bx) and Ultra Disk on Azure, both come at a premium cost and may still not deliver enough IOPS to support many critical workloads. Achieving the required IOPS for many critical workloads would require provisioning more instances on either platform, which results in higher costs.

We all know that database administrators (DBAs) love their SAN, and they want to replicate the SAN characteristics in the cloud. But public cloud storage either lacks SAN features such as advanced clustering schemes, snapshots, clones, thin provisioning, and replication — or the capabilities are available but come at an additional cost and require additional oversight to manage.

It can therefore be quite complex to architect and manage a storage system using native cloud storage that is high-performing, multi-tenant, resilient, and highly available.

Lightbits: Overcoming the Limitations of Native Cloud Storage

The Lightbits Cloud Data Platform offers a solution for organizations that want to overcome the limitations of the public cloud for storage-intensive workloads. Lightbits offers a software-defined, disaggregated block storage system that combines speed, flexibility, and efficiency. Lightbits works with common orchestration environments, including Kubernetes, and plugs directly into Amazon EC2 instances and Azure VMs.

Lightbits enables organizations to migrate their critical workloads with high-I/O requirements to the cloud. It also allows for shifting I/O-intensive workloads without any compromises — enabling DBAs and others to meet or exceed their service level agreements (SLAs). Lightbits enables SAN capabilities in the public cloud, with unified storage provisioning and bulk volume management.

Like an on-premises SAN, customers can have a large pool of throughput and data, which can be shared across multiple workloads that can automatically scale up or down with virtually zero unutilized capacity. The Lightbits architecture eliminates the need to predict future business demand when storage is provisioned and further reduces management time and costs.

In these ways, Lightbits fills in the gaps that native cloud solutions can't deliver. Lightbits delivers the highest possible performance with consistently low latency, comparable to direct attached flash but with resiliency and efficiency. Enterprise-grade features include high availability with multi-zone capabilities and hybrid and multi-cloud models in one scheme. Lightbits also offers customers predictable and lower costs.

Data Services

Lightbits enterprise-class data services include auto-scaling, auto-healing, fast restores, replication, snapshots, clones, data reduction, and thin provisioning. Users can manage multiple storage volumes in a single cluster. This stands in contrast to most public cloud platforms, which only allow the management of single volumes. With Lightbits, cloud users can also create a SAN in the cloud, flexibly pooling storage for high-I/O use cases such as high-performance databases (e.g., Oracle, MongoDB, Postgres), analytics, and transactional workloads.

Flexibility with Hybrid and Multi-Cloud

Customers have embraced Lightbits' ability to deploy hybrid and multi-cloud storage architectures. Lightbits is available on multiple public clouds, allowing users to manage data both on-premises and on public cloud platforms. The Lightbits software license is portable between on-premises storage servers and storage-optimized Amazon EC2 instances or Microsoft Azure VMs — giving customers the ability to migrate storage to the cloud at their pace and on their terms.

Communication protocols also facilitate flexibility. Lightbits uses the network to communicate with application instances. As a result, customers gain more flexibility to run applications on more AWS EC2 instances or Azure VM types.

Efficiency

Using Lightbits makes for more efficient storage than is available with cloud-native options. Users can effectively "right size" their storage for each workload, reducing the number of clients in use. This can then reduce relative application licensing costs.

You can also accelerate application performance without having to add cores or nodes and employ auto-scaling to respond to spikes in demand. The disaggregated architecture enables independent scaling of compute and storage resources. The net effect of these features is a better price/performance efficiency than is available with native cloud storage solutions.

Faster Data Processing, Analytics, and Transactions

Lightbits delivers high-performing cloud storage — higher than native AWS EBS or Azure managed disks. Users get more than one million IOPs per volume — 10 to 20X faster than native cloud solutions. The average latency is <500 µsec. Tail latency is less than 1 msec. This all means consistent low latency at scale.

Lower Risk and Better Business Continuity

Lightbits lowers the risk of over- or under-provisioning storage for critical workloads. With Lightbits, you get nearly 1M IOPS per volume, and auto-scaling ensures that capacity and performance are available to help meet unpredictable demand. You can also establish business continuity with less downtime and improved SLAs.

This outcome is due, in part, to the high availability that occurs through Lightbits' advanced clustering scheme, as well as support for multiple geographic availability zones. Data protection is one of the biggest advantages of multi-availability zone (AZ) support. Customers can replicate copies of their data across AZs to mitigate data loss, should an AZ go down.

The Lightbits Cloud Data Platform is highly resilient as well. Users can provision high-performance persistent volumes to applications, which protects data from failures. It's a better alternative to running workloads using local non-volatile memory (NVMe), which is often short-lived.

Predictable and Lower Costs

Storing data in the cloud is more cost-effective with Lightbits than with native public cloud solutions. Running Lightbits on AWS or Azure, the price is fixed and predictable. You only pay for the AWS instances or the Azure VMs and the Lightbits software licenses.

The public cloud practice of charging for provisioned IOPS — and in some cases throughput and ad hoc data services — can result in unpredictable costs. With Lightbits, thin provisioning, and compression are also included in the license, meaning that you use less capacity.

Conclusion

Native storage options on public cloud platforms – such as AWS and Microsoft Azure – are architected for non-criticial Tier 2 and 3 workloads and are not ideal for critical, storage-intensive workloads. These are ideal platforms to host non-critical workloads and applications, but times are changing and enterprise workloads are changing.

The public crises and geopolitical events of the past few years have put tremendous pressure on organizations to accelerate cloud-first strategies. As such, there's been a large increase in demand for a high-performance, low-latency data platform that delivers SAN-like capabilities on the cloud.

The Lightbits data platform solves the limitations of native cloud storage. It's a robust storage target for AWS Elastic Cloud (AWS EC2) instances and Azure Virtual Machines (VMs) — with unmatched speed, flexibility, and efficiency. It overcomes the limitations of native cloud storage and enables organizations' migration of critical I/O-intensive workloads to the cloud by delivering high performance (high IOPS and the lowest latencies), predictable and lower costs, hybrid cloud flexibility, and enterprise-grade features that are included in the software license.

And just like on-premises SAN, customers can have a large pool of throughput and data that can be shared across multiple workloads, thus eliminating the need for accurately predicting demand when storage is provisioned and costs are minimized. The Lightbits solution offers the ultimate flexibility in deployment methods and delivers on the true promises of the cloud: simplicity, lower costs, lower risk, better efficiency, and increased flexibility.

1830 The Alameda, San Jose, CA 95126 E: info@lightbitslabs.com www.lightbitslabs.com