

# DATA HUB: A MODERN VISION FOR STORAGE

## THE NATURE OF DATA IS CHANGING

Data has come alive, becoming one of the most strategic resources fueling the modern enterprise. A number of colliding trends spur this new age of data. Data sources are multiplying while IT infrastructure is evolving. The nature of compute itself is changing as analytics becomes a core capability in the enterprise data center. Technology providers must evolve the products they offer to meet the needs of this new world of data, and this is especially true in the world of storage.

The volume, velocity, and very nature of data itself is changing. Edge computing, where data is collected and processed well outside the traditional data center, is flooding enterprise storage systems with a non-stop stream of data that needs to be stored and processed. Edge delivers data from remote and branch offices, sensors embedded in an organization's transportation fleet, factories, smart buildings, surveillance systems, and retail beacons tracking customers through stores. These are just a few of the dozens of examples of edge computing driving new data types and sources.

Central to the changing nature of data is the introduction of advanced analytics into enterprise workflows. The underlying technologies for machine learning (ML) have matured and become practical for deployment by enterprise IT and are no longer constrained to scientific computing environments. The availability of affordable graphics processing units (GPUs) coupled with the widespread availability of artificial intelligence (AI) and ML software opens the door for businesses to look at their data in new and innovative ways.

ML allows multiple sources of data to be quickly correlated to recognize patterns, patterns which quickly become business-critical. Retailers and financial institutions, for example, leverage deep learning across transactions and inventory data to identify fraud and stop losses. Logistics companies use AI to correlate micro-targeted weather forecasts, real-time traffic information, and sensors aboard their trains and trucks to guide the routing and scheduling that fuel their business. There are dozens of examples of advanced analytics that both guide and deliver insights to help businesses make better decisions.

While edge computing and analytics offer new capabilities to the enterprise, the foundational infrastructure of the data center itself is evolving. The widespread and affordable availability of public and private cloud services, along with on-premise infrastructure-as-a-service (IAAS), have moved IT to a world where compute has become portable.

It is relatively easy to move a compute workload from one server to another and even between local servers and the cloud. However, complications arise when those workloads have significant data dependencies. Data gravity is real, and it is expensive and time-consuming to move large blocks of data between different infrastructures. This becomes problematic when workloads demand data be local and accessible. Data and the systems storing that data must be re-architected for the hybrid-cloud world.

There is a tremendous amount of innovation that is elevating data into a strategic differentiator. Delivering on the promise of the evolving world of data requires a fresh look at how and where data is stored and shared between the workloads that depend on its availability.

## LIMITATIONS OF TODAY'S STORAGE ARCHITECTURES

A typical data center spreads data across the enterprise in discrete silos. Data may exist in enterprise data warehouse systems, storage arrays tied to application-specific servers, cloud-based repositories, or even exist in software-defined data pools supporting virtualized or converged infrastructure. These data silos are inhibitors to modern analytics-driven workflows.

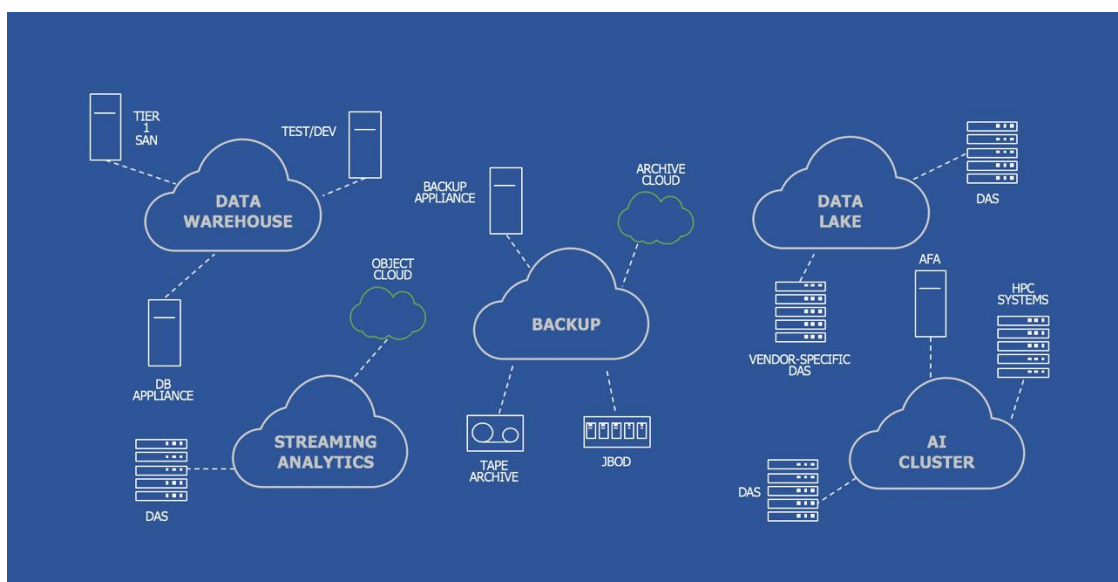
Different compute models, whether business intelligence applications, analytics, edge storage, or virtual desktops, require different capabilities from its associated storage infrastructure. While these disparate systems have different needs, they often leverage the same sets of data. It is this imbalance that forces a fresh look at enterprise storage systems.

AI-driven analytics, for example, have specific requirements both in consuming collections of data and how that data is provided to the servers executing the analytics workloads. The nature of analytics demand data be available as a consistent whole, not fractured in data silos, and be delivered to the compute engines with high-speed and low-latency.

The unstructured nature of this data, along with the need for disparate datasets to co-exist, requires a reconsideration of how data itself is stored. The challenge of data silos as they exist in the traditional enterprise is that those silos inhibit the ability to combine the sets of data needed into a cohesive whole for analytics. Silos also provide challenges in maintaining data integrity across the enterprise and delivering on a cohesive path for both data protection and governance.

As analytics becomes a driving force, data lakes will replace some data silos. Data lakes allow an enterprise-wide view of the totality of its structured and unstructured data. While data lakes enable any application to access underlying data cohesively and consistently, the technology has held back enterprises in practice because the underlying architecture is not built to deliver data in real-time and in a multi-dimensional way.

**FIGURE 1: DATA TODAY LIVES IN A WORLD OF SILOS**



*Source: Pure Storage*

The inhibitor to this vision of enterprise-wide data lakes is the limitation of the underlying storage systems that contain the data. Storage technology is capable of delivering unprecedented levels of performance. Solid-state drives (SSDs) are replacing spinning hard drives for hot and warm data. Interconnects such as non-volatile memory express (NVMe) take the latency out of the storage path. High-speed storage networking over ethernet delivers exceptionally high levels of throughput. Storage is faster, denser, and more affordable than ever before.

At the same time, storage systems are built today to serve specific applications, not to simultaneously deliver vastly different capabilities to a variety of different workloads sharing a pool of data.

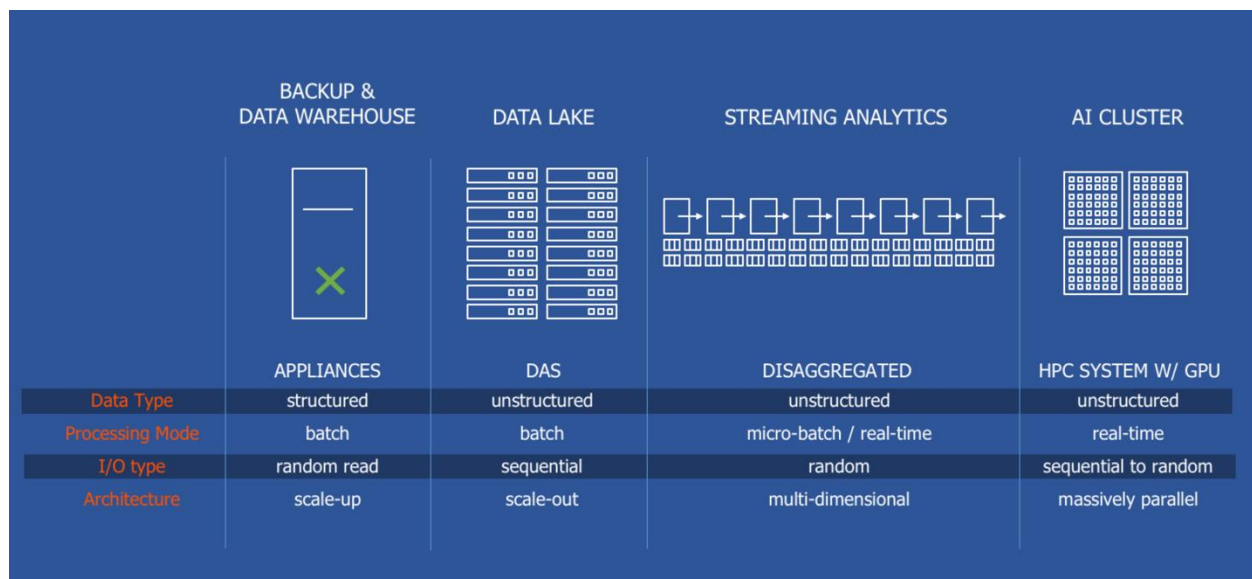
The limitations of storage are artificial and a result of architectural inertia. The foundational elements all exist, waiting to be bundled into a new architecture that can serve the rapidly evolving needs of the enterprise compute and data analysis.

## THE DATA HUB: ALIGNING DATA AND STORAGE

Data has become strategic due to the rise of analytics. Data has become plentiful due to the promise of analytics coupled with new data sources such as edge computing. Enterprises often turn to data lakes to provide a consistent view of their data to the various processing workflows that need that data, only to be disappointed with the complexity and performance of traditional solutions.

The challenge of deploying this vision on traditional storage systems is that these systems are not designed to deliver multiple streams of the same data with the varying levels of performance required by each of the workflows. Traditional storage systems are designed to be tuned to the needs of a single set of applications.

FIGURE 2: DIFFERENT STORAGE NEEDS FOR EACH SILO



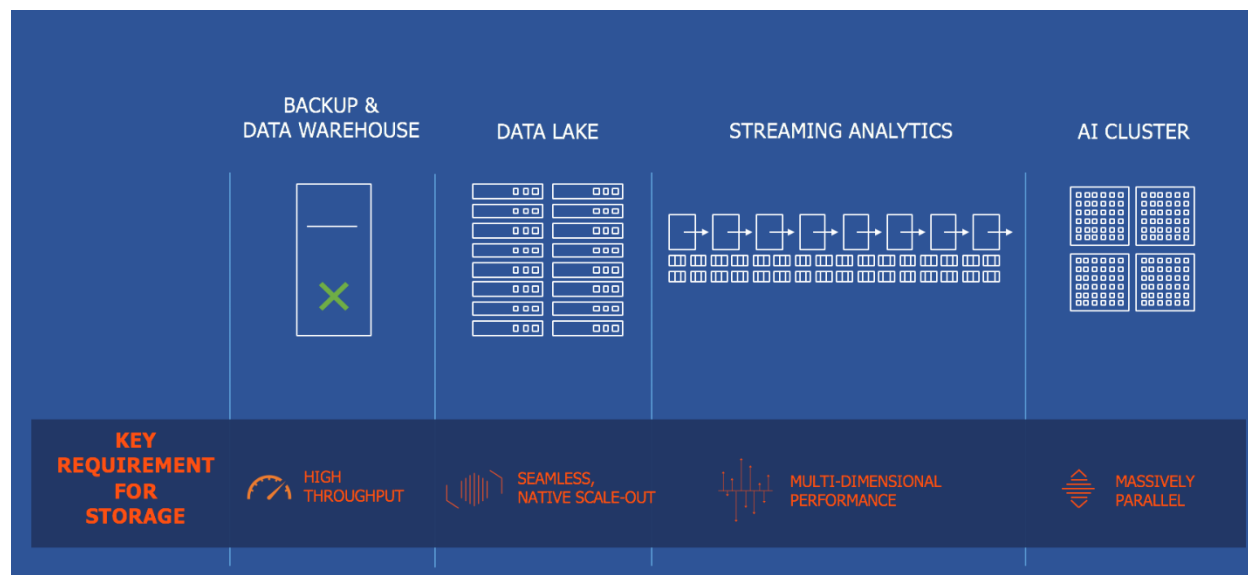
Source: Pure Storage

New storage and data architectures are needed to address the evolving needs of enterprise compute, one that addresses the limitations previously described. We call this architecture a data hub.

The data hub must deliver on a number of promises in order to fulfill the vision of a data-driven enterprise. It must produce high throughput file and object storage and provide application-specific performance characteristics (e.g., latency, throughput, and input/output operations per second), while offering those application-specific capabilities to multiple clients simultaneously. The data hub should also provide the ability for native scale-out.

Layered atop the Data Hub can be the usual set of storage capabilities demanded by the enterprise, including replication, snapshots, compression, and encryption. It is the ability to provide multi-dimensional performance that differentiates the Data Hub from traditional storage technologies.

**FIGURE 3: THE DATA HUB UNIFIES DATA**



Source: Pure Storage

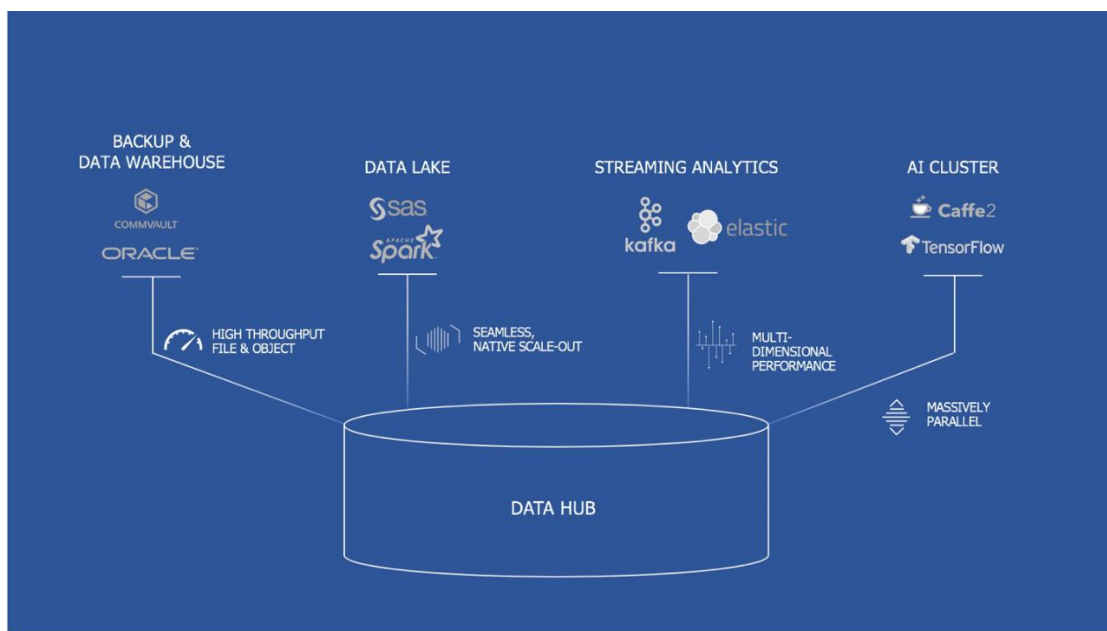
A reimagining of the storage system is required to deliver the capabilities of a data hub. No longer will a storage array be built from a server motherboard and a backend array of drives. Instead, the data hub requires multiple parallel compute and storage elements that can be individually partitioned and tuned for various workloads that are attached. Software orchestration becomes critical beyond the basic hardware configuration.

The industry has attempted to address these needs with various flavors of convergence, where compute, storage, and networking are combined into closely-coupled resource pools. While hyperconverged infrastructure is well-suited for many workloads, and may even be a consumer of data contained within a data hub, the approach suffers when attempting to provide a consistent view of data across multiple application domains.

Convergence breaks down when applications that share data are spread throughout the organization or even span on-premises and cloud-based services. Co-location of data and compute become more difficult as hybrid-cloud becomes the new normal and new architectures such as edge computing begin to move resources across the landscape. Coordinating access to an enterprise's data then becomes a long-term problem.

The data hub architecture offers the ability to collectively leverage an enterprise's data in a flexible and scalable manner. The data hub's software coordinates the delivery of data to applications as that data is needed and with the performance characteristics required. This becomes a true enabler of the software-defined data center, providing IT practitioners the ability to deliver capabilities that are data-centric defined by the needs of the applications. Practitioners are not limited by the inherent inflexibilities of convergence and other architectures that rely on the closely-coupled pooling of resources.

**FIGURE 4: THE DATA HUB, POWERED BY FOUR ESSENTIAL FEATURES**



Source: Pure Storage

## CONCLUSION

An enterprise moves at the speed of its data and that data is accelerating. It's flowing in from more sources than ever before and being consumed by analytics engines to offering increasingly valuable insights to the business. Fully embracing the new world of data requires enterprises to re-examine data architectures and for technology providers to rethink how storage systems are designed.

The technology industry has the opportunity to come together to define a new storage architecture that meets the new realities of data. Storage silos have no place in the current world of IT, while the limitations of legacy storage systems and architectures challenge effective deployment of data lakes into modern environments.

Storage providers must deliver solutions that provide multi-dimensional capabilities with application-specific performance capabilities, high-throughput file and object storage, and native scale-out.

The data hub is a different way of thinking about storage architecture. All of the enabling pieces are available today. The technology company that delivers to this vision will provide the infrastructure required to allow enterprises to embrace data-driven analytics fully.

In order to embrace the modern world of data, enterprises will demand a different approach to data and how that data is stored and delivered.

## IMPORTANT INFORMATION ABOUT THIS PAPER

### *CONTRIBUTORS*

[Steve McDowell](#), Senior Analyst, [Moor Insights & Strategy](#)

[Patrick Moorhead](#), Founder, President, and Principal Analyst, [Moor Insights & Strategy](#)

### *PUBLISHER*

[Patrick Moorhead](#), Founder, President, and Principal Analyst, [Moor Insights & Strategy](#)

### *INQUIRIES*

[Contact us](#) if you would like to discuss this report, and Moor Insights & Strategy will respond promptly.

### *CITATIONS*

This paper can be cited by accredited press and analysts, but must be cited in-context, displaying author's name and title and "Moor Insights & Strategy." Non-press and non-analysts must receive prior written permission from Moor Insights & Strategy for any citations.

### *LICENSING*

This document, including any supporting materials, is owned by Moor Insights & Strategy. This publication may not be reproduced, distributed, or shared in any form without Moor Insights & Strategy's prior written permission.

### *DISCLOSURES*

This paper was commissioned by Pure Storage. Moor Insights & Strategy provides research, analysis, advising, and consulting to many high-tech companies mentioned in this paper. No employees at the firm hold any equity positions with any companies cited in this document.

### *DISCLAIMER*

The information presented in this document is for informational purposes only and may contain technical inaccuracies, omissions, and typographical errors. Moor Insights & Strategy disclaims all warranties as to the accuracy, completeness, or adequacy of such information and shall have no liability for errors, omissions, or inadequacies in such information. This document consists of the opinions of Moor Insights & Strategy and should not be construed as statements of fact. The opinions expressed herein are subject to change without notice.

Moor Insights & Strategy provides forecasts and forward-looking statements as directional indicators and not as precise predictions of future events. While our forecasts and forward-looking statements represent our current judgment on what the future holds, they are subject to risks and uncertainties that could cause actual results to differ materially. You are cautioned not to place undue reliance on these forecasts and forward-looking statements, which reflect our opinions only as of the date of publication for this document. Please keep in mind that we are not obligating ourselves to revise or publicly release the results of any revision to these forecasts and forward-looking statements in light of new information or future events.

©2018 Moor Insights & Strategy. Company and product names are used for informational purposes only and may be trademarks of their respective owners.