
Actifio Big Data Director

Virtual Data Pipeline for Unstructured Data

Contact Actifio Support

As an Actifio customer, you can get support for all Actifio products through the Support Portal at <http://support.actifio.com/>.

Copyright, Trademarks, and other Legal Matter

© 2015 Actifio, Inc. All rights reserved.

Actifio®, AnyIT®, Dedup Async®, and VDP® are registered trademarks of Actifio, Inc. Manage Data Simply™, Virtual Data Pipeline™, Protection and Availability Storage Platform™, PAS™, Copy Data Storage Platform™, CDS™, and Actifio Sky™ are trademarks of Actifio, Inc. All other brands and product names for goods and/or services mentioned herein are trademarks or property of their respective owners.

Actifio believes the information in this publication is accurate as of its publication date. Actifio reserves the right to make changes to information published in this document, including without limitation specifications and product descriptions, at any time and without notice. This document supersedes and replaces all information supplied prior to the publication hereof.

THE INFORMATION IN THIS PUBLICATION IS PROVIDED "AS IS." ACTIFIO, INC. MAKES NO REPRESENTATIONS OR WARRANTIES OF ANY KIND WITH RESPECT TO THE INFORMATION IN THIS PUBLICATION, AND SPECIFICALLY DISCLAIMS IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.

This software and the associated documentation are proprietary and confidential to Actifio. Use, copying, and distribution of any Actifio software described in this publication requires an applicable software license. Any unauthorized use or reproduction of this software and the documentation may be subject to civil and/or criminal liability.

Actifio strives to produce quality documentation and welcomes your feedback. Please send comments and suggestions to docs@actifio.com.

October 2015

Contents

Challenges Protecting Unstructured Data	1
Actifio BDD Addresses These Challenges	2
What is Actifio Big Data Director?	2
How Does it Work?	3
Datasets	3
Actifio SLAs	3
Initial Data Capture	3
The Second and All Subsequent Data Captures	4
Optional Inline Deduplication	4
File Recovery	5
Actifio BDD Vs. NDMP	5
Summary	6
About Actifio	6

Actifio Big Data Director

Virtual Data Pipeline for Unstructured Data

October 2015

The simplicity of the file and folder model and ease of access has made Network Attached Storage (NAS) the technology of choice for storing large amounts of unstructured data.

The introduction of scale-out storage for NAS systems has raised the bar for both scale and performance. With scale-out storage for NAS systems, customers can expand their storage as needed and now routinely store hundreds of terabytes or multiple petabytes of data and manage the data from a single pane of glass. Protecting this amount of data has presented a formidable challenge to legacy backup tools.

Network Data Management Protocol (NDMP) was designed over a decade ago and is a common protocol used to backup (protect) NAS environments. It was designed when terabytes were the exception, not the norm.

This paper identifies the challenges associated with protecting data stored on NAS devices and how Actifio addresses those challenges with its Big Data Director solution.

Challenges Protecting Unstructured Data

Some of the most common data protection challenges that face NAS storage include:

- **The Scan Curse:** To perform incremental backups, current protection solutions rely on scanning the entire file system to determine changed files. Scanning terabytes and petabytes of data is simply not viable.
- **Recovery Speed:** Recovering file(s) and folder(s) from a NAS share is like finding a needle in a haystack. Recovery is a multi-step process that comes at the expense of speed: Proprietary incremental and full data dump files must be recovered; often times to proprietary hardware. Recovered, incremental and full backups must be stitched together. Once stitched together, the entire backup must be searched for the file(s) and folder(s) to be recovered.

Since most restores are done from the most recent backup, as a work around, users will keep recent snapshot copies of their data from which they will perform restores. Though effective, this approach does not scale and creates unnecessary operational and capital (OPEX and CAPEX) expenses.
- **Periodic Full Backups:** NDMP requires periodic full backups to limit the complexity and time needed to stitch together full and incremental backups. Recoveries become more complex and time consuming as the chain of incremental backups between full backups grows.
- **Consumption of System Resources:** Full file system scans, periodic full backups, keeping multiple copies of full backups on hand for fast restores, and the stitching together of incremental backups consume system resources (CPU, memory, etc.).
- **Vendor Lock-in:** To perform a recovery, NDMP requires a NAS server of the same type and in some instances, similar versions of firmware. This locks users in to a particular vendor and often leads to higher costs.
- **Complexity:** A backup strategy has to balance the needs of Recovery Time Objectives and Recovery Point Objectives. Requirements include local and remote storage as well as long and short term storage of data. To accommodate all of this, existing protection solutions necessitate the use of multiple point tools, often from multiple vendors, leading to a backup infrastructure that is complex to procure, deploy and operate.

It became apparent to Actifio that NAS storage was in need of a modern, radically simple approach to protecting data.

Actifio BDD Addresses These Challenges

Actifio's purpose-built Big Data Protector (BDD) coupled with Actifio's modern approach to data management eliminates the challenges encountered in NAS environments.

Specifically BDD:

- Eliminates file system scanning curse
- Provides instant access to unstructured data regardless of size
- Enables a true incremental forever without periodic fulls and the overhead associated with constructing synthetic fulls
- Minimizes the load on production NAS systems
- Breaks the storage vendor lock-in by allowing storage systems from different vendors in both primary and disaster recovery sites
- Captures and maintains file data in file system format along with ACLs
- Uses a simple SLA driven approach to protecting unstructured data
- Eliminates the need for multiple point solutions to enable short and long term retention at local and remote sites

What is Actifio Big Data Director?

Actifio's Big Data Director (BDD) solution consists of:

Big Data Director Node(s): Actifio BDD nodes provide the front end that allows you to scale ingest capacity and to enhance optional inline deduplication.

Each BDD node is configured with 2X10G Ethernet ports for reading and sharing file data and up to 4X8G Fibre Channel connections for SAN connectivity to an Actifio CDS appliance. Up to 8 BDD nodes can be paired with a single Actifio CDS appliance.

BDD nodes handle NAS vendor API integration for efficient change file tracking, ACL parsing, and file based sharing.

An Actifio CDS Appliance: An Actifio appliance is used as the back end to an Actifio BDD solution. The Actifio BDD leverages the proven Virtual Data Pipeline (VDP) technology built into an Actifio CDS appliance. The Actifio CDS appliance provides the SLA framework, block-level deduplication and replication services for file data.

The entire Actifio BDD solution is managed from a single pane of glass via the Actifio CDS appliance's user interface.

How Does it Work?

Protecting data with Actifio BDD is simple, radically so: Define a dataset, then apply an Actifio SLA to the dataset.

Datasets

As the name implies, a dataset is simply a set of data. A dataset can be made up of an entire NAS share, a piece of a NAS share or a group of NAS shares. You define the start paths, as well as the file and folder include and exclude patterns that make up the dataset.

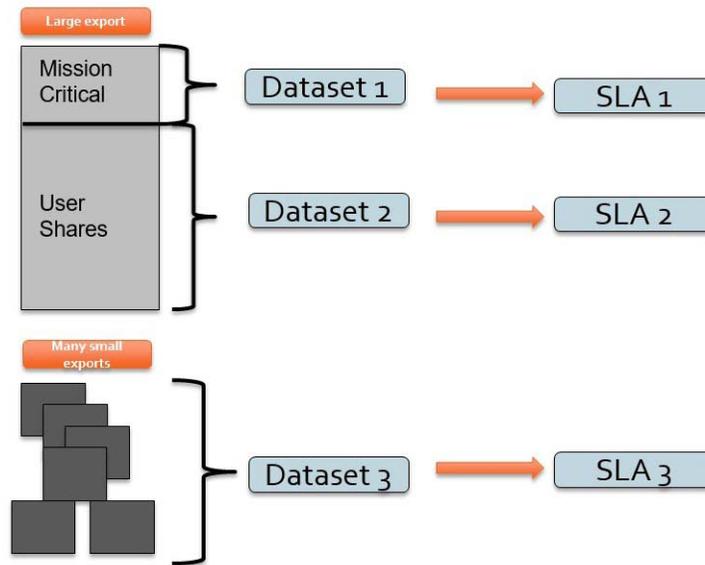
Datasets do not make changes to the layout of the NAS share's production data.

Actifio SLAs

An Actifio SLA defines capture schedules, as well as optional deduplication, and replication requirements. SLAs are applied to datasets.

For example, in the following figure there are three dataset/SLA pairs:

1. A dataset consisting of mission critical data that is paired with an aggressive SLA that protects the dataset every 6 hours and replicates the dataset to another Actifio appliance.
2. A dataset that does not contain mission critical data and is paired with an SLA that protects the dataset daily and does not replicate the dataset to another Actifio appliance.
3. A dataset made up of multiple NAS shares that is paired with a single SLA that protects all of the datasets weekly and does not replicate the dataset to a remote Actifio appliance.



Initial Data Capture

The first time a dataset is captured, the Actifio BDD captures the entire dataset according to the SLA to which the dataset is paired.

As seen in the following illustration, Actifio BDD makes the vendor specific API call to take a snapshot of the data set and mount the snapshot. The Actifio BDD then mounts the snapshot and then captures the entire snapshot.



The Second and All Subsequent Data Captures

After a dataset has had its initial full data capture, going forward the Actifio SLA applied to the dataset will only perform incremental data captures made up of changed files.

When the incremental capture is complete it is combined with the previous full data capture to create a new, up-to-date copy the entire dataset.

To perform incremental data captures, the Actifio BDD integrates with the native API's from the NAS vendor to obtain a list of changes files.

As seen in the following illustration, the Actifio BDD makes the vendor specific API call to take a snapshot of the dataset and fetches only the changed files. The Actifio BDD then mounts the snapshot and captures only the changed files.



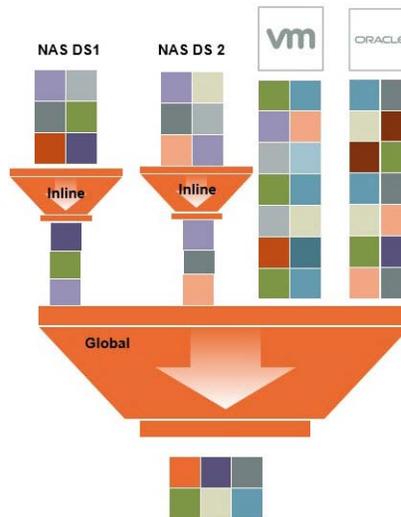
Optional Inline Deduplication

The Actifio BDD has an Inline Deduplication option that can be applied to datasets. This deduplication is in addition to the global deduplication provided by the Actifio CDS appliance to which the BDD is connected.

The inline deduplication is designed for speed and operates at a higher granularity than the Actifio CDS appliance. Inline deduplication is particularly valuable in those environments where space is a premium.

Enabling inline deduplication on datasets minimizes the amount of storage needed in the snapshot pool. This is particularly useful for large datasets.

The following figure illustrates an environment where two NAS datasets will first be deduplicated inline by the Actifio BDD. The data sets will then be deduplicated along with a VM and an Oracle database via an Actifio CDS appliance's global deduplication engine.



File Recovery

Actifio BDD File recovery is performed through an operation called Mount & Export. Unlike block data, file data needs to be shared as an NFS or CIFS share to be usable. An Actifio BDD node acts as a filer during a Mount and Export operation.

Captured datasets, which are stored on Actifio CDS appliance VDP LUNs, are shared as an NFS or CIFS share. Regardless of the size of the dataset, the Mount and Export operation is completed and made available almost instantly.

ACL information is recovered along with the data set. This allows you add the Actifio BDD to a Microsoft Active Directory/ LDAP domain that will honor the ACL information.

Actifio BDD Vs. NDMP

Actifio BDD was designed to eliminate the trade offs customers are forced to make when implementing a NDMP based backup solution. Specifically, Actifio BDD eliminates:

- The recovery time imposed on users of NDMP
- Vendor lock-in that comes from the NDMP proprietary dump formats
- The need for heterogeneous storage systems at both primary and disaster recovery sites
- Periodic full backups required by the NDMP's incremental architecture
- Merging full and incremental backup's to restore and recover data.

Summary

Actifio Big Data Director brings a fresh, new, and modern approach to unstructured data management. The Actifio BDD eliminates the challenges that customers face using traditional approaches thereby simplifying unstructured data management, reducing CAPEX and OPEX.

The following table highlights how Actifio BDD modernizes the management of unstructured data:

Feature	Benefit	Description
Incremental Forever	Reduce backup window	An Initial full backup, then incremental forever captures eliminates the need for periodic full backups. Multiple data capture streams can be run simultaneously.
Change Data Tracking	Reduced: Backup window, performance impact on NAS filers, and operational complexity	Integrates with vendor APIs to quickly determine changed files and avoids the scan curse.
Instant Recovery	Low RTO Simple and easy for operations to achieve business resiliency	Data is present instantly without moving the data from Actifio BDD to a NAS filer. The data is writable with scalable performance that is as good as the underlying storage presented
Multiple Virtual Copies	Can be used for data analytics, test and development and reduce storage costs	Users can provision multiple virtual copies instantly. These virtual copies are writable and scalable
Dissimilar Storage Support	Eliminates vendor lock-in and reduces storage costs	Customers can use any storage from their HCL. Gain independence from storage vendors.
Efficient Dedup and replication	Reduces storage and network bandwidth costs	Global deduplication across all data protected, at very small block sizes. Achieves significant dedup compression even for small files. This is important for NAS filers where there could be millions of small files.
SLA Architect	Simple and easy to operate from a centralized management location	Create policies for data sets on NAS filers. Admins can specify how often to capture data, how long to retain at primary site, how much to replicate to a remote site, how long to retain it at a remote site, and whether to rehydrate data at remote a site for instant recoveries.

About Actifio

Actifio is radically simple copy data management. Our copy data management lets businesses manage and recover anything instantly, for up to 90 percent less. Actifio eliminates siloed data protection applications, virtualizing data management to deliver an application-centric, SLA-driven solution that decouples the management of data from storage, network, and server infrastructure. Actifio has helped liberate IT organizations and service providers of all sizes from vendor lock-in and the management challenges associated with exploding data growth. Actifio is headquartered in Waltham, Massachusetts, with offices around the world. For more information, please visit www.actifio.com or email info@actifio.com.