

Incremental Backup for S3 Storage

Key Benefits

- > Fast incremental backup
- > *Low RTO and RPO with continuous backup*
- > *Isolated backup protection vs ransomware*
- > *Prioritized recovery options*
- > *Works with existing S3 Storage buckets*
- > *HA deployment options*
- > *Storage Pools for multiple targets*

S3 storage is evolving

S3-compatible object storage has evolved from low-cost cold storage to a next generation platform designed to meet the evolving needs of organizations managing massive volumes of unstructured data. On-site S3 storage products like Dell ObjectScale, Dell ECS, IBM iCOS, NetApp StorageGrid, Hitachi HCP, Scalify Ring and Minio Alstor are now being used as primary storage. In addition, there are a number of S3 compatible cloud providers including: AWS, Wasabi and Oracle. As a result, this data needs to be backed up and protected just like any other primary storage system.

Why You Can't Easily List "New Files" in S3

S3 compatible object storage systems are designed for massive scalability and simplicity, but that comes with trade-offs:

- **Stateless operations:** S3 APIs are designed to be stateless and simple. This makes them highly scalable but limits features like incremental change tracking.
- **Flat Namespace, No File System Semantics:** S3 stores objects in a flat namespace within buckets. There's no concept of folders or file system metadata like.
- **No Built-in Change Log or Index:** S3 doesn't maintain a native index of recently added or modified objects. To find new files, you must **list all objects** and compare timestamps or versions — which can be slow and expensive for large buckets.
- **Eventual Consistency (in some systems):** Some S3-compatible systems offer **eventual consistency** for listing operations, meaning newly added objects might not appear immediately in a list. This makes real-time change tracking unreliable without application add-ons or external tooling.

Due to these limitations determining new, modified or deleted items is computationally expensive and time consuming. In use cases like backup, large object stores with more than 200 million objects may take more than 24 hours to process. Most organizations would find this situation unacceptable.

Some S3 object storage systems, like Dell ECS do utilize internal event notifications to help with the new object tracking that can be passed to an external tool for processing. However, there are a number of hard “requirements” in order to use this capability and they include:

- ECS version 3.8.0.1 or newer is used
- Copy-to-Cloud must be enabled at bucket creation
- Metadata Search must be enabled at bucket creation
- All ECS access nodes require a minimum of 172GB of RAM

Awareness of the requirements at bucket creation often trip up customers, especially if they have rolled their ECS CAS to S3 buckets to be more open. The additional tracking and computations are detracting from the primary use of the object storage system just to send a list to an external backup system.

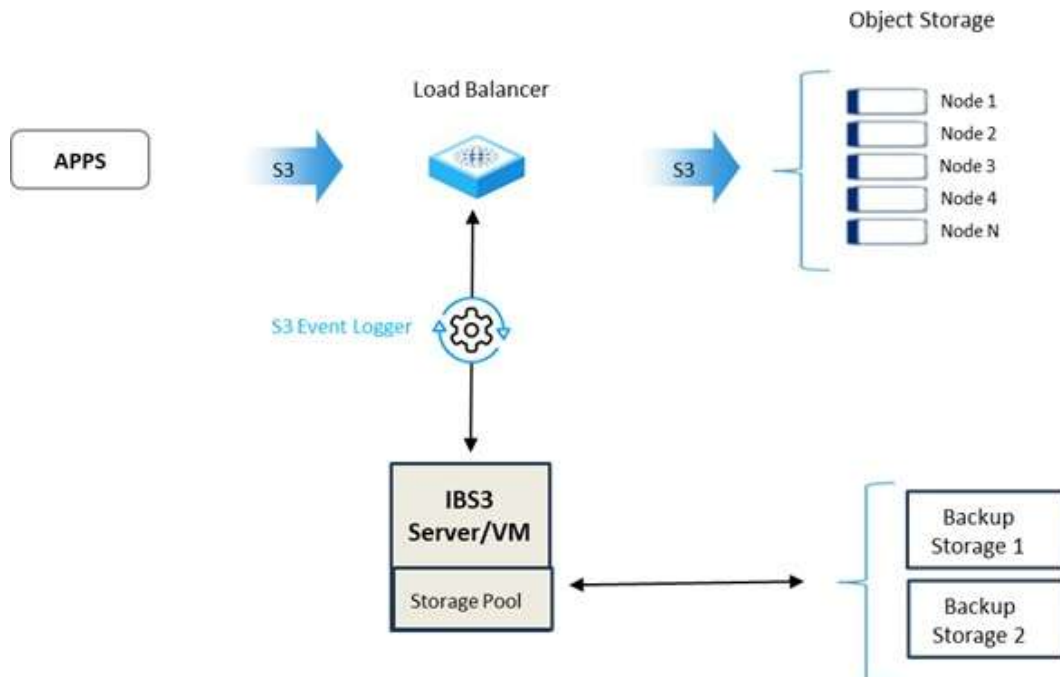
Load Balancers are Required to enable Scalable S3 Storage Performance

The ability to scale performance independent of capacity is a key selling point for S3 object storage. To increase the aggregate throughput more “access nodes” are added to the S3 object storage system along with a Load Balancer to evenly distribute the traffic.



A New Approach

Introducing our new software Incremental Backup for S3 object and cloud storage. It is unique in that it receives traffic events from HA Proxy load balancer instead of the S3 object storage system. It creates and maintains list of all S3 objects. This list contains all objects, new, or, changed for any given time period. Our software analyzes this list and creates a manifest containing the incremental objects that need to be backed up. Then our backup process runs and writes those objects to another storage system. The backup process can run continuously or be scheduled.



Isolated Backup for Protection against Ransomware

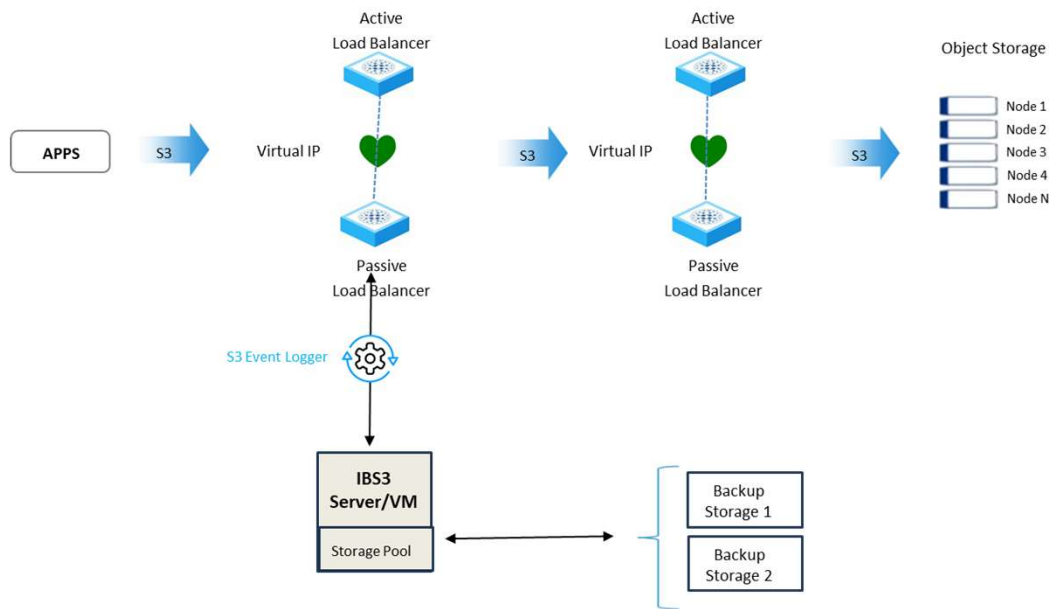
Bad actors deploying ransomware prioritize deleting backups to prevent organizations from recovering their data. Our “touchless” approach backs up data from S3 storage systems outside the context of the storage system itself, therefore making the backups invisible to the bad actors.

Traditional backup software operates on a scheduled basis and this can create potential protection time gaps. Our backup process can operate continuously providing lower RTO and RPO for customers requiring higher levels of protection.

Flexible recovery options include: a single object, a list of objects, a date range or entire bucket. This enables your system administrators to prioritize a recovery process if needed.

Flexible Deployment Options

Every organization has different requirements and our Incremental S3 object and cloud storage software can be deployed in a variety of configurations to meet your unique needs. Our solution consists of a customized HA Proxy Load Balancer and the Incremental Backup S3 server. The Load Balancer can be deployed as a single load balancer, active-passive pair or active-active pair that directly interact with the S3 storage or in front of existing Load Balancers. The backup software can run on a single server or active passive pair. Note some S3 object storage vendors want HTTPS traffic terminated at the load balancer and just HTTP traffic to the S3 access nodes and this is a supported option as well as HTTPS straight through.



DataTrust Solutions

Proven software solutions and services for organizations in health care, finance, government and compliance driven market segments. Our customers range from small medical clinics to Fortune 500 enterprises. Common to all our customers is the need to protect, move or recover their data in a verified manner.

For more information please contact us at:

+1 720-675-9070

info@DataTrustSolutions.com.