Unleash the Power of Petabyte Object Storage with StorReduce and ActiveScale™

StorReduce high-speed deduplication brings scale, simplicity, and dramatically lower costs to hybrid and private cloud backup.

March 2018

Table of Contents

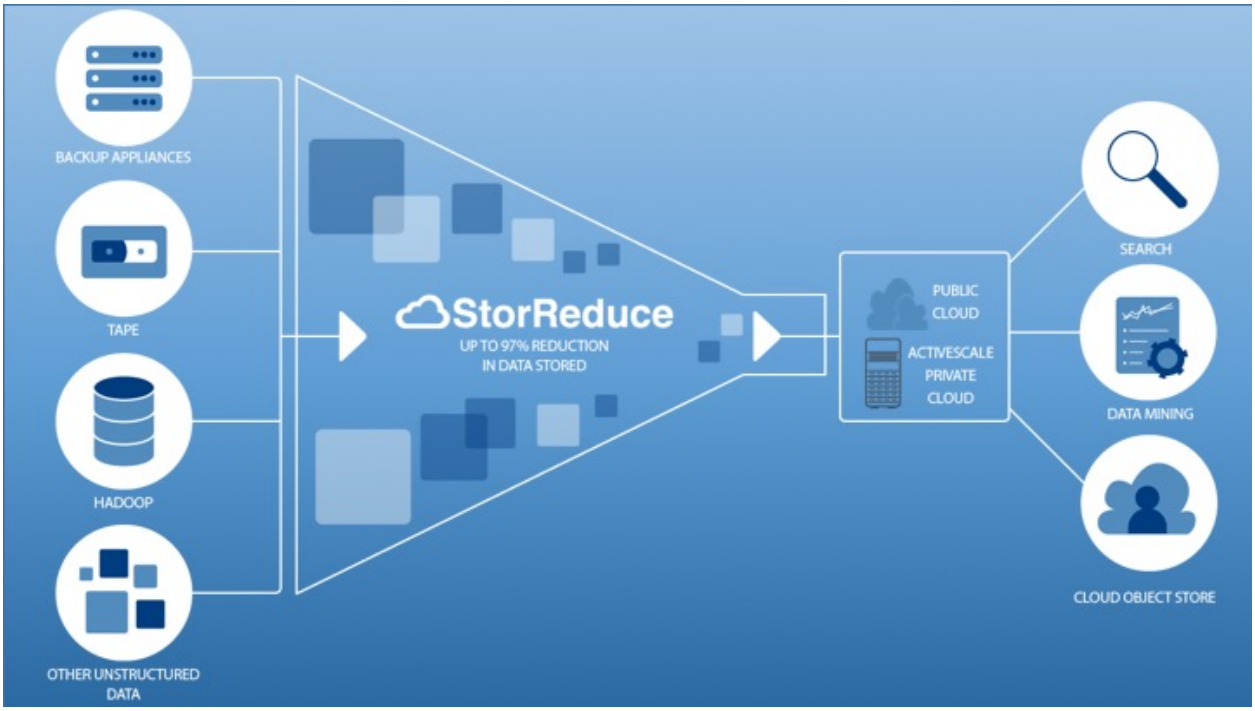
Introduction

Natural disasters. Security breaches. Fat fingers. Catastrophic disruptions to critical business services can come in many forms. That's why, whether due to regulatory requirements or just common sense, every responsible enterprise thinks long and hard about data protection. Unfortunately, the current approaches on which many organizations rely—storing primary backups on purpose-built backup appliances and using magnetic tape for long-term retention—are far from ideal.

Conventional backup appliances carry high price tags. And, as organizations store more and more data, they constantly need to add appliances—locking themselves into a never-ending capital investment cycle with ever-growing power, cooling, and management costs. These appliances also typically become "hardware silos," offering limited flexibility to move and access stored data. Even worse: they don't fully address the problem they're intended to solve. An ill-timed system outage can easily lead to data loss, undermining the rationale for investing in backup appliances in the first place.

Meanwhile, tape backups, while theoretically less expensive, carry their own hidden costs in the ongoing need to handle, transport, and store physical tapes. They're vulnerable to human error—like being lost, stolen, or accidentally rewritten. And, in the event an organization needs to use them to recover from an outage, they can expect slow restore times and significant downtime while they wait.

Now, there's a better way to store massive amounts of backup data. By combining advanced deduplication with high-density object storage, you can slash data protection capital and operating costs and better protect your business. And you can get rid of backup appliances and tape forever.



ActiveScale with StorReduce Software

By combining two powerful technologies, ActiveScale™ object storage and StorReduce deduplication, organizations can now easily and affordably transition from purpose-built backup appliances and tape to high-performance cloud data protection.

ActiveScale builds on a heritage of innovation that goes back to the world's first hard disk drive, providing a high-density object storage solution designed for large enterprises handling massive amounts of data for long periods. Deployed with StorReduce inline deduplication software, organizations can quickly move huge volumes of data to ActiveScale private or hybrid cloud, eliminating redundant data along the way.

With StorReduce scale-out deduplication software, you can implement cloud-based object storage for primary and secondary backups at petabyte scales. And, you can slash storage costs by up to 70% by replacing backup appliances and tape with private or hybrid cloud. Using StorReduce with ActiveScale gives you best-of-breed scalability, durability, and recoverability for your data. At the same time, you can now actually *use* that backup data for searching, analytics, and even machine learning via standard cloud APIs.

This paper provides a technical overview of the StorReduce solution. It highlights key features and capabilities, and details the StorReduce architecture. And it presents several use cases to illustrate how organizations can use StorReduce with ActiveScale object storage to realize concrete benefits. (For more information on ActiveScale itself, visit https://www.hgst.com/storreduce)

StorReduce Overview

StorReduce is a specialized cloud deduplication solution designed to meet the unique requirements of companies using object storage (public cloud or private object store, or a hybrid combination) for huge volumes of data. StorReduce sits between your applications and ActiveScale object storage, transparently performing inline data deduplication. It reduces storage and bandwidth requirements by as much as 97% and can reduce data transfer times by just as much.

StorReduce is fast and massively scalable. When using multiple servers deployed as a scale-out cluster, it can achieve sustained write speeds of 10 Gigabytes per second or higher and manage hundreds of petabytes of data. All data is managed in a single deduplication pool. And, data deduplicated with StorReduce can always be accessed on-cloud or on-premises using the Amazon Simple Storage Service™ (S3) REST API.

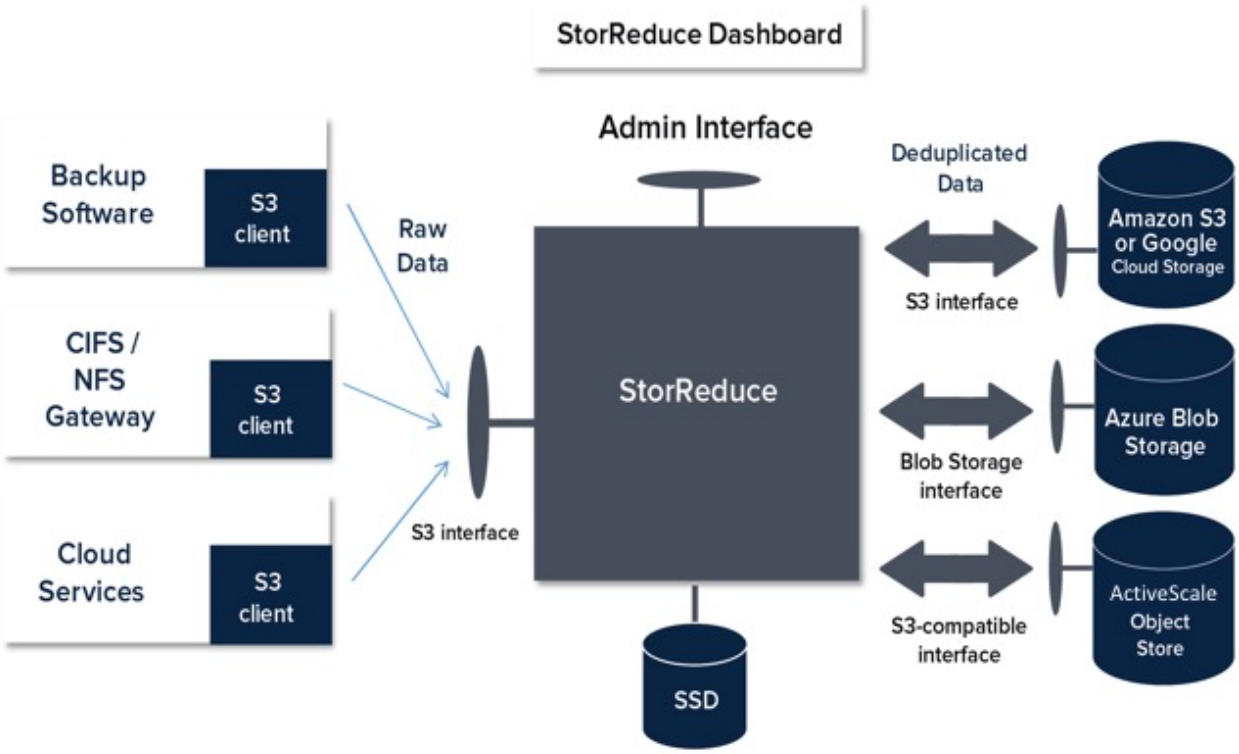


DIAGRAM 1: STORREDUCE ARCHITECTURE

Unlike other data backup and deduplication solutions, StorReduce delivers:

- **Massive scalability** with an architecture designed for high throughput and low latency at petabyte scale
- **Up to 70% storage cost savings** by migrating data off expensive, complex backup appliances and legacy tape
- **Improved resiliency** with a solution that runs on commodity hardware with no single point of failure
- **Faster performance** through inline deduplication that removes the risk of data loss without slowing backup times
- **No more storage silos** with a single namespace across your hybrid or on-premises cloud storage
- **Ability to actually use migrated backup data by looking inside backup archive files, and providing** a cloud-native interface that enables you to use services like Search (to search inside all migrated backups) and data mining
- **Increased flexibility** with the ability to replicate deduplicated data between cloud regions and cloud-based infrastructures, or between cloud-based and on-premises environments

- **A big data-ready solution** that offers scalability for large data sets and very fast throughput, making it ideal for cloud-hosted workloads that now require less cloud storage and can utilize Hadoop and services
- **High availability** via clustering, with the ability to deploy clusters in multiple Regions to ensure that data access is maintained even in the unlikely event of a complete failure of a cloud-based datacenter
- **More value from your backup data,** with the ability to clone it onto ActiveScale object storage, where it can be used for additional workloads such as dev/test and compliance

Additionally, StorReduce delivers the following key benefits:

- **Fast Sustained Throughput:** Achieve writes speeds of 10 gigabytes per second or more using a scale-out cluster, while adding less than 50ms of latency. Each individual StorReduce server can deduplicate data at speeds of 2 gigabytes per second or faster. And, StorReduce can deliver sustained throughput 24/7—unlike some solutions, it does not stage data in buffers that can fill up.
- **Cloud-native Storage:** Make all deduplicated data immediately accessible to cloud services via StorReduce's S3 REST API. Store data on all major public cloud providers—Amazon S3 or S3IA, Microsoft Azure Blob Storage, or Google Cloud Storage.
- **Resiliency by design:** The system stores index information required to re-hydrate the data in multiple locations, both in object storage and on each StorReduce server. Extensive use of checksums and hashes ensures *data integrity* during both deduplication and re-hydration.

All of these advantages derive from StorReduce's unique architecture—designed from the ground up for performance, scale, and flexibility. Key StorReduce architectural features include:

- **Scale-out Clusters:** Increase data storage capacity and write throughput almost linearly by simply adding more StorReduce servers to a cluster. At the same time, you maintain a single deduplication pool and a single namespace for the data.
- **Backup Software Integration/Data Management:** StorReduce works with any existing data management or backup software that is compatible with Amazon S3, including Veritas NetBackup, Backup Exec, Commvault, and EMC Networker. Veeam is supported via "gateways" until they provide native S3 support.
- **Object Clone:** StorReduce can create writable "virtual clones" of all the data in a storage bucket, without using additional storage space. Create unlimited virtual clones for testing or distribution of data to different teams in an organization. Copy-on-write semantics allow each individual clone of the data to diverge as new versions of individual objects are written.
- **Read-Only Endpoints:** You can deploy additional StorReduce servers or clusters on-cloud or on-premises as read-only endpoints. Uploaded data is only stored once but is immediately available at each endpoint location. With this capability, your migrated backup workloads are no longer just a cost center; you can now repurpose them on the cloud for development, test, QA, disaster recovery, and for use by other cloud-based services.

- **Data Replication:** StorReduce can replicate data between regions, between cloud vendors, or between public and private cloud to increase data resiliency. The system only transfers unique data, providing up to 30 times faster transmission and up to 30 times lower bandwidth and storage costs.
- **High Availability:** When configured as a scale-out cluster StorReduce provides automatic failover in the event of server failure. You can set up the system with any degree of redundancy you choose between servers.
- **Write Speed Throttling:** Even massive data backup doesn't have to impede the performance of other business applications. Set a maximum write speed to prevent StorReduce from using too much bandwidth when sharing an Internet connection with other infrastructure.

StorReduce is also built to ensure that sensitive data stays protected—in motion and at rest. Key security features include:

- **Encryption:** StorReduce supports encryption of data before it lands in the object store. It integrates with key management systems like Amazon KMS or KMIP-compatible hardware security modules for storage of the cryptographic keys. Data can be encrypted on-premises before being sent to the cloud, or you can use cloud encryption-at-rest services. StorReduce *always* encrypts data in-transit.
- **Secure User Account and Key Management:** You can assign users or servers individual user accounts within StorReduce to restrict data access. Multiple access keys can be created and managed as needed for each user account.
- **Secure Policy-based Access Control:** You can express enterprise security policies using the StorReduce policy engine, using Amazon's Identity & Access Management (IAM) policy language.

StorReduce Architecture

From an architectural perspective, StorReduce acts as an intelligent, transparent layer between your client applications and ActiveScale object storage and public cloud, enabling applications to use your storage more efficiently and cost-effectively. (See Diagram 1.) It operates transparently to applications, allowing them to store and retrieve data as they normally would, but with best-in-class data deduplication and high throughput.

StorReduce can be deployed as either a single server or a scale-out cluster of servers. A single server can be expanded to become a cluster by adding more servers at any time. In either case, client applications see a single logical server.

StorReduce server software can be deployed in the public cloud (such as an Amazon EC2 instance), or on-premises in a virtual machine (VM) or Docker container, or installed natively on Linux.

The StorReduce server provides similar functionality to object store or cloud storage solutions, including object management, user accounts, access keys, access control policies and a Web-based management interface (the StorReduce Dashboard).

The following sections describe StorReduce architectural components and features in more detail.

S3 Client Software

StorReduce is designed to be easy to implement in your existing data backup infrastructure, without having to reengineer the way applications interact with storage. It works with client software that supports Amazon's *S3 REST interface* for object storage. This includes clients designed to work with Amazon S3, as well as those designed to work with Google Cloud Storage via the XML API (which is S3-compatible). For your applications, the only difference from conventional S3 object storage is that client software is configured to talk to StorReduce instead of directly to object storage, using access keys provided by the StorReduce server.

Practically any S3 client software used in conventional object storage environments can be used with StorReduce. This includes on-premises backup software (including Veritas NetBackup) as well as custom software written to use the S3 REST interface. You can also use any cloud-based services designed to work with Amazon S3 or Google Cloud Storage with StorReduce, as these also act as S3 clients.

StorReduce automatically translates S3 client requests into whichever protocol is needed by the underlying object storage providers—including Microsoft Azure. It provides an ***S3-compatible interface onto Azure Blob Storage***, allowing backup products and other software designed to work with S3 to be used with Azure Blob Storage without modification.

CIFS/NFS Client Software

StorReduce also supports file-sharing interfaces such as Common Internet File System (CIFS) and Network File System (NFS) via gateway software and appliances. In these scenarios, the gateway exposes a file share interface and converts requests into calls to Amazon's S3 protocol.

StorReduce Servers

Each StorReduce server runs on its own physical or virtual machine, with local SSD storage recommended. Each StorReduce server can handle tens of petabytes (tens of millions of Gigabytes) of raw data, depending on the deduplication ratio achieved and the amount of SSD storage available for index information. (See the section "Local SSD" below.)

With StorReduce, you can create multiple storage buckets and perform ***global deduplication*** across all of them.

When using ***public cloud***, the StorReduce server runs on an Amazon EC2 instance, Azure Virtual Machine, or Google Compute Engine. VM images are available through the AWS Marketplace, Azure Marketplace, and Google Cloud Marketplace with the server pre-installed, allowing quick and easy setup. You can also easily install StorReduce on any Linux virtual machine using Red Hat Package Manager (RPM).

StorReduce also supports on-premises deployments. For **migration of on-premises data to the cloud**, or for private object store deployments, the StorReduce server can be run on-premises on a physical or virtual machine, or under Docker. Pre-built virtual appliance (OVA file) and Docker container images are available.

StorReduce is designed from the ground up for resiliency and flexibility. The architecture allows **multiple StorReduce servers** to be run against the same back-end object storage service, for redundancy, load-sharing, and increased storage volume. This capability forms the basis for StorReduce scale-out clusters, as well as allowing multiple servers to act as endpoints providing access to the same data from different locations. For example, an on-premises StorReduce cluster might be used to deduplicate and upload backup data, with a second in-cloud StorReduce cluster providing immediate access to this data for cloud services as the data is uploaded.

StorReduce servers feature two interfaces:

- **S3 Interface:** The StorReduce server exposes an S3-compatible REST interface for object storage. This highly scalable interface supports most S3 interface calls, including:
    - Object GET/PUT/POST/DELETE (including multiple-object delete)
    - Multipart uploads (including listing and deleting uploads)
    - Digital signature verification
    - Bucket create/delete/rename
    - Setting/reading bucket policies for access control.

- **Admin Interface:** StorReduce servers expose a separate REST interface for use by the Web-based dashboard. This admin API is served on a separate port to allow firewalls to restrict network-level access, and can be served over HTTPS. The admin API is available for use by other client applications as well as the StorReduce dashboard, and supports manipulation of user accounts, access policies and index snapshots, as well as providing a replica of the S3 API for use by management tools.

Local SSD Index Storage

StorReduce stores index information on local SSD storage. The amount of raw data a StorReduce server can handle depends on the amount of fast local storage available and the deduplication ratio achieved for the data.

The use of local SSD storage rather than RAM for index information allows tens of petabytes of data to be managed by a single StorReduce server. Typically, the amount of SSD storage required is less than 0.2% of the amount of data put through StorReduce, for a standalone server.

The StorReduce server treats local SSD storage as ephemeral. All information stored in local storage is also sent to object storage and can be recovered later if required (see later section).

Object Storage

The StorReduce server uses ActiveScale object storage in private or hybrid cloud scenarios for all persistent data. StorReduce acts as an S3 client, making use of the object storage API to store all its data in a single account. StorReduce is architected to work with any S3-compatible public object storage solution, including Amazon S3, as well as with Azure Blob Storage.

The StorReduce server makes use of Object Storage to store the following types of data:

- **Deduplicated user data:** Raw data is deduplicated using state-of-the-art algorithms and then compressed. Typically, this requires as little as 3% of the object storage space that the raw data would have required, depending on the type of data stored.
- **System Data:** Information about buckets, users, access control policies, and access keys is also stored in backend object storage, making it available to all StorReduce servers in a given deployment.
- **Index snapshots:** Data for rapidly reconstructing index information on local storage can also be stored in backend object storage.

Performance

StorReduce is the product of groundbreaking cloud storage innovations that deliver unprecedented scalability, high throughput, and low latency. The internal architecture and code are finely tuned for data deduplication and optimized to ensure superior performance even when running in a public cloud environment.

A single StorReduce server is capable of sustained write speeds of 2 Gigabytes per second or more, depending on the deduplication ratio, CPU cores available, and network connection available. These real-world speeds have been achieved on Amazon EC2 instances and within customer data centers using the StorReduce virtual appliance.

A StorReduce cluster is capable of throughput that scales almost linearly with the number of servers in the cluster. Distributing the load over all servers in the clusters, the system can achieve throughputs of 10 Gigabytes per second or more, while maintaining a single deduplication pool. By deploying a StorReduce cluster in front of an ActiveScale private object store installation, you can dramatically reduce your storage requirements while maintaining very high throughput rates.

The best way to substantially speed up throughput and decrease transfer bandwidth to cloud-based storage is to run a StorReduce server or cluster on-premises, deduplicating data prior to sending it into the cloud, and reading deduplicated data from the cloud and reconstituting it locally. By taking advantage of StorReduce's fast throughput, you can reduce migration times from years to weeks and greatly reduce migration costs. (For a direct independent comparison of StorReduce vs. a well-known migration vendor, visit http://www.storreduce.com/case-studies/apn-spectrumdata/.)

StorReduce also keeps latency to a minimum, typically adding less than 50ms of latency even when running in the cloud. For most situations, this added latency makes no difference at all to end users and does not affect throughput.

StorReduce maintains an index of user data on fast local storage. Each StorReduce server keeps its own independent index, typically on locally attached SSD. In a scale-out cluster, index data is spread over the servers in the cluster and can be configured with any required degree of redundancy.

All index data can be rebuilt from the log of transactions stored in object storage. For large data sets, however, it can take a long time to rebuild the index from scratch. To speed this up, the server or cluster periodically takes a snapshot of the index information and stores this in object storage.

When a StorReduce server starts up, if an index needs to be rebuilt, the server will:

1. Load the last index snapshot from object storage
2. Replay subsequent transactions to bring the index up to date.

Note: When stopping a StorReduce server running on an Amazon EC2 instance, Amazon will delete all data on that machine's SSD instance storage. When the machine is started again, the index must be rebuilt as described above. For this reason, it is recommended to leave production EC2 StorReduce servers running rather than stopping and starting them. Note that when *restarting* an EC2 instance, the instance storage is preserved.

StorReduce Demonstration Cluster

If you'd like to see how StorReduce performs in a real-world deployment, StorReduce operates a nine-server cluster on Amazon EC2 instances to provide real-time performance demonstrations. This cluster is continuously (24x7) ingesting data at ¾ of a petabyte per day (8.8 Gigabytes per second). This cluster is simultaneously able to recover data at 4.8 Gigabytes per second (double the speed of the largest Data Domain). StorReduce clusters can scale to 31 servers, with speeds scaling almost linearly. A similar-sized cluster on physical hardware will operate at even faster speeds, as physical hardware can typically provide more CPU resources than public cloud instances.

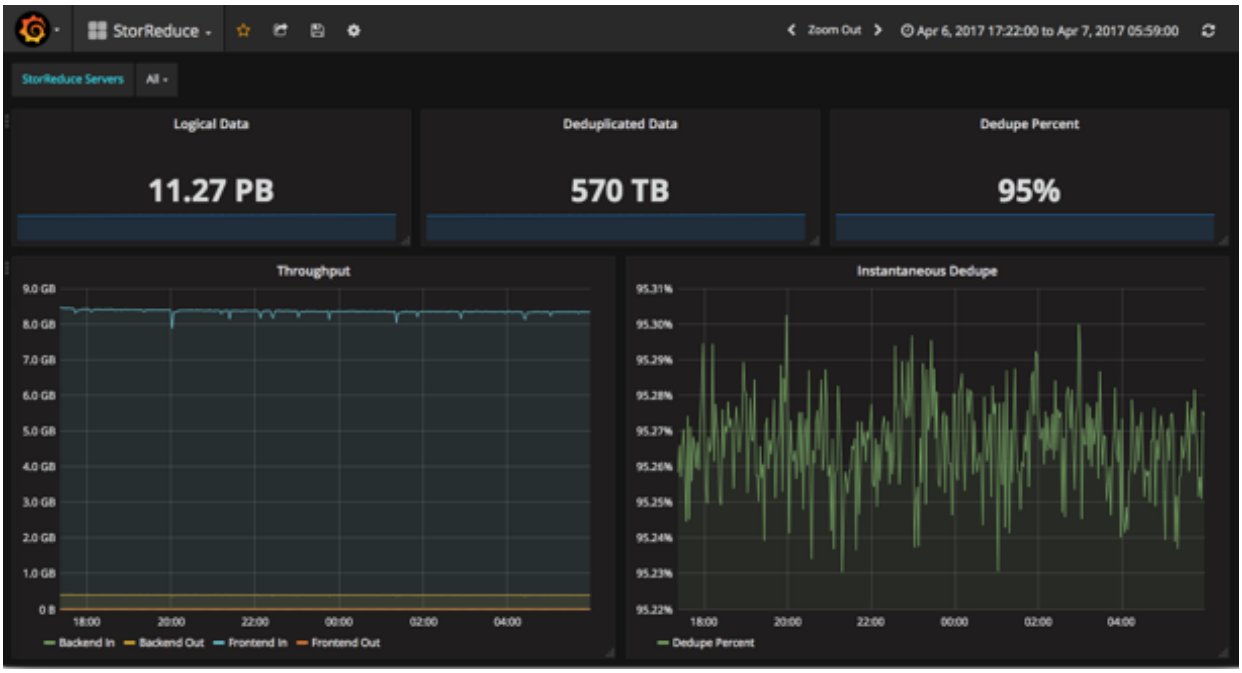


DIAGRAM 2: STORREDUCE DEMONSTRATION CLUSTER DASHBOARD

The screengrab above shows the StorReduce demonstration cluster ingesting data at 8.8 Gigabytes per second while achieving a 95% deduplication rate. StorReduce can provide live demonstrations using this system on request.

Using Multiple StorReduce Servers

StorReduce offers a number of advantages when using multiple StorReduce servers and is designed to make adding servers a quick, easy process. Because StorReduce maintains a log of all transactions on object storage, multiple servers can watch this transaction log to keep their independent indices up to date. This means that you can set up new servers to talk to an existing Object Storage service, and they will automatically populate their local index data from Object Storage.

The sections below detail some of the capabilities and advantages that organizations can realize when using multiple StorReduce servers.

Scale-out Clusters

One of the best ways to improve the redundancy and resiliency of your data backup infrastructure is through StorReduce's scale-out cluster capability. A scale-out cluster uses multiple StorReduce servers to distribute load and provide automatic failover within a single data center or region.

In a scale-out cluster, data is divided into shards, with each shard assigned to a primary "write" server and, optionally, one or more secondary "read" servers. For each shard, the secondary server(s) stand ready to take over as primary server for that shard in the event of a problem. Failover happens automatically, coordinated via an open-source distributed key-value store.

Any S3 protocol request can be processed by any StorReduce server, so load can be distributed by DNS round-robin or a load balancer deployed in front of the cluster. The server processing a given request handles any required coordination with other servers in the cluster.

This architecture allows the amount of data managed and the throughput to scale up almost linearly as more servers are added to the cluster. Intra-cluster communications are highly optimized and can be handled on a separate network interface to maximize throughput.

For more information on deployment options for scale-out clusters, please contact StorReduce.

Making Data Available in Multiple Locations with Read-Only Endpoints

Large enterprises need to be able to access data from multiple locations—and they should be able to do it without having to navigate multiple complex data silos. StorReduce was designed to allow StorReduce servers or scale-out clusters to be deployed as read-only endpoints, each of which can read and re-hydrate data. In this way, the same content can be fetched from multiple StorReduce servers in different locations, each having the same view of the content, and all locations updated in real-time.

To implement this, one common deployment scenario is to have a StorReduce server running on-premises, deduplicating data as it is sent to the cloud. A second StorReduce server runs in the cloud as a read-only endpoint, providing real-time access to the data (in re-hydrated form) to cloud-based applications and services via its S3 interface. This model works particularly well for moving backups to cloud.

Creating Multiple Copies of the Data with Data Replication

StorReduce also makes it easy and cost-effective to create and store multiple copies of data. Similar to the read-only endpoint scenario, a StorReduce server can be configured to automatically replicate data from its primary object store to one or more other object store locations. Whenever the StorReduce server detects any changes, it automatically copies them to the other location(s). And, since the server only replicates deduplicated data, the cost of that data transfer is a tiny fraction of conventional data transfers.

A StorReduce replication server can be used to replicate data:

- Across multiple regions in the same cloud
- Across multiple cloud providers (including Amazon S3, Azure Blob Storage, and Google Cloud Storage)
- From private to public cloud, or vice versa (e.g. ActiveScale hybrid cloud solutions).

Object Clone

Enterprises face many scenarios where it's necessary for users or groups to have their own, separate copies of a data set. StorReduce makes it easy and efficient to create such copies through object cloning.

A StorReduce server can "clone" all objects in a storage bucket, creating any number of virtual copies of large data sets. Each clone is writable on an object-by-object basis. StorReduce uses copy-on-write semantics, so clones do not take any extra object storage space until data is written to the clone. Even then, data is fully deduplicated between clones.

By combining StorReduce user accounts and access control policies with object cloning, teams within an organization can each be provided with their own writable copy of a data set, without being able to see each other's changes or new data. And enterprises can do this while still maintaining a single deduplication pool across the entire system.

Object cloning also enables simpler and more flexible testing and QA, since software can be tested against a cloned bucket. The software maintains full control over all data in its bucket, without being able to interfere with any other clone or the original data.

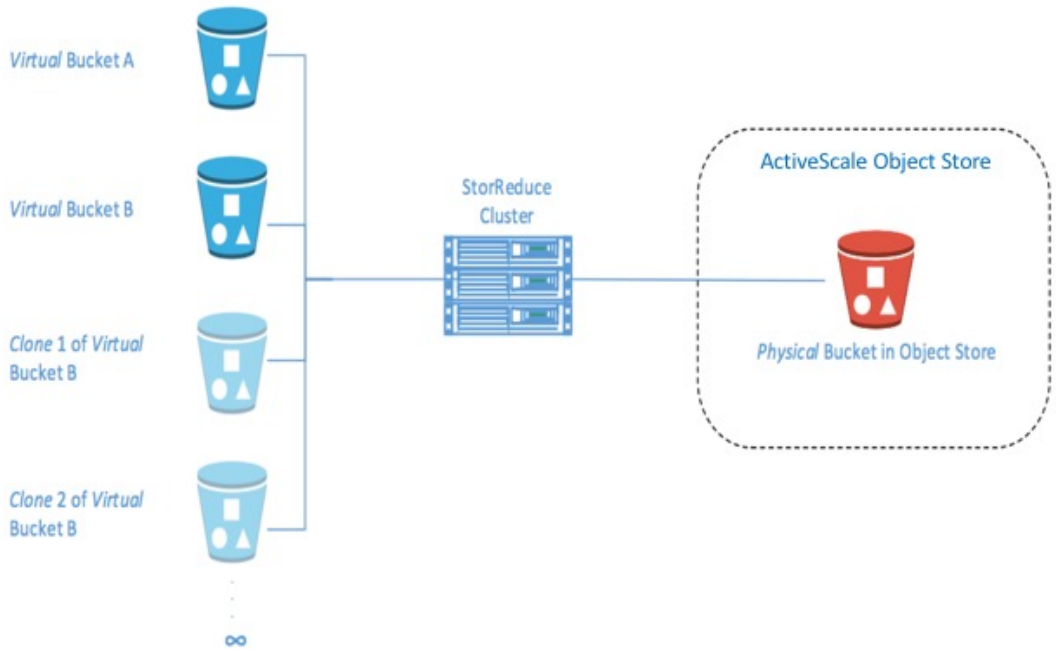


DIAGRAM 3: CLONING OF OBJECT DATA

Security

Security is essential to any viable cloud storage solution, and StorReduce is designed from the ground up to protect data at rest and in motion. In addition to using the security of your underlying object storage service, StorReduce provides the following security capabilities:

- **Client-Side Data Encryption:** StorReduce supports client-side data encryption, performed in real-time as data is passed through the StorReduce server. This allows data to be encrypted on-premises before being sent to the cloud. Key management can be performed using Amazon's KMS (Key Management Service) or via a KMIP compatible key server.
- **Cloud Encryption-at-Rest:** As an alternative to client-side encryption, StorReduce supports the use of encryption-at-rest services from cloud service providers, such as Amazon's S3 Server-Side Encryption for Data at Rest.
- **In-transit Data Encryption:** The StorReduce server requires HTTPS encryption for dashboard requests by default and can be configured to require HTTPS encryption for all S3 API requests and object storage requests. You can upload and set server certificates through the StorReduce dashboard.
- **User Account Management:** StorReduce maintains a set of user accounts for each StorReduce deployment. You can use these user accounts to provide people with limited access to the StorReduce dashboard, or to provide people or programs with limited access to the S3 API. You can always revoke individual user accounts to instantaneously cut off access.
- **Client Access Keys:** Each user account can have multiple access keys used for accessing the S3 API. These work in the same way as access keys managed by Amazon's IAM service. Individual access keys can be revoked using the StorReduce dashboard.
- **Digital Signatures:** All requests from S3 clients must be digitally signed using a secret access key tied to a StorReduce user account. StorReduce accepts AWS version 2 and version 4 signatures.
- **Data Segregation:** StorReduce supports the creation of multiple storage buckets for data segregation, with different access rights for each bucket.
- **Policy-Based Access Control:** Enterprise security policies can be expressed using the StorReduce policy engine, which supports Amazon's IAM policy language. You can apply access control to buckets using bucket policies (compatible with Amazon S3 IAM bucket policies).
- **Amazon AWS Role Credentials:** A StorReduce server running in an EC2 instance can make use of AWS IAM Roles to securely obtain credentials for accessing its underlying object storage, enabling automatic key rotation.

Use Cases

StorReduce can deliver significant business and technical advantages in a number of different use cases. The following sections detail some of the ways that enterprises can use StorReduce to improve data backup, simplify operations, unlock new value from their existing data, and more.

Using StorReduce for Primary Backups Straight to Object Storage: Removing Purpose-Built Backup Appliances

One of the most important steps an enterprise can take to slash capital and operational costs is eliminating purpose-built backup appliances and backing up data directly to cloud. StorReduce's scale-out architecture and highly optimized design, along with leading object storage solutions like ActiveScale, make it easy. By combining StorReduce with solutions like ActiveScale, organizations can achieve faster speeds for both ingest and recovery than the largest purpose-built backup appliances.

Using StorReduce from an on-premises installation of a private object store, LAN-speed recovery of primary backups becomes fast and simple. StorReduce also optimizes recovery of primary backups from a public cloud such as Amazon Web Services, Google Cloud, or Microsoft Azure. For these scenarios, enterprises can recover *in full in the cloud* or *recover back to on-premises*. The latter case requires a LAN-speed connection such as AWS Direct Connect or its equivalent. Or, enterprises can use a hybrid cloud installation with archival data storing to public cloud, while primary backups store to private object storage.

StorReduce can deploy in the head nodes of private object stores, further reducing data center footprint and costs.

Backup/Tape Migration to Cloud

Shifting from tape to cloud storage can drive major cost savings. According to 451 Research's Enterprise Storage Survey 2016, "Moving data off tape to (mainly) public cloud object storage ...for backup and disaster recovery ... is the largest swing in most enterprises IT budgets for 2017, with an estimated 12% drop in on-premises IT spend over this period." In the same survey, 35% of enterprises said they plan to redesign their backup and/or disaster recovery infrastructures to include cloud over the next year.

Deduplication plays a big role in those savings. Tape archives generally contain periodic full backups with multiple copies of the same data sets, which can be reduced to a single copy with deduplication. By shifting to cloud storage with deduplication, enterprises can reduce the amount of data they store down to as little as 1/30th of current data volumes.

StorReduce makes migrating tape or disk-based backup simple. By installing StorReduce software on-premises, enterprises can initiate a very fast migration of their large tape archives and backup appliance data to cloud, while minimizing bandwidth during the transfer.

Once organizations are using cloud, StorReduce also makes it possible to realize much more value from backup data. Enterprises can deploy a StorReduce read-only endpoint to make the data available on-cloud for development, testing, quality assurance, and disaster recovery.

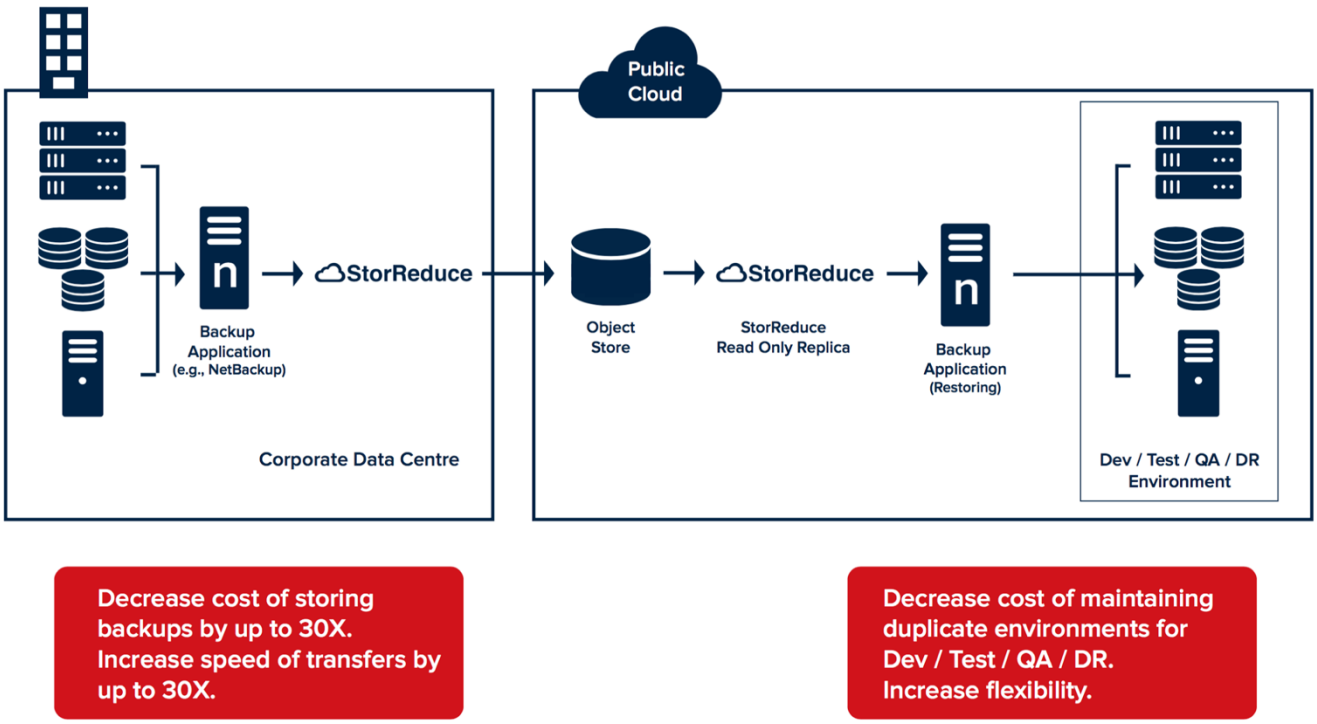


DIAGRAM 4: ON-PREMISES BACKUP TO CLOUD MIGRATION

Cloning Large Data Sets on Cloud or in Object Stores

StorReduce Object Clone makes it simple for enterprises to repeatedly produce fully isolated, copy-on-write (COW) clones of buckets containing millions of objects and petabytes of data, at virtually zero cost. Cloned buckets contain a comprehensive snapshot of the data at a single point in time. Enterprises can use that data for a wide range of purposes without worrying that changes and deletions from developers, researchers, and experiments will cause data corruption.

Object Clone benefits any kind of data that can be stored in the cloud, whether generated from Internet of Things appliances, financial trading, life sciences, multimedia (images, videos, music/audio etc.), or even backup workloads. Object Clone benefits data that deduplicates, as well as data that does not deduplicate. It brings state-of-the-art volume cloning capabilities to object storage and extends the utility of an organization's object storage investment well beyond basic data backup.

With Object Clone, organizations can:

- **Protect Against "Fat Fingers" or Malicious Deletion of Data in Object Store:** Administrators can now efficiently make time-based clones of entire buckets and assign read-only access for "locking in" protected data at critical points in time.
- **Capitalize on Big Data, Internet of Things, and Research Use Cases:** Enterprises can clone petabyte-scale data sets so that researchers can work independently of each other in their own scratch areas.
- **Streamline IT Operations:** Object Clone allows organizations to test new versions of software against their data set in isolation.
- **Optimize Development and Software Testing:** Organizations can clone test data sets so that developers and testers can work with the whole data set, not just a small subset, in isolation. They can roll the state back after failures and retest rapidly.
- **Improve Software Quality Assurance:** Software teams can take a clone of an entire system at the point that it fails a test and hand it back to the developer.
- **Simplify Hadoop and Big Data Implementations:** Organizations can use Object Clone to take point-in-time snapshots of the state of a cluster and roll back to a previous state after a problem.

Replacing Traditional Storage Solutions with Object Storage

Object storage within the data center can provide significant cost and scalability advantages over traditional storage solutions. Private object store products such as ActiveScale can provide protected, redundant storage spread over multiple locations for a fraction of the cost of traditional solutions.

Data deduplication is an essential part of any modern storage solution, and StorReduce can handle the high throughput and data volumes required of modern private object store solutions. By using a StorReduce scale-out cluster in front of ActiveScale object store, enterprises can further reduce storage costs while still maintaining very high throughput rates (10 Gigabytes per second or more).

On-Cloud Deduplication for Backups

Many organizations would like to move their entire IT infrastructures to cloud and eliminate the expense and hardware silos of on-premises purpose-built backup appliances. By doing so, they can reduce total cost of ownership (TCO), improve disaster recovery, eliminate hardware refresh cycles, and improve the scalability of recovery. StorReduce can play a significant role in these cost savings, reducing the cost of storing primary backups on-cloud down to as little as 3%.

To enable deduplication of backups on-cloud, StorReduce software can be installed directly from leading Cloud Marketplaces (Amazon, Azure, Google Cloud), and integrate with most on-cloud backup offerings, including Veritas NetBackup. By deduplicating backup data inline, StorReduce removes up to 97% of the data that organizations are currently paying to store.

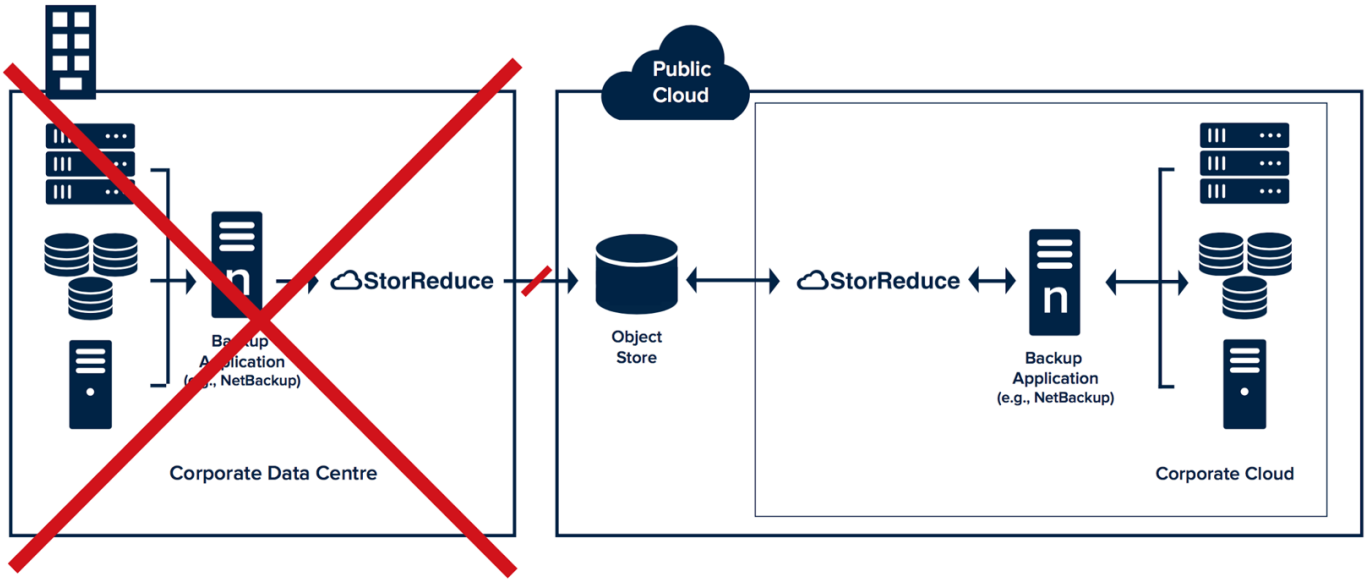


DIAGRAM 5: ON-CLOUD DEDUPLICATION

Moving or Replicating Data Between Clouds

Many organizations use more than one public cloud or have data in a hybrid cloud system. If a multi-cloud strategy isn't going to add huge new layers of complexity, however, organizations need to be able to quickly and affordably replicate or move their data from one cloud to another while minimizing storage costs.

StorReduce makes it easy to use multiple clouds with automated cloud-to-cloud replication. Any organization with data in more than one cloud can install StorReduce software on each cloud, and then replicate data. By doing so, they can reduce data size to as little as 3% and reduce bandwidth and data transfer costs by up to 30 times. The data they move is accessible by any cloud service. In addition, by reducing data volume, StorReduce enables organizations to affordably keep data in two clouds or two cloud regions to satisfy redundancy or compliance requirements. Even within a single public cloud, organizations can use the same capability to replicate data from region to region.

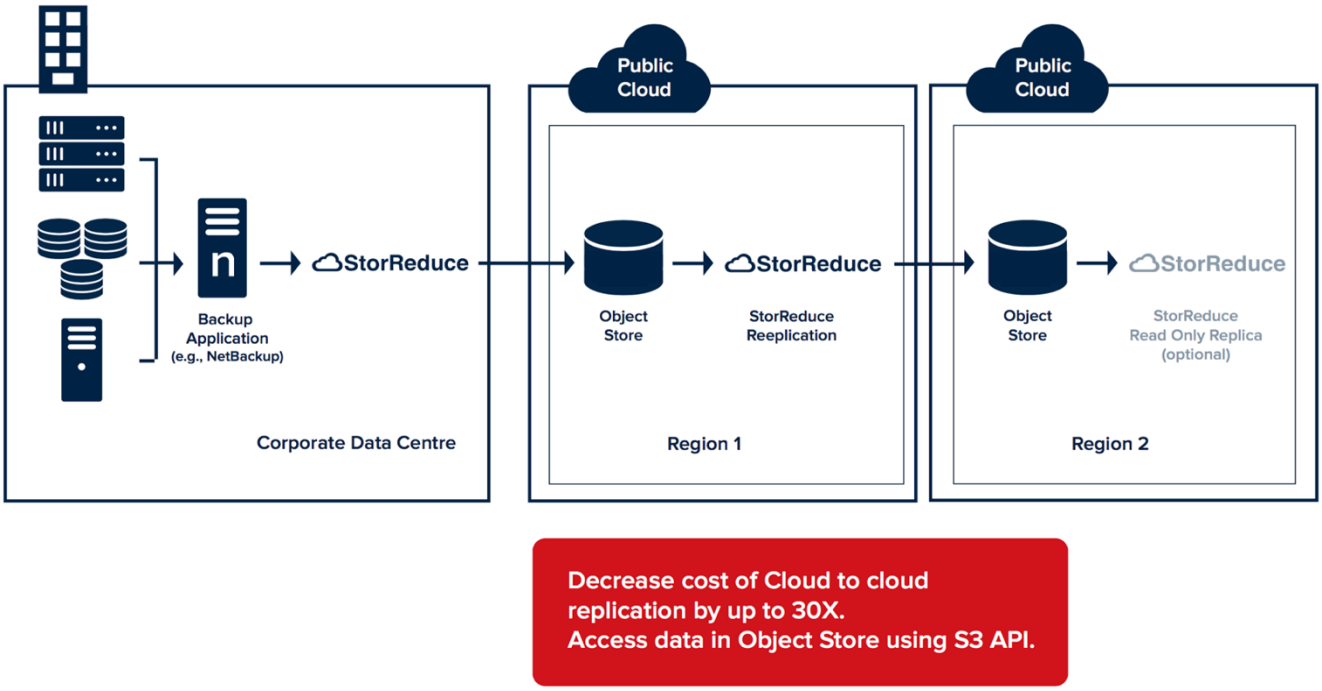


DIAGRAM 6: CLOUD-TO-CLOUD DATA REPLICATION

Optimizing Big Data/Hadoop Storage

More organizations are using big data and Hadoop every day. With state-of-the-art deduplication and storage efficiencies, StorReduce can make big data analysis tools much more efficient and affordable.

Enterprises can use StorReduce in conjunction with products like WANdisco Fusion for backups, snapshots, and cloning of storage for Hadoop clusters and other big data solutions.

Is Your Storage Keeping Up with Your Business?

A sound data protection strategy has never been more important. But too many enterprises are wrestling with soaring storage capital costs, complex hardware appliances, and siloed data. Now, StorReduce inline deduplication software, combined with high-density ActiveScale object storage, makes it easy to capitalize on the power of cloud storage. Better protect your business. Efficiently manage massive amounts of data. And slash your storage costs by up to 70%.

To learn more, visit www.hgst.com/storreduce