

SOLUTION BRIEF

## Big Data and Analytics – A Data Lake Infrastructure for the Future

Highly scalable Big Data and Analytics (BDA) storage solutions that break data silos and help organizations turn data into a source of competitive advantage

### Highlights

ActiveScale™ data lake solutions help enterprises utilize organizational data to create a competitive advantage by discovering new insights, improving operations and providing better customer experiences.

### Customer Challenges

- Multiple DAS storage silos
- DIY Fatigue
- Lock-ins from expensive proprietary data lake solutions
- Scaling-related opex challenges
- Compromise with eventual consistency

### Solution Features

- Enterprise class reliability, availability and serviceability
- Complete BDA portfolio: scalable capacity tier with scalable performance tier
- S3A connector, rateless erasure coding, and strong consistency
- Encryption for end-to-end data protection

### Solution Benefits

- Unified repository
- Ease of setup and management
- Lower acquisition costs and operating expenses
- Ability to run in-place analytics and linearly scalable capacity with no performance degradation
- Transforms global workloads, speeding time to market, innovation and competitive advantage

Organizations across industry verticals are striving to increase operational efficiencies and provide a superior customer experience leveraging data generated across their value chain. Data generated from millions of sensors, video feeds, genomic sequencing or machine logs is helping enterprises improve their supply chains, do predictive maintenance and design better products across verticals. However, this petabyte-scale Big Data deluge is posing challenges for business and IT managers on how to manage their data storage and analytics infrastructure as well as capture value while staying within the capital and operating expenditure limits.

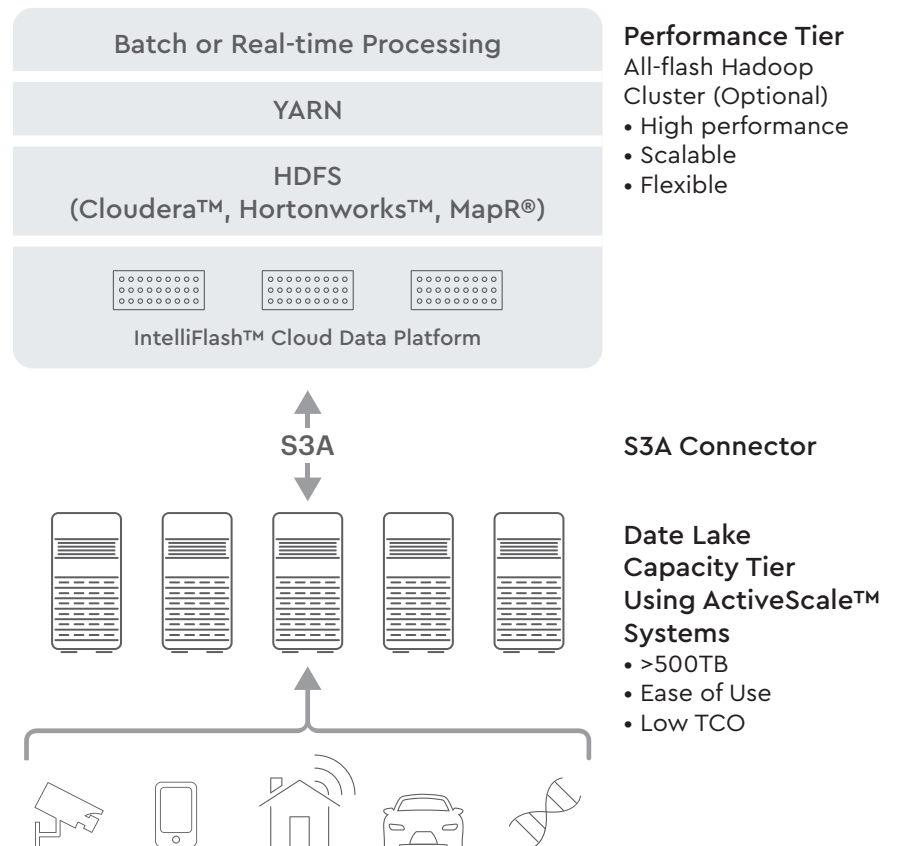


Figure 1: Big Data and Analytics (BDA) Data Lake Solution

## Solution Overview

ActiveScale object storage provides enterprises with a platform to build a petabyte scalable data lakes with one of the lowest TCOs and without compromising on performance, scalability and reliability. The solution differs from existing data lake approaches due to its ability to run in-place analytics with the agility to combine the high performance HDFS analytics tier with the high capacity tier. The ability to move data back and forth via an S3A connector between the two tiers makes the most optimal use of resources. The solution delivers the power of disaggregated performance scaling as well as disaggregated capacity scaling to our customers.

## Solution Details

**Break Data Silos:** The highly-scalable data lake solution is built using an ActiveScale object store. The data lake serves as a unified high-capacity repository to store all of your enterprise Big Data, thereby breaking inefficient data silos that plague many enterprises. Designed bottom up with BDA needs in mind, our data lake expertise delivers at scale deployments with lifecycle management tools, system availability protection with rateless erasure coding, and geo-spreading capabilities that our needed for business continuity.

**Run In-place Analytics:** The solution allows you to run analytics in place without worrying about data movement or ETL overheads. This immensely increases efficiency because there is no overhead of data movement and no need to maintain redundant data copies. This delivers a solution with low capex and opex and business managers can deliver on their promise of return on investment (ROI) much faster.

**Future-proof Big Data Investment:** Strong consistency is a standard feature as part of the ActiveScale object store which is important for certain enterprise applications and cloud-native applications. Most importantly, there is no lock-in via a proprietary file system or expensive data warehouse. This helps future proof your BDA investments and brings peace of mind to data center architects and business managers. These leaders can now focus more on capturing value from organizational data and how to leverage it as a source of competitive advantage.

## Conclusion

Leveraging our enterprise ready and proven infrastructure solutions enables business and IT managers to focus on capturing business and operational insights. This enables them to play a pivotal role in their organization's digital transformation by helping create value and high ROI through in the following ways:

- Improve business processes for operational efficiency
- Reduce cost
- Expedite new product development
- Gain insights into new services
- Provide better consumer experience

Enterprises who are experiencing dramatic operating expenses utilizing a public cloud will be pleasantly surprised to know that our ActiveScale object storage can help you build out a petabyte-scale data lake on private or hybrid cloud with a total cost of ownership that is comparable or lower than leading public cloud providers.

To learn more about ActiveScale visit:  
<http://www.hgst.com/products/systems>

# Western Digital®