COHESITY

WHITE PAPER

Cohesity's OST and SpanFS: Advancing cyber resilience for NetBackup deployments

An inside look at how DirectIO, the upcoming OST-SpanFS integration, delivers breakthrough security, speed, and TCO advantages to NetBackup

TABLE OF CONTENTS

Executive summary

Cohesity's flagship platform, **Cohesity Data Cloud**, is expanding to include tighter integration with **Cohesity NetBackup**. This combination will be enabled by DirectIO - an enhanced version of the widely-used NetBackup OST protcol. DirectIO allows NetBackup data to be stored on **SpanFS**, the file system that underpins Cohesity Data Cloud and more. When complete, DirectIO will deliver superior business outcomes for NetBackup customers, including: rapid cyber recovery, linear scalability, simple replication, and improved total cost of ownership. In addition, DirectIO will enable more advanced use cases for NetBackup deployments, including enhanced threat detection, cyber vaulting, and AI-powered data insights.

This white paper details the technical characteristics of the new DirectIO protocol and the upcoming integration between NetBackup and SpanFS.

Editor's note:
For business context, we recommend reading "***Cohesity Data Cloud: A Unified Platform for Superior Cyber Resilience and Economic Outcomes.***"

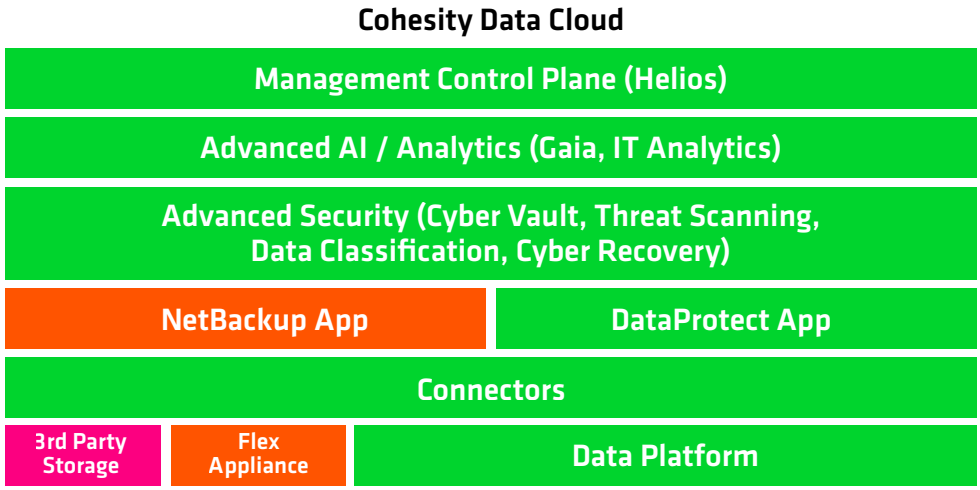
Introduction

The market challenges for data protection, security, and cyber resilience are varied and dynamic. In response to these challenges, Cohesity is bringing Cohesity NetBackup into its flagship platform, Cohesity Data Cloud.

How will NetBackup join with Cohesity Data Cloud? With DirectIO, a new and enhanced version of the widely-used OST protocol, that enables NetBackup to connect to SpanFS, the file system that powers Cohesity Data Cloud.

This white paper describes the business advantages—and technical implementation of—DirectIO.

As you read through the following sections, keep the target architecture of Cohesity Data Cloud in mind—it's depicted in the figure below. This unified architecture provides benefits to both NetBackup and DataProtect deployments.



Terms to know

There are three important terms to know in this white paper. Let's define them now.

- **DirectIO** – This is a new and enhanced protocol based on the widely-used OST protocol, currently under development by Cohesity. When generally available, data protected by NetBackup may be stored in Cohesity's file system. This connection unlocks exciting new capabilities for NetBackup deployments. DirectIO goes beyond traditional OST features; we will describe these in more detail.
- **Cohesity SpanFS** – SpanFS is the distributed, web-scale file system used in Cohesity Data Cloud today. The system is designed to consolidate and manage secondary storage for use cases such as backups, file shares, object storage, test/development, and analytics. Unlike traditional file systems optimized for specific scenarios, SpanFS supports multiple protocols (NFS, SMB, S3, and soon OST) simultaneously on the same data volume, enabling seamless access across on-premises, cloud, and edge environments.
- **Cohesity SmartFiles** – An object storage service for unstructured data, built atop SpanFS. With SmartFiles, data stored in the OST format can be accessed via SpanFS. In the context of this whitepaper, SmartFiles is a key enabling technology between NetBackup and other Cohesity Data Cloud capabilities.

Expect these business outcomes from DirectIO

DirectIO, currently in development, will connect the NetBackup application to the file system that powers Cohesity Data Cloud. In turn, this innovation enables the rich capabilities of the Cohesity Data Cloud to be available to NetBackup users.

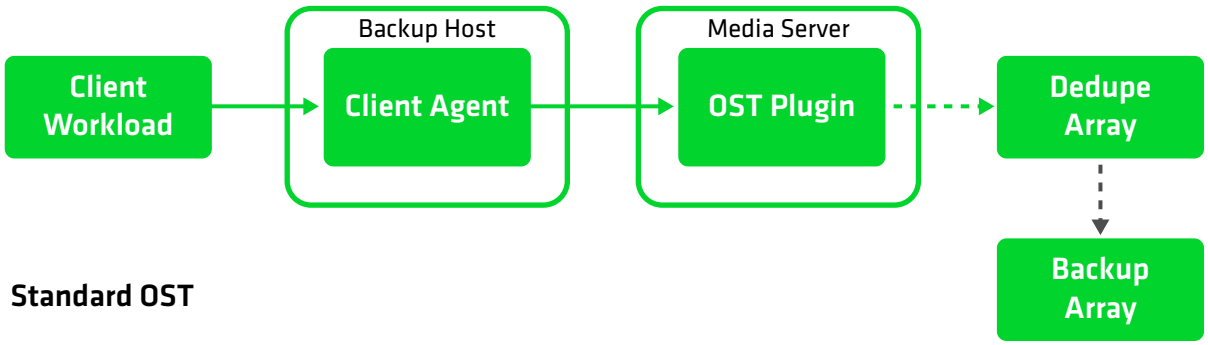
The integration is designed to enable these outcomes:

- Provide future-proof data storage that works with all DataProtect and NetBackup applications.
- Greater cyber recovery speed and scale.
- Improved cyber response capabilities with enhanced threat scanning and threat hunting.
- More flexibility for deployment topologies with a greater set of cyber vaulting options.
- Linear scalability, with constant performance as new nodes are added.
- The best price/performance combination of any OST-based implementation.
- The lowest TCO option for NetBackup deployments.
- Greater CapEx flexibility, including "pay-as-you-grow."
- More options for hyperconverged hardware from trusted vendors (such as Cisco, HPE, and Dell), including white-box options from Cohesity.

Our aim for the initial release of DirectIO is to deliver an industry-leading feature set, the lowest total cost of ownership, as well as simple scaling and upgrade processes. We also plan to conduct rigorous benchmarking. Results will be made available for maximum transparency.

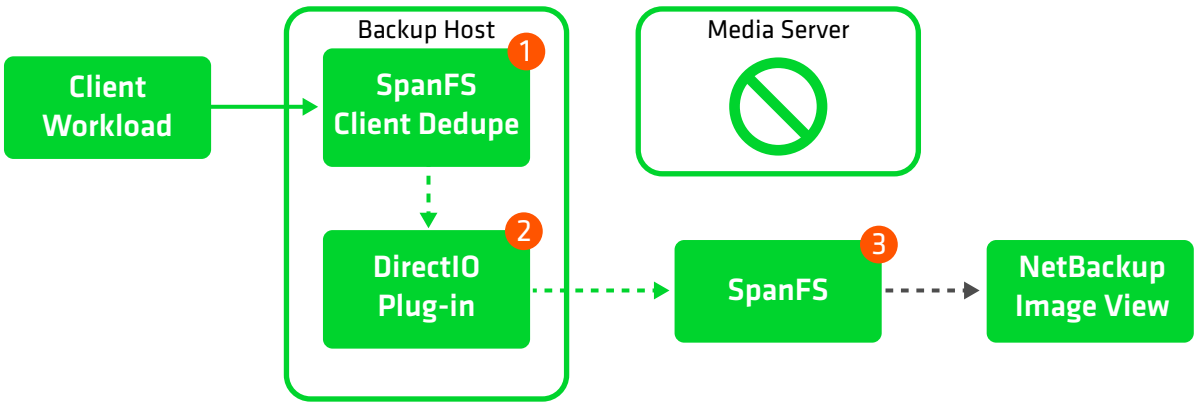
DirectIO: the anatomy of the new NetBackup-SpanFS integration

Let's start with an overview of the standard OST implementations in NetBackup today, shown in the diagram below. Multiple third-party vendors use this tried-and-true architecture to provide backup data storage for NetBackup deployments. This workflow uses a bifurcated approach with a client agent resident on the backup host and a third-party media server with an OST plugin.



Standard OST

The future state, with DirectIO, will feature a modern approach with three new components, shown below.



Let's describe this new approach in detail, in particular, how the three components will work in concert to bring NetBackup into Cohesity Data Cloud closer together.

1. SpanFS client deduplication

- This client will perform deduplication for Linux, Windows, and NAS workloads with more to follow. This is a new capability for for SpanFS, offered exclusively with the new DirectIO integration. It uses Cohesity's best-in-class easy to scale deduplication technology.
- DirectIO deduplication works directly from the original host to optimize the payload at the source, reducing bandwidth requirements. This minimizes the data throughout the system, delivering you the lowest cost, the highest optimization, and superior performance.

2. DirectIO

- Today, OST capabilities include optimized de-duplication, optimized synthetic protection, auto image replication, immutability, and inline indexing for granular recovery.
- DirectIO will build on the features above with SmartFiles views of backup image content for secondary processing (read-only) from the stored backup data. This approach will make features like instant access for VMware workloads and universal shares (for dump and sweep operations) highly scalable and immediately available at backup time.

3. SpanFS

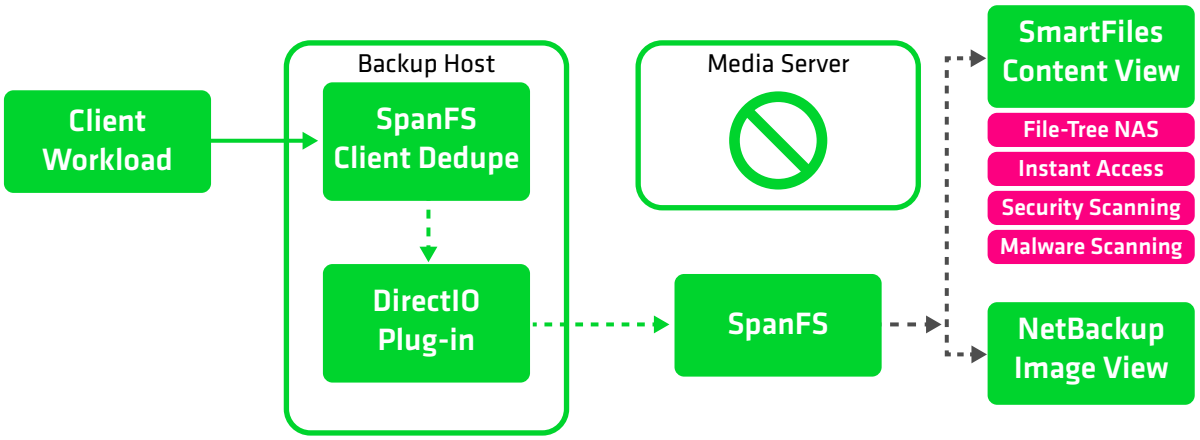
- The SpanFS file system is integral for performing additional operations on backup data without changing its native format. NetBackup will "see" data stored in SpanFS as backup sets (backed up images), and DataProtect will "see" NetBackup data images as stored in SmartFiles, allowing both applications to access protected data. In this future state, data can be accessed in either NetBackup or DataProtect formats **even though only a single optimized image of that data has been stored**.

The three components will converge previously disparate capabilities. This results in a single image stored in SpanFS, which can be accessed in four different ways:

1. NetBackup users will be able to both index and recover data.
2. DataProtect users can index and recover data in the familiar SpanFS format.
3. NetBackup users can index and recover data in the SpanFS format.
4. Data from both NetBackup and DataProtect will be accessible to additional services, including advanced security and analytics.

Scenario 1 is available today. Scenarios 2 and 3 will be unique in the industry. The diagram below illustrates this future state.

Any data stored in SpanFS will be available to additional services that are part of Cohesity Data Cloud today and in the future.



Key use cases for DirectIO

The new integration brings backup data originating from NetBackup into SpanFS. This data can then be retrieved in one of two ways. First, it can be accessed as an image by NetBackup. Second, it can be accessed as files and folders via the SmartFiles content view. This enables many new and exciting capabilities, listed below.

Back up to tape

Let's say that you'd like to send NetBackup files to a tape archive for compliance reasons. This would be done with a familiar workflow, whereby the files are retrieved in an OST format and subsequently exported to tape by NetBackup.

Cyber vault

Administrators can increase cyber resilience by adding a cyber vault to their NetBackup deployment. Here, **Cohesity FortKnox** is deployed alongside NetBackup to offer additional protection against destructive cyberattacks. The NetBackup customer can adhere to the 3-2-1 pattern of data protection.

Custom AI projects

Consider a scenario where you'd like to give an LLM access to enterprise data for an AI project. Data captured via NetBackup can be exposed to this LLM with the SmartFiles content view.

Advanced security: Threat scanning, threat hunting, and data classification

Administrators have long used advanced security scanning natively within NetBackup to reduce their risk. With the new converged platform, users will have additional security capabilities at their fingertips. Cohesity Data Cloud includes several key features for use in prevention and protection ("peacetime") and response and recovery ("wartime"). Because the data will be stored in SpanFS, users can scan NetBackup data for threats using built-in threat feeds, third-party feeds (like CrowdStrike, Mandiant), and custom YARA rules.

Similarly, the platform's **data classification** capabilities for proactive scenarios (pre-breach to identify sensitive data that's not adequately protected) and reactive scenarios (post-breach to assess risk of exfiltrated data) can be used with NetBackup deployments.

Looking ahead to the initial release of DirectIO

Now that we've described the underpinnings of the new integration, let's look ahead at the scope of the "MVP" release, scheduled for later 2025.

A summary of our MVP capabilities of the product are as follows:

- Client dedupe engine (for Red Hat Linux and Windows workloads)
- OST interface (for Red Hat Linux and Windows workloads)
- Supported workloads:
  - Linux & Unix File Systems
  - Windows File Systems
  - VMware snapshots
  - Oracle
  - NAS
- SpanFS as an immutable (WORM) storage target

Post-MVP releases will add support for additional workloads.

Conclusion

The **prior white paper** notes that Cohesity has "created a true 'sum of the parts' solution." NetBackup customers are expected to retain everything they love about that application, while also benefiting from the unique capabilities of the Cohesity Data Cloud.

DirectIO, the new OST-SpanFS integration delivers on this promise. Long-time NetBackup customers will enjoy the benefits of a future-proof platform that combines best-in-class data protection, cyber resilience, and AI-driven analytics at exabyte scale. Best of all, you can enjoy all these new capabilities without compromise—your data is always available in your required format.

A note about forward-looking statements

This document includes forward-looking statements that are subject to risks, uncertainties, and assumptions. You should not rely upon forward-looking statements as predictions of future events. All statements other than statements of historical fact could be deemed forward looking. Forward-looking statements include statements concerning new or planned products and features or service availability, and technological developments.

Although we believe that the expectations reflected in the forward-looking statements are reasonable, we cannot guarantee that the future results, performance or events reflected in the forward-looking statements will be achieved.

Any unreleased services or features referenced in this document are not currently available and may not be made generally available on time or at all, as may be determined in our sole discretion. Any such referenced services or features do not represent promises to deliver, commitments, or obligations of Cohesity, Inc. and may not be incorporated into any contract. Customers should make their purchase decisions based upon services and features that are currently generally available.

Learn more at Cohesity

COHESITY

cohesity.com
1-855-926-4374
2625 Augustine Drive, Santa Clara, CA 95054

2000062-001-EN 8-2025