StorReduce Technical White Paper
Scale-out data deduplication, cloning and replication with a single name space for Object Storage

See also at http://storreduce.com/docs:
- StorReduce Quick Start Guide
- StorReduce FAQ
- Use Case Reports
- StorReduce Blog at storreduce.com/blog

Published May 2017, v5.0
StorReduce is a specialized cloud deduplication solution, designed to meet the unique requirements of companies using object storage (public cloud or private object store) for large volumes of data. StorReduce sits between your applications and the object store, transparently performing inline data deduplication. Storage and bandwidth requirements are reduced by as much as 30 times, and data transfer times can be reduced by the same amount.

StorReduce is fast and massively scalable. Sustained write speeds of 10 gigabytes/s or higher are possible using multiple servers deployed as a scale-out cluster, and hundreds of petabytes of data can be managed. Data deduplicated with StorReduce can always be accessed on-cloud or on-premises using the S3 API.

Diagram 1: StorReduce Architecture
StorReduce Overview

Key Characteristics

- **Reduces cost**: Reduces storage and data transmission costs up to to 97%.
- **Fast sustained throughput**: Writes speeds of 10 gigabytes/s or more using a scale-out cluster, adding less than 50ms of latency. Each individual StorReduce server can deduplicate data at speeds of 2 gigabytes/s or more. StorReduce is capable of sustained throughput 24/7 – it does not stage the data in buffers that can fill up.
- **Massively Scalable**: A scale-out cluster can manage almost unlimited amounts of data (hundreds of petabytes or more) in a single deduplication pool. Each individual server can manage tens of petabytes of data.
- **Cloud-native**: Deduplicated data is immediately accessible to cloud services via StorReduce’s S3 REST API. Data can be stored on Amazon S3 or S3IA, Microsoft Azure Blob Storage or Google Cloud Storage.
- **Private Object Store Compatible**: Data can be also stored in any S3-compatible private cloud object store including Cloudian HyperStore, IBM COS, HDS HCP, EMC ECS, Scality and HGST Active Archive & Active Scale.
- **Resilient**: Index information required to re-hydrate the data is stored in multiple locations, both in object storage and on each StorReduce server. Extensive use of checksums ensures data integrity during both deduplication and re-hydration.
- **Software-only solution**: No hardware purchases required, free from cost and lock-in.

Key Architectural Features

- **Scale-out clusters**: Data storage capacity and write throughput increase almost linearly by adding more StorReduce servers to a cluster, while maintaining a single deduplication pool and a single namespace for the data.
- **Backup Software Integration / Data Management**: Works with existing data management or backup software that is compatible with Amazon S3, including Veritas NetBackup, Backup Exec, Commvault Simpana, Veeam and EMC Networker.
- **Object Clone**: StorReduce can create writable ‘virtual clones’ of all the data in a storage bucket, without using additional storage space. Unlimited virtual clones can be created 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**: Additional StorReduce servers or clusters can be deployed on-cloud or on-premises as read-only endpoints. Uploaded data is only stored once but is immediately available at each endpoint location, enabling migrated backup workloads to
be re-purposed on the cloud for development, test, QA, disaster recovery and to be used by cloud-based services.

- **Data Replication**: StorReduce can replicate data between regions, between cloud vendors, or between public and private cloud, providing increased data resilience. Only the unique data is transferred, providing up to 30 times speedup on transmission and up to 30 times reduction in bandwidth and storage costs.
- **High Availability**: When configured as a scale-out cluster StorReduce provides automatic failover in the event of server failure, and can be set up with any desired degree of redundancy between servers.
- **Write Speed Throttling**: A maximum write speed can be set, to prevent StorReduce using too much bandwidth when sharing an Internet connection with other infrastructure.

**Key Security Features**

- **Encryption**: StorReduce supports encryption of data before it lands in the object store. StorReduce 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 cloud encryption-at-rest services can be used. Data is *always* encrypted in-transit.
- **Secure User Account and Key Management**: Users or servers can be given individual user accounts within StorReduce, allowing data access to be restricted. Multiple access keys can be created and managed as needed for each user account.
- **Secure Policy-based Access Control**: Enterprise security policies can be expressed using StorReduce policy engine, using Amazon’s IAM policy language.
StorReduce Architecture

StorReduce sits between client programs (wanting to store and retrieve data) and the Object Storage service they are using - See Diagram 1. It transparently provides best-of-class data deduplication and high throughput.

StorReduce can be deployed as either a single server, or as a scale-out cluster of servers. In either case client applications see a single logical server.

StorReduce server software can be deployed on public cloud (e.g. Amazon EC2), or on-premises in a VM, Docker container, or installed natively on Linux.

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

S3 Client Software

StorReduce works with client software that supports Amazon’s S3 REST interface for object storage. This includes clients designed to work with Amazon S3 and those designed to work with Google Cloud Storage via the XML API (which is S3-compatible). Client software is configured to talk to StorReduce instead of directly to object storage, using access keys provided by the StorReduce server.

S3 client software includes on-premises backup software (including Veritas NetBackup) as well as custom software written to use the S3 REST interface. Cloud-based services designed to work with Amazon S3 or Google Cloud Storage can also be used with StorReduce - these also act as S3 clients.

StorReduce translates S3 client requests into whichever protocol is needed by the underlying object storage providers. StorReduce 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

File sharing interfaces (particularly CIFS and NFS) are supported via gateway software such as QStar that exposes a file share interface and converts requests into calls to Amazon’s S3 protocol.

StorReduce Server(s)
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. For lower data volumes, magnetic disk can be used instead of SSD (see 'local SSD' section below).

StorReduce supports the creation of multiple storage buckets, with global deduplication performed across all buckets.

For public cloud the StorReduce server runs on Amazon EC2, Azure Virtual Machines or Google Compute Engine. VM images are available through AWS Marketplace, Azure Marketplace and Google Cloud Marketplace with the server pre-installed, allowing quick and easy setup. StorReduce can also easily be installed on any Linux virtual machine, using rpm.

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.

The architecture is designed to allow 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 server might be used to deduplicate and upload backup data, with a second in-cloud StorReduce server providing immediate access to this data for cloud services as the data is uploaded.

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: A separate REST interface is exposed 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 optionally also be served over HTTPS on port 443. 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, index snapshots as well as providing a replica of the S3 API for use by management tools.

Local SSD or Magnetic Disk Storage

Each StorReduce server stores index information on local storage, and requires fast access to that data to achieve high throughput for deduplication. 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.
For large data volumes StorReduce uses local SSD storage for this index, allowing tens of petabytes of data to be managed by a single StorReduce server. Typically, the amount of SSD storage required is less than 0.1% of the amount of data put through StorReduce, for a standalone server.

For relatively low data volumes StorReduce can be run using magnetic disk instead of SSD, and will use available RAM to cache the information stored on magnetic disk. This works for up to around 100TB of data (before deduplication), depending on the deduplication ratio achieved.

Local SSD or magnetic storage is treated as ephemeral by the StorReduce server. 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 an Object Store (public cloud or a private object store) for all persistent data. It acts as an Amazon S3, Azure Blob Storage or Google Cloud Storage client, making use of the corresponding object storage API to store all its data in a single account. StorReduce is architected to work with any S3-compatible object storage solution, including Amazon S3, as well as with Azure Blob Storage.

For private or hybrid cloud deployments StorReduce can use any S3-compatible private object store. These include IBM Cloud Object Storage, Cloudian HyperStore, Hitachi Content Platform, HGST Active Scale, EMC ECS, Scality and others.

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 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 (see later section).
The StorReduce server is optimized for scalability, high throughput and low latency. The internal architecture and code are highly optimized for data deduplication, and to ensure that performance is maintained 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 speeds have been achieved on Amazon EC2 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. Throughputs of 10 Gigabytes per second or more are possible by distributing the load over all the servers in the cluster, while maintaining a single deduplication pool. Deploying a StorReduce cluster in front of a private object store installation can cut storage requirements while maintaining very high throughput rates.

Running a StorReduce server or cluster on-premises can significantly speed up throughput and decrease transfer bandwidth to cloud-based storage by deduplicating data prior to sending it into the cloud, and by reading deduplicated data from the cloud and reconstituting it locally. StorReduce’s fast throughput reduces migration times from years to weeks and greatly reduces the cost. See direct independent comparison of StorReduce vs a well known migration vendor at http://www.storreduce.com/case-studies/apn-spectrumdata/.

Latency is kept to a minimum, typically less than 50ms of additional latency even when StorReduce is running in the cloud. For most situations, this 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 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 then 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 Amazon EC2, 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

StorReduce operates a 9-server cluster on Amazon EC2 to provide real-time performance demonstrations. This cluster is continuously (24x7) ingesting data at ¾ PB per day (8.8 GB/s). This cluster is simultaneously able to recover data at 4.8 GB/s (double the largest Data Domain). StorReduce clusters can scale to 31 servers, and speeds scale almost linearly. A similar sized cluster on physical hardware will operate at even faster speeds, as physical hardware is typically able to provide more CPU resources than public cloud.

Diagram 2: StorReduce Demostration Cluster Dashboard

The screengrab above shows the StorReduce demonstration cluster ingesting data at 8.8 GB/s whilst achieving a 95% deduplication rate. StorReduce is able to provide live demonstrations using this system on request.
**Multiple StorReduce Servers**

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. New servers can be set up to talk to an existing Object Storage service and they automatically populate their local index data from Object Storage.

**Scale-out Clusters**

A Scale-out Cluster uses multiple StorReduce servers to distribute load and provide automatic failover within a single data centre or region.

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 is automatic, coordinated using etcd.

Any S3 protocol request can be processed by any StorReduce server, so load can be distributed by DNS round-robin or load balancer 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.

**Read-Only Endpoints (making data available in multiple locations)**

A StorReduce server or scale-out cluster can be deployed as a read-only endpoint, allowing reading and re-hydration of data. This allows the same content to be fetched from multiple StorReduce servers in different locations, each with the same view of the content updated in real-time.

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 running in the cloud as a read-only endpoint can provide real-time access to the data (in re-hydrated form) to cloud-based applications and services via its S3 interface. This works particularly well for moving backups to cloud.
Data Replication (multiple copies of data)

A StorReduce server can be configured to automatically replicate data from its primary object store to one or more other object store locations. Any changes seen by this StorReduce server will be copied to the other location(s). Because only deduplicated data is replicated, data transfer charges are reduced to a tiny fraction of the regular cost.

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, Google Cloud Storage)
- from private to public cloud, or vice versa (hybrid cloud, including IBM COS, Cloudian, HCP and HGST).
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. Copy-on-write semantics mean that clones do not take any extra object storage space until data is written to the clone, and 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 and will not be able to see each other's changes or new data. This can be achieved while maintaining a single deduplication pool across the entire system.

Software can be tested against a cloned bucket and the software will have full control over all data in its bucket, without being able to interfere with any other clone or the original data.

**Diagram 3: Cloud-to-cloud Data Replication**
Security

Security is extremely important for any cloud storage solution. As well as leveraging the security of your underlying object storage service, StorReduce provides the following 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 by default requires HTTPS encryption for dashboard requests, and can be configured to require HTTPS encryption for all S3 API requests and object storage requests. Server certificates can be uploaded and set through the StorReduce dashboard.

**User Account Management:** StorReduce maintains a set of user accounts for each StorReduce deployment. User accounts can be used to provide people with limited access to the StorReduce dashboard, or to provide people or programs with limited access to the S3 API. Individual user accounts can be revoked 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, supporting Amazon’s IAM policy language. Access control can be applied 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.
1. Using StorReduce for Primary Backups Straight to Object Storage: Removing Purpose Built Backup Appliances like Data Domain

StorReduce’s scale-out architecture and highly optimized design, combined with leading Object Storage (such as IBM COS, Cloudian, HGST Active Archive, ECS, Hitachi Content Platform, AWS S3, Google Cloud Storage, …) enables StorReduce to achieve greater speeds on both ingest and recovery than the largest purpose built backup appliances: Data Domain systems, Veritas Appliances, ExaGrid, HPE StoreOnce, NEC HYDRAstor, Quantum DXi’s.

LAN speed recovery of primary backups is simple and fast with StorReduce from an on-premises installation of a private object store. For recovery of primary backups from a public cloud such as Amazon Web Services, Google Cloud or Microsoft Azure:

- a company can recover in full in the cloud,
- or for recovery back on-premises:
  - a LAN speed connection such as an AWS Direct Connect or its equivalent is required, or
  - a hybrid cloud installation with archival data storing to public cloud and primary backups storing to private object storage is appropriate.

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

2. Backup / Tape Migration to Cloud

‘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,’ according to 451 Research’s Enterprise Storage Survey 2016. 35% of enterprises surveyed plan to redesign their backup and/or disaster recovery infrastructures to include cloud over the next year.

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. This has the potential to reduce the amount of data stored down to as little as 1/30th.

For tape or disk-based backup migration, StorReduce software can be installed on-premises for a CAPEX-free, very fast migration of an enterprise’s large tape archives and backup appliance data to Cloud. Installing StorReduce on-premises minimizes bandwidth during the transfer.

A StorReduce read replica can be deployed to make the data available on-cloud for development, test, quality assurance and disaster recovery.
3. Clone Large Data Sets on Cloud or in Object Stores

StorReduce Object Clone enables you to repeatedly produce fully isolated, copy-on-write (COW), clones of buckets containing millions of objects and petabytes of data, at virtually zero cost. Your cloned bucket contains a comprehensive snapshot of your data at a single point in time without the worry of changes and deletions from developers, researchers and experiments causing data corruption.

Object Clone benefits any kind of data that can be stored in the cloud - whether it be 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 is like NetApp FlexClone but for object storage and for use cases extending past backup.

For more information about how StorReduce works with Veritas NetBackup, see the 'Configuring NetBackup 7.7 with StorReduce' guide, available at storreduce.com/docs.
With Object Clone you 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.
- **Big Data, Internet of Things, Research** – Clone petabyte scale data sets so that researchers can work independently of each other in their own scratch areas.
- **IT Operations** – Test new versions of software against your data set in isolation.
- **Developers, Software Testers** – Clone your test data sets so that developers and testers can work with the whole data set, not just a small subset, in isolation. Roll the state back after failures and retest rapidly.
- **Software Quality Assurance** – Take a clone of an entire system at the point that it fails a test and hand it back to the developer.
- **Hadoop, Big Data** – Take point in time snapshots of the state of your cluster and rollback after problems.

4. **Replace 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 IBM COS, Cloudian HyperStore, Hitachi HCP and HGST Active Scale can provide protected, redundant storage spread over multiple locations for a fraction of the cost of more 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 an object store product, storage costs can be further reduced while still maintaining very high throughput rates (10 Gigabytes per second or more).

5. **On-Cloud Deduplication for Backups**

Many customers now wish to move their entire IT infrastructures to Cloud and remove the expense of their on-premises purpose built backup appliances such as EMC Data Domains. This can provide better TCOs, improved disaster recovery, eliminate hardware refresh cycles and improve scalability of recovery. StorReduce can reduce 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 your chosen Cloud Marketplace and is integrated with most on-cloud backup offerings, including Veritas NetBackup. It deduplicates backup data inline to remove up to 97% of the data stored.
6. Moving or Replicating Data Between Clouds

Many organizations make use of more than one public cloud or have data in a hybrid cloud system, and want to quickly and affordably replicate or move their data from one cloud to another while minimizing storage costs.

StorReduce includes automated Cloud to Cloud replication. Any customer with data in more than one cloud can install StorReduce software on each cloud, then replicate data. Data size can be reduced to as little as 3%, and bandwidth and data transfer costs can be reduced by up to 30 times. The data moved is accessible by any cloud service. In addition, StorReduce’s reduction of the data volume enables organizations to affordably keep data in two clouds or two cloud regions to satisfy redundancy or compliance requirements.

Data can be replicated from region to region within a single public cloud in the same way.
7. Big Data / Hadoop Storage

StorReduce can be used to provide deduplication and storage for Big Data analysis tools.

In conjunction with products like WANdisco Fusion, StorReduce can be used for backups, snapshots and cloning of storage for Hadoop clusters and other Big Data solutions.

Please contact StorReduce for more information about deduplication and object cloning for Big Data storage.